# THE EFFECTS OF MUSIC ON COGNITION AND ACTION

EDITED BY: Marta Olivetti Belardinelli, Franco Delogu, Elvira Brattico and Cunmei Jiang

PUBLISHED IN: Frontiers in Human Neuroscience, Frontiers in Psychology and Frontiers in Neuroscience

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-730-6
DOI 10.3389/978-2-88974-730-6

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

# THE EFFECTS OF MUSIC ON COGNITION AND ACTION

Topic Editors:
**Marta Olivetti Belardinelli,** Sapienza University of Rome, Italy
**Franco Delogu,** Lawrence Technological University, United States
**Elvira Brattico,** Aarhus University, Denmark
**Cunmei Jiang,** Shanghai Normal University, China

**Citation:** Belardinelli, M. O., Delogu, F., Brattico, E., Jiang, C., eds. (2022). The Effects of Music on Cognition and Action. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-730-6

# Table of Contents

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 November 2019
doi: 10.3389/fpsyg.2019.02577

# Head Movement Synchrony and Idea Generation Interference – Investigating Background Music Effects on Group Creativity

**Sarinasadat Hosseini[1], Xiaoqi Deng[1], Yoshihiro Miyake[1] and Takayuki Nozawa[2]***

[1] Department of Computer Science, Tokyo Institute of Technology, Kanagawa, Japan, [2] Research Institute for the Earth Inclusive Sensing, Tokyo Institute of Technology, Tokyo, Japan

Previous studies have indicated that divergent idea integration is an effective way to foster extraordinary creativity in groups. This study posits that background music (BGM) may aid in eliciting this phenomenon. Here to describe the effectiveness of BGM on group creativity, we hypothesized and suggested different mechanisms that genre and valence attributes of BGM would lead to extraordinary creativity. The temporal co-ordination of head movement synchrony (HMS) was investigated as a non-verbal cue and we found significant HMS response levels to idea generation. While the HMS as response did not depend on the quality of the prior ideas; it led to higher divergence and originality in the successively generated ideas. Results of this study showed the dominant contribution of upbeat positive valence (UP) music, relative to other genres, in HMS leading to divergent ideas. Following this, the potential role of upbeat music in enhancing participant sociability and positive valence in enhancing cooperation level was discussed. Upbeat positive music may decrease judgmental behavior during creative group tasks and inspire participants to share divergent perspectives. The use of such music can encourage participants to share new perspectives and integrate ideas. It may also provide a potential explanation for the enhancing effect of upbeat positive music on creative outcomes in groups.

**Keywords: background music, communication, group, creativity, head movement synchrony, idea generation, idea sharing**

OPEN ACCESS

***Edited by:***
*Marta Olivetti,*
*Sapienza University of Rome, Italy*

***Reviewed by:***
*Dan Zhang,*
*Tsinghua University, China*
*Enrico Cupellini,*
*Independent Researcher, Rome, Italy*

***Correspondence:**
*Takayuki Nozawa*
*nozawa.t.ac@m.titech.ac.jp*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology*

***Received:*** *07 July 2019*
***Accepted:*** *31 October 2019*
***Published:*** *15 November 2019*

***Citation:***
*Hosseini S, Deng X, Miyake Y and Nozawa T (2019) Head Movement Synchrony and Idea Generation Interference – Investigating Background Music Effects on Group Creativity. Front. Psychol. 10:2577. doi: 10.3389/fpsyg.2019.02577*

## INTRODUCTION

Creativity is the development of influential original ideas (Paulus and Nijstad, 2003). Group creativity can be considered a clear illustration of goal-oriented social interaction. Group idea generation and brainstorm sessions are commonly used tools to foster the creative process in educational and industrial settings (Elias et al., 2011). Although brainstorming and group creativity demonstratively facilitate creative outcomes; barriers such as social anxiety and social loafing arguably reduce the efficiency of such processes (McCauley, 1998). The hindered performance can be interpreted by the motivation aspect during teamwork. Group creativity is a group task, and group task outcomes will be based on the output of two major factors – (i) the motivation of individuals to cooperate to perform the task (Slavin, 1983) and (ii) the capacity of individual group members to complete the task. The motivation to co-operate is necessary for a group to function as a system that is more than a mere collection of people placed together, and can be viewed as constituting a system of social interactions (Hare et al., 1955).

Here in this study, we introduced background music (BGM) as a factor to foster group creativity. Despite many studies have been conducted on the effects of music, as music comprises an important part of our daily life, most studies have focused on individual aspects (Rauscher et al., 1995; Thompson et al., 2001; Waterhouse, 2006). Although the impact of music on individual creativity is still under investigation and not based on a unified perspective (Schellenberg, 2015), one recent article highlighted how positive valence classical pieces enhance divergent creativity (Ritter and Ferguson, 2017). On the other hand, studies support the potential effect of music on social interaction – and thus goal-orientated group achievements – including the impact of music on task involvement and the capacity to cope with perceived stress (Fried and Berkowitz, 1979),which would affect communication (Schabracq and Cooper, 2000), and reinforcing social interaction (Cross and Morley, 2009). Consequently, while the potential impact of music on group coordination, and thus group creativity, in varying ways have been indicated, there has been scarce research examining the mechanisms that may underlie music's impact on group creativity outcomes. Building on our previous work on music (Hosseini et al., 2019), this study firstly confirms the effects of music on group creativity indices, with a larger sample. More importantly, this study explores the temporally refined process that underlies the observed effects of music on group creativity, by positing dynamic non-verbal behavior as a possible determinant of creative interaction. In the following further explanation on non-verbal behavior and its possible connection to group creativity is provided.

As social animals, humans share and receive static and dynamic experiences of non-verbal cues incorporated in their facial features and body movements (Knapp et al., 2013). In group communication, due to the contribution of cooperation and interaction, non-verbal cues have always played an important role to convey information (Mehrabian, 2017). Accordingly, interacting individuals unintentionally synchronize their non-verbal behavior along many levels of social interaction (Latif et al., 2014), leading to feeling of connectedness and cooperation (Marsh et al., 2009), which in turn may bring significant and positive impacts on group creativity (Chen et al., 2008). Thus, dynamic non-verbal cues arguably represent a source of substantial information on the quality of social interactions and therefore group creativity. Specifically, head movement synchrony (HMS) is a common phenomenon in various human communications, and have been indicated to be a marker of successful communication (Ramseyer and Tschacher, 2014). Regarding the music impact on non-verbal cues, noting the enhancing effect some rhythms can have on group synchronization (Brown, 2000), and therefore cooperation (Wiltermuth and Heath, 2009), previous study by our group has also showed the enhancing effect of music (of all combinations and kinds) compared to no music on HMS (Hosseini et al., 2019). The correlation between HMS and group creativity and possible response of HMS to idea generation had not been addressed in our previous study. To address this deficiency, we aimed to first clarify the relationship between HMS response and idea generation.

Understanding group creativity requires recognizing that both the capacities to contribute divergent ideas and to integrate those ideas are wellspring of group creativity. Recent studies on creative outcomes have noted the importance of divergent input for group creativity, as divergent resources stimulate variety in output, as exposure to a greater diversity of ideas enhances the likelihood that participants will produce innovative outcomes (Harvey, 2014). Contrastingly, idea integration during creative process is also important (Xue et al., 2018), as groups that can integrate divergent input have the potential of engendering extraordinary creativity (Eisenbeiss et al., 2008; Harvey, 2014). Consequently, as the second aim we explored potential contribution of HMS response to the quality of the successive idea generation, especially on the divergent/convergent aspect. In a similar fashion with a recent study that examined how dynamic non-verbal movement may emergently affect quality of speech (Paxton and Dale, 2017), the present study suggests that dynamic HMS, which emerges through exchanges of non-verbal cues in interaction, may likely modulate the quality of shared ideas in a bottom-up manner. Here, we also wanted to introduce the BGM capacity to systemize divergent idea sharing. Elaborating the dynamic aspect of idea generation with the idea quality importance aspect, as the second part of our hypothesis, we assumed HMS would help to induce divergent ideas, while upbeat genre, relating to extraversion, along with positive valence, associated with cooperation, and results in successive divergent idea sharing.

For these two purposes, we augmented the Alternative Uses Task (AUT) for this study as the creativity task. The AUT is originally a measure of individual divergent creativity levels (Guilford, 1967). This study used an adapted version of the AUT, that was expanded to investigate group creativity. The adapted AUT involves communication between participants and requesting participants to discuss their ideas so that each group would reach a common idea via a discussion decision-making approach. We recorded the head movement of participants during discussion for further assessments of HMS along with the records of participants' answers during the AUT discussion, to acquire raw data on idea generation; as the measure timing of idea generation. The timing of idea generation is important as it allows us to investigate if any possible mutual coordination between HMS and idea generation was present. This data also would aid in examining the potential determining effect of the HMS on the quality of the ideas as well as also us to observe the effect of music types.

## MATERIALS AND METHODS

### Ethics Statement

The Human Subjects Research Ethics Review Committee of the Tokyo Institute of Technology approved this study procedure. All participants were briefed on the experimental procedure and provided written informed consent prior to participation in the experiment. Participants were paid 3000 Yen as a reward for their time and effort after the experiment.

## Participants

Forty international students were recruited from the Tokyo Institute of Technology (23 males, 17 females; age = 26 ± 3.83) via flyers and online surveys to participate in this experiment. They were from various nationalities. All participants were right-handed with normal or corrected-to-normal vision. Participants answered two online preparatory surveys and were paired in groups of two or dyads (6 female–female; 10 female–male; 4 male–male) based on the results of those surveys. Information regarding the online preparatory surveys is provided in the next section.

## BGM Music Pieces

A previous study have referred to the association between unconscious aspects of personality and music preference (Cattell and Saunders, 1954). However, given that such studies were based on a limited selection of music genres, Rentfrow and Gosling (2003) proposed several major dimensions of music – reflective and complex, intense and rebellious, upbeat and conventional, and energetic and rhythmic – to formulate a systematic account of how music may impact individual behavior. Thereafter, this account organized several well-known music genres along each dimension and compared the correlations between music preference in each dimension, Big Five personality trait scores, and cognitive ability. Their results indicated that the reflective and complex dimension was positively related to openness and self-perceived intelligence, while negatively related to social ability and self-perceived rejection of old fashion ideals. In contrast, the upbeat and conventional dimension was positively associated with extraversion, conscientiousness, and self-perceived physical attractiveness. Furthermore, in this study, the classical genre was related to the reflective dimension, while pop, country, and sound track music genres were related to the upbeat dimension (Rentfrow and Gosling, 2003). In the present study, we fixed our music genre dimension of interest as reflective and complex (hereafter labeled as reflective) versus upbeat and conventional (labeled as upbeat), and selected classical tracks to represent reflective dimension while music tracks in the pop, country and sound track genres to represent upbeat dimension.

On the other hand, positive valence in musical pieces has been reported to correlate with divergent creativity while negative valence was found to relatively hinder creativity levels in previous studies (Thompson et al., 2001). In the present study to determine the valence of musical pieces we referred to the general key of pieces and assigned pieces with major keys as having positive valence while labeling pieces with minor keys as having negative valence.

Combining the genre and valence dimensions, we derived four music categories – reflective negative valence (RN), reflective positive valence (RP), upbeat negative valence (UN), and upbeat positive valence (UP), and compared their effects on group creativity.

## Dyadic Group Formation

In the preparation phase, participants answered two online preparatory surveys. The contents of the first survey concerned participants' available time to participate in the experiment and their contact details. The second survey was concerned with their subjective evaluation of 100 music pieces, consisting of 25 pieces each from each musical category (RN, RP, UN, and UP), presented in a randomized order. As there is a rhythmic effect of music on body movement, the selected music pieces were controlled for tempo (Kellaris and Kent, 1991) to ensure that the pieces were within the range of [95–105] BMP (beats per minute) on average. A recent discussion over the effect of music familiarity and likability on modulating brain region activity motivated us to control for the familiarity and likability attributes in the presented music stimulus (Pereira et al., 2011). Thus, this study only used tracks that unfamiliar but likable for the both members in each dyad. Participants were asked to rate their perceived mood regarding musical pieces to ensure our music valence selection related to engendering or inducing a mood in the participants. To do so participants were asked to listen to the first fifteen seconds of one hundred instrumental music pieces and rate them in terms of likability on a five-point scale (1 "really disliked" to 5 "really liked"), familiarity on a five-point scale (from 1 "unfamiliar" to 5 "familiar"), and perceived mood individually on a five-point scale (from 1 "felt sad" to 5 "felt happy"). Participants who were available on the same experimental time slot, and who rated at least one matching track from each of the reflective and upbeat genres as low on familiarity (unfamiliar) and likable, and very low on mood (representing sad mood) for negative valence pieces and very high on mood (happy mood represented) for positive valence ones, were placed in dyads and called to participate in the experiment. Furthermore, we controlled the intensity of selected pieces to control for participants perceived arousal (Ilie and Thompson, 2011).

## Experimental Procedures

Participants sat face-to-face with a table in between them and a television screen on their side (experiment setting is depicted in **Figure 1**). The experimental procedure consisted of two sessions. The first session included three trials and was followed by a ten-minute break and the second session with two trials. Each trial began with an auditory (beep) cue, lasted for a duration of 6 min while the participants completed a group AUT problem, and ended with another auditory cue. AUT is typically a measure of individual creativity. To adapt the AUT for the group creativity measurement, participants were asked to share and discuss the feasibility of their ideas over the alternative usage of familiar objects between each other. On each starter auditory cue, participants were asked to look at the TV screen where name of the AUT stimuli or objects ("1 m of cotton rope," "An egg," "A plank of wood," "A tennis ball," and "A pair of socks") were presented. Participants had to try discuss as many ideas as possible and to try be as original as possible when imagining and discussing alternative uses for the objects presented. In each trial, a musical piece (from one of the four music categories selected in the preparation phase) or no music was played as

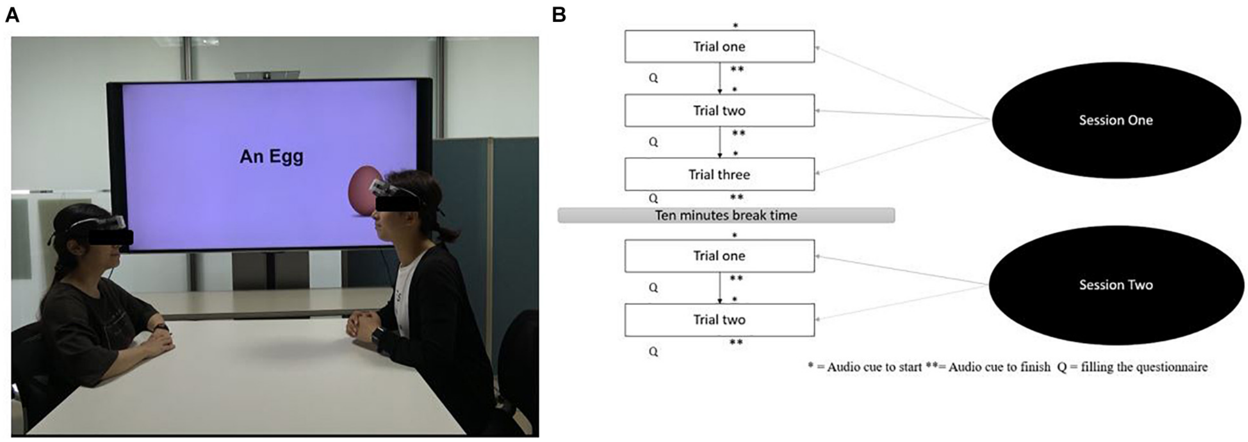



**FIGURE 1 |** A snapshot of experimental setting **(A)** and the overview of experimental procedures **(B)**.

BGM. The order of the objects and music conditions presented in the group AUT task were independently randomized over the dyads. Participants then rated their emotional state (within 50 s) in terms of pleasantness after each trial (the questionnaire results were irrelevant to this particular study).

The experimenter conducted a practice trial with the upbeat positive condition as BGM before the start of the first session to ensure the participants understood the contents of the experiment and that measurement devices were working well. It should be noted that the track used for practice trial was a pre-defined track and differed from the tracks used during main experiment. After the practice trial participants rested for 5 min to reduce any potential effects from this trial on their mood. The task procedure is depicted in **Figure 1**. To evaluate creative task performance, the experimenter recorded the utterances made during the group AUT using a video camera (FDR-AX40; SONY).

## Assessment of Indices of Creativity Task Performance

We addressed the indices of creativity in three categories of fluency, originality, and index of the convergent (IOC) (Larey and Paulus, 1999). Fluency was the total number of ideas that dyads mentioned during each trial. Idea originality was the function of occurrence of an idea for each object over all groups, where those that occurred in fewer groups indicated higher originality (Dippo and Kudrowitz, 2013).

Firstly, let us denote the frequency count of idea *ID* for an object *OB* over all participant groups as $n(OB,\ ID)$. Then, the raw originality score of *ID* for *OB* is defined by

$$OriginalityRaw\ (OB,\ ID)\ =\ max(n(OB,\ \cdot))\ -\ n(OB,\ ID)$$

Here, $function(n(OB,\ \cdot))$ indicates the function [$maximum,\ mean,\ and\ SD\ (as\ shown\ below)$] of frequency of idea(s) over all groups. Note that the distribution of {*OriginalityRaw* (*OB, ID*) | *ID* in the set of all ideas for object *OB* generated by the all groups} can be very different between objects. For example, an *OB* might have lower max ($n(OB,\ \cdot)$) than other objects, meaning less common ideas for that particular *OB*. Then, the range of *OriginalityRaw (OB, ID)* would be narrower. This difference in distribution is problematic, as though the objects were assigned to different BGM conditions over participant groups, we want to probe into the difference in idea originality caused by different BGMs while controlling (and thus excluding) the influence of object assignment. Therefore, we calculate the *standardized originality score* of each idea as

$$OriginalityZ\ (OB,\ ID) = \frac{OriginalityRaw(OB,\ ID)\ -\ mean(OriginalityRaw(OB,\ \cdot))}{SD(OriginalityRaw(OB,\ \cdot))}$$

Then, the *mean originality score* of the ideas generated in the trial by a group *GR* for an object *OB* is given by

$$Originality(GR,\ OB)\ =\ mean\_GR(OriginalityZ(OB,\ ID\_GR))$$

Here, the mean is calculated only over the ideas generated by the group *GR*.

Index of the convergent is the index of cooperation and idea integration. IOC for a trial is defined as the total number of times a generated idea by a participant remained in the same category as a previously mentioned idea by the other participant in the dyad over the total number of ideas mentioned during that trial.

To decide on the categories, we referred to the general treatment of the objects – e.g., a plank of wood, may produce ideas like bracelets and earrings which were then placed into a category called "accessories."

## Data Analyses

### Head Movement Synchronization (HMS) in Communication

To assess the data of head movement and detect head movement synchronization (HMS), a small wireless accelerometer (TSND121; ATR Promotions) with a sampling rate of 10 Hz was attached to a functional Near-Infrared Spectroscopy device (HOT-1000; Hitachi High-Technologies; the results of brain signals were not a matter of interest in this particular study) in the middle of a participant's forehead. The head movement time series was analyzed using the same method used in prior studies (Thepsoonthorn et al., 2016; Yokozuka et al., 2018). After extracting the raw head movement data from each participant and applied a short-time frequency analysis (window width = 1280 ms and window Increment = 100 ms) (Fraisse, 1982) to acceleration norms, we assessed the amplitude of head movement within a frequency band of 1.0–5.0 Hz. By applying Spearman's rank correlation, we calculated head motion correlations with time lags within the range of −500 to 500 ms. We used the head movement correlations in two formats. First, we averaged the value of head movement correlations in all frequency bands and time lags as the HMS during that trial. Second, for each second of the duration of each trial, we summed up all HMS in all frequency bands and if the result indicated the occurrence of HMS (i.e., a summation greater than zero), we coded that second as 1; otherwise, the HMS in that second was coded as 0. By doing so, a sequence of 360 s (6 min) of HMS responses was formed for each trial and we called the acquired vector $HMS_t$.

## Temporal Coordination of Idea Generation and HMS

We hypothesized that HMS takes place more frequently in response to idea generation, reflecting the coordinated bodily expressions (e.g., agreement, evaluation, and co-laughter) to the ideas. If this is the case, we would observe higher chance of HMS within a short time window just after each idea generation than in other moments. This temporal coordination between idea generation and the responding HMS was assessed in the following steps. Step 1 involves the idea generation sequence. Step 2 involves the execution of a surrogate idea generation sequence. Step 3 involves the time lag-based co-occurrence of idea generation and HMS response. Finally, step 4 involves surrogate-based $z$-score of this co-occurrence. We further used $z$-score based analysis to test if the co-occurrence of idea generation and HMS falls out of the null distribution based on the surrogate data, and identified the time lag that provided the most robust co-occurrence detection in terms of the highest departure from the null distribution. Here, selection of this time window setting is important as the selection of an optimal time window setting, through surrogate-based $z$-scores, enables us to identify the most robust temporal coordination. To be more descriptive, the co-occurrence captured with other time windows might be more contaminated by chance or other factors such as the preceding ideas. With the choice of the current analysis, we can claim that the detected idea-to-HMS co-occurrences are least affected with such contaminating factors. We detail each step in the following.

In Step 1, the time series of idea generation $I_t$ is related to the idea generation times during the discussion between dyads in each of the trials. Based on the recorded voices of the participants during the experiment, for each second during 6 min of one trial, if any of the participants mentioned an idea at $t_i$, the idea is noted with $t_i$. Based on this information, the idea generation binary sequence was constructed as such below:

$$I_t = \begin{cases} 1 & t = t_i \\ 0 & otherwise \end{cases}$$

In Step 2, we first captured the time gap between two idea generation times. Where the interval denotes the sequence of this time. With this in mind, we consider that $N$ is the total number of ideas, $t_i$ represents the output time of one idea, and $t_{i+1}$ represents the output time of the following idea. Thus, the interval sequence is shown in the following equation:

$$Intervals = \{t_{i+1} - t_i \; ; \qquad i = 0,\ 1, \ldots, N\}$$

Here we used surrogate-based data, in order to characterize group creativity. Surrogating the signals helps in the robust characterization of the system's behavior and function of any dynamical units. Consequently, the use of surrogate data testing enables robust statistical evaluations to ensure that the observed results are not by chance but are true characteristics of the dynamics. To make the surrogate data, the order of the interval sequence was randomly permutated 1000 times, while maintaining the distribution of the intervals between ideas generated the same as the original distribution. A 1000 surrogate idea generation time series thus obtained is denoted by $SI_t$. **Figure 2** depicts a sample of the idea generation sequence and a sample of its surrogate sequence.

In Step 3, set the idea generation sequence as a reference, we captured the HMS within $w$ seconds of the generation of a particular idea. For this purpose, we assessed ten time-window settings of HMS sequences within [0, $w$] ($w$ = 1–10) seconds from idea generation. Taking $HMS_t$ as the original sequence of the head movement time series for each trial, the $w$ seconds time window adopted sequence of head movement synchrony $HMSW_t$ was formulated as:

$$HMSW_t = \begin{cases} 1 & HMS_t,\ HMS_{t+1},\ \ldots,\ HMS_{t+w} = 1 \\ 0 & otherwise \end{cases}$$

**Figure 3** shows the $HMSW_t$ within ten-time window settings.

Furthermore, to assess the co-occurrence of idea generation and HMS response, we calculated the dot product of $I_t$ and $HMSW_t$ as $co\_oc_w$, along with the dot product of $SI_t$ and $HMSW_t$ as $Sco\_oc_w$. These can be summarized as:

$$co\_oc_w = (I_t, HMSW_t)\,,$$
$$Sco\_oc_w = (SI_t, HMSW_t)$$

Based on these steps we calculated surrogate-based $z$-scores of the summary co-occurrence of idea generation and HMS below:

$$z - score = \frac{co\_oc_w - mean(Sco\_oc_w)}{SD(Sco\_oc_w)}$$

In the Step 4, we averaged the surrogate-based $z$-scores of the co-occurrence over the five BGM conditions (including no

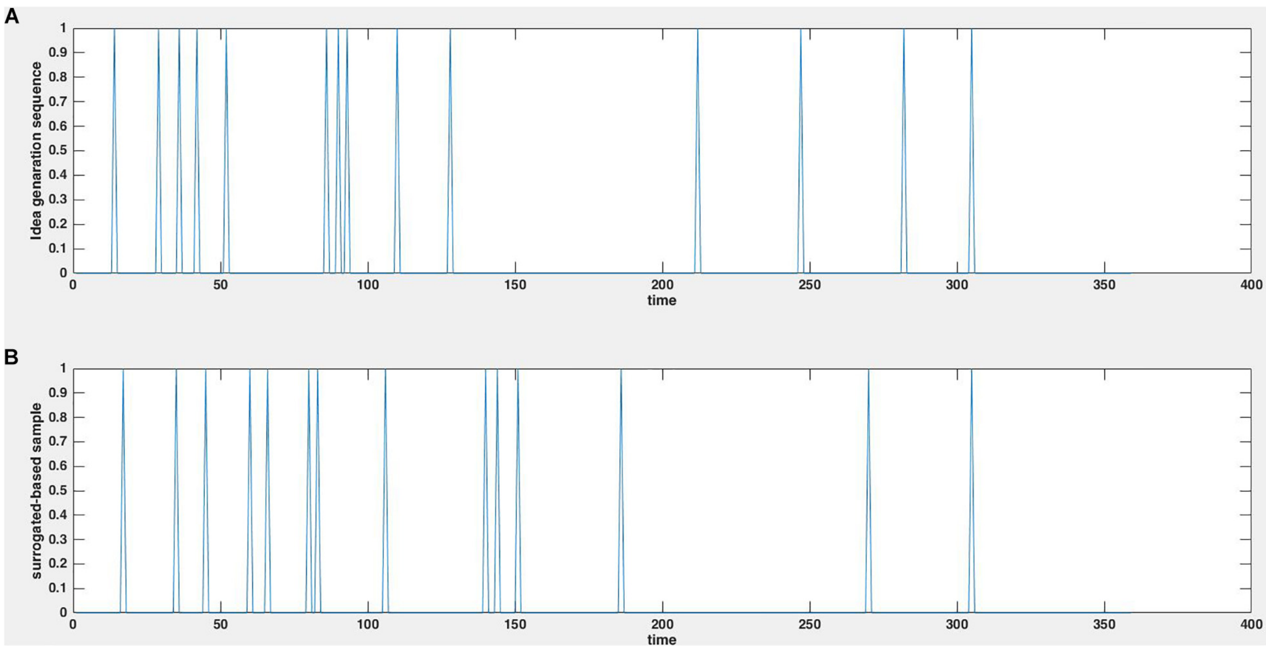



**FIGURE 2 |** Example of an original idea generation sequence **(A)** and its surrogate **(B)**, which was generated by randomizing the order of idea-to-idea intervals.

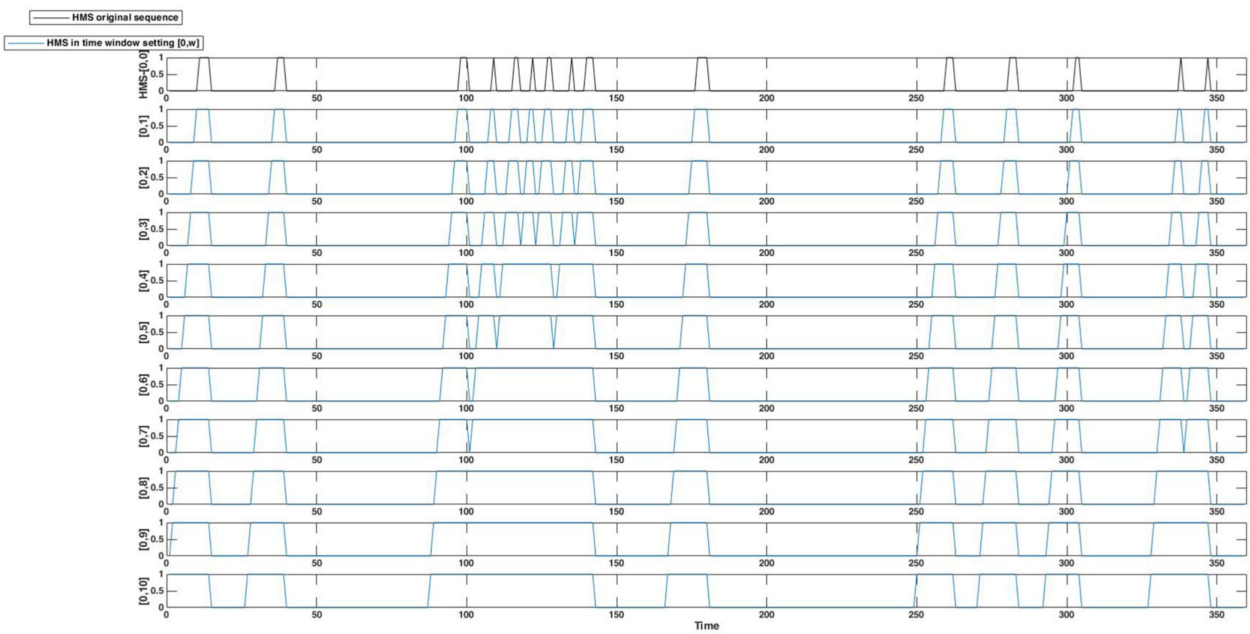



**FIGURE 3 |** An example of original HMS **(top)** and its window-adopted versions, with 10 time window settings [0, *w*] (*w* = 1–10).

music) for each dyad as the following:

$$mean - z - score = \frac{\sum_{i=1}^{5}(z - score)_i}{5}$$

Finally, to test whether idea generation was accompanied by lagged HMS beyond chance, we performed a one-sample *t*-test on the mean-*z*-scored co-occurrence values of the 20 dyads for each window setting *w*. Also, comparing the *t*-values we identified the window setting *w* with the highest *t*-value as the optimal time window that captures the co-occurrence of idea generation and HMS response the most robustly. From this point onward, we only used the time window setting with the highest *t*-value to capture HMS from idea generation.

### Relationship Between HMS and Convergent/Divergent Nature of Ideas

Recent theoretical and practical advances regarding sharing different kinds of ideas to facilitate group creativity motivated us to consider relationship between HMS and the nature of ideas (Harvey, 2014).

Harvey (2014) tried to identify factors that facilitate both divergent and convergent creativity in groups by modulating groups and their individual tasks. Here, we tried to use music as a means to improve group creativity outcomes by fostering divergent idea sharing and the integration of diverse ideas.

For this purpose, two idea attribution vectors, $C_j$ and $O_j$, were introduced, where $C_j$ represents the convergence of each idea denoted by $I_j$, and $O_j$ represents the originality of idea $I_j$. As defined above, originality $O_j$ is negatively proportional to the count of the ideas being generated for each stimuli object by different groups. Here the constant factor of $O_j$ does not change the later result due to the normalization applied below. Given that the idea occurrence count was $n$, groups was $ID$, and object was $OB$, and the total ideas generated during each trial was $N$, $C_j$and $O_j$ were formulated as below:

$$C_j = \begin{cases} 1 \\ 0 \end{cases} \quad I(t)\ is\ Convergent\ otherwise \qquad for\ j = 1, 2, \ldots, N$$

$$O_j = -n \qquad forj = 1, 2, \ldots, N$$

Furthermore, we introduced a binary idea-HMS co-occurrence series $Vco\_oc_{wj}$ as the logical product of the elements of $I_t$ and $HMSW_t$ (i.e., for an idea $j$ with $I_{ti} = 1$, if $HMSW_{t_i} = 1$ then $Vco\_oc_{wj} = 1$, otherwise $Vco\_oc_{wj} = 0$). **Figure 4** shows the $Vco\_oc_{wj}$ for the time window [0,2] in contrast to the idea generation and original HMS. In the next step, we calculated the zero centered normalized vectors of $Vco\_oc_{wj}$, $C_j$ , and $O_j$. For a vector $V_j$, the normalized vector $norm\_V_j$ was given by:

$$norm\_V_j = \frac{V_j - mean(V)}{SD(V)}$$

The aim of this step was to ensure that the times where participants generate a convergent/original idea and HMS reflect this idea generation accurately are treated as times when participants formulate divergent/non-original ideas and there is no HMS reflected. This normalization was conducted to ensure equally balanced contribution from existence and non-existence or high and low values in the idea-HMS co-occurrence and the attributions of the ideas in calculating their temporal relationship measures (defined as below), irrespective of the possible imbalance in the idea-HMS co-occurrence, convergence, and originality within each trial. Furthermore, to observe whether the co-occurrence of HMS depend on the quality of the generated ideas and to test if such a dependence is out of chance level, we first permuted the orders of $norm\_C_j$ and $norm\_O_j$ vectors resampling it 1000 times to observe whether the order mattered, and called these permutation-based normalized vectors as $Pnorm\_C_j$ and $Pnorm\_O_j$, respectively. In the next step we calculated the $C\_co\_oc_w$ as a dot product of idea-HMS co-occurrence series and idea-convergence series, and $O\_co\_oc_w$ as a dot product of idea-HMS co-occurrence series and idea-originality series:

$$C\_co\_oc_w = (norm\_Vco\_oc_{wj},\ norm\_C_j)$$

$$O\_co\_oc_w = (norm\_Vco\_oc_{wj},\ norm\_O_j)$$

These variables represent how much HMS occurrences depend on the convergence and originality of the ideas, respectively. **Figure 5** depicts these steps. We repeated the same calculation of the dot product between the $norm\_Vco\_oc_{wj}$ and all the 1000 permuted normalized vectors $Pnorm\_C_j$ and $Pnorm\_O_j$, denoting them as $PC\_co\_oc_w$ and $PO\_co\_oc_w$, respectively. In the following step we assessed permutation-based $z$-scores using the following equations:

$$Convergent - z - score = \frac{C\_co\_oc_w - mean(PC\_co\_oc_w)}{SD(PC\_co\_oc_w)}$$

$$Original - z - score = \frac{O\_co\_oc_w - mean(PO\_co\_oc_w)}{SD(PO\_co\_oc_w)}$$

These $z$-scores express in each trial how much occurrences of the HMS as response to ideas have been affected by the convergence and originality of the ideas, respectively.

We also wanted to observe if the HMS response to idea generation would affect the nature of the successive ideas generated. Therefore, in a parallel session, in the two last steps, we shifted the idea-HMS co-occurrence vector $norm\_Vco\_oc_{wj}$ by one idea behind and evaluated its relationship with the convergence and originality of the successive ideas. To do so, first we re-calculated $C\_co\_oc_{wj}$ and $C\_co\_oc_{wj}$ using the shifted $norm\_Vco\_oc_{wt}$ vector. Secondly, calculating and using the $PSHC\_co\_oc_w$ and $PSHO\_co\_oc_w$ for the respective permutation-based convergent and originality vectors, we derived the $z$-scores with the same formula as depicted above but using the results of the shifted vector and named the additional measures as $SH - Convergent - z - score$ and $SH - Original - z - score$. The formulas are depicted below:

$$SH\_C\_co\_oc_w = (SH - norm\_Vco\_oc_{wj}, norm\_C_j)$$

$$SH\_O\_co\_oc_w = (SH - norm\_Vco\_oc_{wj},\ norm\_O_j)$$

$$SH - Convergent - z{-}score = \frac{SH\_C\_co\_oc_w - mean(PSHC\_co\_oc_w)}{SD(PC\_co\_oc_w)}$$

$$SH\_Original - z - score = \frac{SH\_O\_co\_oc_w - mean(PSHO\_co\_oc_w)}{SD(PSHO\_co\_oc_w)}$$

It should be noted that the trials that contained no HMS within the identified optimal time window from any of idea generation (three trials), and the trials with no convergent ideas (17 trials), were excluded from the analysis.

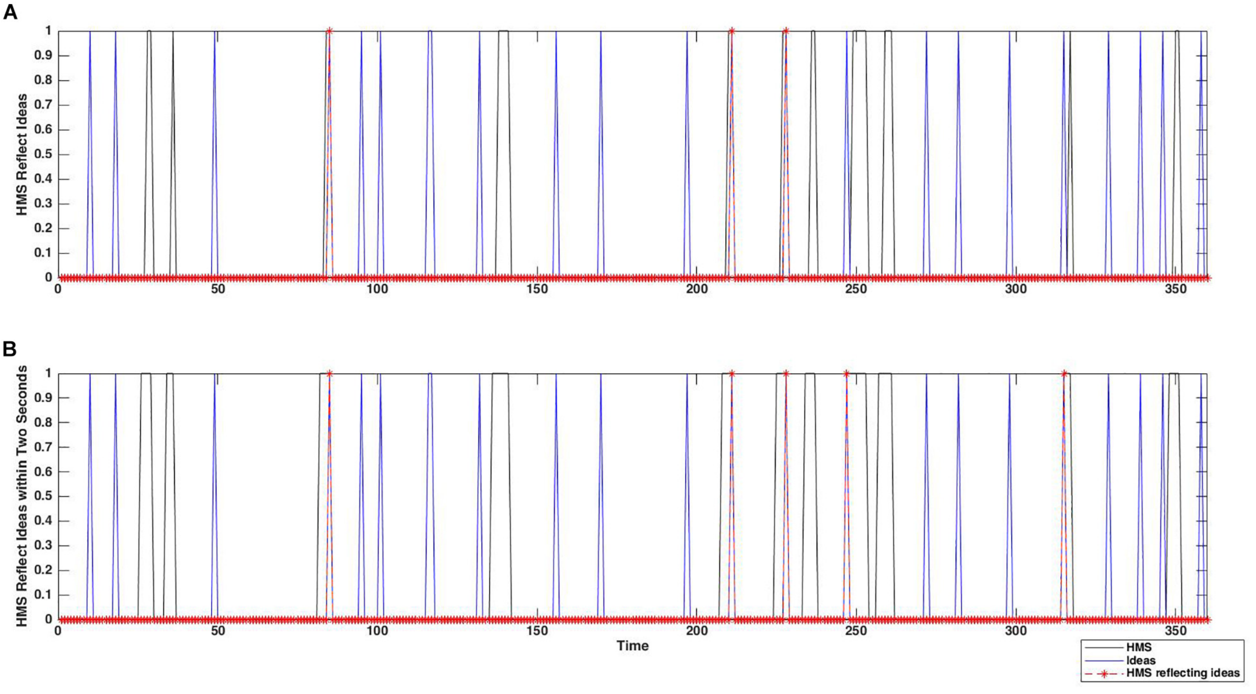



**FIGURE 4 |** An example of co-occurrence relationship between idea and HMS sequences, with the original HMS series **(A)** and the window-adopted HMS series with time window [0,2] **(B)**.

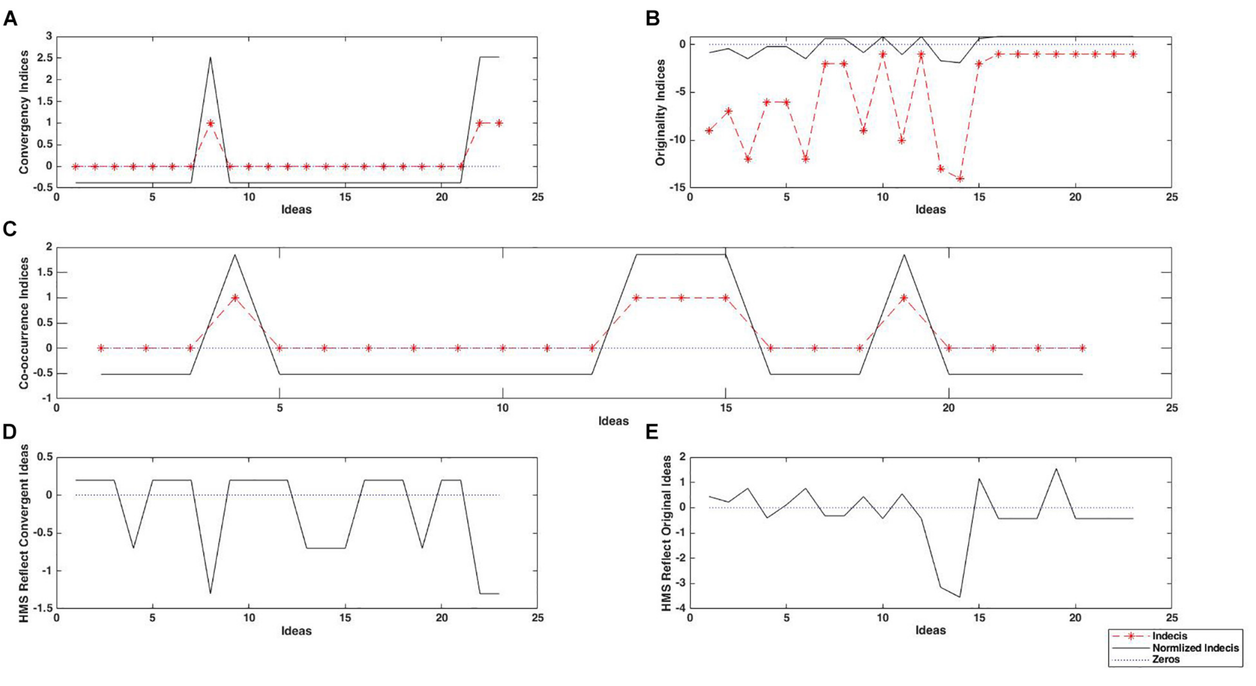



**FIGURE 5 |** Examples of convergence attribution vector **(A)**, originality attribution vector **(B)**, and the idea-HMS co-occurrence vector **(C)** from a single trial. In each panel, black solid lines represent raw scores, and red dashed lines represent normalized scores. By multiplying the normalized idea-HMS co-occurrence vector with the attribution vectors, relationship between the occurrence of HMS and ideas' convergence **(D)**, originality **(E)** are evaluated, respectively. By summing up these values over ideas, we obtained summary measures $C\_co\_oc_w$ and $O\_co\_oc_w$, which represent how much HMS occurrences depend on the convergence and originality of the ideas, respectively.

# RESULTS

## Effects of Music Type on Creativity Outcomes

To observe the possible effects of the music conditions on creativity performance, we used R statistical computing software. A one-way repeated measures ANOVA performed on the five music conditions showed significant differences across conditions on fluency ($F(4,76) = 3.566$, $p = 0.010$), originality ($F(4,76) = 2.93$, $p = 0.025$), and IOC ($F(4,76) = 4.95$, $p = 0.001$). Two-way repeated measures ANOVA on fluency, using genre and valence as factors, while excluding the data of the no-music condition indicated significant main effects on valence ($F(1,19) = 7.81$, $p = 0.012$). Yet no significant effect was observed on the interaction term (i.e., valence and genre) and the main effect of genre with respect to fluency. On the other hand, a significant main effect in genre ($F(1,19) = 10.324$, $p = 0.005$) with respect to originality was observed. In addition, there was also a marginally significant interaction between the two factors ($F(1,19) = 3.30$, $p = 0.085$) but no main effect of valence with respect to originality. Contrastingly, a two-way repeated measures ANOVA (using the same aforementioned terms) on the IOC score revealed a significant main effect of genre ($F(1,19) = 10.49$, $p = 0.004$) and valence ($F(1,19) = 4.82$, $p = 0.041$), with upbeat genre tracks associated with higher IOC scores. No significant interaction (i.e., valence and genre) was observed on IOC. Observing these results, separate pairwise $t$-tests were performed between positive-upbeat music and no music control condition, and significant enhancing effects were observed on fluency ($t(14) = 6.39$, $p = 0.02$), originality ($t(19) = 10.03$, $p = 0.005$), and IOC ($t(19) = 8.40$, $p = 0.009$). The creativity performance indices are illustrated in **Figure 6**.

## Effect of Music Type on HMS

To test the effect of music of non-verbal communication we evaluated HMS during each trial of each dyad. To examine the effect of music on HMS, a one-way repeated measures ANOVA was conducted on five conditions. A significant difference was observed across the conditions ($F(4,76) = 2.58$, $p = 0.043$; **Figure 7**). In addition, a two-way repeated measures ANOVA (factors – genre and valence) was conducted and a significant effect was observed on valence ($F(1,19) = 4.60$, $p = 0.045$), but not on genre or the interaction term with respect to music type.

## Temporal Coordination of Idea Generation and HMS

In order to test our hypothesis that HMS takes place more frequently in response to idea generation, we identified occurrence of HMS within the time windows of $[0, w]$ ($w = 1$ to 10) seconds from idea generation, and evaluated how such co-occurrence (allowing time delay $w$ for response) is above chance level, using the permutation-based surrogate data analysis (see section "Materials and Methods" for detail). For each time window, one-sample $t$-test was conducted on the 20 surrogate-based $z$-scores for co-occurrence (obtained one for each dyad, by averaging the $z$-scores over the music conditions). The results revealed the highest level of temporal coordination between idea generation and HMS with the time window setting of [0,2] seconds ($t(1,19) = 7.00$, $p < 0.0001$; **Figure 8**), indicating that this time window is optimal in capturing HMS as response to idea generation. For the remainder of our results we used this time window setting. Although we expected that HMS would more likely follow idea generation as response, following the reviewer's suggestion, we also explored the possibility of HMS-preceding temporal coordination with idea generation. Using "negative" time windows HMS within the time windows of $[-w, 0]$ ($w = 1$–10) with reference to idea generation as zero, the same analysis showed much lower degree of co-occurrence between HMS and idea generation (**Supplementary Figure S1**). This result further supports our hypothesis that significant proportion of HMS took place as responses to idea generation.

Furthermore, to check whether the temporal coordination between idea generation and HMS show any difference between the BGM conditions. Excluding data of sessions with no co-occurrence between HMS and idea generation, a one-way repeated measures ANOVA was conducted on the five music conditions and showed no significant difference over conditions ($F(4,73) = 0.92$, $p = 0.45$; **Figure 9**). Subsequently, a two-way repeated measures ANOVA (genre and valence) was conducted and a significant main effect was observed on genre ($F(1,18) = 5.73$, $p = 0.028$), but not on valence and the interaction term.

## Nature of Previously Generated Ideas and HMS Responses

To examine whether the HMS depended on the quality of previously generated ideas, we averaged the $Convergent - z - score$ and $Original - z - score$ over the five BGM (including the no-music condition). A one-sample $t$-test on the 20-mean permutation-based $z$-scores illustrated no general tendency of HMS as a function of either the convergent ($t(1,19) = -0.10$, $p = 0.92$) or original ($t(1,19) = -0.22$, $p = 0.82$) nature of previously generated ideas. To examine any potential difference with respect to the conditions, a one-way repeated measures ANOVA on the five BGM conditions was conducted. The results showed no significant difference over BGM conditions in respect to convergence ($F(4,59) = 0.78$, $p = 0.545$). Yet, a significant difference was observed on originality ($F(4,68) = 2.74$, $p = 0.035$; **Supplementary Figure S2**).

## HMS and the Nature of Successive Generated Ideas

Averaging the $SH - Convergent - z - score$ and $SH - Original - z - score$ gained from shifted $norm\_co\_oc_{wt}$ over five BGM conditions, we could assess the impact of

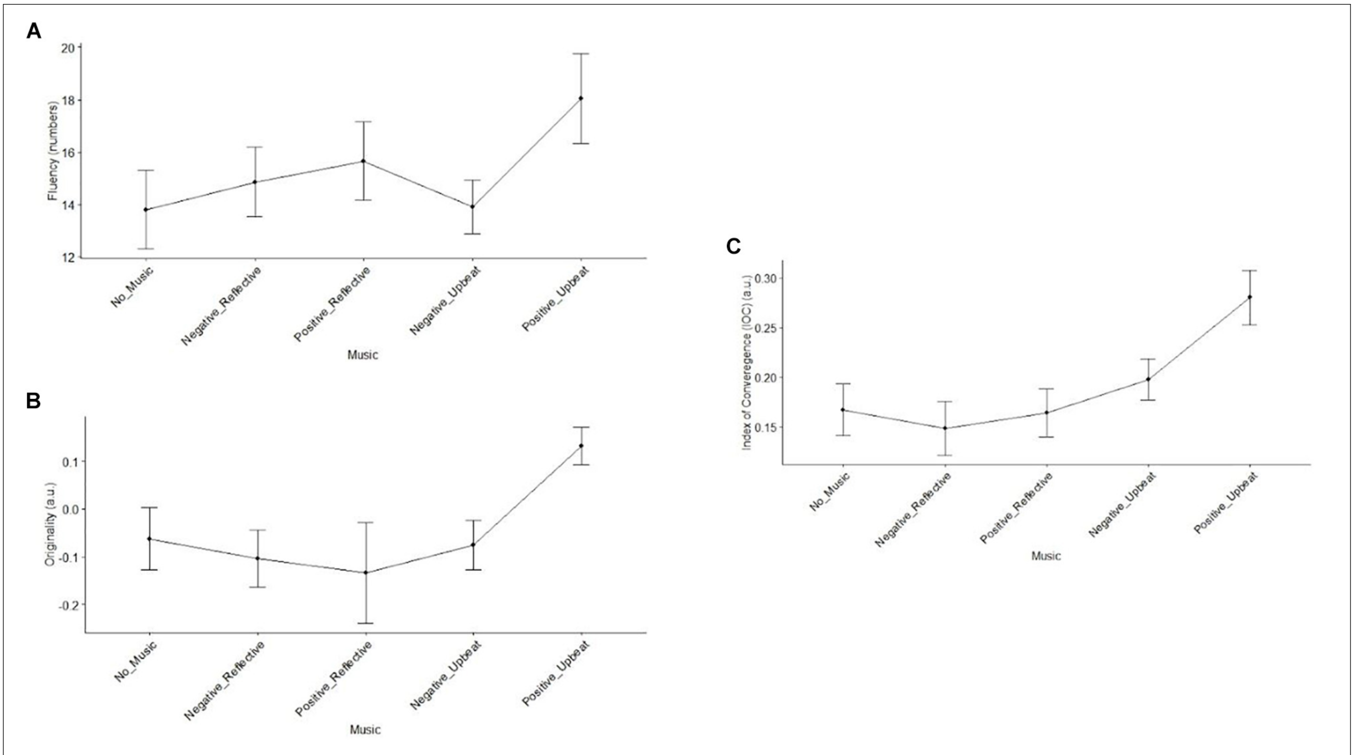



**FIGURE 6 |** Effects of music types on the group creativity outcomes: fluecny **(A)**, originality **(B)**, and IOC **(C)**. Error bars represent standard error of the mean.

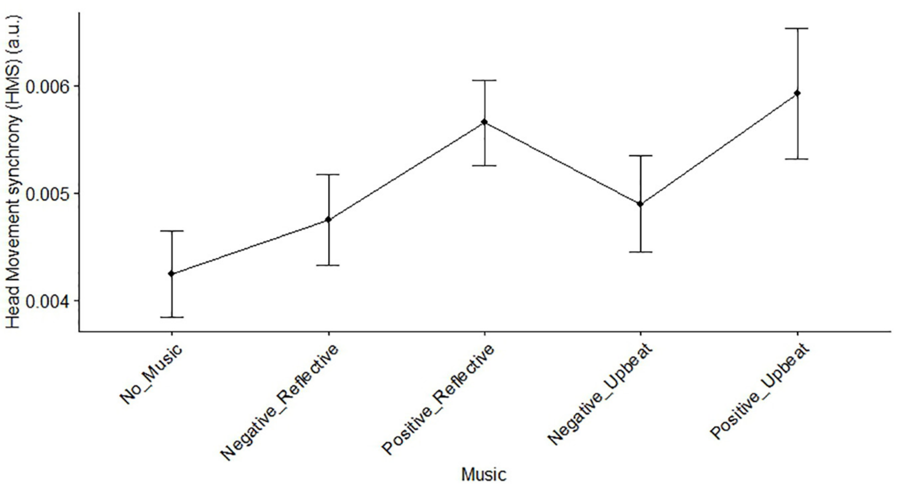



**FIGURE 7 |** Effect of music types on HMS. Error bars represent standard error of the mean.

HMS response on the nature of successive generated ideas. A one-sample $t$-test on the 20-mean shifted permutation-based $z$-scores showed the general tendency of HMS associated with the successive divergent idea generation ($t(1,19) = -2.44$, $p = 0.024$) along with a quasi-significant general tendency of HMS associated with successive original idea generation ($t(1,19) = 2.02$, $p = 0.058$). A one-way repeated measures ANOVA was conducted on music condition and showed significant differences on the convergent ($F(4,59) = 6.51$, $p = 0.00021$; **Figure 10**) but not the originality attribute ($F(4,68) = 1.15$, $p = 0.342$). In addition, a two-way repeated measures ANOVA (i.e., genre and valence) was conducted on the convergent attribute. The two-way repeated measurement ANOVA revealed significant main effects on genre ($F(1,12) = 10.05$, $p = 0.008$) and a significant interaction between the valence and genre ($F(1,12) = 10.63$, $p = 0.07$). However, no main effect of valence on the convergent attribute was observed ($F(1,12) = 0.41$, $p = 0.536$).

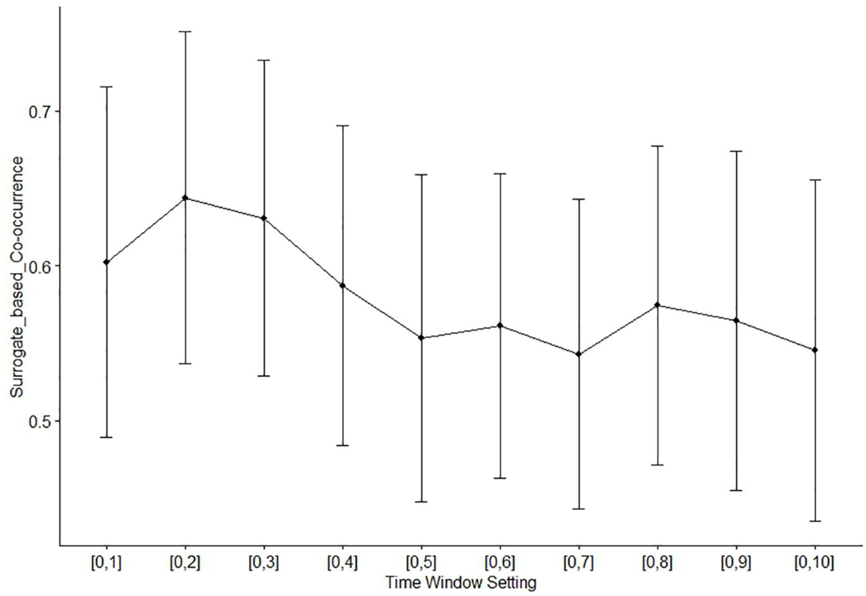



**FIGURE 8 |** Surrogate-based *z*-scores of the idea-HMS co-occurrence, with 10 time window settings [0, *w*] (*w* = 1–10). Error bars represent standard error of the mean.

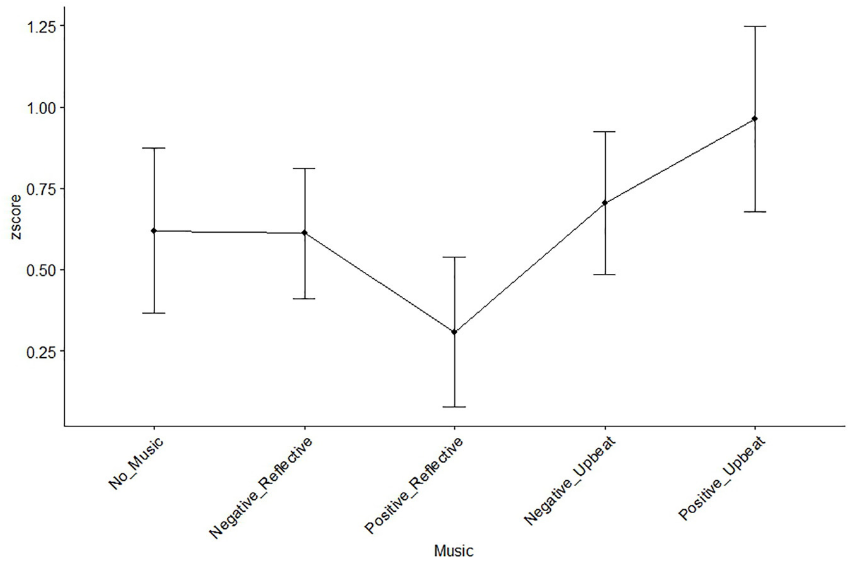



**FIGURE 9 |** Effect of music on temporal coordination between HMS and idea generation. Error bars represent standard error of the mean.

# DISCUSSION

## BGM Effect on Group Creativity

To confirm the previous results regarding effects of BGM on group creativity indices (Hosseini et al., 2019) with a larger sample, we investigated whether listening to a specific type of music as compared to no music control condition might enhance group creativity. The first part of our results indicated a significant main effect of BGM on group creativity in terms of indices of idea fluency, originality and IOC. The main results indicated that the music condition enhanced the fluency of group creativity outcome relative to the control (no music) condition and upbeat genre enhanced the group creativity in terms of originality and IOC. The supplementary results indicate the effect of positive valence in respect to increasing the total frequency of ideas mentioned, while upbeat genre improved the originality of the ideas generated by participant dyads in addition to the IOC as the tendency of the groups to integrate ideas. The effect of positive valence music on fluency could arguably be explained by the constructive effect of positive mood in enhancing ideation

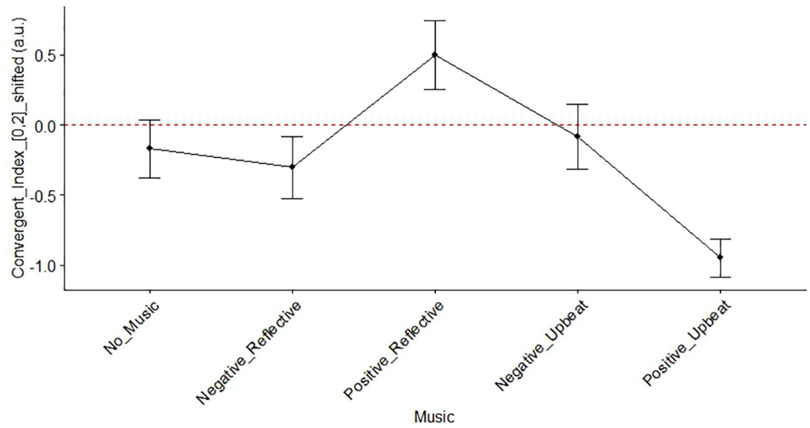



**FIGURE 10 |** The degree to which a HMS response to idea generation affected the convergence/divergence attribute of the successive ideas generated, for the different music conditions. The horizontal axis shows the score of *SH − Convergent − z − score* in the main text. Error bars represent standard error of the mean. *SH − Convergent − z − score*.

component of creativity, such that participants could generate more ideas along with the mood alteration (Harvey, 2014). In terms of originality, no main effect of valence was observed, though contrastingly, a significant main effect was observed for genre. The findings on IOC provided evidence for the benefit of listening to upbeat genres as BGM on idea integration. In conclusion, upbeat positive music have facilitated group creativity outcome, with upbeat genre helping the originality of ideas and positive mood leading to higher scores in the dimension of fluency in ideation.

By comparing a non-verbal communication measure (HMS) of group communication process between the conditions, this study indicated that exposure to BGM enhances synchrony of head motions, which is in accordance with an observation made in a real-world situation (Ellamil et al., 2016).

The overall results of both creativity outcome as well as HMS, suggest that positive valence has an effect on cooperation and integrated communication within dyads, whilst upbeat genre might facilitate the idea integration, leading into the original idea built up. Furthermore, the significant effect of positive valence tracks on HMS enhancement leads us to discuss over the enhancing effect of positive mood on connection between dyads. With such kind of a consideration, we further investigated the relationship between HMS and idea generation during group creativity dynamics, and how the relationship was affected by the types of BGM.

## Generated Ideas and HMS Responses

Examining the HMS response in relation to previously generated ideas, the contribution of non-verbal social behavior during conversation, and the synthesis of interpersonal movement synchrony responses in conversation, has produced great insights into conversation in previous studies. To be specific, movement synchrony is recognized as a contributing factor to rapport (Lakens and Stel, 2011), which can lead to cooperation, during social exchanges. Our results on the temporal coordination provide support for the presence of a relationship between idea generation and HMS response, which was more prominent with the upbeat genre. This finding could be possibly explained by the desire to connect during social interaction (Horowitz et al., 2006), where synchronized head movements may be an embodied expression of agreement with the generated ideas. Also, the prominent association of genre gained in the results might suggest upbeat genres encouraged dyads to participate in the task and behave task related at a higher level. On the other hand, when all BGM conditions were combined, we did not find significant general relationship between HMS and the quality of the previously generated ideas. Consequently, it might be valid to suggest that during creativity related experiments, participants would likely show some levels of understanding and agreement on the task (Shockley et al., 2003) and engagement in the task, regardless of the quality of the generated ideas.

## HMS and Successive Idea Generation

In this study, as a part of testing the second hypothesis, we investigated how the HMS responses would induce participants to share divergent perspectives and original ideas in the successive idea generation. The findings indicate the significant impact of HMS in encouraging dyads to generate more divergent and original ideas. In other words, the chance of such idea generation would increase after dyads expressed some level of agreement on the previously generated ideas. In general, divergent and original ideas are valuable resources during creative group tasks, as has been indicated by the importance of sharing divergent ideas that has been reported in numerous previous articles. Previous research on creativity suggest that group creativity is stimulated by diverse membership (Runco, 1993). Also, previous research on the individual differences in understanding indicate the role of how integrating the diversity of group members enhances decision making (Maznevski, 1994). In sum, it is important to identify what kind of element in the creative communication can promote sharing of more divergent and original ideas. Hence, the relationships between HMS and the nature of generated ideas, specifically the divergence of ideas, are arguably the most important findings produced in this study.

In the other part of our second hypothesis, we were also interested in the potential effect of music type in enhancing this effect further. As for this point, the differential contribution of successive divergent idea generation during upbeat genre BGM (relative to reflective music and a control condition) suggests the role of this genre in facilitating divergent idea sharing. To support this finding, we can refer to the distinct characteristics of the upbeat genre on the embodied extraversion of dyads during group creativity which have facilitated divergent idea sharing. This complementary result provide an interoperation that upbeat positive valence encouraged to share more divergent idea following agreement on the previous idea between group participants. Previously we mentioned the effect of music with a positive valence on enhancing individual creativity and cooperation (Hosseini et al., 2019). The current findings provides mechanistic description of how the positive upbeat BGM enhanced interpersonally coordinated head motions, possibly reflecting increased expression of agreement, and in turn led to generation and sharing of more divergent ideas.

## General Discussion

Altering mood is one of the most studied effects of music in the literature (Gerrards-Hesse et al., 1994; Balch and Lewis, 1996), while the literature on the group creativity suggests there are numerous disadvantages (e.g., social anxiousness and social loafing) to group activities that will likely impact group creativity. In addition to these widespread production blocks, behaviors such as judging and evaluative behavior have been noted as limitations that can impede the creativity of group sessions (Sternberg and Lubart, 1991). Previous studies have demonstrated the effects of nominal groups (Paulus and Nijstad, 2003) or virtual groups (Elias et al., 2011) to address the issue. However, this paper utilized BGM in an attempt improve creativity during group sessions. Here we suggest that music would decrease the stress level during the group creativity task by influencing participants moods (Fried and Berkowitz, 1979; North and Hargreaves, 1999; Ratcliffe et al., 2013). Contrastingly, positive mood results in the increased generation of ideas and cooperation, while the upbeat genre increased motivations to share ideas and incorporate diverse input into the creative group session. As such, this resulted in higher levels of total creative performance in the fluency and originality of the ideas produced.

## Limitations

The primary limitation of our study is that our participant sample was based on a variety of cultures and nationalities. This would induce different body movements in synchronicity (Ekman, 1977) so much that such synchronized body movement might not be always observed during idea sharing. Furthermore, the potential effect of language as a barrier while sharing ideas could be considered a limitation to this study. Furthermore, this study did not control for gender (Pearsall et al., 2008) nor familiarity of groups members and future studies should ensure they account for these two effects as it may bring further insight into how BGM influences group creativity.

# CONCLUSION

As observed in the findings of this BGM has the capacity to improve group creativity performance – with findings that indicated higher amounts of shared ideas and increased originality relative to no music in the positive-upbeat condition. Upbeat positive music embodied this enhancing behavior the most. We observed that positive valence can enhance HMS and purposed the effect of a positive mood to facilitate cooperation level. We also stated that an upbeat genre might have enhanced the level of idea integration that was encouraged by engendering divergent perspective sharing. Our first claim was due to a higher level of IOC observed during upbeat music and the fact the HMS during idea generation affected the quality of successive ideas produced. For this purpose, we first addressed the temporal coordination of HMS and idea generation. Significant levels of synchronized head movements contributed to the ideas generated during the group creativity task. The upbeat genre music appeared to embody this effect the most, given the results.

In terms of HMS being related to specific kinds of ideas, a general tendency toward agreement on ideas leading to subsequent further sharing of divergent and original ideas was observed. This tendency of sharing divergent ideas was most significant during upbeat music with a positive valence. In the first part of this study, we argued for the possible enhancing effect this music condition could have on group creativity in terms of both fluency and originality. The HMS corresponding responses arguably represented a form of embodied cognition in relation to the sharing divergent resources after agreement on ideas was established during the most upbeat and positive valence tracks. This may illuminate the effect of this BGM type in respect to its impacts of divergent idea generation and integration to induce extraordinary group creativity.

# DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

# ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Subjects Research Ethics Review Committee of the Tokyo Institute of Technology. The patients/participants provided their written informed consent to participate in this study.

# AUTHOR CONTRIBUTIONS

SH designed and conducted the experiment, analyzed the data, and wrote the manuscript. XD helped with conducting the experiment. YM provided conceptual advice on the experiment. TN designed and supervised the study and experimental design, as well as provided advice on the analytical methods, results, and overall manuscript. All authors discussed the results and commented on the manuscript.

## FUNDING

This study was partially supported by the KAKENHI (Grant Numbers JP15H01771 and JP17H01753) from JSPS/MEXT, Japan and the Center of Innovation Program from Japan Science and Technology Agency (JST), Japan.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02577/full#supplementary-material

## REFERENCES

Balch, W. R., and Lewis, B. S. (1996). Music-dependent memory: the roles of tempo change and mood mediation. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1354–1363. doi: 10.1037//0278-7393.22.6.1354

Brown, S. (2000). Evolutionary models of music: from sexual selection to group selection. *Perspect. Ethol.* 13, 231–281. doi: 10.1007/978-1-4615-1221-9_9

Cattell, R. B., and Saunders, D. R. (1954). Musical preferences and personality diagnosis: i. a factorization of one hundred and twenty themes. *J. Soc. Psychol.* 39, 3–24. doi: 10.1080/00224545.1954.9919099

Chen, M.-H., Chang, Y.-C., and Hung, S.-C. (2008). Social capital and creativity in R&D project teams. *R&D Manag.* 38, 21–34. doi: 10.1111/j.1467-9310.2007.00494.x

Cross, I., and Morley, I. (2009). "The evolution of music: theories, definitions and the nature of the evidence," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch, and C. Trevarthen. Oxford: Oxford University Press, 61–81.

Dippo, C. and Kudrowitz (2013). "Evaluating the alternative uses test of creativity," in *Proceedings of the National Converence on Undergraduate Research (NCUR)*, (Washington, DC: NCUR).

Eisenbeiss, S. A., vanKnippenberg, D., and Boerner, S. (2008). Transformational leadership and team innovation: integrating team climate principles. *J. Appl. Psychol.* 93, 1438– 1446. doi: 10.1037/a0012716

Ekman, P. (1977). "Biological and cultural contributions to body and facial movement,". In *The Anthropology of the Body*. Cambridge, MA: Academic Press Inc, 34–84.

Elias, E., Chamakiotis, P., Howard, T. J., Dekoninck, E., and Culley, S. J. (2011). "Can a virtual design environment enhance group creativity and the use of stimuli?," in *Proceedings of the 3rd International Conference on Research into Design Engineering,ICORD 11*, Bangalore.

Ellamil, M., Berson, J., Wong, J., Buckley, L., and Margulies, D. S. (2016). One in the dance: musical correlates of group synchrony in a real-world club environment. *PLoS One* 11:e0164783. doi: 10.1371/journal.pone.0164783

Fraisse, P. (1982). Rhythm and tempo. *Psychol. Music* 1, 149–180. doi: 10.1016/b978-0-12-213562-0.50010-3

Fried, R., and Berkowitz, L. (1979). Music hath charms. and can influence helpfulness 1. *J. Appl. Soc. Psychol.* 9, 199–208. doi: 10.1111/j.1559-1816.1979.tb02706.x

Gerrards-Hesse, A., Spies, K., and Hesse, F. W. (1994). Experimental inductions of emotional states and their effectiveness: a review. *Br. J. Psychol.* 85, 55–78. doi: 10.1016/j.encep.2006.08.003

Guilford, J. P. (1967). Creativity: yesterday, today and tomorrow. *J. Creat. Behav.* 1, 3–14.

Hare, A. P., Bales, R. F., and Borgatta, E. (1955). *Small Groups: Studies in Social Interaction*. New York, NY: Knopf.

Harvey, S. (2014). Creative synthesis: exploring the process of extraordinary group creativity. *Acad. Manag. Rev.* 39, 324–343. doi: 10.5465/amr.2012.0224

Horowitz, L. M., Wilson, K. R., Turan, B., Zolotsev, P., Constantino, M. J., Henderson, L., et al. (2006). How interpersonal motives clarify the meaning of interpersonal behavior: a revised circumplex model. *Personal. Soc. Psychol. Rev.* 10, 67–86. doi: 10.1207/s15327957pspr1001_4

Hosseini, S., Hattori, Y., Miyake, Y., and Nozawa, T. (2019). "Music valence and genre influence group creativity," in *Proceedings of the HCI International 2019 Conference*, Copenhagen.

Ilie, G., and Thompson, W. F. (2011). Experiential and cognitive changes following seven minutes exposure to music and speech. *Music Percep. Interdiscip. J.* 28, 247–264. doi: 10.1525/mp.2011.28.3.247

Kellaris, J. J., and Kent, R. J. (1991). Exploring tempo and modality effects, on consumer responses to music. *Adv. Con. Res.* 181, 243–248.

Knapp, M. L., Hall, J. A., and Horgan, T. G. (2013). *Nonverbal Communication in Human Interaction*. Massachusetts, MA: Cengage Learning.

Lakens, D., and Stel, M. (2011). If they move in sync, they must feel in sync: movement synchrony leads to attributions of rapport and entitativity. *Soc. Cogn.* 29, 1–14. doi: 10.1521/soco.2011.29.1.1

Larey, T. S., and Paulus, P. B. (1999). Group preference and convergent tendencies in small groups: a content analysis of group brainstorming performance. *Creat. Res. J.* 12, 175–184. doi: 10.1207/s15326934crj1203_2

Latif, N., Barbosa, A. V., Vatiokiotis-Bateson, E., Castelhano, M. S., and Munhall, K. G. (2014). Movement coordination during conversation. *PLoS One* 9:e105036. doi: 10.1371/journal.pone.0105036

Marsh, K. L., Richardson, M. J., and Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Top. Cogn. Sci.* 1, 320–339. doi: 10.1111/j.1756-8765.2009.01022.x

Maznevski, M. L. (1994). Understanding our differences: performance in decision-making groups with diverse members. *Hum. Relations* 47, 531–552. doi: 10.1177/001872679404700504

McCauley, C. (1998). Group dynamics in janis's theory of groupthink: backward and forward. *Organ. Behav. Hum. Decis. Process.* 73, 142–162. doi: 10.1006/obhd.1998.2759

Mehrabian, A. (2017). *Nonverbal Communication*. Abingdon: Routledge.

North, A. C., and Hargreaves, D. J. (1999). Can music move people? the effects of musical complexity and silence on waiting time. *Env. Behav.* 31, 136–149. doi: 10.1177/00139169921972038

Paulus, P. B., and Nijstad, B. A. (2003). *Group Creativity: Innovation Through Collaboration*. England: Oxford University Press.

Paxton, A., and Dale, R. (2017). Interpersonal movement synchrony responds to high-and low-level conversational constraints. *Front. Psychol.* 8:1135. doi: 10.3389/fpsyg.2017.01135

Pearsall, M. J., Ellis, A. P., and Evans, J. M. (2008). Unlocking the effects of gender faultlines on team creativity: is activation the key? *J. Appl. Psychol.* 93, 225–234. doi: 10.1037/0021-9010.93.1.225

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., Brattico, E., et al. (2011). Music and emotions in the brain: familiarity matters. *PloS One* 6:e27241. doi: 10.1371/journal.pone.0027241

Ramseyer, F., and Tschacher, W. (2014). Nonverbal synchrony of head- and body-movement in psychotherapy: different signals have different associations with outcome. *Front. Psychol.* 5:979. doi: 10.3389/fpsyg.2014.00979

Ratcliffe, E., Gatersleben, B., and Sowden, P. T. (2013). Bird sounds and their contributions to perceived attention restoration and stress recovery. *J. Env. Psychol.* 36, 221–228. doi: 10.1016/j.jenvp.2013.08.004

Rauscher, F. H., Shaw, G. L., and Ky, K. N. (1995). Listening to Mozart enhances spatial-temporal reasoning: towards a neurophysiological basis. *Neurosci. Lett.* 185, 44–47. doi: 10.1016/0304-3940(94)11221-4

Rentfrow, P. J., and Gosling, S. D. (2003). The do re mi's of everyday life: the structure and personality correlates of music preferences. *J. Personal. Soc. Psychol.* 84, 1236–1256. doi: 10.1037/0022-3514.84.6.1236

Ritter, S. M., and Ferguson, S. (2017). Happy creativity: listening to happy music facilitates divergent thinking. *PloS One* 12:e0182210. doi: 10.1371/journal.pone.0182210

Runco, M. A. (1993). Operant theories of insight, originality, and creativity. *Am. Behav. Sci.* 37, 54–67. doi: 10.1177/0002764293037001006

Schabracq, M. J., and Cooper, C. L. (2000). The changing nature of work and stress. *J. Manag. Psychol.* 15, 227–241.

Schellenberg, E. G. (2015). Music training and speech perception: a gene-environment interaction. *Ann. N. Y. Acad. Sci.* 1337, 170–177. doi: 10.1111/nyas.12627

Shockley, K., Santana, M.-V., and Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 326–332. doi: 10.1037/0096-1523.29.2.326

Slavin, R. E. (1983). When does cooperative learning increase student achievement? *Psychol. Bull.* 94, 429–445. doi: 10.1037//0033-2909.94.3.429

Sternberg, R. J., and Lubart, T. I. (1991). An investment theory of creativity and its development. *Hum. Devel.* 34, 1–31. doi: 10.1159/000277029

Thepsoonthorn, C., Yokozuka, T., Miura, S., Ogawa, K., and Miyake, Y. (2016). Prior knowledge facilitates mutual gaze convergence and head nodding synchrony in face-to-face communication. *Sci. Rep.* 6:38261. doi: 10.1038/srep38261

Thompson, W. F., Schellenberg, E. G., and Husain, G. (2001). Arousal, mood, and the mozart effect. *Psychol. Sci.* 12, 248–251. doi: 10.1111/1467-9280.00345

Waterhouse, L. (2006). Multiple intelligences, the mozart effect, and emotional intelligence: a critical review. *Educ. Psychol.* 41, 207–225. doi: 10.1207/s15326985ep4104_1

Wiltermuth, S. S., and Heath, C. (2009). Synchrony and cooperation. *Psychol. Sci.* 20, 1–5. doi: 10.1111/j.1467-9280.2008.02253.x

Xue, H., Lu, K., and Hao, N. (2018). Cooperation makes two less-creative individuals turn into a highly-creative pair. *NeuroImage* 172, 527–537. doi: 10.1016/j.neuroimage.2018.02.007

Yokozuka, T., Ono, E., Inoue, Y., Ogawa, K.-I., and Miyake, Y. (2018). The relationship between head motion synchronization and empathy in unidirectional face-to-face communication. *Front. Psychol.* 9:1622. doi: 10.3389/fpsyg.2018.01622

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 11 February 2020
doi: 10.3389/fnhum.2020.00032

# Multimodal Recognition of Emotions in Music and Facial Expressions

*Alice Mado Proverbio*, Elisa Camporeale and Alessandra Brusa*

*Department of Psychology, University of Milano-Bicocca, Milan, Italy*

The aim of the study was to investigate the neural processing of congruent vs. incongruent affective audiovisual information (facial expressions and music) by means of ERPs (Event Related Potentials) recordings. Stimuli were 200 infant faces displaying Happiness, Relaxation, Sadness, Distress and 32 piano musical pieces conveying the same emotional states (as specifically assessed). Music and faces were presented simultaneously, and paired so that in half cases they were emotionally congruent or incongruent. Twenty subjects were told to pay attention and respond to infrequent targets (adult neutral faces) while their EEG was recorded from 128 channels. The face-related N170 (160–180 ms) component was the earliest response affected by the emotional content of faces (particularly by distress), while visual P300 (250–450 ms) and auditory N400 (350–550 ms) responses were specifically modulated by the emotional content of both facial expressions and musical pieces. Face/music emotional incongruence elicited a wide N400 negativity indicating the detection of a mismatch in the expressed emotion. A swLORETA inverse solution applied to N400 (difference wave Incong. – Cong.), showed the crucial role of Inferior and Superior Temporal Gyri in the multimodal representation of emotional information extracted from faces and music. Furthermore, the prefrontal cortex (superior and medial, BA 10) was also strongly active, possibly supporting working memory. The data hints at a common system for representing emotional information derived by social cognition and music processing, including uncus and cuneus.

**Keywords: N400, emotions, multimodal processing, music, facial expressions, cross-modal, neuroaesthetics, neuroscience of music**

OPEN ACCESS

***Edited by:***
*Elvira Brattico,*
*Aarhus University, Denmark*

***Reviewed by:***
*Johanna Maksimainen,*
*University of Jyväskylä, Finland*
*Diana Omigie,*
*Max-Planck-Gesellschaft (MPG), Germany*

****Correspondence:***
*Alice Mado Proverbio*
*mado.proverbio@unimib.it*

***Specialty section:***
*This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience*

***Received:*** *17 September 2019*
***Accepted:*** *23 January 2020*
***Published:*** *11 February 2020*

***Citation:***
*Proverbio AM, Camporeale E and Brusa A (2020) Multimodal Recognition of Emotions in Music and Facial Expressions. Front. Hum. Neurosci. 14:32. doi: 10.3389/fnhum.2020.00032*

## INTRODUCTION

Neuroaesthetics studies have outlined how music stimulation can clearly convey distinct emotional sensations to the listeners (e.g., joy, sadness, fear), particularly on the basis of piece musical structure, tonality, style and perceptual characteristics (e.g., Peretz, 1998, 2001; Vieillard et al., 2008; Brattico et al., 2011; Koelsch, 2011a, 2014; Bogert et al., 2016; Proverbio et al., 2016; Proverbio and De Benedetto, 2018). A meta-analysis of 41 musical studies has demonstrated that typical emotions such as happiness, sadness, anger, threat and tenderness can be easily decoded with above-chance accuracy by any listener (Juslin and Laukka, 2003). Nordströma and Laukka (2019) have recently shown that emotion recognition in music is a rather fast process. In their study above-chance accuracy was observed for musical stimuli lasting $\leq$100 ms for anger, happiness, neutral, and sadness recognition, and for $\leq$250 ms stimuli for the recognition of more complex emotions. Harmonic structure and timing of a musical fragment are promptly processed and interpreted by auditory regions of our brain. It is known that fast rate music tends to convey joyful sensations, whereas slow music tends to transmit sad feelings (Khalfa et al., 2005). Again, Vieillard et al. (2008) demonstrated that while happy excerpts, with a fast tempo and in a major mode were rated as arousing and pleasant, sad/melancholic excerpts, with a slow tempo and in minor mode, were judged as low in arousal. Threatening excerpts, expressed in a minor mode with intermediate tempo, were rated as arousing and unpleasant, while peaceful excerpts characterized by a slow tempo and major mode were perceived as little arousing and pleasant (Vieillard et al., 2008). Going even further, Vuilleumier and Trost (2015) were able to provide the neural underpinnings of complex music-induced emotions (such as, for example, wonder, transcendence, or nostalgia). They found that aesthetic emotions elicited by music shared core features with basic emotions that can be mapped onto independent dimensions of valence (positive or negative) and arousal (high or low) observed in other affective contexts. These findings highlighted the multidimensional nature of emotional reactions that is object of the present investigation.

In an interesting fMRI study (Trost et al., 2012) the neural mechanisms subserving the arising of positive (high arousal) and negative (low arousal) music evoked sensations were investigated indicating the role of the left striate, insula, motor and sensory areas in positive emotions, and of right striate, orbitofrontal (OBF) and ventromedial (vmPFC) prefrontal cortex and hippocampus in negative emotions. Particularly relevant seemed to be the nigro-striatal reward mechanism [including ventral striatum, ventral tegmental area (VTA), caudate and accumbens nuclei, as well as OBF and vmPFC] for mediating music-evoked pleasurable sensations (Blood and Zatorre, 2001; Menon and Levitin, 2005; Ishizu and Zeki, 2011; Salimpoor et al., 2011). On the other hand, amygdala and insula would be more involved in the processing of negative emotions (Gosselin, 2005; Gosselin et al., 2007).

Overall, the neuroaesthetics literature provides robust evidence that a certain variety of music-induced emotions can be reliably recognized by human listeners, regardless of education, exposure and familiarity to music style (e.g., Laukka et al., 2013; Sievers and Polansky, 2013; Egermann et al., 2015). This hints at a biological root for the neural mechanism devoted to the comprehension of music-evoked emotions (see Proverbio et al., 2019 for a discussion on this topic). In this vein, musical stimuli have been used in several multimodal studies where they were compared with emotions derived from the visual modality, and particularly with facial affect (e.g., Spreckelmeyer et al., 2006; Jeong et al., 2011; Hanser et al., 2015; Baranowski and Hecht, 2017). For example, Logeswaran and Bhattacharya (2009) examined how music-elicited emotions can influence subsequent vision-elicited emotional processing, by having neutral, happy and sad faces preceded (primed) by short excerpts of musical stimuli (happy and sad). The results showed that prior listening to a happy (or sad) music enhanced the perceived happiness (or sadness) of a face, regardless of specific facial expression. Similarly, Kamiyama et al. (2013) examined the integrative process between emotional facial expressions and musical excerpts by using an affective priming paradigm. Happy or sad musical stimuli were presented after happy or sad facial images. Participants had to judge the affective congruency of the presented face-music pairs, and incongruent musical targets elicited a larger N400 component than congruent pairs. In this study only two affective expressions (positive vs. negative) were used, so that the decision was dichotomous, thus having a 50/50 probability to be correct. In our paradigm, overt attention was diverted to the detection of neutral adult faces, in order to tap at inner mechanism of extraction and representation of complex emotions in music by using the N400 paradigm, and four different emotional hues.

N400 is a precious tool for investigating semantic and conceptual representations even in unaware subjects. It reflects the semantic analysis of visual (or auditory information) after about 400 ms from the presentation of the stimulus. The N400 peak could represent the process of accessing to semantic information in long-term memory, and/or the integration of this information with previous knowledge. Therefore, N400 amplitude would be greater in anomalous (incongruent) semantic contexts since integration requires more time and cognitive resources with respect to congruent scenarios (Lau et al., 2008). N400 has been effectively used to investigate the semantic processing of language (Kutas and Federmeier, 2011), pictures (Barrett and Rugg, 1990), pictures and words (Nigam et al., 1992), sounds and musical gestures (Proverbio et al., 2015a), skilled motor actions (e.g., basketball, Proverbio et al., 2012), sign language (Proverbio et al., 2015b), words and music (Daltrozzo and Schön, 2009; Goerlich et al., 2011), space and music (*e.g.*, low tones associated with *basement*, or ascending pitch steps associated with *staircase* (Koelsch et al., 2004; Zhou et al., 2014). Overall, the literature suggests the existence of an amodal and shared conceptual system, indexed by N400 response (to incongruity) in different domains, such as language, music, pictures, actions.

To address this issue, in this study EEG/ERPs were recorded in a multimodal audiovisual task featuring visual perception of facial expressions and listening to emotional music. We hypothesized that emotionally incongruent pairs of facial expressions and musical pieces would elicit N400 like responses to emotional mismatch, provided that music was able to clearly convey its emotional meaning in a "abstract" form, even if music was task irrelevant and unattended by viewers.

We also expected that perceptual N170 to faces was modulated by their emotional content, and particularly by distress (Batty and Taylor, 2003; Proverbio et al., 2006; Sun et al., 2017). In addition, we hypothesized that later P300 response was found greater in amplitude to the more arousing (e.g., distress) than less arousing (e.g., relaxation) emotion, since this ERP response has been proven to reflect the degree of emotion-induced arousal (Carretié and Iglesias, 1995; Polich, 2007). As for the auditory stimulation, the ERP literature predicted an effect of musical content on fronto/central N400 response (Koelsch, 2011b). We hypothesized that N400 was greater in response to negative than positive music, and that possibly showed a right hemispheric asymmetry for the processing of negative music, and vice versa, as predicted by neuroimaging studies (e.g., Schmidt and Trainor, 2001).

Visual N170 and P300, auditory N400 and multimodal N400 responses were therefore quantified and analyzed in all individual subjects.

## MATERIALS AND METHODS

### Stimuli Validation

Twenty non-musician University fellows participated in stimulus validation. They were 8 males and 12 females with a mean age of 29.2 years. Stimuli to be evaluated were visual (facial expressions) and auditory (music tracks), and were presented simultaneously, in different combinations. The visual stimuli consisted of faces of children showing emotional facial expressions of different types; the musical stimuli instead consisted of instrumental tracks transmitting different emotional sensations.

Facial expression were paired with musical excerpts so that they were emotionally congruent or incongruent (half of the times). The emotional value and effectiveness of the combinations was specifically assessed through validation.

#### Visual Stimuli (Faces)

The participants were presented with 200 faces of children (**Figure 1**) (apparent age < 15 months); the images were iso-luminant B/W photographs of infant faces. Each face exhibited a clear-cut, spontaneous, emotional expression caught by the camera, namely: Joy (*N* = 50), Relaxation (*N* = 50), Sadness (*N* = 50), Distress (*N* = 50). The pictures were obtained by a previous study (Proverbio et al., 2007), which validated the universality and comprehensibility of the expressions (see **Figure 1** for some examples). The pictures showed only the head of the babies which were all alike. All pictures had a size of 270 × 300 pixels.


**FIGURE 1 |** Examples of baby faces displaying the four types of emotions.

### Auditory Stimuli (Music)

Auditory stimulation consisted of 32 music tracks lasting 12 s and validated in the preliminary study by Vieillard et al. (2008). Musical pieces might belong to the following categories, because of their emotional content: Joy (*N* = 8), Relaxation (*N* = 8), Sadness (*N* = 8), Distress (*N* = 8). For any given 12 s musical excerpt 6 different faces were presented.

The music tracks were generated by a computer using the piano tone; they clearly differed for tonality (minor or major) and rhythm (fast or slow) to elicit a different level of arousal in the listener. The happy excerpts were written in a major mode at an average tempo of 137 at metronome (Beat per Minute, BPM), with the melodic line lying in the medium–high pitch range. The sad excerpts were written in a minor mode at an average slow tempo of 46 BPM. The peaceful excerpts were composed in a major mode, had an intermediate tempo (74 BPM). The distress excerpts were composed with minor chords on the third and sixth degree. Although most distress musical pieces were regular and consonant, a few had irregular rhythms and were dissonant. Further information can be found in Vieillard et al. (2008). In our study musical fragments had an average duration of 12.38 s, were normalized with Audacity to −1 Db, and leveled at −70 Db.

The four emotional categories were selected for several reasons: (1) they were clearly comprehensible, (2) were both positive and negative in polarity, (3) were both mild (relaxation and sadness) and strong (distress and joy) in intensity and (4) because a validated set of the same types of facial expressions and emotional music existed.

On the basis of the emotional dimension of the stimuli, 8 types of face-music pairings couplings were created, thus resulting in the emotional congruence or incongruence of the audiovisual stimulation (face/music): joy-distress, relaxation-sadness, sadness-relaxation, distress-joy for the incongruent conditions, and joy-joy, relaxation-relaxation, sadness-sadness, distress-distress for the congruent conditions. Distress baby expressions were paired to "distress" (frightening) musical excepts since stimulus and response are strongly associated (e.g., babies react to fear with tears). Stimulus validation was carried out to ascertain the degree of congruence/incongruity of pairs, and the effectiveness of the face-music couplings. The judges were seated in front of a PC wearing headphones: they were shown Power Point presentation lasting about 25 min, featuring 200 visual stimuli paired to 32 musical fragments, coupled according to their valence and congruence dimensions. To each stimulus pair, the subjects had evaluate their degree of emotional congruence by means of a 4-points Likert scale, where 1 = very incongruent, 2 = incongruent, 3 = congruent, 4 = very congruent.

A repeated measures one-way ANOVA was carried out on the mean scores attributed to the stimuli as a function of the experimental condition (two levels, congruent, incongruent). The ANOVA yielded the significance of condition factor [$F(1.19) = 262.87$, $p < 0.001$], with an average score of 2.03

(incongruent; *SD* = 0.05) for incongruent pairing, and 3.5 (halfway between congruent and very congruent; *SD* = 0.05) for congruent pairings. This preliminary finding confirmed the assumption that both music pieces and facial expressions conveyed a clearly understandable emotional meaning, and that their incongruent mixing was clearly detected by judges.

## EEG Study

### Participants

Twenty University students (9 males and 11 females) ranging in age from 20 to 26 years participated in the study. Participants were recruited through *Sona System* (a system for recruiting students who earn credit for their Psychology courses by participating in research studies), received academic credits for their participation and provided written informed consent. The data of four participants were excluded because of excessive EEG/EOG artifacts. The final sample comprised sixteen participants (eight males, eight females), aging on average 22.3 years (*SD* = 1.9). All participants had normal or corrected-to-normal vision. They were strictly right-handed as assessed by the Oldfield Inventory and reported no history of drug abuse or neurological or mental disorders. Experiments were conducted with the understanding and written consent of each participant according to the Declaration of Helsinki (BMJ 1991; 302: 1194), with approval from the Ethics Committee of University of Milano-Bicocca (protocol: RM-2019-176).

Data were protected according to EU Regulation 2016/679: article 2/2016 and article 9/2016, concerning the processing of sensitive data and protection of personal data. Participants were informed that their EEG/behavioral data would have been stored in anonymous and aggregate format (combined with those of other participants) for scientific purposes only and for a period not exceeding 5 years.

### Procedure

The participants were seated inside an anechoic and electrically shielded cubicle about 120 cm away from a PC monitor placed outside the cabin. The images representing children's emotional faces were presented on a VGA monitor, connected to a compatible IBM-PC computer, located outside the cabin. The onset of auditory stimuli was synchronized with that of visual stimuli at the beginning of each experimental sequence, through an external PC (MacBook Air, Apple) controlling audio clips administration according to the established order of presentation. Participants were instructed to gaze at the center of the screen where a small dot served as a fixation point to avoid any eye or body movement during the recording session. All stimuli were presented in random order at the center of the screen in 4 different, randomly mixed, short runs lasting approximately 1.5 min (plus a training initial run). Stimulus presentation was controlled by *EEvoke* stimulation software (*ANT Software*, Enschede, Netherlands). Each run consisted in the presentation of 50 infant pictures and 8 musical traces. Infant faces were shown for 1000 ms, with an ISI (inter-stimulus Interval) ranging from 600 to 800 ms. Each musical fragment lasted 12 s, and represented a significant musical phrase.

Each participant was provided with written experimental instructions. Before subjecting participants to the actual experimental task, a practical training of sequences was conducted, in which visual and auditory stimuli were used that were not reproduced during the experiment. The training run lasted about 45 s and included the randomly mixed presentation of 24 infants faces (displaying the four types of facial expressions), 4 musical stimuli (one for each of the 4 emotions), and 3 adult faces acting as rare targets.

To keep the subject focused on visual stimulation participants were instructed and trained to respond as accurately and quickly as possible by pressing a response key with the index finger of the left or right hand when they spotted the face of an adult individual. Rare targets were casually intermixed and had a 10% of probability.

### EEG Recording and Analysis

The EEG was recorded and analyzed using *EEProbe* recording software (*ANT Software*, Enschede, Netherlands). EEG data were continuously recorded from 128 scalp sites according to the 10–5 International System. Sampling rate was 512 Hz. Horizontal and vertical eye movements were additionally recorded, and linked ears served as the reference lead. Vertical eye movements were recorded using two electrodes placed below and above the right eye, while horizontal movements were recorded using electrodes placed at the outer canthi of the eyes, via a bipolar montage. The EEG and electro-oculogram (EOG) were filtered with a half-amplitude band pass of 0.016–100 Hz. Electrode impedance was maintained below 5 KOhm. EEG epochs were synchronized with the onset of video presentation and analyzed using *ANT-EEProbe software*. Computerized artifact rejection was performed prior to averaging to discard epochs in which amplifier blocking, eye movements, blinks or excessive muscle potentials occurred. The artifact rejection criterion was a peak-to-peak amplitude exceeding 50 $\mu$V and resulted in a rejection rate of ~5%. Event-related potentials (ERPs) from 100 ms before to 1000 ms after stimulus onset were averaged off-line. ERP averages were computed as a function of facial expression (regardless of auditory stimulation), for the, therefore so-called, visual N170 and visual P300 responses. Further ERP averages were computed as a function of music emotional content (regardless of visual stimulation) for the, therefore so-called, auditory N400 response. Finally, ERP averages were computed to audiovisual stimuli (regardless of stimuli specific content) as a function of congruence in emotional content (e.g., happy music vs. happy faces = congruent stimulation; distress music vs. happy faces = incongruent stimulation) for the multimodal N400 response.

ERP components were measured when (in time) and where (at which scalp sites on the basis of scalp topography) they reached their maximum amplitudes, and also according to the available literature (e.g., for acoustic potentials central sites were closely monitored). N170 was measured in between 160–180 ms at occipito/temporal sites (P9-P10 and PPO9h-PPO10h), according to previous literature (e.g., Proverbio et al., 2006; Sun et al., 2017). Visual P300 was measured in between

250 and 450 ms at anterior and fronto/central sites (FC1-FC2 and C1-C2). Auditory N400 was measured in between 350-550 ms at fronto-central sites (C1-C2, C3-C4, FCC3h-FCC4h). Multimodal N400 was quantified in between 300–500 ms (N400) at midline prefrontal and inferior frontal sites (Fpz, F5, F6). For each ERP component mean area values underwent distinct repeated-measures ANOVAs whose factors of variability were 3 within-groups factors: Emotion (Joy, Pain, Relaxation, Sadness), Electrode (2 or 3 levels depending on the ERP component of interest), Hemisphere (left, right). Tukey *post hoc* comparisons among means were performed. The effect size for the statistically significant factors was estimated using partial eta squared ($\eta_p^2$). The alpha inflation due to multiple comparisons was controlled by means of Greenhouse–Geisser epsilon correction.

Difference Waves (DWs) were also computed by subtracting the ERP waveforms related to congruent stimuli, from those elicited by incongruent stimuli, with the aim of investigating the extraction of the emotional significance of auditory and visual stimuli, regardless of the categories of belonging. The latency of the DW responses considered was between 300 and 500 ms.

### Behavioral Data

With regard to behavioral data, the percentages of correct responses and response times to targets were analyzed, but not subjected to ANOVA. Accuracy in detecting adult faces was 100% for all of the participants; the average response time was 538 ms (*SD* = 54). This finding suggests that participants were paying close attention to the stimulation provided.

### SwLORETA Source Reconstruction

Low-resolution electromagnetic tomography (LORETA) was applied to visual N170 and to the difference-waves incongruent-congruent in the N400 time window. LORETA (Pasqual-Marqui et al., 1994) is a discrete linear solution to the inverse EEG problem and corresponds to the 3D distribution of neuronal electrical activity that has a maximal similar orientation and strength (i.e., maximally synchronized) between neighboring neuronal populations represented by adjacent voxels. In this study we used an improved version (Palmero-Soler et al., 2007) of the standardized weighted LORETA. The data were automatically re-referenced to the average reference as part of the LORETA analysis. A realistic boundary element model (BEM) was derived from a T1-weighted 3D MRI data set through segmentation of the brain tissue (Zanow and Knösche, 2004). The source reconstruction solutions were then projected onto the 3D MRI of the Collins brain provided by the Montreal Neurological Institute. The synchronization and coherence tomography incorporates a standard dipole modeling. The probabilities of source activation based on Fisher's *F*-test were provided for each independent EEG source, the values of which are indicated in the 'unit' scale (the greater, the more significant). Both the segmentation and generation of the head model were performed using the ASA software program *Advanced Neuro Technology* (ANT, Enschede, Netherlands). SwLORETA source reconstruction was applied to the ERP waveforms related to the visual N170 (between 160 and 180 ms) as well as to the DWs related to the multimodal N400 (between 300 and 500 ms).

# RESULTS

## Visual N170 (160–180 ms)

The ANOVA performed on N170 mean amplitude values recorded to visual stimuli (independent of auditory stimulation) showed the significance of emotion [$F(3,45) = 4.47$, $p < 0.008$; $\varepsilon = 1$; $\eta_p^2 = 0.23$], with significantly greater N170 responses to distress (−1.68 μV, *SD* = 0.73) than other facial expressions (Joy = −1.0 μV, *SD* = 0.71; Relax = −1.01 μV, *SD* = 0.67; Sadness = −0.68 μV, *SD* = 0.69), as confirmed by *post hoc* comparisons ($p < 0.005$) and as visible in **Figures 2A,B**. SwLORETA applied to N170 to faces (independently of emotional content) identified as most active areas the right hemispheric occipito/temporal sites (see **Table 1**). SwLORETA solutions relative to N170 generators are visible in **Figure 2C**

## Visual P300 (250–450 ms)

The ANOVA performed on P300 mean amplitude values recorded to visual stimuli (independent of auditory stimulation) showed the significance of electrode factor [$F(1,15) = 15.57$, $p < 0.001$; $\varepsilon = 1$; $\eta_p^2 = 0.51$] with greater P300 amplitudes over central (C1-C2 = −0.68 μV, *SD* = 0.83) than fronto/central sites (FC1-FC2 = −1.22 μV, *SD* = 0.78), as proved by *post hoc* comparisons ($p < 0.001$). Further significant was the interaction of emotion × electrode × hemisphere [$F(3,45) = 2.83$, $p < 0.049$; = 0.85, adjusted *p*-value = 0.05; $\eta_p^2 = 0.19$]. *Post hoc* comparisons showed that P300, especially at right central sites was much greater ($p < 0.0001$) to distress expressions than any other expression (C1 = −0.16 μV, *SD* = 3.24; C2 = −0.08 μV, *SD* = 3.38; FC1 = −0.67 μV, *SD* = 3.11; FC2 = −0.52 μV, *SD* = 3.42), and smaller ($p < 0.0001$) to sad expression than any other expression (C1 = −1.62 μV, DS = 4.25; C2 = −1.4 μV, *SD* = 4.56; FC1 = −2.2 μV, DS = 3.93; FC2 = −1.8 μV, *SD* = 4.12). These P300 modulation as a function of facial expression *per se* can be appreciated by looking at waveforms of **Figure 3**.

## Auditory N400 (350–550 ms)

N400 response was measured on ERP waveforms averaged as a function of auditory emotional content (regardless of facial expressions). The ANOVA yielded the significance of hemisphere [$F(1,15) = 8.73$, $p < 0.009$; $\varepsilon = 1$; $\eta_p^2 = 0.36$] with greater N400 amplitudes over the left (−1.57 μV, *SD* = 0.48) than right hemisphere (−1.23 μV, *SD* = 0.49). Also significant the interaction between emotion and hemisphere [$F(3,45) = 4.51$, $p < 0.007$; $\varepsilon = 0.92$, adjusted *p*-value = 0.0095; $\eta_p^2 = 0.23$]. *Post hoc* comparisons showed that N400 was greater to negative [distress ($p < 0.001$) and sadness ($p < 0.001$)] than positive emotional musical fragments, especially over the right hemisphere (Joy: LH = −0.92 μV, *SD* = 0.56; RH = −0.78 μV, *SD* = 0.6; Relax: LH = −1.54 μV, *SD* = 0.5; RH = −0.89 μV, *SD* = 0.53; Distress: LH = −1.87 μV, *SD* = 0.49; RH = −1.56 μV, *SD* = 0.5; Sadness: LH = −1.95 μV, *SD* = 0.67; RH = −1.72 μV, *SD* = 0.65), as can be appreciated by looking at waveforms of **Figure 4**.

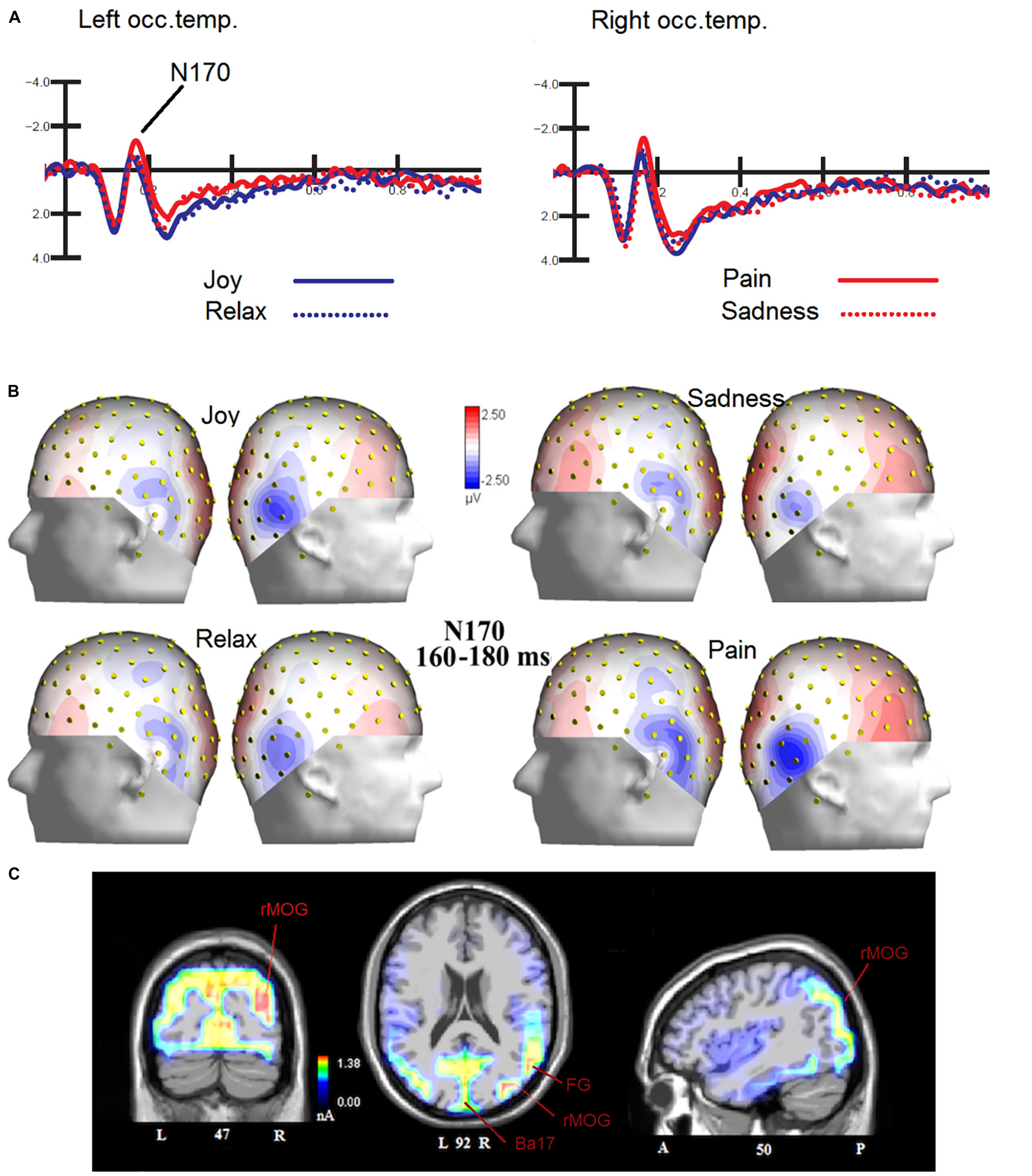



**FIGURE 2 | (A)** Grand-average ERP waveforms elicited by the four types of facial expressions (regardless of music content) as recorded at P9 and P10 sites. **(B)** Isocolor topographic maps of surface N170 voltage measured in between 160-180 ms over the left and right hemispheres, in response to the four types of facial expressions**. (C)** Coronal, Axial and sagittal views of swLORETA activations during face processing at N180 latency range (160–180 ms). The different colors represent differences in the magnitude of the electromagnetic signal (in nA). Numbers refer to the displayed brain slice: L, left hemisphere; R, right hemisphere. A, anterior; P, posterior.

Also significant the electrode factor [$F(2,30) = 18.31$, $p < 0.001$; $\varepsilon = 1$; $\eta_p^2 = 0.55$], showing greater N400 amplitudes over fronto/central ($p < 0.001$) than central sites, as proved by *post hoc* comparisons (C1-C2 = −1.54 μV, *SD* = 0.54; C3-C4 = −1.05 μV, *SD* = 0.44; FCC3h-FCC4h = −1.62 μV, *SD* = 0.47).

**TABLE 1 |** List of active electro-magnetic dipoles (along with their Talairach coordinates) explaining the surface voltage recorded between 160 and 180 ms post-stimulus (N170 latency range) to faces.

| Magn. | T-x [mm] | T-y [mm] | T-z [mm] | Hem. | Lobe | Gyrus | BA |
|---|---|---|---|---|---|---|---|
| 1.54 | 40.9 | −79.2 | 12.7 | R | O | Middle Occipital | 19 |
| 1.38 | 11.3 | −88.3 | 3.0 | R | O | Lingual | 17 |
| 1.31 | −28.5 | −79.2 | 12.7 | L | O | Middle Occipital | 19 |

*Magn., magnitude of the signal in nA, Hem., hemisphere. L, left; R, right. BA, Brodmann area. O, occipital.*

## Audiovisual N400 (300–500)

ERPs were also averaged as a function of congruence of emotional content of the audiovisual information, regardless of specific emotion conveyed by music and faces. The ANOVA performed on the amplitude values of anterior N400 showed the strong significance of congruence factor [$F(1,15) = 5.53$, $p < 0.03$; $\varepsilon = 1$; $\eta_p^2 = 0.27$], with larger amplitudes to incongruent (−2.48 μV, *SD* = 0.44) than congruent stimulation (−1.70 μV, *DS* = 0.44) as clearly visible in **Figure 5A**. Also significant the factor electrode [$F(2,30) = 4.18$, $p < 0.025$; = 0.96, adjusted p value = 0.027; $\eta_p^2 = 0.22$], showing greater amplitudes ($p < 0.02$) at midline prefrontal than inferior frontal areas (Fpz = −2.56 μV, *SD* = 0.40; F5 = −1.97 μV, *SD* = 0.46 μV; F6 = −1.74, *SD* = 0.45 μV), as was proven by *post hoc* comparisons and visible in topographical maps of **Figure 5B**.

To locate the possible neural sources of the N400 response, a swLORETA source reconstruction was performed on the difference waves obtained by subtracting the ERPs elicited by the congruent from those elicited by the incongruent condition in the 300–500 ms time window. **Table 2** shows the electromagnetic dipoles that significantly explained the surface difference voltages, while the inverse solution is displayed in **Figure 6**. Overall, the localization of intracranial sources highlighted the contribution of areas extracting and comparing facial and musical affect, particularly the left inferior temporal gyrus (BA20) and the left and right superior temporal gyri (BA38); also active were regions involved in the processing of emotional music (such as the cuneus bilaterally and the left inferior parietal lobule), and the medial prefrontal cortex (BA10).

## DISCUSSION

The aim of this study was to investigate how the conceptual incongruity between facial expressions and musical pieces that expressed different emotions was implicitly processed by unaware participants. The more general assumption was indeed that music was able to clearly convey emotional meanings

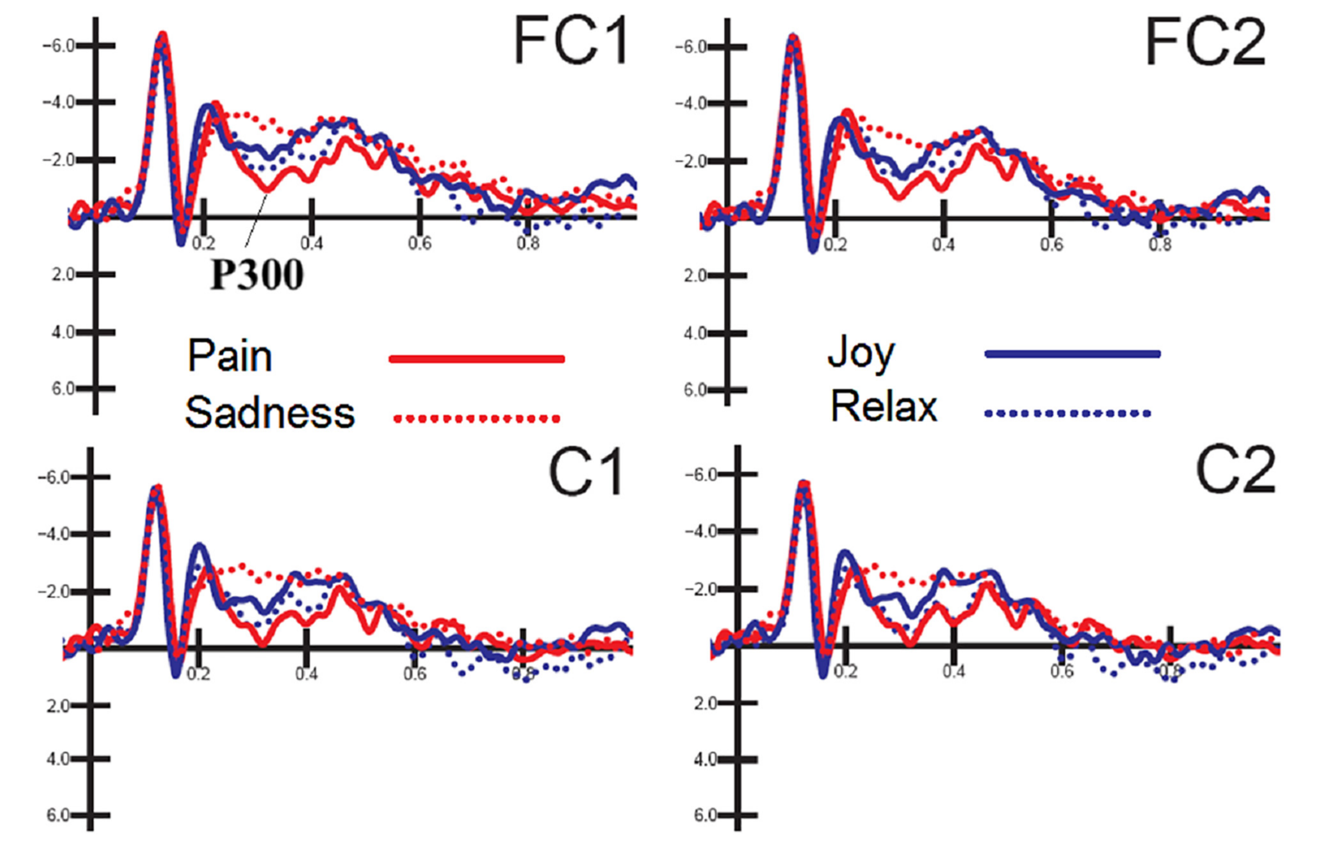



**FIGURE 3 |** FACES. Grand-average ERP waveforms elicited by the four types of facial expressions (regardless of music content) as recorded at left and right fronto/central sites.

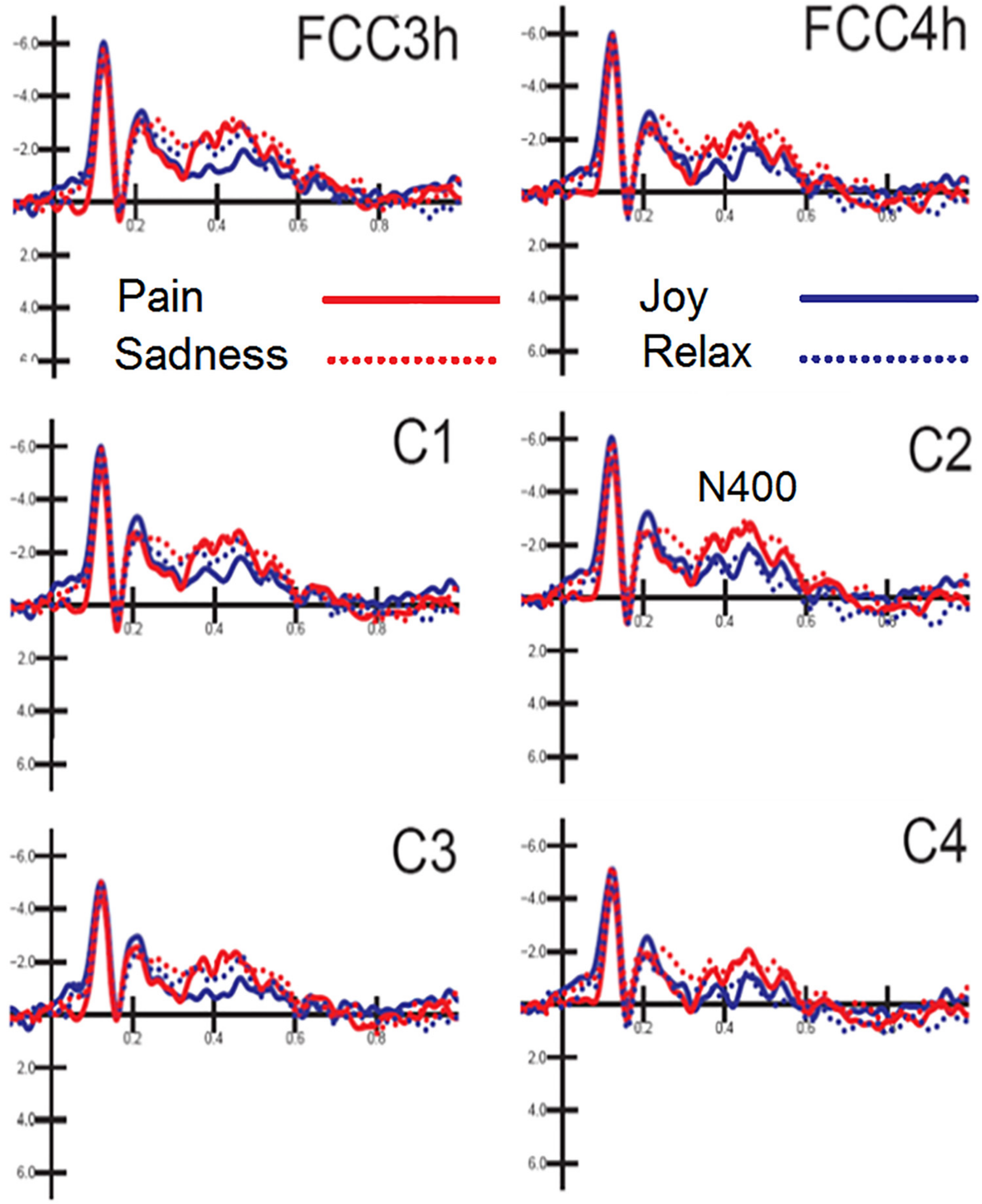



**FIGURE 4 |** MUSIC. Grand-average ERP waveforms elicited by the four types of emotional content in musical pieces (regardless of facial expressions) as recorded at left and right fronto/central and central sites.

(Panksepp and Bernatzky, 2002; Brattico and Pearce, 2013), so that the concurrent presentation of incongruent information might trigger a N400 response to semantic incongruence, as the one recorded in linguistic (Kutas and Federmeier, 2011), motor (Proverbio and Riva, 2009), musical (Koelsch, 2011b) or perceptual (Barrett and Rugg, 1990) contexts. To this end, pictures depicting emotional facial expressions in infants (Proverbio et al., 2007) and musical fragments belonging to the same four emotional categories as the faces (Vieillard et al., 2008) were created and validated. Stimulus validation demonstrated that music pieces and facial expressions conveyed a clearly understandable emotional meaning, and that their incongruent mixing was clearly detected by subjects.

## Emotional Content of Facial Expressions

The analyses carried out on the visual N170 component showed differences in the processing of the emotions expressed by the faces, specifically N170 was greater in amplitude to distress than other expressions. This result fits with previous literature showing larger N170s to negative than positive or neutral expressions (Batty and Taylor, 2003; Proverbio et al., 2006; Sun et al., 2017). To identify the most active dipoles

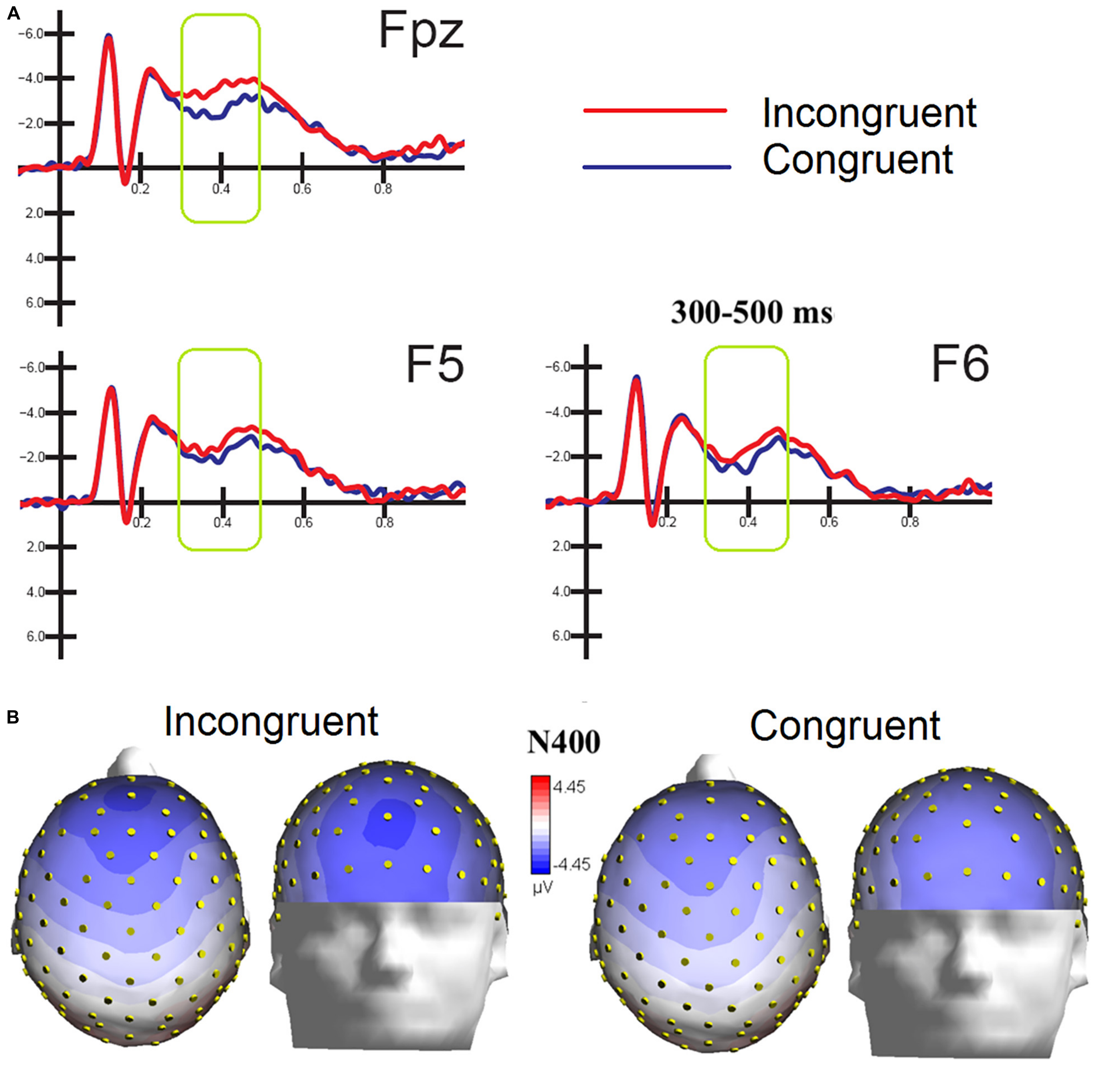



**FIGURE 5 | (A)** MULTIMODAL. Grand-average ERP waveforms elicited by congruent and incongruent audiovisual stimuli (regardless of specific emotional content of music and faces) as recorded at left and right fronto/central an central sites. **(B)** Top and front view of Isocolor topographical maps of surface voltage recorded in between 300 and 500 ms in response to emotionally congruent and incongruent audiovisual stimulation.

during the coding of facial expressions, a swLORETA inverse solution (*standardized weighted Low Resolution Electromagnetic Tomography*) was applied to the surface potentials recorded in the time window between 160 and 180 ms: the right middle Occipital Gyrus (MOG, BA 19) was the most active dipole, along with left MOG and primary visual cortex. This findings agrees with previous literature (Haxby et al., 2000; Gobbini and Haxby, 2007; Fusar-Poli et al., 2009) showing the role of the so-called "occipital face area" in the processing of faces, and especially face details. The emphasis on the local element (i.e., a single face element, such as a wrinkle, a skin fold, eye sockets, etc…) might possibly be related to the task, which consisted in analyzing face age (see also Wiese et al., 2012) to detect and respond to adult faces. Althought N170 it is generally larger over the right hemisphere, it is notable that it was more prominent over the right to negative emotions.

The later visual P300 (between 250 and 450 ms), averaged as a function of facial expression, was strongly modulated by emotional content, being it larger to the most arousing emotions (distress), with an intermediate amplitude for joy and relaxed/neutral expressions and smaller for sadness displays.

**TABLE 2 |** List of active electro-magnetic dipoles (along with their Talairach coordinates) explaining the surface difference-voltage (incongruent – congruent) recorded between 300 and 500 ms post-stimulus (N400 latency range) to audiovisual stimuli.

| Magn. | T-x [mm] | T-y [mm] | T-z [mm] | Hem. | Lobe | Gyrus | BA | Presumed function |
|---|---|---|---|---|---|---|---|---|
| 7.92 | −58.5 | −8.7 | −21.5 | L | T | Inferior temporal | 20 | Regions extracting and comparing facial and musical affect (e.g., Aubé et al., 2015; Pehrs et al., 2014) Harmonic processing (Peretz, 1998, 2001) |
| 7.39 | −38.5 | 9.1 | −27.5 | L | T | Superior temporal | 38 | |
| 4.29 | 60.6 | 5.3 | 2.7 | R | T | Superior temporal | 22 | |
| 7.13 | −48.5 | −33.7 | −23.6 | L | T | Fusiform | 20 | |
| 5.87 | 11.3 | −98.5 | 2.1 | R | O | Cuneus | | Processing of frightening music (Koelsch et al., 2018) |
| 5.58 | −8.5 | 57.3 | −9.0 | L | F | Superior frontal | 10 | Working memory, Music processing (Bogert et al., 2016) |
| 5.49 | 1.5 | 65.3 | 7.9 | R | F | Medial frontal | 10 | |
| 4.41 | −48.5 | −61.8 | 41.2 | L | P | Inferior parietal lobule | 39 | Processing of frightening music (Koelsch et al., 2018) |
| 4.32 | 21.2 | −8.0 | −28.9 | R | Limbic | Uncus | 36 | Emotion |

*Magn., magnitude of the signal in nA; Hem., hemisphere; L, left; R, right. T, temporal; O, occipital; F, frontal; P, parietal.*

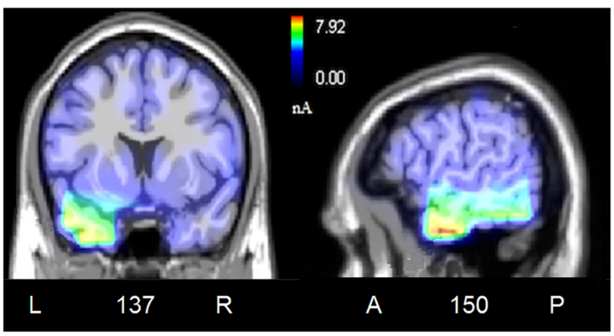



**FIGURE 6 |** Coronal, and sagittal views of active sources during the processing of incongruent audiovisual information (300–500 ms). The different colors represent differences in the magnitude of the electromagnetic signal (in nAm). The numbers refer to the displayed brain slice in sagittal view: A, anterior; P, posterior. The images highlight the strong activation of inferior and superior temporal cortex in combining the emotional meanings of music and facial expressions.

This gradient in neural response possibly reflects an effect of face emotional intensity and induced arousal level (Carretié and Iglesias, 1995; Polich, 2007). This finding is consistent with the previous literature showing larger P300s to emotional than neutral expressions (Lang et al., 1990; Proverbio et al., 2006).

## Hemispheric Asymmetry for Positive vs. Negative Emotions

As for the processing of music content, auditory N400, recorded over the fronto/central area in between 350 and 550 ms, was greater to negative (distress and sadness) than positive emotional musical fragments, especially over the right hemisphere. This right-sided hemispheric asymmetry for processing negative emotions agrees with previous neuroimaging literature showing how the processing of positive and joyful music mostly engage the left frontal cortex, whereas sad or fearful music would mostly engage the right frontal cortex (Schmidt and Trainor, 2001). Again neurological studies in patients with unilateral brain lesions provided evidence of a dominance of the left hemisphere for positive emotions and of the right hemisphere for negative emotions (Ahern and Schwartz, 1979; Reuter-Lorenz and Davidson, 1981; Rodway et al., 2003; Gainotti, 2019). Consistently Davidson (1992) provided evidence that the left prefrontal cortex (PFC) would be more involved in positive and enjoyable stimuli inducing an approach to appetitive stimuli, whereas the right PFC would be more involved in processing aversive and negative stimuli.

## Multimodal Processing of Affective Information

The analyses of grand-average ERP waveforms computed as a function of stimulus congruity, therefore reflecting a multimodal processing, showed a strong N400 effect for the presentation of incongruent pairs. N400 amplitude would be greater in anomalous (incongruent) semantic contexts since integration requires more time and cognitive resources with respect to congruent scenarios (Lau et al., 2008). The present findings are in agreement with previous ERP literature combining musical and facial information and finding larger N400s to emotionally incongruent pairs (Kamiyama et al., 2013), as well as to other types of incongruent multimodal information. For example N400 has been found larger to incongruent pairs of pictures and words (Nigam et al., 1992), sounds and musical gestures (Proverbio et al., 2015a), words and music (Daltrozzo and Schön, 2009; Goerlich et al., 2011), space and music (Koelsch et al., 2004; Zhou et al., 2014). The above studies suggest the existence of an amodal and shared conceptual system, indexed by N400 response to incongruity in different multimodal domains. The present data showed that emotional content of music (appropriately selected) can induce rather distinctive meanings, as clearly comprehensible as facial expressions.

To assess which were the cortical areas responsible for extracting and processing the emotional significance of audiovisual stimuli, difference-waves were computed by subtracting the ERPs elicited by the congruent stimuli from those elicited by incongruent stimuli, and analyzed in the N400 time window (between 300 and 500 ms). The inverse swLORETA solution (*standardized weighted Low Resolution Electromagnetic Tomography*) applied to the N400 response showed as most active dipole the left inferior temporal gyrus (ITG, BA 20), which is anatomically adjacent to the medial temporal gyrus (MTG) identified as the neural generator of linguistic N400 (McCarthy et al., 1995). Very active were also found the bilateral Superior Temporal Gyrus (STG, BA 38/22) and the left Fusiform Gyrus (FG, BA 20), specialized in the coding of the emotional content of music and faces (e.g., Pehrs et al., 2014; Aubé et al., 2015) and harmonic processing (STG, Peretz, 1998, 2001). Indeed, multimodal audiomotor neurons located in the posterior superior temporal sulcus (pSTS) and in the medial temporal gyrus (MTG) respond both to sounds and to visual images of objects and animals (Wright et al., 2003). Again, a specific region located posteriorly and ventrally to STS, named the *temporal visual speech area* (TVSA), seems particularly responsive to auditory and visual speech stimuli as it is also at the basis of audiovisual McGurk illusion (Bernstein and Liebenthal, 2014; Proverbio et al., 2018).

The STG's role in integrating audiovisual information and in the extraction of affective properties from both sensory modalities has been explored by Aubé et al. (2015) in an fMRI study aimed at investigating the neural correlates of processing specific basic emotions (fear, sadness and joy), expressed through music, vocalizations and facial expressions. They found that the STG was deeply involved in the response to both happy and frightening music, the activation signal being modulated by intensity and arousal. Similarly, in a study in which faces were presented during listening of strongly emotional music (namely, *Pathetic symphony* by Tchaikovsky) it was found that the medial prefrontal cortex (BA10) and the STG were strongly involved the combined processing of facial and musical affective information (Proverbio and De Benedetto, 2018). Again, a recent neuroimaging study (Pehrs et al., 2014) have investigated how the brain integrates the visual information of a movie with its musical soundtrack into a coherent percept. At this purpose dynamic kissing scenes from romantic comedies were presented during fMRI scanning. The kissing scenes were either accompanied by happy music, sad music or no music. The presence of music enhanced activation signals in multisensory integration network consisting of fusiform gyrus, amygdala and anterior superior temporal gyrus (aSTG). Again Baumgartner et al. (2006a,b) explored brain activity during the cross-modal presentation of affective images (sad, fearful) in two conditions, alone or with emotionally congruent musical traces. They found that the main areas involved in the cross-modal (multisensory) integration between emotional images and music tracks were the MTG and the temporal pole. The findings outlined above are generally coherent with the present data, except for the amygdala activation that cannot unfortunately be detected through EEG signals via LORETA.

Other brain areas found to be active in the processing of emotional audiovisual content, in our study, were the right cuneus and the left inferior parietal lobule (IPL, BA 39). Intriguingly Koelsch et al. (2018) found that both regions were active during the processing of fearful music. Also active according to swLORETA were the left superior frontal gyrus and the right medial frontal gyrus, commonly active during music processing (Bogert et al., 2016; Proverbio and De Benedetto, 2018) and reflecting stimulus coding and working memory.

On the basis of the data obtained in this study it is possible to conclude that the extraction and integration of the emotional content of multimodal stimuli takes place automatically (on task-irrelevant information) in a very short time, after about 400 ms from the presentation of the stimuli. This result confirms the extraordinary ability of music to communicate emotions clearly, and distinctively as emotional facial expressions. A discussion about the biological bases of such an innate ability can be found in a recent electrophysiological study comparing the comprehension of spontaneous vocalizations (e.g., laughs and crying) vs. instrumental music (Proverbio et al., 2019). Both stimulation types involved brain areas shaped for processing the human voice and its affective modulations. Indeed it has been suggested that music universality derives from the existence of a common neural mechanism for the comprehension of the emotional content of music, vocalizations (e.g., laughter and crying) and speech prosody (Panksepp and Bernatzky, 2002; Proverbio and Santoni, 2019; Proverbio et al., 2019), mostly relying on fronto/temporal areas. The literature suggests that music and vocalizations use similar patterns of acoustic cues to express emotions (Juslin and Laukka, 2003; Paquette et al., 2013), which might explain some universal and pretty innate brain reaction to music, regardless of cultural factors such as: education, familiarity or aesthetic taste.

The present data show how the brain is highly capable of integrating emotional information coming from different sensory modalities, to form a coherent conceptual representation comparable to semantic meaning of information, and how this mainly involves the MTG and the STG, the superior and medial frontal gyri, uncus, parietal and limbic areas. A right hemispheric dominance for processing negative (distress and sadness) vs. positive emotions (joy and relaxation), was also found at anterior areas, as indexed by N400 response to music.

One study's limitation might be the somewhat limited sample size, comprising 20 participants in the EEG recording session but only 16 after the EEG artifact rejection procedure. The merits of this study, compared to the previous ones are to have clarified the communicative power of music and facial expressions with a symmetrical and balanced mode of stimulation, with ultra-validated stimuli and with an implicit paradigm to detect the automatic mechanisms of extraction of emotional meaning, without directing or conveying subjective interpretations. The data show a certain universality of some musical parameters in inducing specific emotional sensations.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author in case of a scientific cooperation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Committee of University of Milano-Bicocca. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

AP conceived and planned the experiment. EC and AB prepared the stimuli, carried out the EEG recordings, and performed the statistical analyses. AP interpreted the data and took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript.

## FUNDING

This work was supported by the 28882 2018-ATE-0003 grant entitled "Neural encoding of the emotional content of speech and music" from University of Milano-Bicocca.

## ACKNOWLEDGMENTS

The authors are extremely grateful to Roberta Adorni, as well as to Francesco De Benedetto and Andrea Orlandi, for their technical assistance.

## REFERENCES

Ahern, G. L., and Schwartz, G. E. (1979). Differential lateralization for positive versus negative emotion. *Neuropsychologia* 17, 693–698. doi: 10.1016/0028-3932(79)90045-9

Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., and Armony, J. L. (2015). Fear across the senses: brain responses to music, vocalizations and facial expressions. *Soc. Cogn. Affect. Neurosci.* 10, 399–407. doi: 10.1093/scan/nsu067

Baranowski, A. M., and Hecht, H. (2017). The auditory kuleshov effect: multisensory integration in movie editing. *Perception* 46, 624–631. doi: 10.1177/0301006616682754

Barrett, S. E., and Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain Cogn.* 14, 201–212. doi: 10.1016/0278-2626(90)90029-n

Batty, M., and Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cogn. Brain Res.* 17, 613–620. doi: 10.1016/s0926-6410(03)00174-5

Baumgartner, T., Esslen, M., and Jäncke, L. (2006a). From emotion perception to emotion experience: emotions evoked by pictures and classical music. *Int. J. Psychophysiol.* 60, 34–43. doi: 10.1016/j.ijpsycho.2005.04.007

Baumgartner, T., Lutz, K., Schmidt, C. F., and Jäncke, L. (2006b). The emotional power of music: how music enhances the feeling of affective pictures. *Brain Res.* 1075, 151–164. doi: 10.1016/j.brainres.2005.12.065

Bernstein, L. E., and Liebenthal, E. (2014). Neural pathways for visual speech. *Front. Neurosci.* 8, 380–386.

Blood, A. J., and Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11818–11823. doi: 10.1073/pnas.191355898

Bogert, B., Numminen-Kontti, T., Gold, B., Sams, M., Numminen, J., Burunat, I., et al. (2016). Hidden sources of joy, fear, and sadness: explicit versus implicit neural processing of musical emotions. *Neuropsychologia* 89, 393–402. doi: 10.1016/j.neuropsychologia.2016.07.005

Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., et al. (2011). A functional MRI study of happy and sad emotions in music with and without lyrics. *Front. Psychol.* 2:308. doi: 10.3389/fpsyg.2011.00308

Brattico, E., and Pearce, M. (2013). The neuroaesthetics of music. *Psychol. Aesthet. Creat. Arts* 7, 48–61. doi: 10.1037/a0031624

Carretié, L., and Iglesias, J. (1995). An ERP study on the specificity of facial expression processing. *Int. J. Psychophysiol.* 19, 183–192. doi: 10.1016/0167-8760(95)00004-c

Daltrozzo, J., and Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *J. Cogn. Neurosci.* 21, 1882–1892. doi: 10.1162/jocn.2009.21113

Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. *Brain Cogn.* 20, 125–151. doi: 10.1016/0278-2626(92)90065-t

Egermann, H., Fernando, N., Chuen, L., McAdams, S., Egermann, H., Fernando, N., et al. (2015). Music induces universal emotion-related psychophysiological responses: comparing Canadian listeners to congolese pygmies. *Front. Psychol.* 7:1341. doi: 10.3389/fpsyg.2014.01341

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., et al. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *J. Psychiatry Neurosci.* 6, 418–432.

Gainotti, G. (2019). The role of the right hemisphere in emotional and behavioral disorders of patients with frontotemporal lobar degeneration: an updated review. *Front. Aging Neurosci.* 19:55. doi: 10.3389/fnagi.2019.00055

Gobbini, M. I., and Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia* 45, 32–41. doi: 10.1016/j.neuropsychologia.2006.04.015

Goerlich, K. S., Witteman, J., Aleman, A., and Martens, S. (2011). Hearing feelings: affective categorization of music and speech in alexithymia, an ERP study. *PLoS One* 6:e19501. doi: 10.1371/journal.pone.0019501

Gosselin, N. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain* 128, 628–640. doi: 10.1093/brain/awh420

Gosselin, N., Peretz, I., Johnsen, E., and Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. *Neuropsychologia* 45, 236–244. doi: 10.1016/j.neuropsychologia.2006.07.012

Hanser, W. E., Mark, R. E., Zijlstra, W. P., and Vingerhoets, A. J. J. M. (2015). The effects of background music on the evaluation of crying faces. *Psychol. Music* 43, 75–85. doi: 10.1177/0305735613498132

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233. doi: 10.1016/s1364-6613(00)01482-0

Ishizu, T., and Zeki, S. (2011). Toward a brain-based theory of beauty. *PLoS One* 6:e21852. doi: 10.1371/journal.pone.0021852

Jeong, J. W., Diwadkar, V. A., Chugani, C. D., Sinsoongsud, P., Muzik, O., Behen, M. E., et al. (2011). Congruence of happy and sad emotion in music and faces modifies cortical audiovisual activation. *Neuroimage* 54, 2973–2982. doi: 10.1016/j.neuroimage.2010.11.017

Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770

Kamiyama, K. S., Abla, D., Iwanaga, K., and Okanoya, K. (2013). Interaction between musical emotion and facial expression as measured by event-related potentials. *Neuropsychologia* 51, 500–505. doi: 10.1016/j.neuropsychologia.2012.11.031

Khalfa, S., Schon, D., Anton, J.-L., and Liégeois-Chauvel, C. (2005). Brain regions involved in the recognition of happiness and sadness in music. *Neuroreport* 16, 1981–1984. doi: 10.1097/00001756-200512190-00002

Koelsch, S. (2011a). Towards a neural basis of processing musical semantics. *Phys. Life Rev.* 8, 89–105.

Koelsch, S. (2011b). Toward a neural basis of music perception – a review and updated model. *Front. Psychol.* 2:110. doi: 10.3389/fpsyg.2011.00110

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., and Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nat. Neurosci.* 7, 302–307. doi: 10.1038/nn1197

Koelsch, S., Skouras, S., and Lohmann, G. (2018). The auditory cortex hosts network nodes influential for emotion processing: an fMRI study on music-evoked fear and joy. *PLoS One* 13:e0190057. doi: 10.1371/journal.pone.0190057

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Ann. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Lang, S. F., Nelson, C. A., and Collins, P. F. (1990). Event-related potentials to emotional and neutral stimuli. *J. Clin. Exp. Neuropsychol.* 12, 946–958. doi: 10.1080/01688639008401033

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., and Beller, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion* 13, 434–449. doi: 10.1037/a0031388

Logeswaran, N., and Bhattacharya, J. (2009). Crossmodal transfer of emotion by music. *Neurosci. Lett.* 455, 129–133. doi: 10.1016/j.neulet.2009.03.044

McCarthy, G., Nobre, A. C., Bentin, S., and Spencer, D. D. (1995). Language-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. *J. Neurosci.* 15, 1080–1089. doi: 10.1523/jneurosci.15-02-01080.1995

Menon, V., and Levitin, D. J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *Neuroimage* 28, 175–184. doi: 10.1016/j.neuroimage.2005.05.053

Nigam, A., Hoffman, J. E., and Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *J. Cogn. Neurosci.* 4, 15–22. doi: 10.1162/jocn.1992.4.1.15

Nordströma, H., and Laukka, P. (2019). The time course of emotion recognition in speech and music. *J. Acoust. Soc. Am.* 145, 3058–3074. doi: 10.1121/1.5108601

Palmero-Soler, E., Dolan, K., Hadamschek, V., and Tass, P. A. (2007). swLORETA: a novel approach to robust source localization and synchronization tomography. *Phys. Med. Biol.* 52, 1783–1800. doi: 10.1088/0031-9155/52/7/002

Panksepp, J., and Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behav. Proc.* 60, 133–155. doi: 10.1016/s0376-6357(02)00080-3

Paquette, S., Peretz, I., and Belin, P. (2013). The "Musical Emotional Bursts": a validated set of musical affect bursts to investigate auditory affective processing. *Front. Psychol.* 4:509. doi: 10.3389/fpsyg.2013.00509

Pasqual-Marqui, R. D., Michel, C. M., and Lehmann, D. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *J. Psychophysiol.* 18, 49–65. doi: 10.1016/0167-8760(84)90014-x

Pehrs, C., Deserno, L., Bakels, J.-H., Schlochtermeier, L. H., Kappelhoff, H., Jacobs, A. M., et al. (2014). How music alters a kiss: superior temporal gyrus controls fusiform–amygdalar effective connectivity. *Soc. Cogn. Affect. Neurosci.* 9, 1770–1778. doi: 10.1093/scan/nst169

Peretz, I. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. *Cognition* 68, 111–141. doi: 10.1016/s0010-0277(98)00043-2

Peretz, I. (2001). Cortical deafness to dissonance. *Brain* 124, 928–940. doi: 10.1093/brain/124.5.928

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Proverbio, A. M., Brignone, V., Matarazzo, S., Del Zotto, M., and Zani, A. (2006). Gender and parental status affect the visual cortical response to infant facial expression. *Neuropsychologia* 44, 2987–2999. doi: 10.1016/j.neuropsychologia.2006.06.015

Proverbio, A. M., Calbi, M., Manfredi, M., and Zani, A. (2015a). Audio-visuomotor processing in the Musician's brain: an ERP study on professional violinists and clarinetists. *Sci. Rep.* 4:5866.

Proverbio, A. M., Crotti, N., Manfredi, M., Adorni, R., and Zani, A. (2012). Who needs a referee? How incorrect basketball actions are automatically detected by basketball players' brain. *Sci. Rep.* 2:883.

Proverbio, A. M., and De Benedetto, F. (2018). Auditory enhancement of visual memory encoding is driven by emotional content of the auditory material and mediated by superior frontal cortex. *Biol. Psychol.* 29, 164–175. doi: 10.1016/j.biopsycho.2017.12.003

Proverbio, A. M., De Benedetto, F., and Guazzone, M. (2019). Shared neural mechanisms for processing emotions in music and vocalizations. *Eur. J. Neurosci.* doi: 10.1111/ejn.14650 [Epub ahead of print].

Proverbio, A. M., Gabaro, V., Orlandi, A., and Zani, A. (2015b). Semantic brain areas are involved in gesture comprehension: an electrical neuroimaging study. *Brain Lang.* 147, 30–40. doi: 10.1016/j.bandl.2015.05.002

Proverbio, A. M., Matarazzo, S., Brignone, V., Zotto, M. D., and Zani, A. (2007). Processing valence and intensity of infant expressions: the roles of expertise and gender. *Scand. J. Psychol.* 48, 477–485. doi: 10.1111/j.1467-9450.2007.00616.x

Proverbio, A. M., Orlandi, A., and Pisanu, F. (2016). Brain processing of consonance/dissonance in musicians and controls: a hemispheric asymmetry revisited. *Eur. J. Neurosci.* 44, 2340–2356. doi: 10.1111/ejn.13330

Proverbio, A. M., Raso, G., and Zani, A. (2018). Electrophysiological indexes of incongruent audiovisual phonemic processing: unraveling the McGurk effect. *Neuroscience* 385, 215–226. doi: 10.1016/j.neuroscience.2018.06.021

Proverbio, A. M., and Riva, F. (2009). RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neurosci. Lett.* 459, 142–146. doi: 10.1016/j.neulet.2009.05.012

Proverbio, A. M., and Santoni, S. (2019). *ERP Makers of Valence Coding in Emotional Speech Processing (November 5, 2019). iScience, ISCIENCE-D-19-01212.* Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3480697 (accessed November 5, 2019).

Reuter-Lorenz, P., and Davidson, R. J. (1981). Differential contribution of the two cerebral hemispheres to the perception of happy and sad faces. *Neuropsychologia* 19, 609–613. doi: 10.1016/0028-3932(81)90030-0

Rodway, P., Wright, L., and Hardie, S. (2003). The valence-specific laterality effect in free viewing conditions: the influence of sex, handedness, and response bias. *Brain Cogn.* 53, 452–463. doi: 10.1016/s0278-2626(03)00217-3

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat. Neurosci.* 14, 257–262. doi: 10.1038/nn.2726

Schmidt, L. A., and Trainor, L. J. (2001). Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cogn. Emot.* 15, 487–500. doi: 10.1080/0269993004200187

Sievers, B., and Polansky, L. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proc. Natl. Acad. Sci. U.S.A.* 110, 70–75. doi: 10.1073/pnas.1209023110

Spreckelmeyer, K. N., Kutas, M., Urbach, T. P., Altenmüller, E., and Münte, T. F. (2006). Combined perception of emotion in pictures and musical sounds. *Brain Res.* 1070, 160–170. doi: 10.1016/j.brainres.2005.11.075

Sun, L., Ren, J., and He, W. (2017). Neural correlates of facial expression processing during a detection task: an ERP study. *PLoS One* 12:e0174016. doi: 10.1371/journal.pone.0174016

Trost, W., Ethofer, T., Zentner, M., and Vuilleumier, P. (2012). Mapping aesthetic musical emotions in the brain. *Cereb. Cortex* 22, 2769–2783. doi: 10.1093/cercor/bhr353

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., and Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cogn. Emot.* 22, 720–752. doi: 10.1080/02699930701503567

Vuilleumier, P., and Trost, W. (2015). Music and emotions: from enchantment to entrainment. *Ann. N. Y. Acad. Sci.* 1337, 212–222. doi: 10.1111/nyas.12676

Wiese, H., Klothm, N., Güllmar, D., Reichenbach, J. R., and Schweinberger, S. R. (2012). Perceiving age and gender in unfamiliar faces: an fMRI study on face categorization. *Brain Cogn.* 78, 163–168. doi: 10.1016/j.bandc.2011.10.012

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., and McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb. Cortex* 13, 1034–1043. doi: 10.1093/cercor/13.10.1034

Zanow, F., and Knösche, T. R. (2004). ASA-advanced source analysis of continuous and event-related EEG/MEG signals. *Brain Top.* 16, 287–290. doi: 10.1023/b:brat.0000032867.41555.d0

Zhou, L., Jiang, C., Delogu, F., and Yang, Y. (2014). Spatial conceptual associations between music and pictures as revealed by N400 effect: music conveys spatial concepts. *Psychophysiology* 51, 520–528. doi: 10.1111/psyp.12195

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 11 February 2020
doi: 10.3389/fnins.2020.00067

# SWS Brain-Wave Music May Improve the Quality of Sleep: An EEG Study

*Dongrui Gao[1,2,3†], Siyu Long[3†], Hua Yang[4†], Yibo Cheng[3], Sijia Guo[3], Yue Yu[2], Tiejun Liu[2,3], Li Dong[2,3], Jing Lu[2,3]* and Dezhong Yao[2,3]**

*[1] School of Computer Science, Chengdu University of Information Technology, Chengdu, China, [2] The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, University of Electronic Science and Technology of China, Chengdu, China, [3] Center for Information in Biomedicine, School of Life Sciences and Technology, University of Electronic Science and Technology of China, Chengdu, China, [4] Department of Composition, Sichuan Conservatory of Music, Chengdu, China*

**Aim:** This study investigated the neural mechanisms of brain-wave music on sleep quality.

**Background:** Sleep disorders are a common health problem in our society and may result in fatigue, depression, and problems in daytime functioning. Previous studies have shown that brain-wave music generated from electroencephalography (EEG) signals could emotionally affect our nervous system and have positive effects on sleep. However, the neural mechanisms of brain-wave music on the quality of sleep need to be clarified.

**Methods:** A total of 33 young participants were recruited and randomly divided into three groups. The participants listened to rapid eye movement (REM) brain-wave music (Group 1: 13 subjects), slow-wave sleep (SWS) brain-wave music (Group 2: 11 subjects), or white noise (WN) (Control Group: 9 subjects) for 20 min before bedtime for 6 days. EEG and other physiological signals were recorded by polysomnography.

**Results:** We found that the sleep efficiency increased in the SWS group but decreased in REM and WN groups. The sleep efficiency in the SWS group was ameliorated [$t(10) = -1.943$, $p = 0.076$]. In the EEG power spectral density analysis, the delta power spectral density in the REM group and in the control group increased, while that in the SWS group decreased [$F(2,31) = 7.909$, $p = 0.005$]. In the network analysis, the functional connectivity (FC), assessed with Pearson correlation coefficients, showed that the connectivity strength decreased [$t(10) = 1.969$, $p = 0.073$] between the left frontal lobe (F3) and left parietal lobe (C3) in the SWS group. In addition, there was a negative correlation between the FC of the left frontal lobe and the left parietal lobe and sleep latency in the SWS group ($r = -0.527$, $p = 0.064$).

**Conclusion:** Slow-wave sleep brain-wave music may have a positive effect on sleep quality, while REM brain-wave music or WN may not have a positive effect. Furthermore, better sleep quality might be caused by a decrease in the power spectral density of the delta band of EEG and an increase in the FC between the left frontal lobe and the left parietal lobe. SWS brain-wave music could be a safe and inexpensive method for clinical use if confirmed by more data.

**Keywords: sleep, brain-wave music, electroencephalography, neural plasticity, power spectra analysis**

OPEN ACCESS

***Edited by:***
*Cunmei Jiang,*
*Shanghai Normal University, China*

***Reviewed by:***
*Weijun Li,*
*Liaoning Normal University, China*
*Rui Zhang,*
*Zhengzhou University, China*

****Correspondence:***
*Jing Lu*
*lujing@uestc.edu.cn*
*Dezhong Yao*
*dyao@uestc.edu.cn*

*†These authors have contributed equally to this work*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *12 October 2019*
***Accepted:*** *16 January 2020*
***Published:*** *11 February 2020*

***Citation:***
*Gao D, Long S, Yang H, Cheng Y, Guo S, Yu Y, Liu T, Dong L, Lu J and Yao D (2020) SWS Brain-Wave Music May Improve the Quality of Sleep: An EEG Study. Front. Neurosci. 14:67. doi: 10.3389/fnins.2020.00067*

# INTRODUCTION

Sleep disorders, affecting up to 30% of adults, are a common health problem in our society and may result in fatigue, depression, and problems in daytime functioning (Chang et al., 2012). Pharmacological treatment is helpful for people suffering from sleep disorders but also has side effects, and some of these individuals could possibly turn to psychological treatment.

As a non-pharmacological treatment, music can affect sleep disorders, as shown in a number of studies. Experiments with subjects listening to music before sleep have revealed that listening to soft music shortens the duration of deep sleep and prolongs the duration of deep sleep (Chang et al., 2012; Chen et al., 2014). In addition, subjects who listened to music had a longer sleep duration, greater sleep efficiency, shorter sleep latency, less sleep disturbance, and less daytime dysfunction as assessed by the Pittsburgh sleep quality index (PSQI) questionnaire (Tan, 2004). Similar results in an assessor-blinded randomized controlled trial (RCT) design showed a positive impact on sleep perception and quality of life (Jespersen et al., 2019). In electroencephalography (EEG) studies using time-frequency analysis methods, Kusumandari et al. (2018) demonstrated that music stimulation improved sleep quality.

Electroencephalography contains a wealth of information about brain activity. Scale-free brain-wave music, generated from EEG signals according to the power law of both EEG and music, possesses the characteristics of both music and EEG, and may contain physiological information that music alone may not (Wu et al., 2010, 2014; Lu et al., 2012). In recent years, brain-wave music has been shown to improve some clinical symptoms, such as pain (Levin, 1998). Levin's (1998) work, which used a combination of behavioral data and power spectral density, showed that brain-wave music incorporates factors of music therapy and biological feedback (Huang et al., 2016). Brain-wave music has been applied in the treatment of orofacial pain, and the results showed that the brain-wave music and cognitive behavioral therapy (CBT) group had lower levels of pain perception than the control group. In addition, the brain-wave music group showed lower EEG complexity and slower waves (Zhuang et al., 2009). Brain-wave music can also provide us with a new way to examine alterations in brains across various populations. The brain-wave music of healthy subjects and epilepsy patients clearly revealed differences in the two brain states, in that the brain music from the epilepsy patients was composed of unusual variations (Yao et al., 2016). Classic studies have allowed us to further explore neural mechanisms. Sleep staging and the PSQI questionnaire have been used to evaluate sleep quality in previous studies, and the results of the behavioral data showed that brain-wave music has a positive effect. However, the neural activities underlying the improvement in the quality of sleep by brain-wave music still need to be clarified. Therefore, our motivation of this study is to uncover this neural mechanism described above.

As representative sleep stages, rapid eye movement (REM) sleep repairs advanced cognitive function, and N3 stage sleep, also called slow-wave sleep (SWS) or deep sleep, can relieve fatigue (Griessenberger et al., 2013). We generated two types of scale-free brain-wave music as music stimulation, one from the REM stage and the other from the SWS stage. Deep sleep can predict sleep satisfaction and is a representative indicator of sleep quality (Riedel and Lichstein, 1998), so for EEG analysis, we mainly analyzed the power spectrum of EEG during deep sleep, and explored the neural mechanisms of these two brain-wave music on sleep promotion from the perspective of EEG.

# MATERIALS AND METHODS

## Participants

The study was implemented at a sleep center at the Clinical Hospital of Chengdu Brain Science Institute, University of Electronic Science and Technology of China (UESTC).

To observe the effect of brain-wave music on sleep, participants with a regular habit of staying up late were enrolled in the experiment. We recruited 36 right-handed subjects who had the sub-healthy sleep quality (PSQI scores should be between 4 and 8) from UESTC, and three of them gave up in midway through the experiment. The data of the remaining 33 participants (16 females; mean = 21.4 ± 5.6 years of age) were finally included in our experiment. All subjects gave informed consent for participation and received compensation.

## Music Stimulation

In this study, we translated EEG into brain-wave music with the paradigm shown in **Figure 1** (Wu et al., 2009; Lu et al., 2012). We used two pieces of brain-wave music in the experiment. One piece was REM brain-wave music, which was generated from EEG during the REM sleep. Another piece was SWS brain-wave music, which was generated from EEG during the SWS sleep. Some musical notes for each piece of music are shown in **Figure 2**.

## Electroencephalography Data Acquisition

A total of 6 Ag/AgCl electrodes (F3, F4, C3, C4, O1, and O2) that obtain signals related to sleep and other physiological signals from 10 to 20 system were selected for EEG recording by using an Alice 5 LDx system (Philips Respironics, PA, United States; Lucey et al., 2016; Gozal et al., 2019). A montage of the six electrodes

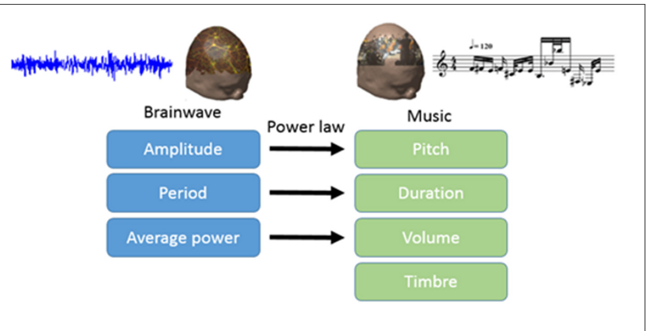



**FIGURE 1 |** Paradigm for translating electroencephalography (EEG) into brain-wave music.


**FIGURE 2 |** Illustration of the brain-wave music for the experiment [printed by Sibelius (2019)]. The notes from the REM brain-wave music at the beginning are shown in **(A)**, and the notes from the SWS brain-wave music at the beginning are shown in **(B)**.

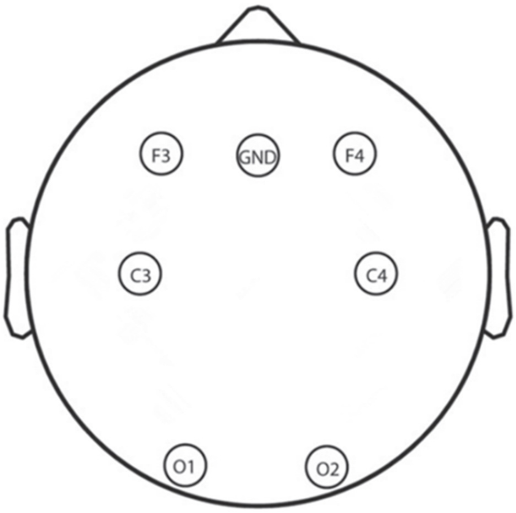



**FIGURE 3 |** Electroencephalography electrode layout.

used in this study is shown in **Figure 3**; these electrodes were chosen to cover the main portion of the brain so that we could calculate the power spectral density and build the whole brain network to examine neural mechanisms. The bilateral mastoids were linked as the reference, and all other electrodes were kept below 10 kΩ. The EEG signals were sampled at 200 Hz which and filtered between 0.5 and 30 Hz with a bandpass filter.

## Experimental Procedures

To avoid the effects of group differences in the initial state of sleep, participants were divided into three groups randomly. In reference to a compelling study about lullaby-accelerated falling asleep in children, the participants listened to REM brain-wave music (Group 1: 13 subjects, mean age = 21.69 ± 2.31), SWS brain-wave music (Group 2: 11 subjects, mean age = 21.77 ± 5.23), or white noise (WN; Control Group: 9 subjects, mean age = 19.78 ± 3.22 years) for 20 min before bedtime for 6 days (Patterson, 2011). EEG and other physiological signals were recorded by polysomnography on the first day and last day. The whole experiment lasted 8 days (**Figure 4**).

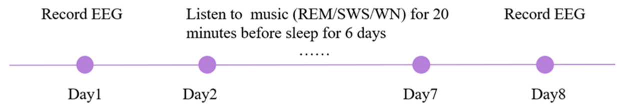



**FIGURE 4 |** Illustration of the experimental procedure.

## Data Analysis

According to previous studies, the period of N3 stage in the proportion of total sleep time and sleep latency can predict sleep satisfaction and are representative indicators of sleep quality (Riedel and Lichstein, 1998). A sleep latency of less than 15 min is rated as an appropriate measure for indexing good sleep quality (Ohayon et al., 2017). Meanwhile, sleep efficiency is also correlated to sleep quality (Jankelowitz et al., 2005), and a sleep efficiency of more than 85% is judged as an appropriate indicator of good sleep quality (Buysse et al., 1991; Åkerstedt et al., 1994; Ohayon et al., 2017). Therefore, we chose these three indicators for our behavioral data analysis.

Sleep stage assessment in the first session was based on EEG, electro-oculography (EOG), electrocardiography (ECG), and electromyography (EMG), according to the American Academy of Sleep Medicine (AASM) criteria and the identified EEG signals of deep sleep. Sleep is divided into five stages: W, R, N1, N2, and N3 in the AASM criteria, and EEG is obtained in the deep sleep stage, where fatigue is effectively relieved (Danker-Hopfe et al., 2009). According to the AASM, the delta wave accounts for more than 20% of a frame during the N3 stage. Moreover, total sleep time, sleep efficiency, sleep latency, and percentage of time in each sleep stage were calculated (Suzuki et al., 2019). Yue Yu, a physician of sleep medicine, extracted the EEG signal either from the N3 stage alone or from eight sets of N2 data, which is similar to N3 (the delta wave accounts for more than 15% of a frame), as the deep sleep data when the participant lacked the N3 stage according to the AASM criteria.

Deep sleep EEG was preprocessed by the reference electrode standardization technique (REST) with zero reference (Yao, 2001; Yao et al., 2019) and 0.5–30 Hz bandpass filtering under the Webrain platform[1]. Consider that the delta band (0.5–4 Hz) is the dominant frequency of EEG at N3 stage and related to the quality of sleep (Danker-Hopfe et al., 2009), we calculated power spectral density and brain network connectivity in delta band after preprocessing. The results were made clear through correlations among the total sleep time, sleep efficiency, sleep latency, percentage of time in each sleep stage, and EEG data (Guevara and Corsi-Cabrera, 1996; Chennu et al., 2016; Comsa et al., 2019).

## Statistical Analysis

Both the time of data collection and the intervention used (REM, SWS, and WN) were factors. All EEG data processing was based

[1] http://webrain.uestc.edu.cn/

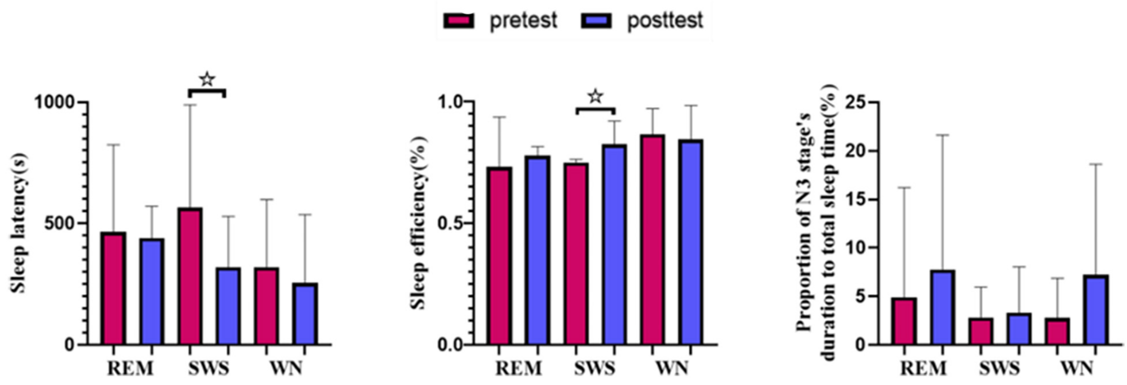



**FIGURE 5 |** Sleep variables from polysomnography (PSG). The sleep latency in the SWS group decreased significantly [$t(10) = 2.441$, $p = 0.031$] after listening to music, and the sleep efficiency in the SWS group increased significantly [$t(10) = -1.943$, $p = 0.076$]. * means there is a significant difference or a marginally significant difference between two groups.

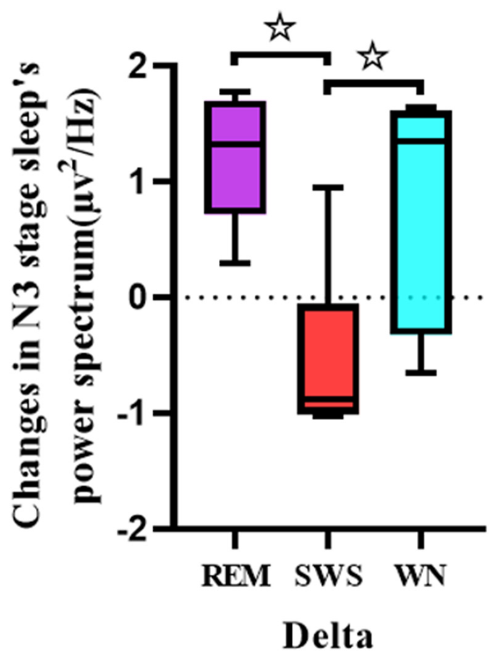



**FIGURE 6 |** Comparison of each group's whole brain power spectral density in the delta band of the EEG during the deep sleep stage (posttest – pretest). For the delta band of the EEG, there was a significant difference between the REM group and the SWS group ($p = 0.005$) and between the SWS group and the WN group ($p = 0.024$). * means there is a significant difference or a marginally significant difference between two groups.

on the EEG of N3 stage sleep or deep sleep. Comparison of the slight differences in the power spectral density among the three groups were assessed with ANOVA. *T*-tests were performed during the analysis.

## Strategy for Removing Outliers

Data from subjects who were not asleep or did not have N2 and N3 stages, as assessed by the sleep recording data, were considered outliers according to the AASM criteria. To obtain clear EEG data, we carefully eliminated some of the data obtained through bad channels.

# RESULTS

We found that the sleep latency in the SWS group decreased by 38.45% [$t(10) = 2.441$, $p = 0.031$] after listening to music. Although the sleep latency in the WN group and REM group subjects also decreased after the intervention, the differences were not significant. The sleep efficiency (sleep efficiency $= \frac{sleep\ time}{bed\ time}$) in the SWS group increased by 3.98% [$t(10) = -1.943$, $p = 0.076$], while in the other two groups, the sleep efficiency decreased. The percentage of sleep time spent in stage N3 increased in all three groups but not to a statistically significant degree (**Figure 5**).

One-way ANOVA was used to analyze the differences between the pretest and posttest whole brain power spectral density in the delta band of the deep sleep EEG among the REM, SWS, and WN groups (**Figure 6**). The SWS group's whole brain power spectral density decreased, while the other two groups showed increases in the delta band [$F(2,31) = 7.909$, $p = 0.005$].

We further analyzed the EEG power spectral density topographic maps in the delta band of the deep sleep stage for the three groups. The power spectral density for the whole brain in the delta band increased in the REM and WN groups, while the completely opposite effect was observed in the SWS group (**Figure 7**).

We investigated the functional connectivity (FC) of different regions of the brain via graph theory, which consists of nodes and edges. In our analysis, the scalp electrodes were defined as the nodes, and the Pearson correlation coefficients between nodes were defined as the edges. We also calculated the correlation coefficients between the F3–C3 and F3–O1 connectivities and sleep latency (**Figure 8**). The results suggested that sleep latency was inversely correlated with F3–C3 connectivity ($r = -0.527$, $p = 0.064$), meaning that a larger increase in the connectivity for F3–C3 could lead to a larger decrease in sleep latency.

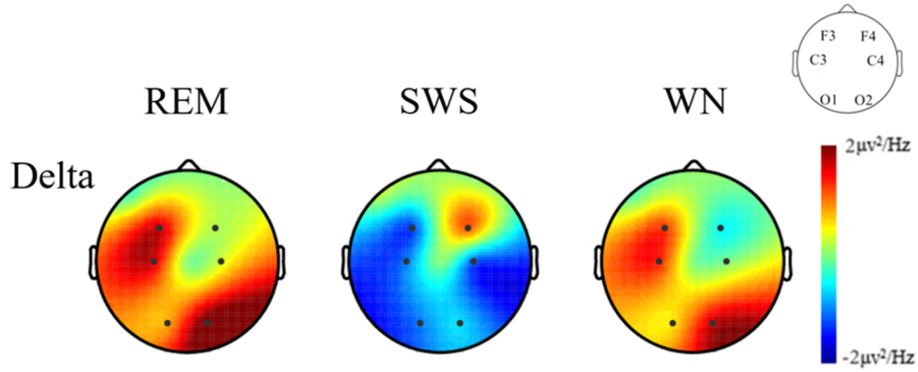



**FIGURE 7 |** The EEG power spectral density topographic maps of the deep sleep stage for the REM, SWS, and WN groups in the delta band, respectively (posttest – pretest). In the REM group and the WN group, the power spectral density at the F4 and C4 channels did not change after music listening. The power spectral density at the O1, C3, F3, and O2 channels all increased after music listening, of which the O1 channel increased the least and O2 increased the most. In the SWS group, the power spectral density at the F4 channel did not change, and the power spectral density at the F3, C3, C4, O1, and O2 channels decreased after the experiment, with the C4 channel decreasing the most. The power spectral density increased in F4 but decreased in other channels in the SWS group.

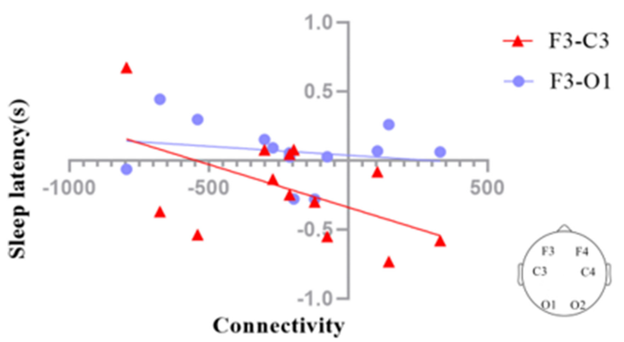



**FIGURE 8 |** Correlation coefficients between the EEG connectivities of the deep sleep stage and sleep latency in the SWS group (posttest – pretest). The connectivity between F3 and C3 (F3–C3) during deep sleep was marginally significantly correlated ($p = 0.064$) with sleep latency (red triangles), while the connectivity between F3 and O1 (F3–O1) during deep sleep was not significantly correlated with sleep latency (blue dots).

## DISCUSSION

Some previous studies have shown that listening to the subjects' own brain-wave music could improve the quality of sleep (Levin, 1998), while another study found that listening to healthy subjects' brain-wave music might be more useful (Yao et al., 2016). There is no agreement on what kinds of brain-wave music can improve the quality of sleep. REM sleep could repair advanced cognitive function, and SWS sleep could relieve fatigue (Griessenberger et al., 2013). Therefore, we chose these two kinds of EEG in different periods of sleep and generated the brain-wave music from them. As a result, our study found that SWS brain-wave music could improve sleep quality but REM brain-wave music could not. In order to understand its mechanism, we did further analysis on both behavioral and EEG data.

Sleep latency can be interpreted as a sense of sleepiness before going to bed and is a very important part of sleep quality (Chen et al., 2014). Music intervention before bedtime may facilitate relaxation as a person falls asleep (Steelman, 1990; Updike, 1990; White, 1992). It was found that listening to SWS brain-wave music at bedtime can shorten sleep latency, which is consistent with Levin's (1998) experimental results. In another study, listening to sedating music did not significantly alter sleep latency (Higuchi et al., 2005). Therefore, SWS brain-wave music may have a better effect with regard to relaxation than sedating music.

Normally, the delta band brain-wave is generated during sleep and relaxed conditions (Kumarahirwal and Londhe, 2013). A lower power in low-frequency band EEG indicates better sleep, especially deep sleep (Svetnik et al., 2017). It was found that the power spectral density in the SWS group decreased in the delta band, which is consistent with experimental results in subjects using benzodiazepines and zolpidem (Monti et al., 2000; Bastien et al., 2003). A study also found that with increasing age, the activity in the delta band decreases in power, which may be related to an attenuation of homeostatic sleep pressure and to an increase in cortical activation during sleep (Carrier et al., 2001). Therefore, the decrease in the delta brain waves in our study may have been indicative of an elevated sleep propensity and a relief from homeostatic sleep pressure in the SWS group (Esposito and Carotenuto, 2014). The power spectral density in the SWS group increased, while that in the other two groups decreased in the delta band (posttest – pretest). There were significant differences between the SWS and REM groups and between the SWS and WN groups. We could conclude that SWS brain-wave music had a positive effect.

However, in the delta band, the power spectral density of the REM and WN groups increased, and there was no significant difference between the two groups, indicating that REM brain-wave music and WN have similar effects on the EEG power spectrum. In Alexander's study, he found that the EEG power density in the low-frequency range (delta band) was an indicator of a progressively decreasing process during sleep (Borbély et al., 1981). It seemed that with the deepening of sleep, the power of delta frequency band decreased simultaneously. In our experiment, we found that after REM brain-wave music or WN listening, the power of delta band increased during sleep (posttest – pretest), and suggesting that these two kinds of music may have a negative effect on the deepening of sleep.

Overnight sleep deprivation leads to reduced activation of the frontal and parietal lobes (Chee and Tan, 2010). A meta-analysis showed brain activation in the right prefrontal cortex and medial frontal cortex was significantly reduced following sleep deprivation compared to rested wakefulness and that the activation in the frontoparietal attention network was reduced following acute total sleep deprivation compared to normal resting (Ma et al., 2015). These findings suggested that the decrease in this connectivity may be related to increased sleepiness and a greater likelihood of falling asleep. Chee et al. (2006) found that activation of the left frontal parietal lobe after normal sleep was negatively correlated with the performance accuracy decreases observed between normal sleep conditions and sleep deprivation over 24 h. In another study, Zou et al. (2018) found that the FC in the left frontoparietal network showed strong a correlation with REM sleep percentage. It appears that the activity of the left frontal and parietal lobes is highly correlated with various aspects of sleep. In our experiment, we found that the connectivity of the left frontal (F3) and parietal (C3) lobes was linked with sleep latency, so we speculate that the connectivity of the left frontal and parietal lobes may affect sleep latency.

This study has some limitations. First, insomniac patients should be recruited to determine the therapeutic effect of SWS brain-wave music on sleep in the future. Secondly, only two representative types of brain-wave music were selected in this experiment. Whether other types of music can promote sleep and the neural mechanisms remains to be further studied. Finally, further study should also consider the different effects on improving sleep quality between the brain-wave music and other types of music, such as classical music.

## CONCLUSION

This was an exploratory study on how SWS brain-wave music affects deep sleep. We suggest that SWS brain-wave music can decrease the delta band EEG power spectral density, shorten sleep latency and significantly correlate the F3–C3 connectivity, and sleep latency to improve sleep quality. However, REM brain-wave music and WN may not improve the quality of sleep.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the School of Life Science and Technology at the University of Electronic Science and Technology of China. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

DG, SL, HY, YC, JL, and DY designed the experiments. DG, SL, YC, and YY performed the experiments and collected the data. DG, SL, YC, SG, TL, LD, and JL analyzed the data. SL, YC, JL, and DY interpreted the results of experiments. All authors wrote and revised the manuscript and approved the submitted version.


## FUNDING

This study was supported by grants from the National Natural Science Foundation of China (Nos. 31600798 and 61761166001) and the Fundamental Research Funds for the Central Universities (No. ZYGX2016J188).


## REFERENCES


Bastien, C. H., LeBlanc, M., Carrier, J., and Morin, C. M. (2003). Sleep EEG power spectra, insomnia, and chronic use of benzodiazepines. *Sleep* 26, 313–317. doi: 10.1093/sleep/26.3.313

Borbély, A. A., Baumann, F., Brandeis, D., Strauch, I., and Lehmann, D. (1981). Sleep deprivation: effect on sleep stages and EEG power density in man. *Electroencephal. Clin. Neurophysiol.* 51, 483–493. doi: 10.1016/0013-4694(81)90225-X

Buysse, D. J., Reynolds, C. F. III, Monk, T. H., Hoch, C. C., Yeager, A. L., and Kupfer, D. J. (1991). Quantification of subjective sleep quality in healthy elderly men and women using the pittsburgh sleep quality index (PSQI). *Sleep* 14, 331–338. doi: 10.1093/sleep/14.4.331

Carrier, J., Land, S., Buysse, D. J., Kupfer, D. J., and Monk, T. H. (2001). The effects of age and gender on sleep EEG power spectral density in the middle years of life (ages 20–60 years old). *Psychophysiology* 38, 232–242. doi: 10.1111/1469-8986.3820232

Chang, E. T., Lai, H. L., Chen, P. W., Hsieh, Y. M., and Lee, L. H. (2012). The effects of music on the sleep quality of adults with chronic insomnia using evidence from polysomnographic and self-reported analysis: a randomized control trial. *Int. J. Nurs. Stud.* 49, 921–930. doi: 10.1016/j.ijnurstu.2012.02.019

Chee, M. W. L., Chuah, L. Y. M., Venkatraman, V., Chan, W. Y., Philip, P., and Dinges, D. F. (2006). Functional imaging of working memory following normal sleep and after 24 and 35 h of sleep deprivation: correlations of fronto-parietal activation with performance. *Neuroimage* 31, 419–428. doi: 10.1016/j.neuroimage.2005.12.001

Chee, M. W. L., and Tan, J. C. (2010). Lapsing when sleep deprived: neural activation characteristics of resistant and vulnerable individuals. *Neuroimage* 51, 835–843. doi: 10.1016/j.neuroimage.2010.02.031

Chen, C. K., Pei, Y. C., Chen, N. H., Huang, L. T., Chou, S. W., Wu, K. P., et al. (2014). Sedative music facilitates deep sleep in young adults. *J. Altern. Complement. Med.* 20, 312–317. doi: 10.1089/acm.2012.0050

Chennu, S., O'Connor, S., Adapa, R., Menon, D. K., and Bekinschtein, T. A. (2016). Brain connectivity dissociates responsiveness from drug exposure during propofol-induced transitions of consciousness. *Plos Compu. Biol.* 12:e1004669. doi: 10.1371/journal.pcbi.1004669

Comsa, I. M., Bekinschtein, T. A., and Chennu, S. (2019). Transient topographical dynamics of the electroencephalogram predict brain connectivity and behavioural responsiveness during drowsiness. *Brain Topogr.* 32, 315–331. doi: 10.1007/s10548-018-0689-9

Danker-Hopfe, H., Anderer, P., Zeitlhofer, J., Boeck, M., Dorn, H., Gruber, G., et al. (2009). Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. *J. Sleep Res.* 18, 74–84. doi: 10.1111/j.1365-2869.2008.00700.x

Esposito, M., and Carotenuto, M. (2014). Intellectual disabilities and power spectra analysis during sleep: a new perspective on borderline intellectual functioning. *J. Intellect. Disabil. Res.* 58, 421–429. doi: 10.1111/jir.12036

Gozal, L. K., Gutierrez-Tobal, G. C., Martin-Montero, A., Poza, J., Alvarez, D., del Campo, F., et al. (2019). Spectral EEG differences in children with obstructive sleep apnea. *Am. J. Respir. Crit. Care Med.* 199, A7375. doi: 10.1164/ajrccm-conference.2019.199.1_MeetingAbstracts.A7375

Griessenberger, H., Heib, D. P. J., Kunz, A. B., Hoedlmoser, K., and Schabus, M. (2013). Assessment of a wireless headband for automatic sleep scoring. *Sleep Breath* 17, 747–752. doi: 10.1007/s11325-012-0757-4

Guevara, M. A., and Corsi-Cabrera, M. (1996). EEG coherence or EEG correlation? *In. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* 23, 145–153. doi: 10.1016/S0167-8760(96)00038-4

Higuchi, S., Motohashi, Y., Liu, Y., and Maeda, A. (2005). Effects of playing a computer game using a bright display on presleep physiological variables, sleep latency, slow wave sleep and REM sleep. *J. Sleep Res.* 14, 267–273. doi: 10.1111/j.1365-2869.2005.00463.x

Huang, R., Wang, J., Wu, D., Long, H., Yang, X., Liu, H., et al. (2016). The effects of customised brainwave music on orofacial pain induced by orthodontic tooth movement. *Oral Dis.* 22, 766–774. doi: 10.1111/odi.12542

Jankelowitz, L., Reid, K. J., Wolfe, L., Cullina, J., Zee, P. C., and Jain, M. (2005). Cystic fibrosis patients have poor sleep quality despite normal sleep latency and efficiency. *Chest* 127, 1593–1599. doi: 10.1378/chest.127.5.1593

Jespersen, K. V., Otto, M., Kringelbach, M., Van Someren, E., and Vuust, P. (2019). A randomized controlled trial of bedtime music for insomnia disorder. *J. Sleep Res.* 28:e12817. doi: 10.1111/jsr.12817

kerstedt, T., Hume, K. E. N., Minors, D., and Waterhouse, J. I. M. (1994). The meaning of good sleep: a longitudinal study of polysomnography and subjective sleep quality. *J. Sleep Res.* 3, 152–158. doi: 10.1111/j.1365-2869.1994.tb00122.x

Kumarahirwal, M., and Londhe, N. D. (2013). Power spectrum analysis of EEG signals for estimating visual attention. *Int. J. Comput. Appl.* 42, 34–40. doi: 10.5120/5769-7993

Kusumandari, D. E., Suhendra, M. A., Amri, M. F., Simbolon, A. I., Rizqyawan, M. I., Wardono, P., et al. (2018). Comparison of EEG sleep characteristic with music and aromatherapy stimuli. *J. Phys.*1080:012050. doi: 10.1088/1742-6596/1080/1/012050

Levin, Y. I. (1998). "Brain music" in the treatment of patients with insomnia. *Neurosci. Behav. Physiol.* 28, 330–335. doi: 10.1007/BF02462965

Lu, J., Wu, D., Yang, H., Luo, C., Li, C. Y., and Yao, D. Z. (2012). Scale-free brain-wave music from simultaneously EEG and fMRI recordings. *Plos One* 7:e49773. doi: 10.1371/journal.pone.0049773

Lucey, B. P., Mcleland, J. S., Toedebusch, C. D., Boyd, J., Morris, J. C., Landsness, E. C., et al. (2016). Comparison of a single-channel EEG sleep study to polysomnography. *J. Sleep Res.* 25, 625–635. doi: 10.1111/jsr.12417

Ma, N., Dinges, D. F., Basner, M., and Rao, H. Y. (2015). How acute total sleep loss affects the attending brain: a meta-analysis of neuroimaging studies. *Sleep* 38, 233–240. doi: 10.5665/sleep.4404

Monti, J. M., Alvarino, F., and Monti, D. (2000). Conventional and power spectrum analysis of the effects of zolpidem on sleep EEG in patients with chronic primary insomnia. *Sleep* 23, 1075–1084.

Ohayon, M., Wickwire, E. M., Hirshkowitz, M., Albert, S. M., Avidan, A., Daly, F. J., et al. (2017). National Sleep Foundation's sleep quality recommendations: first report. *Sleep Health* 3, 6–19. doi: 10.1016/j.sleh.2016.11.006

Patterson, E. C. (2011). *The Effect of Lullaby Music Versus Rain Sounds on Inducing Sleep in the First 20 Minutes of Daycare Naptime*. Tallahassee, FL: Florida State University Libraries.

Riedel, B. W., and Lichstein, K. L. (1998). Objective sleep measures and subjective sleep satisfaction: how do older adults with insomnia define a good night's sleep? *Psychol. Aging* 13, 159–163. doi: 10.1037/0882-7974.13.1.159

Steelman, V. M. (1990). Intraoperative music therapy. *AORN J.* 52, 1026–1034. doi: 10.1016/s0001-2092(07)69164-9

Suzuki, Y., Rompré, P., Mayer, P., Kato, T., Okura, K., and Lavigne, G. J. (2019). Changes in oxygen and carbon dioxide in the genesis of sleep bruxism: a mechanism study. *J. Prosthodont. Res.* S1883-1958, 30446–30448. doi: 10.1016/j.jpor.2019.04.012

Svetnik, V., Snyder, E. S., Ma, J. S., Tao, P. N., Lines, C., and Herring, W. J. (2017). EEG spectral analysis of NREM sleep in a large sample of patients with insomnia and good sleepers: effects of age, sex and part of the night. *J. Sleep Res.* 26, 92–104. doi: 10.1111/jsr.12448

Tan, L. P. (2004). The Effects of background music on quality of sleep in elementary school children. *J. Music Ther.* 41, 128–150. doi: 10.1093/jmt/41.2.128

Updike, P. (1990). Music therapy results for ICU patients. *Dimens. Cri. Care Nurs.* 9, 39–45. doi: 10.1097/00003465-199001000-00013

White, J. M. (1992). Music therapy: an intervention to reduce anxiety in the myocardial infarction patient. *Clin. Nurse Spec.* 6, 58–63. doi: 10.1097/00002800-199200620-00002

Wu, D., Li, C., Yin, Y., Zhou, C., and Yao, D. (2010). Music composition from the brain signal: representing the mental state by music. *Intell. Neurosci.* 2010:4. doi: 10.1155/2010/267671

Wu, D., Li, C. Y., Liu, J., Lu, J., and Yao, D. Z. (2014). Scale-free brain ensemble modulated by phase synchronization. *J. Z. Univer. Sci. C Comput. Electron.* 15, 821–831. doi: 10.1631/jzus.C1400199

Wu, D., Li, C.-Y., and Yao, D.-Z. (2009). Scale-free music of the brain. *PLoS One* 4, e5915–e5915. doi: 10.1371/journal.pone.0005915

Yao, D., Chen, M., Lu, J., and Guo, D. (2016). *Neural Mass Model-Based scale-Free EEG Music*. Berlin: Springer, 455–460.

Yao, D., Qin, Y., Hu, S., Dong, L., Bringas Vega, M. L., and Valdés Sosa, P. A. (2019). Which reference should we use for EEG and ERP practice? *Brain Topogr.* 32, 530–549. doi: 10.1007/s10548-019-00707-x

Yao, D. Z. (2001). A method to standardize a reference of scalp EEG recordings to a point at infinity. *Physiol. Meas.* 22, 693–711. doi: 10.1088/0967-3334/22/4/305

Zhuang, T., Zhao, H., and Zheng, T. (2009). A study of brainwave entrainment based on EEG brain dynamics. *Comput. Inform. Sci.* 2:80. doi: 10.5539/cis.v2n2p80

Zou, Q. H., Zhou, S. Q., Xu, J., Su, Z. H., Li, Y. Z., Ma, Y. D., et al. (2018). Dissociated resting-state functional networks between the dream recall frequency and REM sleep percentage. *Neuroimage* 174, 248–256. doi: 10.1016/j.neuroimage.2018.03.015

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

GENERAL COMMENTARY
published: 01 February 2021
doi: 10.3389/fnins.2021.609169

# Commentary: SWS Brain-Wave Music May Improve the Quality of Sleep: An EEG Study

*Jennifer M. Johnson*[1,2]* *and Simon J. Durrant*[2,3]

[1] *School of Health and Social Care, University of Lincoln, Lincoln, United Kingdom,* [2] *Lincoln Sleep Research Centre, University of Lincoln, Lincoln, United Kingdom,* [3] *School of Psychology, University of Lincoln, Lincoln, United Kingdom*

**Keywords: slow-wave sleep, delta power, music, auditory cognitive neuroscience, sleep**

**A Commentary on**

**SWS Brain-Wave Music May Improve the Quality of Sleep: An EEG Study**
*by Gao, D., Long, S., Yang, H., Cheng, Y., Guo, S., Yu, Y., et al. (2020). Front. Neurosci. 14:67. doi: 10.3389/fnins.2020.00067*

Sleep is important for maintaining health and general well-being. Improving sleep is becoming more important due to the growing prevalence of sleep disorders, with non-pharmacological sleep interventions increasing in popularity (de Niet et al., 2009; Ngo et al., 2013b). When comparing interventions, music-based were the most successful for improving subjective sleep quality (de Niet et al., 2009), using a range of music types including classical, jazz, and sedative (Chan et al., 2010; Chen et al., 2014; Shum et al., 2014), as well as sounds including white (Afshar et al., 2016) and pink noise (Zhou et al., 2012). The effect of music on specific sleep stages, however, is less well-known with relatively few studies to date using objective sleep quality measures (Cordi et al., 2019).

Sleep consists of rapid eye-movement (REM) and non-rapid eye-movement (NREM) components, with NREM comprising various stages including slow-wave sleep (SWS) (Rechtschaffen and Kales, 1968), which is dominated by slow-wave activity (SWA) (0.5–4 Hz) consisting of delta and slow waves (<2 Hz) (Dijk et al., 1993). To date, music interventions have increased the amount of SWS (Chen et al., 2014; Cordi et al., 2019) and REM sleep (Chang et al., 2012), without changing delta power during SWS (Lazic and Ogilvie, 2007). To our knowledge, Gao et al. (2020) are the first to explore the impact of brain-wave music created from EEG during REM sleep.

Gao et al. (2020) explored the impact of SWS music ($n = 11$), REM sleep music ($n = 13$), and white noise ($n = 9$) played for 20 min before bedtime for 6 days on objective measures of sleep quality, including spectral power. The brain-wave music was created using the amplitude, period and average power of each sleep stage and translated into music pitch, duration, volume, and timbre using power law. Using this method, they aimed to compare the effects of each type of music on sleep quality and neural activation. The key finding was that after SWS brain-wave music, delta power significantly reduced, which the authors interpreted as a positive effect on sleep quality.

This is an important area of research, but in this case we offer an alternative interpretation of the results. The authors argue that lower delta power is indicative of improved sleep and SWS, citing previous research (Svetnik et al., 2017) and interpret the reduction in delta power as a reduced homeostatic pressure for delta power, as found in older adults (Landolt et al., 1996; Landolt and Borbély, 2001). However, Svetnik et al. (2017) actually found a reduction in delta power in younger adults (the age group in this study) was associated with insomnia, i.e., poorer sleep quality. Furthermore, while reduced delta activity could indicate a reduction in homeostatic sleep pressure, this is unlikely to be the case given that sleep latency after SWS music was significantly lower, and sleep efficiency somewhat higher, which are both associated with increased homeostatic sleep



***Edited by:***
*Josef P. Rauschecker,*
*Georgetown University, United States*

***Reviewed by:***
*Laura Verga,*
*Max Planck Institute for Psycholinguistics, Netherlands*

****Correspondence:***
*Jennifer M. Johnson*
*jjohnson@lincoln.ac.uk*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *22 September 2020*
***Accepted:*** *11 January 2021*
***Published:*** *01 February 2021*

***Citation:***
*Johnson JM and Durrant SJ (2021) Commentary: SWS Brain-Wave Music May Improve the Quality of Sleep: An EEG Study. Front. Neurosci. 15:609169. doi: 10.3389/fnins.2021.609169*

pressure (Dijk et al., 2010; Dijk and Landolt, 2019), while the duration of SWS was unchanged. These results suggest that the reduction of delta power is instead indicative of an impairment. This view results from both the positive consequences of increasing delta power and the negative consequences of prolonged delta power reduction (both clinically and cognitively), which is why research has focused on increasing delta power for optimal sleep quality (Marshall et al., 2006; Santiago et al., 2019).

SWS is the sleep stage most commonly associated with sleep quality due to its restorative nature (Åkerstedt et al., 1997; Dijk, 2009). To infer cause and effect, studies have focused on (a) enhancing SWS using stimulation, resulting in improvements to learning and health outcomes (Besedovsky et al., 2017; Johnson and Durrant, 2018); and (b) suppressing SWS, increasing the risk of type 2 diabetes through impaired insulin and glucose (Tasali et al., 2008; Herzog et al., 2013) and negatively affecting cognitive performance (Ferrara et al., 2000). These results collectively suggest that there are benefits to increasing delta power, whilst reducing delta power is problematic for a range of outcomes.

The brain-wave music intervention therefore appears to be unsuccessful and we suggest there are three possible reasons for this. Successful interventions use entrainment (Marshall et al., 2006; Ngo et al., 2013a) which may not be present in the current study. Auditory closed-loop stimulation entrains the sounds to the up-states of the slow-oscillations, with the slope of each individual participants slow-oscillation being measured, which then increases the amplitudes and power (Ngo et al., 2013a). The SWS brain-wave music was not, however, entrained to the SWA of each individual. Related to that, the music stimuli appeared to have no clear metrical structure and the tempo was nominally set to 120 bpm for both conditions with no consistent beat present in reality, in spite of previous findings connecting EEG periodicity to tempo (Fujioka et al., 2012). More generally, the particular mapping of EEG to music parameters is highly questionable, with no inherent relationship between average spectral power and musical timbre, for example. Finally, the stimulation was performed for only 20 min prior to sleep, when successful interventions performed prior to sleep/during sleep onset have been performed for a longer duration, e.g., 90 min (Ngo et al., 2013b). Similarly, performing auditory stimulation during SWS is more successful for improving SWA than prior to sleep/sleep onset (Ngo et al., 2013a,b). We therefore suggest that lack of entrainment, arbitrary EEG-music mapping and unfortunate stimulus timing may have contributed to the lack of success of the intervention.

In conclusion, we believe that this research does not show a positive effect of SWS brain-wave music on sleep quality; quite the contrary. We agree, however, that it is an important question and using music with characteristics specific to individual sleep stages is a positive development. The implementation in this case, however, seems to be flawed and as such has led to inconclusive findings. Future research should, therefore, focus on improving the implementation, incorporating entrainment, appropriate stimulus timing, and a music-EEG mapping grounded in existing evidence from cognitive neuroscience.

## AUTHOR CONTRIBUTIONS

JJ and SD wrote the paper. Both authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

We would like to thank the authors of the original paper.

## REFERENCES

Åkerstedt, T., Hume, K., Minors, D., and Waterhouse, J. (1997). Good sleep - its timing and physiological sleep characteristics. *J. Sleep Res.* 6, 221–229. doi: 10.1111/j.1365-2869.1997.00221.x

Afshar, F., Bahramnezhad, F., Asgari, P., and Shiri, M. (2016). Effect of white noise on sleep in patients admitted to a coronary care. *J. Caring Sci.* 5, 103–109. doi: 10.15171/jcs.2016.011

Besedovsky, L., Ngo, H.-V. V., Dimitrov, S., Gassenmaier, C., Lehmann, R., and Born, J. (2017). Auditory closed-loop stimulation of EEG slow oscillations strengthens sleep and signs of its immune-supportive function. *Nat. Commun.* 8:1984. doi: 10.1038/s41467-017-02170-3

Chan, M. F., Chan, E. A., and Mok, E. (2010). Effects of music on depression and sleep quality in elderly people: a randomised controlled trial. *Complement. Ther. Med.* 18, 150–159. doi: 10.1016/j.ctim.2010.02.004

Chang, E. T., Lai, H. L., Chen, P. W., Hsieh, Y. M., and Lee, L. H. (2012). The effects of music on the sleep quality of adults with chronic insomnia using evidence from polysomnographic and self-reported analysis: a randomized control trial. *Int. J. Nurs. Stud.* 49, 921–930. doi: 10.1016/j.ijnurstu.2012.02.019

Chen, C. K., Pei, Y. C., Chen, N. H., Huang, L. T., Chou, S. W., Wu, K. P., et al. (2014). Sedative music facilitates deep sleep in young adults. *J. Altern. Complement. Med.* 20, 312–317. doi: 10.1089/acm.2012.0050

Cordi, M. J., Ackermann, S., and Rasch, B. (2019). Effects of relaxing music on healthy sleep. *Sci. Rep.* 9:9079. doi: 10.1038/s41598-019-45608-y

de Niet, G. J., Tiemens, B. G., Kloos, M. W., and Hutschemaekers, G. J. M. (2009). Review of systematic reviews about the efficacy of non-pharmacological interventions to improve sleep quality in insomnia. *Int. J. Evid. Based. Healthc.* 7, 233–242. doi: 10.1111/j.1744-1609.2009.00142.x

Dijk, D.-J. (2009). Regulation and functional correlates of slow wave sleep. *J. Clin. Sleep Med.* 5, 13535–13539. doi: 10.5664/jcsm.5.2S.S6

Dijk, D.-J., Hayes, B., and Czeisler, C. A. (1993). Dynamics of electroencephalographic sleep spindles and slow wave activity in men: effect of sleep deprivation. *Brain Res.* 626, 190–199. doi: 10.1016/0006-8993(93)90579-C

Dijk, D.-J., and Landolt, H.-P. (2019). "Sleep physiology, circadian rhythms, waking performance and the development of sleep-wake therapeutics," in *Sleep-Wake Neurobiology and Pharmacology*, eds H.-P. Landolt and D.-J. Dijk (Cham: Springer International Publishing), 441–481. doi: 10.1007/164_2019_243

Dijk, D. J., Groeger, J. A., Stanley, N., and Deacon, S. (2010). Age-related reduction in daytime sleep propensity and nocturnal slow wave sleep. *Sleep* 33, 211–223. doi: 10.1093/sleep/33.2.211

Ferrara, M., Gennaro, L., Casagrande, M., and Bertini, M. (2000). Selective slow-wave sleep deprivation and time-of-night effects on cognitive performance upon awakening. *Psychophysiology* 37, 440–446. doi: 10.1111/1469-8986.3740440

Fujioka, T., Trainor, L. J., Large, E. W., and Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *J. Neurosci.* 32, 1791–1802. doi: 10.1523/JNEUROSCI.4107-11.2012

Gao, D., Long, S., Yang, H., Cheng, Y., Guo, S., Yu, Y., et al. (2020). SWS brain-wave music may improve the quality of sleep: an EEG study. *Front. Neurosci.* 14:67. doi: 10.3389/fnins.2020.00067

Herzog, N., Jauch-Chara, K., Hyzy, F., Richter, A., Friedrich, A., Benedict, C., et al. (2013). Selective slow wave sleep but not rapid eye movement sleep suppression impairs morning glucose tolerance in healthy men. *Psychoneuroendocrinology* 38, 2075–2082. doi: 10.1016/j.psyneuen.2013.03.018

Johnson, J. M., and Durrant, S. J. (2018). The effect of cathodal transcranial direct current stimulation during rapid eye-movement sleep on neutral and emotional memory. *R. Soc. Open Sci.* 5:172353. doi: 10.1098/rsos.172353

Landolt, H. P., and Borbély, A. A. (2001). Age-dependent changes in sleep EEG topography. *Clin. Neurophysiol.* 112, 369–377. doi: 10.1016/S1388-2457(00)00542-3

Landolt, H. P., Dijk, D. J., Achermann, P., and Borbély, A. A. (1996). Effect of age on the sleep EEG: slow-wave activity and spindle frequency activity in young and middle-aged men. *Brain Res.* 738, 205–212. doi: 10.1016/S0006-8993(96)00770-6

Lazic, S. E., and Ogilvie, R. D. (2007). Lack of efficacy of music to improve sleep: a polysomnographic and quantitative EEG analysis. *Int. J. Psychophysiol.* 63, 232–239. doi: 10.1016/j.ijpsycho.2006.10.004

Marshall, L., Helgadóttir, H., Mölle, M., and Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature* 444, 610–613. doi: 10.1038/nature05278

Ngo, H.-V., Martinetz, T., Born, J., and Mölle, M. (2013b). Auditory closed-loop stimulation of the sleep slow oscillation enhances memory. *Neuron* 78, 545–553. doi: 10.1016/j.neuron.2013.03.006

Ngo, H.-V. V., Claussen, J. C., Born, J., and Mölle, M. (2013a). Induction of slow oscillations by rhythmic acoustic stimulation. *J. Sleep Res.* 22, 22–31. doi: 10.1111/j.1365-2869.2012.01039.x

Rechtschaffen, A., and Kales, A. (1968). A *Manual of Standardised Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. *1st Edn.* Los Angeles, CA: Brain Information Service/Brain Research Institute: Elsevier Inc. doi: 10.1016/B978-1-4557-1267-0.00003-5

Santiago, J. C. P., Ngo, H. V., Jickeli, C., Peter, A., and Hallschmid, M. (2019). Intensifying sleep slow oscillations does not improve metabolic control in healthy men. *Psychoneuroendocrinology* 99, 1–7. doi: 10.1016/j.psyneuen.2018.08.028

Shum, A., Taylor, B. J., Thayala, J., and Chan, M. F. (2014). The effects of sedative music on sleep quality of older community-dwelling adults in singapore. *Complement. Ther. Med.* 22, 49–56. doi: 10.1016/j.ctim.2013.11.003

Svetnik, V., Snyder, E. S., Ma, J., Tao, P., Lines, C., and Herring, W. J. (2017). EEG spectral analysis of NREM sleep in a large sample of patients with insomnia and good sleepers: effects of age, sex and part of the night. *J. Sleep Res.* 26, 92–104. doi: 10.1111/jsr.12448

Tasali, E., Leproult, R., Ehrmann, D. A., and Van Cauter, E. (2008). Slow-wave sleep and the risk of type 2 diabetes in humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 1044–1049. doi: 10.1073/pnas.0706446105

Zhou, J., Liu, D., Li, X., Ma, J., Zhang, J., and Fang, J. (2012). Pink noise: effect on complexity synchronization of brain activity and sleep consolidation. *J. Theor. Biol.* 306, 68–72. doi: 10.1016/j.jtbi.2012.04.006

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 24 March 2020
doi: 10.3389/fnins.2020.00247

# Lowered Rhythm Tapping Ability in Patients With Constructional Apraxia After Stroke

**Naomi Kobinata[1,2]*, Hideto Yoshikawa[3], Yuji Iwasaka[4] and Nobuyuki Kawate[1]**

[1] *Department of Rehabilitation Medicine, School of Medicine, Showa University, Tokyo, Japan,* [2] *Department of Rehabilitation, Yoshieikai Hospital, Osaka, Japan,* [3] *Department of Medicine, Tokyo Chidori Hospital, Tokyo, Japan,* [4] *Department of Physical Therapy, Nihon Institute of Medical Science, Saitama, Japan*

Rhythm tapping tasks are often used to explore temporal reproduction abilities. Many studies utilizing rhythm tapping tasks are conducted to evaluate temporal processing abilities with neurological impairments and neurodegenerative disorders. Among sensorimotor and cognitive disorders, rhythm processing abilities in constructional apraxia, a deficit in achieving visuospatial constructional activities, has not been evaluated. This study aimed to examine the rhythm tapping ability of patients with constructional apraxia after a stroke. Twenty-four patients were divided into two groups: with and without constructional apraxia. There were 11 participants in the constructional apraxia group and 13 in the without constructional apraxia group. The synchronization-continuation paradigm was employed in which a person performs a synchronized tapping activity to a metronome beat and continues tapping after the beat has stopped. For statistical analysis, a three-way mixed analysis of variance (2 × 2 × 3) was conducted. The factors were groups (with and without constructional apraxia), tapping tasks (synchronization and continuation), and inter-stimulus intervals (600, 750, and 1000 ms). A significant effect of group factor was found ($F[1,132] = 16.62$; $p < 0.001$). Patients in the without constructional apraxia group were able to more accurately reproduce intervals than those in the constructional apraxia group. Moreover, a significant effect of tapping tasks was found ($F[1,132] = 8.22$; $p < 0.01$). Intervals were reproduced more accurately for synchronization tasks than continuation tasks. There was no significant inter-stimulus interval effect. Overall, these results suggest that there might be a relation between temporal and spatial reproductions in a wide spectrum of processing levels, from sensory perception to cognitive function.

**Keywords: rhythm tapping, constructional apraxia, synchronization, temporal reproduction, spatial reproduction**

OPEN ACCESS

**Edited by:**
*Elvira Brattico, Aarhus University, Denmark*

**Reviewed by:**
*Chia-Liang Tsai, National Cheng Kung University, Taiwan*
*Mark T. Elliott, University of Warwick, United Kingdom*

***Correspondence:**
*Naomi Kobinata*
*n-kobinata@yoshieikai.com*

**Specialty section:**
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

**Received:** *25 July 2019*
**Accepted:** *04 March 2020*
**Published:** *24 March 2020*

**Citation:**
*Kobinata N, Yoshikawa H, Iwasaka Y and Kawate N (2020) Lowered Rhythm Tapping Ability in Patients With Constructional Apraxia After Stroke. Front. Neurosci. 14:247. doi: 10.3389/fnins.2020.00247*

## INTRODUCTION

Rhythm tapping tasks are often used to explore temporal reproduction abilities (see Repp and Su, 2013, for a review). It is often performed as a finger tapping task in synchrony with an external rhythm, usually a steady metronome beat (Repp and Su, 2013). Along with the synchronization paradigm, synchronization-continuation tasks are often used to assess entrainment to an external rhythm (Flach, 2005; Ullén et al., 2008; Avanzino et al., 2013; McPherson et al., 2018).

With synchronization-continuation tasks, individuals tap in synchrony to an external beat and continue tapping after the external beat has stopped (Flach, 2005).

In synchronization tasks, automatic or cognitive control is involved depending on the speed of the external beat (Miyake et al., 2004; Repp and Su, 2013; Bååth et al., 2016). In time perception research, sub-second time processing is automatic and supra-second time processing involves cognitive control (Bååth et al., 2016). For example, Mangels et al. (1998) showed that patients with prefrontal lesions who had difficulty with a non-temporal working memory task also struggled with long duration temporal discrimination (4-s interval) but not with short duration temporal discrimination (400-ms interval). Miyake et al. (2004) conducted a study employing the dual tasks of synchronization tapping and word-memory; they found that with anticipatory tapping, synchronization with a stimulus interval of 1800 to 3600 ms was affected by a word-memory task but not synchronization with a stimulus interval of 1500 ms or less. Similar results were found with dual tasks involving executive control (Bååth et al., 2016).

Neural mechanisms for time measurement support the available behavioral evidence. Measurements of sub-second intervals revealed activity in the bilateral supplementary motor area, left sensorimotor cortex, right cerebellum, right lateral premotor area, left thalamus, left basal ganglia, and right superior temporal gyrus (Lewis and Miall, 2003). In cognitively controlled timing tasks, the right prefrontal and parietal cortices were involved in addition to some parts of the autonomic system (right premotor area and bilateral supplementary motor area) (Lewis and Miall, 2003).

Compared to the synchronization paradigm, synchronization-continuation requires internal pacing without external cues and increases the neural resources required (Serrien, 2008). In Serrien (2008)'s study, electroencephalogram coherence increased in mesial-central connections under the continuation condition. Moreover, Ullén et al. (2008) reported a correlation between tapping stability and the volume of the right prefrontal white matter regions under a continuation condition. These studies show that performing a continuation task requires internal control and increases neural activities. Unlike the synchronization task, the continuous sub-second tapping task requires cognitive control. According to Ullén et al. (2008), intelligence and the stability of continuous sub-second tapping were correlated; also, Holm et al. (2017) reported that executive control and working memory were involved in continuous sub-second tapping.

Many studies utilizing rhythm tapping tasks are conducted to evaluate temporal processing abilities with neurological impairments and neurodegenerative disorders (Freeman et al., 1993; Schwartze et al., 2011, 2016; Avanzino et al., 2013; Roalf et al., 2018). Schwartze et al. (2016) reported that patients with cerebellar lesions display imprecise temporal processing compared to healthy participants in a control group. Similar results were reported with patients with basal ganglia lesions that might have impaired attention-dependent temporal processing (Schwartze et al., 2011). Furthermore, with Parkinson's disease, temporal processing impairments were discussed in association with abnormalities of internal rhythm generation (Freeman et al., 1993) and motor planning impairments (Avanzino et al., 2013). Besides these reports, studies have shown time processing impairments in cases of Huntington's disease (Agostino et al., 2017), Alzheimer's disease (Roalf et al., 2018), mild cognitive impairment (Roalf et al., 2018), attention deficit hyperactivity disorder (Hove et al., 2017), and aphasia (Zipse et al., 2014).

Among sensorimotor and cognitive disorders, rhythm processing abilities in constructional apraxia have not been examined. Constructional apraxia is defined as a deficit in performing visuospatial constructional activities (Cubelli and Della Sala, 2018; Gainotti and Trojano, 2018) such as 2- or 3-dimensional copying or reproducing a drawing from memory and re-arranging patterns by blocks or sticks (Laeng, 2006; Russell et al., 2010); it is caused by cerebrovascular diseases such as stroke or brain damage on either hemisphere or neurodegenerative diseases such as Alzheimer's disease (Mack and Levine, 1981; Trojano et al., 2004; Laeng, 2006; Gainotti and Trojano, 2018). With stroke patients, lesion sites associated with constructional apraxia include the basal ganglia, thalamus, posterior parietal lobule, lingual gyrus, calcarine, insula, temporal gyrus, temporo-parietal junction (Chechlacz et al., 2014), parietal lobes, frontal lobes, and occipital lobes (Cubelli and Della Sala, 2018). Notably, various regions of the brain are involved in the drawing process. Therefore, constructional apraxia is related to a broad range of symptoms including: dysfunctions in visuospatial abilities such as the processing of shapes and the interrelations between different components of objects, perception, attentional allocation to global and local features, executive functions such as planning, and motor mechanisms (Chechlacz et al., 2014; Gainotti and Trojano, 2018).

Based on studies on lowered cognitive abilities with constructional apraxia (Laeng, 2006; Chechlacz et al., 2014; Nagaratnam et al., 2014; Gainotti and Trojano, 2018) and on the involvement of cognitive control such as general intelligence, working memory, and executive control on temporal reproduction (Ullén et al., 2008; Holm et al., 2017), it is likely that patients with constructional apraxia would show lowered temporal processing that requires cognitive control. It is worth examining the automatic temporal processing abilities of patients with constructional apraxia, including impairments in visuospatial perception, given the shared temporal and spatial performance and shared neural resources in sensorimotor synchronization (Doumas and Wing, 2007; Comstock et al., 2018), the common magnitude system in spatial lines and temporal duration representation (De Corte et al., 2017), the left-to-right ordering system (Bonato et al., 2016), and the temporal coding of visual spaces (Rucci et al., 2018).

The current study aims to examine the rhythm tapping ability of patients with constructional apraxia after a stroke. The performance of patients was examined during synchronization and continuation tapping tasks with sub-second stimulus intervals. If the patients demonstrated a lowered ability to synchronize with sub-second stimulus intervals, then their automatic timing process was regarded as lowered. If the patients' sub-second continuation tapping was less accurate than those without constructional apraxia, then a deficit in cognitive control on temporal reproduction was suggested. Based on previous studies (Laeng, 2006; Doumas and Wing, 2007; Ullén et al., 2008; Chechlacz et al., 2014; Nagaratnam et al., 2014; Bonato et al., 2016; De Corte et al., 2017; Holm et al., 2017; Comstock et al., 2018; Gainotti and Trojano, 2018; Rucci et al., 2018), we hypothesized that patients with constructional apraxia would perform less accurately with both sub-second synchronization and continuation tapping tasks than those without constructional apraxia.

## MATERIALS AND METHODS

In this retrospective study, clinical records of stroke patients admitted to a post-acute rehabilitation unit in Japan between November 2012 and February 2015 were queried for results of constructional apraxia tests and finger tapping tasks. The finger tapping tasks performed during this period were conducted to examine the ability of patients to synchronize to auditory stimulation. This study was approved by the ethical committee of Shimousa Hospital and conducted in accordance with the Declaration of Helsinki. The requirement of informed consent was waived. Instead, the patients were provided with the opportunity to opt out after posting the purpose and method of this research.

### Patients

There were 44 eligible patients who performed constructional apraxia tests and finger tapping tasks. Data were excluded from 20 patients according to the following exclusion criteria: a prior stroke episode, bilateral lesions, a strong influence of unilateral neglect on drawing, a disturbance of consciousness, and a failure to complete the assessments. Of the 44 patients, data were analyzed from 24 patients. These 24 patients were divided into two groups: with or without constructional apraxia. Eleven patients were allotted to the constructional apraxia group and 13 were assigned to the without constructional apraxia group. Characteristics of the patients are described in **Tables 1**, **2**. Lesion sites were diverse in both groups. Regarding the lesioned brain hemispheres, nine patients had damage on the right side and two had damage on the left side in the constructional apraxia group. In the without constructional apraxia group, five had damage on the right side and eight had damage on the left side. The proportion of affected dominant hands was determined by the proportion of lesioned right and left hemispheres. In the constructional apraxia group, two had affected dominant hands and nine had unaffected dominant hands. In the without constructional apraxia group, eight had affected dominant hands and five had unaffected dominant hands. The mean motor and cognition subscale values of the Functional Independence Measure (FIM) for the constructional apraxia group were 39.2 and 14.2, respectively, and the values for the without constructional apraxia group were 55.5 and 23, respectively. The FIM consists of 18 items and is grouped into motor and cognition subscales. The value of the total score for the motor subscale is between 13 and 91 and that for the cognition subscale is between 5 and 35 (Hamilton et al., 1994).

### Constructional Apraxia Test

To determine the presence of constructional apraxia, the results of a cube copying test or an intersecting pentagon copying test were used except for a patient who did not have either test result but had performed well on the Rey-Osterrieth complex figure test. This patient was included in the without-constructional apraxia group. To assess the cube copying test results, the scoring method developed by Yorimitsu et al. (2013) was employed that involves a checklist of inadequacies such as a lack of line or depth and a distortion in shape or proportion. All individuals who scored 6 or less out of a 10-point scale were included in the constructional apraxia group. To evaluate the intersecting pentagon copying test results, the scoring method developed by Nagaratnam et al. (2014) was utilized; this is a 10-point scoring method based on the degree of drawing failure. Participants with an intersecting pentagon copying score of 8 or less were included in the constructional apraxia group. Sample drawings from the constructional apraxia group are presented in **Figure 1**.

### Finger Rhythm Tapping

Participants completed a finger tapping task at least once during their hospitalization. The first dataset of individuals who completed the task twice was used. For this study, the synchronization-continuation paradigm was employed in which tapping is synchronized to a metronome beat and the patient continues tapping after the beat has stopped. The metronome sound of a Yamaha electronic keyboard EZ-J210, presented through a speaker, was used. The volume was set at a comfortable level for each participant. At first, patients listened to about 10 metronome beats and tapped in synchrony with the beats about 13 times. After the beat stopped, patients continued to tap at the same interval about 10 times. Patients repeated this procedure for three inter-stimulus intervals: 600 ms (100 metronome beats/min), 750 ms (80 metronome beats/min), and 1000 ms (60 metronome beats/min). There was no practice of the task prior to the assessment. Each condition was measured once without any repetition. In the finger tapping task, inter-stimulus intervals longer than 1000 ms are often used to examine cognitive involvement (Mangels et al., 1998; Miyake et al., 2004; Bååth et al., 2016). In this study, inter-stimulus intervals less than 1000 ms were chosen and the results of synchronization tasks were compared with those of continuation tasks. **Figure 2** shows the procedure.

### Data Acquisition

Recorded metronome beats were used for each condition. The participants wore a plastic finger pick on the index finger of their less affected side and tapped on a hard surface next to the touchpad on a laptop computer. The metronome beats and taps were recorded using audio editing software Sound it! 6 (Internet Co., Ltd., Osaka, Japan).

### Data Analysis

Tap onset and metronome beats were identified with a waveform display using the software. The first 3 synchronization taps were disregarded and 10 taps were used for both the synchronization

**TABLE 1 |** Description of the participants in the constructional apraxia group.

| Patient | Age range | Post-stroke day | Lesioned hemisphere | Dominant hand | Affected side | FIM motor score | FIM cognition score | Diagnosis |
|---|---|---|---|---|---|---|---|---|
| 1 | 66–70 | 53 | Right | Right | Left | 29 | 20 | Corona radiata infarction |
| 2 | 71–75 | 59 | Right | Right | Left | 33 | 11 | Fronto-temporal lobe infarction |
| 3 | 66–70 | 209 | Right | Right | Left | 51 | 16 | Internal capsule and corona radiata infarction |
| 4 | 76–80 | 146 | Right | Right | Left | 81 | 22 | Occipital lobe, thalamic infarction |
| 5 | 21–25 | 30 | Right | Right | Left | 25 | 12 | Putaminal hemorrhage |
| 6 | 61–65 | 36 | Right | Right | Left | 20 | 13 | Internal carotid artery territory infarction |
| 7 | 71–75 | 73 | Right | Right | Left | 60 | 15 | Occipital lobe and thalamic infarction |
| 8 | 71–75 | 35 | Right | Right | Left | 32 | 13 | Temporal lobe and corona radiata infarction |
| 9 | 81–85 | 58 | Left | Right | Right | 24 | 11 | Frontal lobe hemorrhage |
| 10 | 75–80 | 23 | Right | Right | Left | 35 | 17 | Thalamic hemorrhage |
| 11 | 75–80 | 54 | Left | Right | Right | 41 | 6 | Frontal subcortical infarction |
| Mean | 69.8 | 70.6 | | | | 39.2 | 14.2 | |
| Standard deviation | 15.9 | 56.6 | | | | 18.3 | 4.5 | |

*FIM motor, Functional Independence Measure motor subscale (a value between 13 and 91); FIM cognition, Functional Independence Measure cognition subscale (a value between 5 and 35).*

**TABLE 2 |** Description of the participants in the without construction apraxia group.

| Patient | Age range | Post-stroke day | Lesioned hemisphere | Dominant hand | Affected side | FIM motor score | FIM cognition score | Diagnosis |
|---|---|---|---|---|---|---|---|---|
| 1 | 66–70 | 173 | Right | Right | Left | 52 | 30 | Parietal lobe infarction |
| 2 | 71–75 | 81 | Right | Right | Left | 25 | 25 | Pontine and medullary infarction |
| 3 | 76–80 | 31 | Left | Right | Right | 83 | 31 | Parietal lobe infarction |
| 4 | 66–70 | 60 | Left | Right | Right | 31 | 11 | Corona radiata, parietal lobe and cerebellar infarction |
| 5 | 51–55 | 40 | Left | Right | Right | 59 | 14 | Subarachnoid hemorrhage, parietal-occipital lobe, and thalamic infarction, corpus callosum infarction |
| 6 | 71–75 | 84 | Right | Right | Left | 80 | 27 | Frontal subcortical and occipital lobe infarction |
| 7 | 66–70 | 69 | Right | Right | Left | 78 | 25 | Frontal subcortical infarction |
| 8 | 41–45 | 81 | Left | Right | Right | 88 | 28 | Frontal lobe hemorrhage |
| 9 | 46–50 | 23 | Right | Right | Left | 53 | 25 | Brainstem and cerebellar infarction |
| 10 | 81–85 | 25 | Left | Right | Right | 33 | 21 | Watershed infarction |
| 11 | 61–65 | 20 | Left | Right | Right | 46 | 23 | Corona radiata infarction |
| 12 | 56–60 | 45 | Left | Right | Right | 53 | 17 | Thalamic hemorrhage |
| 13 | 71–75 | 22 | Left | Right | Right | 40 | 22 | Putaminal hemorrhage |
| Mean | 65.5 | 58 | | | | 55.5 | 23 | |
| Standard deviation | 11.6 | 42.2 | | | | 21.1 | 6 | |

and continuation phases to calculate the interval reproduction accuracy index for each condition. The interval reproduction accuracy index is the ratio between the finger tapping interval reproduced by the person and the inter-stimulus interval set by the metronome. The method in the Avanzino et al. (2013) study served as a guide for the interval reproduction accuracy index (= the finger tapping interval reproduced by the person/the inter-stimulus interval set by the metronome). Briefly, when the tapping interval and inter-stimulus interval set by the metronome are equal, the interval reproduction accuracy index value equals 1. However, when the tapping interval is longer than the inter-stimulus interval, the index value is more than 1, and when it is shorter, the index value is less than 1.

## Statistics

Statistical analyses were performed with EZR (Saitama Medical Center, Jichi Medical University, version 1.35), a graphical user interface for R (The R Foundation for Statistical Computing,

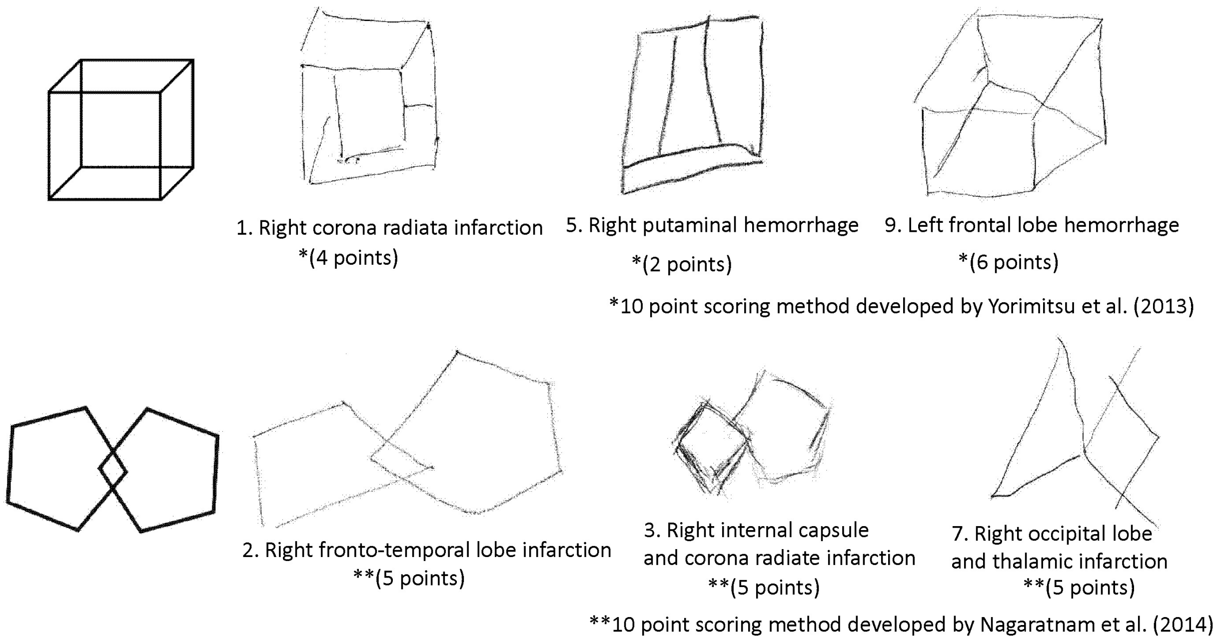



**FIGURE 1 |** Sample drawings of a cube and an intersecting pentagon from the constructional apraxia group.

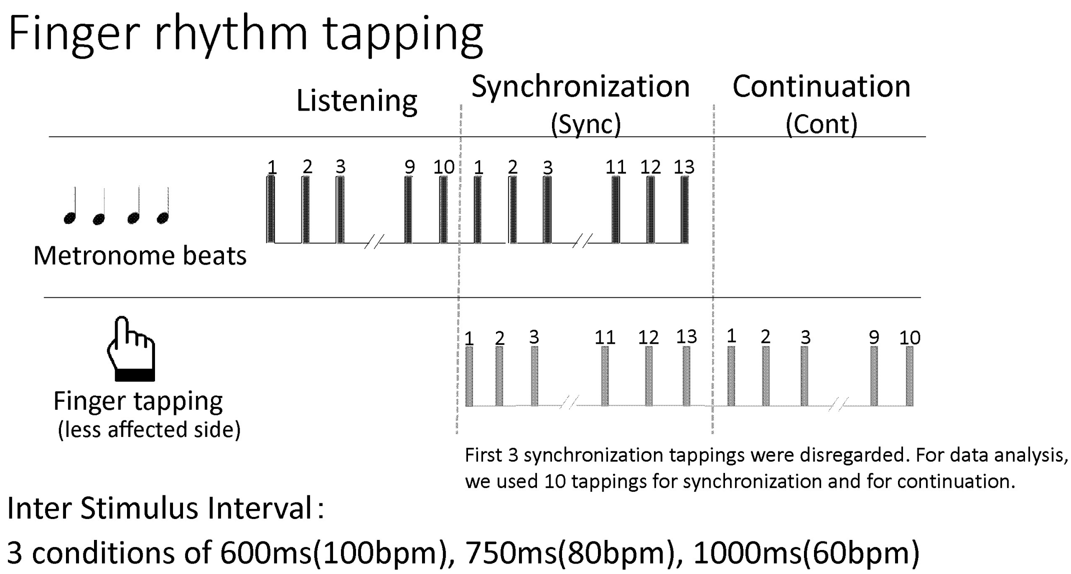



**FIGURE 2 |** The procedure for the finger rhythm tapping task. bpm, beats per minute.

**TABLE 3 |** The mean and standard deviation of the interval reproduction accuracy index.

| | Sync | | | Cont | | |
|---|---|---|---|---|---|---|
| **Interstimulus interval** | **600 ms** | **750 ms** | **1000 ms** | **600 ms** | **750 ms** | **1000 ms** |
| CA | 0.952 ± 0.063 | 0.982 ± 0.028 | 0.979 ± 0.041 | 0.947 ± 0.096 | 0.925 ± 0.132 | 0.881 ± 0.118 |
| w/o CA | 0.993 ± 0.020 | 0.997 ± 0.017 | 0.998 ± 0.015 | 0.988 ± 0.048 | 0.984 ± 0.054 | 0.982 ± 0.069 |

*CA, constructional apraxia; w/o CA, without constructional apraxia; Sync, synchronization; Cont, continuation.*

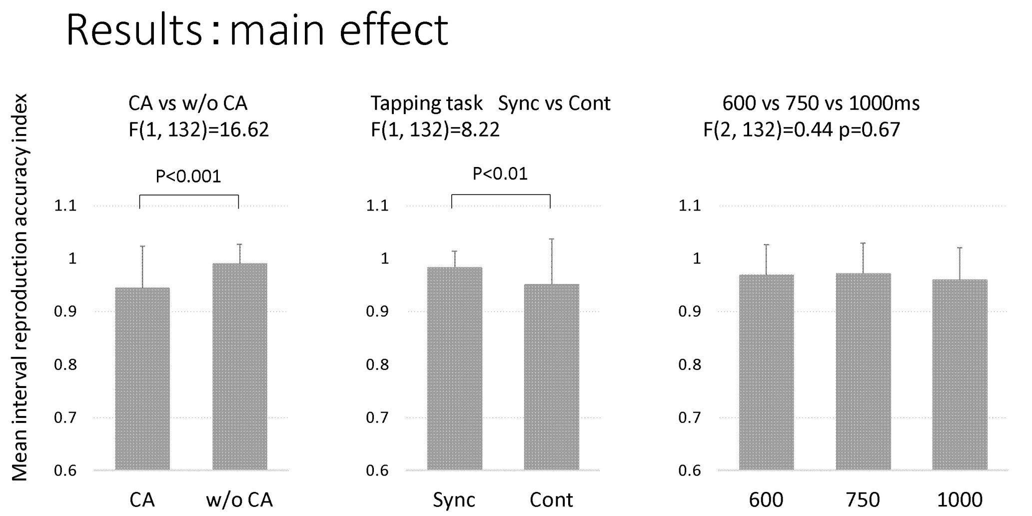



**FIGURE 3 |** The main effect of the factors. The factors included groups (CA, constructional apraxia; w/o CA, without constructional apraxia), tapping tasks (Sync, synchronization; Cont, continuation), and inter-stimulus intervals (600 ms, 750 ms, and 100 ms). Error bars indicate standard deviations.

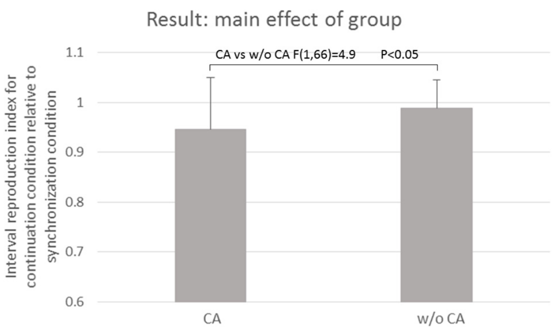



**FIGURE 4 |** The main effect of the group. CA, constructional apraxia; w/o CA, without constructional apraxia. Error bars indicate standard deviations.

version 3.3.2) (Kanda, 2013). A 3-way mixed analysis of variance (ANOVA) ($2 \times 2 \times 3$) was performed for the interval reproduction accuracy index. The factors were groups (with and without constructional apraxia), tapping tasks (synchronization and continuation), and inter-stimulus intervals (600, 750, and 1000 ms). To evaluate how well the continuation performance was maintained from the synchronization condition, a 2-way mixed ANOVA ($2 \times 3$) was performed for the interval reproduction index (the finger tapping interval reproduced by a participant for the continuation condition relative to the interval reproduced for the synchronization condition). When the tapping interval for the continuation and synchronization conditions are equal, the index value is 1. The factors were groups (with and without constructional apraxia) and inter-stimulus intervals (600, 750, and 1000 ms). Also, a *post hoc* statistical power analysis was conducted using the G*Power 3.1.9.4 software (Christian-Albrechts-Universität Kiel, Kiel, Germany) (Faul et al., 2007) with the effect size ($f$) = 0.25, the significance level ($\alpha$) = 0.05, and the power ($1-\beta$) = 0.8 for between-factor comparisons. For demographic differences between groups, both age and days from the onset were compared using a Welch's $t$-test and the proportion of lesioned hemispheres and dominant hands were examined using a Fisher's exact test. In addition, a 1-way analysis of covariance (ANCOVA) was conducted to determine statistically significant differences in the interval reproduction accuracy indexes controlling for lesioned hemispheres between groups (with and without constructional apraxia). Scores for FIM motor and cognition subscales were compared between groups using a Welch's $t$-test. $P$ values $< 0.05$ were considered statistically significant.

## RESULTS

**Table 3** and **Figure 3** show the results of the mean reproduction accuracy index; **Table 3** shows the mean and standard deviation for each condition while **Figure 3** displays the main effects. There was a significant effect of group ($F[1,132] = 16.62$; $p < 0.001$); the without constructional apraxia group was able to more accurately reproduce the intervals than the constructional apraxia group. As expected, tapping tasks had a significant effect ($F[1,132] = 8.22$; $p < 0.01$); intervals were reproduced more accurately for synchronization tasks than continuation tasks. There was no significant effect of inter-stimulus intervals. A 2-way ANOVA revealed a significant main effect of group ($F[1,66] = 4.9$; $p < 0.05$); the without constructional apraxia group was able to more accurately reproduce the interval for the continuation condition relative to the interval reproduced for the synchronization condition. In other words, the without constructional apraxia group was able to maintain the continuation performance from the synchronization condition. The results of the ANOVA analysis are shown in **Figure 4**. There was no significant main effect of inter-stimulus intervals or interaction effect. The *post hoc* statistical power analysis revealed a value of 0.30 with a sample size of 24 for this study. This power analysis also revealed that a sample size of 86 would be needed to detect a medium size effect (f = 0.25; cf. Cohen, 1977) with 0.80 power (1−β) at the 0.05 statistical significance level.

Concerning the demographic characteristics, there were no significant differences in the age or day from onset between the participants. There was a significant between-group difference in the proportion of the lesioned hemispheres. The proportion of right-side brain damage was higher in the constructional apraxia group than in the without constructional apraxia group ($p < 0.05$). Additionally, there was a significant between-group difference in the proportion of the affected dominant hands. The proportion of affected dominant hands was higher in the without constructional apraxia group than in the constructional apraxia group ($p < 0.05$). After controlling for lesioned hemispheres, the difference in interval reproduction accuracy between the groups was significant ($F[1,141] = 12.61$; $p < 0.001$). Also, there was a significant between-group difference in the FIM cognition subscale score ($p < 0.001$). However, there was no significant between-group difference in the FIM motor subscale score.

## DISCUSSION

Patients in the constructional apraxia group were less able to accurately reproduce sub-second intervals than those in the without constructional apraxia group. This result indicates the lowered automatic timing process of patients with constructional apraxia. Regarding cognitive involvement in temporal processing, our results support those of previous studies (Serrien, 2008; Ullén et al., 2008; Holm et al., 2017). The FIM cognition subscale score of the constructional apraxia group was significantly lower than the scores observed in the without constructional apraxia group. In the continuation condition, the constructional apraxia group displayed less accurate temporal reproduction compared to the without constructional apraxia group. These results raise the possibility that spatial and temporal reproduction abilities are related to a wide range of processing levels.

However, it is also possible that a distinctive line between automatic and cognitive processes does not exist, and that cognitive control is also involved in sub-second synchronization tapping. Bååth et al. (2016) stated that synchronization to sub-second intervals requires executive control, although its involvement is less than that observed with synchronization to longer intervals. Notably, the discrete neural resources involved in automatic and cognitive processes are unclear. Constructional apraxia is related to cognitive activities such as working memory, executive control, and general intelligence, as well as spatial perception. In this study, the constructional apraxia group had a significantly lower FIM cognition subscale score than those in the without constructional apraxia group. The involvement of cognitive control might be sufficient to explain why the patients in the constructional apraxia group were less able to accurately reproduce temporal intervals in both the synchronization and continuation tasks.

Another finding in this study is related to the effect of sub-second inter-stimulus intervals. Several intervals were employed (600, 750, and 1000 ms) for the synchronization and continuation paradigms, but there was no significant effect of the inter-stimulus interval. This null finding agrees with previous rhythm tapping studies (see Repp, 2005; Repp and Su, 2013, for a review) and suggests that there are common mechanisms across populations in terms of sub-second inter-stimulus interval differences.

A limitation of this study is the limited statistical power due to the small sample size ($n = 24$). The power analysis revealed that a sample size of 86 would be needed to detect a medium-sized effect with the recommended statistical power. Therefore, it is important to be cautious when interpreting the present results and further study with an increased sample size is required.

## CONCLUSION

This study shows that patients with constructional apraxia display a lowered ability to synchronize and reproduce temporal intervals. Given the lowered temporal and spatial reproduction abilities in patients with constructional apraxia, there might be a relationship between temporal and spatial reproductions in a wide spectrum of processing levels including those for sensory-perception and cognition.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethical committee of Shimousa Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

NKo collected data, performed the statistics, and wrote the manuscript. HY, YI, and NKa contributed with critical feedback to shape the research and assisted with the writing of the manuscript through extensive editing.

## ACKNOWLEDGMENTS

We thank A. Hamasaki, Y. Suganuma, N. Kunisawa, A. Komatsu, A. Yagyu, and M. Fushimi for their useful feedback on this manuscript. We also thank Shimousa Hospital for the support of this experiment. We would like to thank Editage (www.editage.com) for English language editing.

## REFERENCES


Agostino, P. V., Gatto, E. M., Cesarini, M., Etcheverry, J. L., Sanguinetti, A., and Golombek, D. A. (2017). Deficits in temporal processing correlate with clinical progression in Huntington's disease. *Acta Neurol. Scand.* 136, 322–329. doi: 10.1111/ane.12728

Avanzino, L., Pelosin, E., Martino, D., and Abbruzzese, G. (2013). Motor timing deficits in sequential movements in Parkinson disease are related to action planning: a motor imagery study. *PLoS One* 8:e75454. doi: 10.1371/journal.pone.0075454

Bååth, R., Tjøstheim, T. A., and Lingonblad, M. (2016). The role of executive control in rhythmic timing at different tempi. *Psychon. Bull. Rev.* 23, 1954–1960. doi: 10.3758/s13423-016-1070-1

Bonato, M., Saj, A., and Vuilleumier, P. (2016). Hemispatial neglect shows that "before" is "left". *Neural Plast.* 2016:2716036. doi: 10.1155/2016/2716036

Chechlacz, M., Novick, A., Rotshtein, P., Bickerton, W. L., Humphreys, G. W., and Demeyere, N. (2014). The neural substrates of drawing: a voxel-based morphometry analysis of constructional, hierarchical, and spatial representation deficits. *J. Cogn. Neurosci.* 26, 2701–2715. doi: 10.1162/jocn_a_00664

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences.* New York, NY: Academic Press.

Comstock, D. C., Hove, M. J., and Balasubramaniam, R. (2018). Sensorimotor synchronization with auditory and visual modalities: behavioral and neural differences. *Front. Comput. Neurosci.* 12:53. doi: 10.3389/fncom.2018.00053

Cubelli, R., and Della Sala, S. (2018). Constructional apraxia. *Cortex.* 104:127. doi: 10.1016/j.cortex.2018.02.015

De Corte, B. J., Navarro, V. M., and Wasserman, E. A. (2017). Non-cortical magnitude coding of space and time by pigeons. *Curr. Biol.* 27, R1264–R1265. doi: 10.1016/j.cub.2017.10.029

Doumas, M., and Wing, A. M. (2007). Timing and trajectory in rhythm production. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 442–455. doi: 10.1037/0096-1523.33.2.442

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Flach, R. (2005). The transition from synchronization to continuation tapping. *Hum. Mov. Sci.* 24, 465–483.

Freeman, J. S., Cody, F. W., and Schady, W. (1993). The influence of external timing cues upon the rhythm of voluntary movements in Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* 56, 1078–1084.

Gainotti, G., and Trojano, L. (2018). Constructional apraxia. *Handb. Clin. Neurol.* 151, 331–348. doi: 10.1016/B978-0-444-63622-5.00016-4

Hamilton, B. B., Laughlin, J. A., Fiedler, R. C., and Granger, C. V. (1994). Interrater reliability of the 7-level functional independence measure (FIM). *Scand. J. Rehabil. Med.* 26, 115–119.

Holm, L., Karampela, O., Ullén, F., and Madison, G. (2017). Executive control and working memory are involved in sub-second repetitive motor timing. *Exp. Brain Res.* 235, 787–798. doi: 10.1007/s00221-016-4839-6

Hove, M. J., Gravel, N., Spencer, R. M. C., and Valera, E. M. (2017). Finger tapping and pre-attentive sensorimotor timing in adults with ADHD. *Exp. Brain Res.* 235, 3663–3672. doi: 10.1007/s00221-017-5089-y

Kanda, Y. (2013). Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplant.* 48, 452–458.

Laeng, B. (2006). Constructional apraxia after left or right unilateral stroke. *Neuropsychologia.* 44, 1595–1606.

Lewis, P. A., and Miall, R. C. (2003). Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Curr. Opin. Neurobiol.* 13, 250–255.

Mack, J. L., and Levine, R. N. (1981). The basis of visual constructional disability in patients with unilateral cerebral lesions. *Cortex* 17, 515–531. doi: 10.1016/s0010-9452(81)80059-7

Mangels, J. A., Ivry, R. B., and Shimizu, N. (1998). Dissociable contributions of the prefrontal and neocerebellar cortex to time perception. *Brain Res. Cogn. Brain Res.* 7, 15–39.

McPherson, T., Berger, D., Alagapan, S., and Fröhlich, F. (2018). Intrinsic rhythmicity predicts synchronization-continuation entrainment performance. *Sci. Rep.* 8:11782. doi: 10.1038/s41598-018-29267-z

Miyake, Y., Onishi, Y., and Pöppel, E. (2004). Two types of anticipation in synchronization tapping. *Acta. Neurobiol. Exp. (Wars)* 64, 415–426.

Nagaratnam, N., Nagaratnam, K., and O'Mara, D. (2014). Intersecting pentagon copying and clock drawing test in mild and moderate Alzheimer's disease. *J. Clin. Gerontol. Geriatr.* 5, 47–52.

Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* 12, 969–992. doi: 10.3758/bf03206433

Repp, B. H., and Su, Y. H. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). *Psychon. Bull. Rev.* 20, 403–452.

Roalf, D. R., Rupert, P., Mechanic-Hamilton, D., Brennan, L., Duda, J. E., Weintraub, D., et al. (2018). Quantitative assessment of finger tapping characteristics in mild cognitive impairment, Alzheimer's disease, and Parkinson's disease. *J. Neurol.* 265, 1365–1375. doi: 10.1007/s00415-018-8841-8

Rucci, M., Ahissar, E., and Burr, D. (2018). Temporal coding of visual space. *Trends Cogn. Sci.* 22, 883–895. doi: 10.1016/j.tics.2018.07.009

Russell, C., Deidda, C., Malhotra, P., Crinion, J. T., Merola, S., and Husain, M. (2010). A deficit of spatial remapping in constructional apraxia after right-hemisphere stroke. *Brain.* 133, 1239–1251. doi: 10.1093/brain/awq052

Schwartze, M., Keller, P. E., and Kotz, S. A. (2016). Spontaneous, synchronized, and corrective timing behavior in cerebellar lesion patients. *Behav. Brain Res.* 312, 285–293. doi: 10.1016/j.bbr.2016.06.040

Schwartze, M., Keller, P. E., Patel, A. D., and Kotz, S. A. (2011). The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. *Behav. Brain Res.* 216, 685–691. doi: 10.1016/j.bbr.2010.09.015

Serrien, D. J. (2008). The neural dynamics of timed motor tasks: evidence from a synchronization-continuation paradigm. *Eur. J. Neurosci.* 27, 1553–1560. doi: 10.1111/j.1460-9568.2008.06110.x

Trojano, L., Fragassi, N. A., Chiacchio, L., Izzo, O., Izzo, G., Di Cesare, G., et al. (2004). Relationships between constructional and visuospatial abilities in normal subjects and in focal brain-damaged patients. *J. Clin. Exp. Neuropsychol.* 26, 1103–1112. doi: 10.1080/13803390490515522

Ullén, F., Forsman, L., Blom, O., Karabanov, A., and Madison, G. (2008). Intelligence and variability in a simple timing task share neural substrates in the prefrontal white matter. *J. Neurosci.* 28, 4238–4243. doi: 10.1523/JNEUROSCI.0825-08.2008

Yorimitsu, M., Tsukada, M., Watanabe, Y., and Yamada, R. (2013). Development of fixed quantitative scoring method of cube copying test: from the drawings by the inpatients of brain surgery in our hospital. *Higher Brain Funct. Res.* 33, 12–19.

Zipse, L., Worek, A., Guarino, A. J., and Shattuck-Hufnagel, S. (2014). Tapped out: do people with aphasia have rhythm processing deficits? *J. Speech Lang. Hear. Res.* 57, 2234–2245. doi: 10.1044/2014_JSLHR-L-13-0309


**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 16 April 2020
doi: 10.3389/fnhum.2020.00127

# Music Form but Not Music Experience Modulates Motor Cortical Activity in Response to Novel Music

*Patricia Izbicki*†, Andrew Zaman and Elizabeth L. Stegemöller†*

*Department of Kinesiology, Iowa State University, Ames, IA, United States*

External cues, such as music, improve movement performance in persons with Parkinson's disease. However, research examining the motor cortical mechanisms by which this occurs is lacking. Research using electroencephalography in healthy young adults has revealed that moving to music can modulate motor cortical activity. Moreover, motor cortical activity is further influenced by music experience. It remains unknown whether these effects extend to corticomotor excitability. Therefore, the primary aim of this study was to determine the effects of novel music on corticomotor excitability using transcranial magnetic stimulation (TMS) in a pilot study of healthy young adults. A secondary aim of this study was to determine the influence of music experience on corticomotor excitability. We hypothesized that corticomotor excitability will change during music conditions, and that it will differ in those with formal music training. Motor evoked potentials (MEPs) were recorded from the first dorsal interosseous using single-pulse TMS in three conditions: (1) No Music, (2) Music Condition I, and (3) Music Condition II. Both pieces were set to novel MIDI piano instrumentation and part-writing conventions typical of early nineteenth-century Western classical practices. Results revealed Music Condition II (i.e., more relaxing music) compared to rest increased MEP amplitude (i.e., corticomotor excitability). Music Condition II as compared to Music Condition I (i.e., more activating music) reduced MEP variability (i.e., corticomotor variability). Finally, years of formal music training did not significantly influence corticomotor excitability while listening to music. Overall, results revealed that unfamiliar music modulates motor cortical excitability but is dependent upon the form of music and possibly music preference. These results will be used to inform planned studies in healthy older adults and people with Parkinson's disease.

**Keywords: motor cortical excitability, music listening, music training, musicians and non-musicians, music experience**

OPEN ACCESS

***Edited by:***
*Cunmei Jiang,*
*Shanghai Normal University, China*

***Reviewed by:***
*Tong Yue,*
*Southwest University, China*
*Jiancheng Hou,*
*University of Wisconsin–Madison,*
*United States*

****Correspondence:***
*Patricia Izbicki*
*patinc91@gmail.com*

***†ORCID:***
*Patricia Izbicki*
*orcid.org/0000-0002-3269-2668*
*Elizabeth L. Stegemöller*
*orcid.org/0000-0002-5287-2295*

***Specialty section:***
*This article was submitted to Motor Neuroscience, a section of the journal Frontiers in Human Neuroscience*

***Received:*** *11 October 2019*
***Accepted:*** *19 March 2020*
***Published:*** *16 April 2020*

***Citation:***
*Izbicki P, Zaman A and Stegemöller EL (2020) Music Form but Not Music Experience Modulates Motor Cortical Activity in Response to Novel Music. Front. Hum. Neurosci. 14:127. doi: 10.3389/fnhum.2020.00127*

## INTRODUCTION

There is increased interest in the effects and efficacy of using music to improve movement in neurodegenerative disorders, specifically Parkinson's disease (PD). Dance, a combination of music and movement, has shown to improve mobility, gait, and postural instability in persons with PD (Hackney and Earhart, 2010; Foster et al., 2013; Houston and McGill, 2013; Volpe et al., 2013). Music listening and music therapy have been shown to improve motor performance in persons with

PD (Sihvonen et al., 2017). However, it is still unclear how music impacts motor cortical activity. An understanding of the basic mechanisms of how music affects motor cortical activity in healthy young adults provides the foundation for further examination of how music influences movement in healthy older adults and persons with PD.

The phenomenon of music eliciting movement is present in humans and other species, suggesting an evolutionarily conserved trait (Patel et al., 2009). Studies have indicated that listening to music globally activates the cerebral cortex (Menon and Levitin, 2005; Bengtsson et al., 2009). More specifically, motor regions, including the primary motor cortex, supplementary motor area, pre-motor cortex, and basal ganglia, are involved in listening to music (Popescu et al., 2004; Baumgartner et al., 2007; Chen et al., 2008; Bengtsson et al., 2009). Thus, music seems to elicit movement through the coupling of sensorimotor processes in the brain (Janata et al., 2012), suggesting that music may be a tool to modulate motor cortical excitability.

Behavioral studies have shown faster tempo, moderate syncopation, and repetitive rhythm elicit a greater urge to move (i.e., high groove) while slower tempo, excessive syncopation, and non-repetitive rhythm elicit little to no urge to move (Janata et al., 2012; Witek et al., 2014). This suggests that different forms of music may have differential effects on motor cortical activity. Furukawa et al. (2017) have shown that even listening to different piano tones increases somatotopic specific motor cortical excitability when compared to listening to noise in musicians. While a tone does not encompass the complexity of a musical excerpt or represent a change in musical form, this study (along with previous studies) supports the notion that musical form may modulate motor cortical excitability.

Music expertise has also been shown to influence motor cortical activity (Koeneke et al., 2006). Changes in motor cortical plasticity have occurred in both short- and long-term piano learning (Bangert and Altenmüller, 2003). Furthermore, music experience has been shown to play a role in modulating motor cortical activity in response to music. Individuals with previous formal music training have shown greater motor cortical activity as compared to non-musicians while listening to previously learned music (Haueisen and Knösche, 2001). Listening to different piano tones demonstrated increased somatotopic specific motor cortical excitability in musicians but not non-musicians (Furukawa et al., 2017). A recent study using transcranial magnetic stimulation (TMS) while listening to familiar music has also shown that music modulates corticomotor excitability in both musicians and non-musicians (Stupacher et al., 2013). Thus, music listening modulates corticospinal excitability differently between musicians and non-musicians. However, these previous studies used familiar music or learned music. D'Ausilio et al. (2006) have been the only group (to our knowledge) to show that there is increased motor cortical excitability for non-rehearsed or "previously unheard" music in amateur piano players. How changes in motor cortical excitability differ between musicians and non-musicians while listening to novel music remains limited.

A meta-analysis using the activation likelihood estimation approach found that music familiarity increased audio-motor synchronization to rhythm in familiar music vs. unfamiliar music (Freitas et al., 2018). In addition, a recent study has been conducted examining motor cortical activity in response to previously novel music using electroencephalography (EEG). Results found differential responses to musical form over the sensorimotor cortex that was further influenced by music experience, which may be reflective of a decrease in movement variability (Stegemöller et al., 2018a). However, EEG cannot determine specific neuronal activity (i.e., excitability). TMS is a technique that can determine more specific neuronal activity and variability in the motor cortex via motor evoked potential (MEP) amplitude and MEP variability. Previous research has shown that MEP amplitude is inversely related to MEP variability (Kiers et al., 1993; Devanne et al., 1997; Darling et al., 2006) Furthermore, exposure to sensory stimuli (e.g., visual, auditory, olfactory) have been shown to modulate MEP amplitude and/or variability (Furubayashi et al., 2000; Carson et al., 2005; Rossi et al., 2008).

Thus, the aim of this study was to determine the effects of listening to two novel musical pieces on motor cortical excitability of the hand area in the primary motor cortex using TMS. We hypothesized that both pieces will increase motor cortical excitability of the hand area, as measured by motor evoked potential. A second aim of this study was to determine the influence of previous music experience on motor cortical excitability. We hypothesized that motor cortical excitability of the hand area will be different for musicians than non-musicians.

# MATERIALS AND METHODS

## Participants

Twenty healthy young adults were recruited (11 women, mean age ± standard deviation age = 21 ± 2.03). See **Table 1** for detailed demographic information. All participants provided written informed consent to participate in the study as approved by the university Institutional Review Board. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institution and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Inclusion criteria included only healthy young adults between ages 18–40. Exclusion criteria included significant cognitive impairment (Mini Mental State Exam (MMSE) <24) and/or major depression (Beck Depression Inventory (BDI) >18). Exclusion criteria for TMS included any previous adverse reactions to TMS, previous seizure, surgery on blood vessels, brain, or heart, previous stroke, severe vision or hearing loss, metal in head, implanted devices, severe headaches, previous brain-related conditions, brain injury, medications (i.e., antibiotics, antifungal, antiviral, antidepressants, antipsychotics, chemotherapy, amphetamines, bronchodilators, anticholinergics, antihistamines, sympathomimetics), family history of epilepsy, pregnancy, alcohol consumption less than 24 h before study, smoking, and illicit drug use.

**TABLE 1** | Participant demographics and music experience.

| Demographics | Musician | Non-musician |
|---|---|---|
| Age (Mean ± SD) | 21 ± 1.4 | 22 ± 2.6 |
| Gender (%Male, %Female) | 50%, 50% | 40%, 60% |
| Ethnicity (%Caucasian, %Hispanic, %Asian | 80%, 20%, 0% | 90%, 0%, 10% |
| Handedness (%RH, %LH) | 90%, 10% | 100% |
| Music training (Years) | 9.8 ± 4.1 | 1.0 ± 1.8 |
| Instrument (%) | 10% Clarinet, 50% Piano, 10% Trumpet, 10% Viola, 20% Voice | 10% Bass, 10% Saxophone, 10% Voice, 70% NA |

*SD, standard deviation; RH, right hand; LH, left hand.*

## Participant Music Experience

Prior to TMS data collection, all participants orally provided information about their previous music experience. The researchers asked participants to provide total years of formal music training and instrument played. Formal music training was defined as private music lessons on an instrument or voice. Participants were classified as musicians (≥5 years experience, $n$ = 10, mean ± standard deviation = 9.8 ± 4.1 years) or non-musicians (<5 years experience, $n$ = 10, mean ± standard deviation = 1.0 ± 1.7 years). See **Table 1** for detailed information on music experience.

Five years of formal music training was chosen as a cutoff because healthy young adult participants who had more than 5 years of music experience received advanced training (i.e., late middle school, high school, and collegiate level). Furthermore, other studies in children and adult musicians have characterized music experience groups using the number of years of music training (Wong et al., 2007; Hanna-Pladdy and MacKay, 2011; Stegemöller et al., 2018a; Strong and Mast, 2018).

## Music

The music was specifically commissioned for the study by an (Iowa State University) music composition student in order to control for previous experience or familiarization with the music. Both pieces were set to novel MIDI piano instrumentation and part-writing conventions typical of early nineteenth-century Western classical practices. Music Condition I was set in the key of C major in ternary form (ABA'), 4/4 meter, and a quarter note pulse of 140 beats per minute (BPM). Music Condition II was set in the key of G-flat major in through-composed form, 3/4 meter, and a quarter note pulse of 70 beats per minute. The piece contained greater tonal and rhythmic variations than Music Condition I. These are the same pieces that were used in a previous EEG study (see **Supplementary Material**) (Stegemöller et al., 2018b). Participants were asked their preference for the two forms of music based on a Likert scale of 1–10. 1 indicated the participants extremely disliked the form and 10 indicated the participants extremely liked the music.

## Data Collection

For TMS, the motor hot spot, specifically the hand knob area in the primary motor cortex (M1), was located on the contralateral hemisphere (left hemisphere; all participants were right-handed). The location and coil orientation (45 degrees to the left of the longitudinal fissure) was marked, and the coil was held in a constant position by the experimenter with the aid of a coil holder. Resting motor threshold (RMT) (i.e., a MEP at an amplitude of at least 50 μV produced for 5 out of 10 trials or 50% of the time) was then found. RMT was completed in 30 min. Single-pulse TMS intensity was set at 120% of RMT.

Participants were seated in an armchair with their right forearm pronated and rested on the armrest. Participants were asked not to move during TMS. Single-pulse TMS was applied to the M1 dominant hand area using the Magstim Model 200 (Magstim, Whitland, Carmarthenshire). The coil was figure-8 coil (7 cm outer diameter of wings). Coil current was induced approximately perpendicular to the motor homunculus and central sulcus. The waveform was monophasic. Spike2 was used to trigger single-pulse stimulations via a Power 1401 data acquisition board and Spike2 software (Cambridge Electronic Design (CED), Cambridge, United Kingdom). Motor evoked potentials (MEPs) were recorded from the right first dorsal interosseous (FDI) using bipolar surface electromyography (EMG) (Delsys, Boston, MA, United States). Twenty single-pulse stimulations were applied during rest (no music) and while passively and continuously listening to two different music selections. The total number of pulses applied across all conditions was 60. There was 5 min of rest (no TMS) between each stimulation condition. Single-pulses were applied approximately every 5 s (for a total of 1.7 min of stimulation in each condition) and were not specifically timed to the beat of the music. Each non-music and music condition lasted 5 min. Along with the informed consent process, the entire experiment lasted around an hour. The order of the music selections was randomized between participants, and TMS was applied during random sections of each music selection.

## Data Analysis

EMG signals were notch filtered (60 Hz) and high-pass filtered (2nd-order dual-pass Butterworth, 2 Hz cut-off). EMG signals were also DC shifted, and the root mean square of the EMG signal was obtained. Peak-to-peak amplitude (μV) was obtained within 100 ms of the TMS pulse. Background EMG was determined for periods of 1.25–0.25s before the peak maximum amplitude and 0.25–1.25s after the peak maximum amplitude. Background EMG trials > 10 μV were discarded (Majid et al., 2015). For EMG activity before peak amplitude, the number of trials discarded were 8 trials in the rest condition, 0 trials in the Music Condition I, and 15 trials in the Music Condition II. For EMG activity after peak amplitude, the number of trials discarded were 5 trials in the rest condition, 1 trial in the Music Condition I, and 15 trials in the Music Condition II. The raw data for each participant in the background EMG activity and for each condition was natural log transformed to obtain a normal distribution. The primary outcome measure of MEP amplitude was obtained by averaging the natural log transformed MEP trials for each condition (i.e., No Music, Music Condition I, and Music Condition II) in the stimulation parameter (i.e., single-pulse) (Nielsen, 1996; Clark et al., 2004). Coefficient of variation (CV) (standard deviation divided by average) was calculated for each participant in each condition. CV was used as the MEP variability measure (Klein-Flügge et al., 2013).

## Statistical Analysis

Statistical analysis was completed in IBM SPSS Statistics for Windows, Version 25.0 (IBM Corp., Armonk, NY, United States). Normality was assessed using the Shapiro-Wilk test. Analyses were completed to determine if there was any potential influence of music preference and background EMG activity on the main outcome measures. Due to the non-normality of the music preference data, a Wilcoxon signed rank test was used to compare whether there was any overall difference in preference to Music Condition I vs. Music Condition II overall. The Mann-Whitney U test was used to compare whether there were any differences in music preference in musicians vs. non-musicians. Due to the non-normality of the background EMG activity pre- and post- MEP, the Friedman test was conducted to determine differences in EMG background activity among all conditions for both pre-MEP EMG activity and post-MEP EMG activity. To examine differences in peak maximum amplitude of the MEP between the three music conditions, a (two-way) mixed ANOVA was completed. The within factor was music condition (Rest, Music Condition I, and Music Condition II) while the between factor was musician or non-musician. To examine the influence of musical form and music training on MEP amplitude and variability (CV), a (two-way) mixed ANOVA was completed. The within factor was music condition (Rest, Music Condition I, Music Condition II) while the between factor was musician or non-musician. Bonferroni correction was used for *post- hoc* analysis. Significance was set at $\alpha = 0.05$.

# RESULTS

## Music Preference

The Wilcoxon signed-rank test showed that participant ratings for Music Condition II were significantly larger than for Music Condition I ($Z = -2.68$, $p = 0.007$) (mean ± standard deviation: Music Condition I = 5.95 ± 1.32; Music Condition II = 6.90 ± 1.48) (**Figure 1A** and **Table 2**). However, the Mann-Whitney U test showed no significant differences in music preference between musicians and non-musicians for Music Condition I ($U = 43.5$, $p = 0.612$) (mean ± standard deviation: musicians Music Condition I = 6.10 ± 0.994; non-musicians Music Condition I = 5.8 ± 1.62) and Music Condition II ($U = 43.0$, $p = 0.577$) (mean ± standard deviation: musicians Music Condition II = 7.10 ± 1.29; non-musicians Music Condition II = 6.70 ± 1.70) (**Figure 1B** and **Table 2**).

## Background EMG

To confirm that potential differences in MEP amplitude are due to cortical mechanisms rather than an increase in drive to spinal mechanisms, a Friedman test was conducted to compare 1.25 to 0.25s before the peak maximum amplitude among the three conditions as well as 0.25 to 1.25s after the peak maximum amplitude among the three conditions. Results revealed no differences in EMG activity before [$\chi^{2(2)} = 4.80$, $p = 0.091$] or after [$\chi^{2(2)} = 3.90$, $p = 0.142$] peak maximum amplitude (**Figures 2A,B**).

## MEP Amplitude

There was a significant main effect of condition ($F_{(2,36)} = 3.51$, $p = 0.04$), but no significant main effect of group ($F_{(1,18)} = 1.65$, $p = 0.22$). There was no significant interaction effect ($F_{(2,36)} = 3.15$, $p = 0.05$). *Post hoc* tests using Bonferroni correction for the main effect of condition ($p < 0.017$) revealed that MEP amplitude did not differ for Music Condition I compared to rest (4.73 ± 0.51 vs. 4.66 ± 0.39 uV) ($p = 1.00$) or for Music Condition I compared to Music Condition II (4.73 ± 0.51 vs. 4.92 ± 0.54 uV) ($p = 0.06$). Music Condition II compared to rest revealed a significant increase in MEP amplitude (4.92 ± 0.54 vs. 4.66 ± 0.39 uV) ($p = 0.017$). *Post hoc* tests using Bonferroni correction for the interaction effect ($p < 0.005$) are listed in **Table 3**. Results revealed no significant differences in musicians for Music Condition I compared to rest, Music Condition I compared to Music Condition II, and Music Condition II compared to rest. Results revealed no significant differences in non-musicians for Music Condition I compared to rest, Music Condition I compared to Music Condition II, and Music Condition II compared to rest. Results revealed no significant differences in musicians and non-musicians for Rest, Music Condition I, or Music Condition II (**Figure 3**).

## MEP Variability

For MEP amplitude CV, results revealed a significant main effect of condition ($F_{(2,36)} = 4.38$, $p = 0.02$), but no significant main effect of group ($F_{(1,18)} = 1.53$, $p = 0.23$). There was no significant interaction effect ($F_{(2,36)} = 1.79$, $p = 0.18$). *Post hoc* tests using Bonferroni correction for the main effect ($p < 0.017$) revealed that MEP amplitude CV did not differ for the Music Condition I compared to rest (0.13 ± 0.057 vs. 0.13 ± 0.068 uV) ($p = 1.00$). Music Condition I compared to Music Condition II revealed a significant increase in MEP amplitude CV (0.13 ± 0.057 vs. 0.10 ± 0.051 uV) ($p = 0.05$). Music Condition II compared to rest did not reveal a significant difference in MEP amplitude CV (0.10 ± 0.051 vs. 0.13 ± 0.068 uV) ($p = 0.06$) (**Figure 4**).

# DISCUSSION

The primary purpose of this study was to determine the effects of listening to two different forms of novel music on motor cortical excitability in the primary motor cortex using TMS. The secondary purpose of this study was to determine the influence of previous music experience on motor cortical excitability. We hypothesized that (1) motor cortical excitability of the hand area, as measured by motor evoked potential (MEP) amplitude, will differ between musical forms and (2) that both forms of music would increase MEP amplitude. Our findings partially support this hypothesis, revealing a main effect of condition. However, only Music Condition II differed from the rest condition. There was no difference between the two music conditions. For variability, a main effect of condition was also revealed,

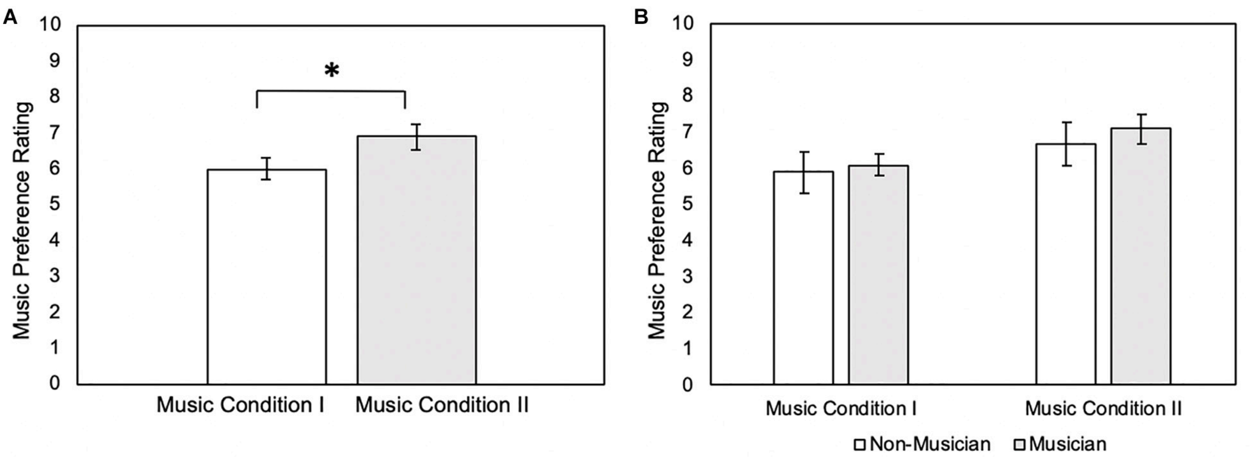



**FIGURE 1 | (A)** Music preference rating of Music Condition I and Music Condition II ($N = 20$). Error bars reflect standard error of mean music preference ratings. $^{*}p = 0.007$. **(B)** Music preference rating of Music Condition I and Music Condition II between non-musicians ($N = 10$) and musicians ($N = 10$). Error bars reflect standard error of mean music preference ratings.

with *post hoc* analyses demonstrating a difference between the two music conditions. For our hypothesis regarding music experience, our findings did not support our hypothesis. No differences were revealed between musicians and non-musicians. To our knowledge, results from this study are the first to show that novel, preferred music selections may have a different motor cortical influence as compared to previous studies using familiar or learned music.

## MEP Amplitude

An interesting finding of this study was an increase in MEP amplitude for Music Condition II as compared to rest. Although participants in this study were not specifically asked if they perceived the music used in the study as relaxing or activating, music in Music Condition I was composed to evoke more of an activated feeling while music in Music Condition II was initially composed to evoke more of a relaxed feeling. Thus, the finding that Music Condition II resulted in an increase in motor cortical excitability seems contradictory to previous literature. Faster tempo, moderate syncopation, and repetitive rhythm have been shown to elicit a greater urge to move (i.e., high groove) while slower tempo, excessive syncopation, and non-repetitive rhythm elicit little to no urge to move (Janata et al., 2012). However, D'Ausilio et al. (2006) showed that there is increased motor cortical excitability for non-rehearsed or previously unheard music in amateur piano players. Additionally, Weigmann demonstrated that less predictable music (i.e., slightly more complex) generated more prediction errors which was reflected as greater pleasure and a greater urge to move. However, the rhythm must still be simple enough and not too complex to see this effect (Weigmann, 2017). This may be reflected in Music Condition II. Thus, the change in motor cortical excitability may be due to unfamiliarity as well as the wider range of rhythmic and harmonic variations found in Music Condition II regardless of the intended perception. This would suggest that future studies examining the effect of musical form on motor cortical activity should consider both participant perception and music composition.

**TABLE 2 |** Means and standard deviations for music preference.

| Condition | Total | Musicians | Non-musicians |
|---|---|---|---|
| Music Condition I | 5.95 ± 1.32 | 6.10 ± 0.994 | 5.8 ± 1.62 |
| Music Condition II | 6.90 ± 1.48 | 7.10 ± 1.29 | 6.70 ± 1.70 |

Another consideration that may have influenced a difference in MEP amplitude for Music Condition II as compared to rest is the higher preference for the Music Condition II music than the Music Condition I in our sample. Listening to pleasurable music stimulates areas of the brain responsible for dopamine production (i.e., nucleus accumbens and ventral tegmental area) in both humans (Menon and Levitin, 2005) and rats (Moraes et al., 2018). These changes in dopamine have been implicated in modulating motor cortical activity (Ziemann et al., 1997; Jenkinson and Brown, 2011) as well as motor cortical plasticity (Calabresi et al., 2007; Molina-Luna et al., 2009). Thus, an increase in preference for Music Condition II may have increased dopamine production, which may modulate motor cortical activity resulting in increased motor cortical excitability of the hand area. However, no measures of dopamine were taken in this study leaving room for continued research to determine the relationship between preferred music, dopamine, and motor cortical activity.

## MEP Variability

An additional finding of this study revealed a decrease in the variability of motor cortical excitability while listening to Music Condition II compared to Music Condition I. This could be due to neural synchrony in motor cortical excitability. An increase in neural synchronization has been shown in individuals listening to music (Bernardi et al., 2017). This decrease in variability

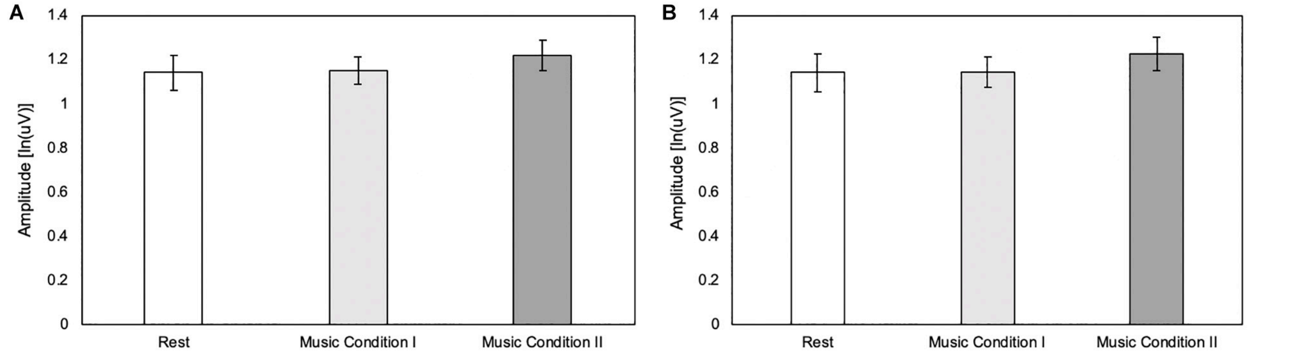



**FIGURE 2 | (A)** Pre-stimulation background EMG 1.25–0.25s before the peak maximum amplitude of the MEP (*N* = 20). Error bars reflect standard error of mean EMG activity. **(B)** Post-stimulation background EMG 0.25–1.25s after the peak maximum amplitude of the MEP (*N* = 20). Error bars reflect standard error of mean EMG activity.

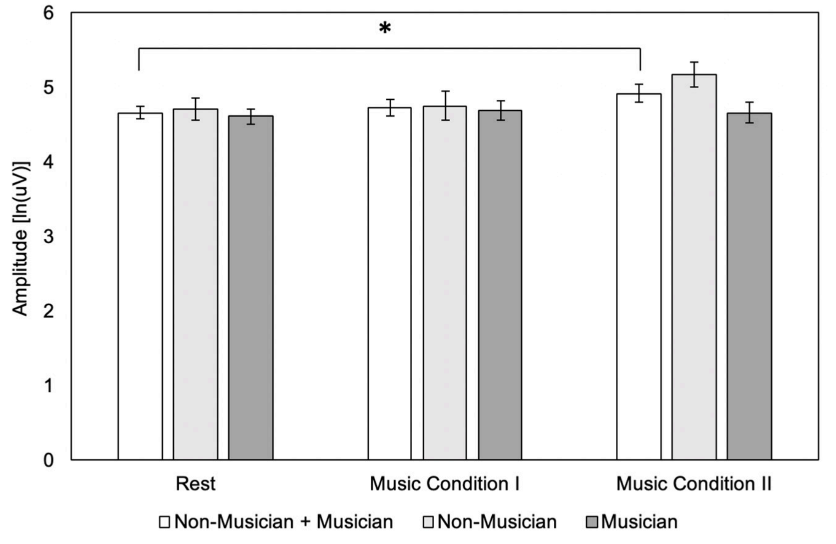



**FIGURE 3 |** Motor evoked potential (MEP) amplitude between conditions (*N* = 20) and groups (*N* = 10) using the single-pulse paradigm (*N* = 20). Error bars reflect standard error. **p* = 0.05.

may transfer to movement performance. In a previous study from our lab, results revealed that repetitive finger movement variability significantly decreased while moving in time with music (Stegemöller et al., 2018a,b). The same two music samples as used in this study were used in this previous study. While the same participants were not tested, perhaps the decrease in MEP variability transfers to a decrease in movement variability. Future studies using TMS while moving with music are needed to confirm this notion. Nonetheless, this study provides continued evidence suggesting that music decreases variability of the motor system.

## MEP Amplitude and Variability in Musicians vs. Non-musicians

The final result of this study revealed no differences between musicians and non-musicians across all conditions. These results are in contrast to previous studies. Other studies have indicated increases in motor cortical excitability in musicians vs. non-musicians without listening to music (Rosenkranz et al., 2007) and while listening to high-groove music (Stupacher et al., 2013). However, there were no novel musical stimuli composed for each of the studies. A recent meta-analysis found familiar music elicited a greater motor pattern of activation as compared to unfamiliar music. Specifically, the ventral lateral nucleus (a motor first-order relay nucleus responsible for receiving input from substantia nigra, internal globus pallidus, and cerebellum) had the second highest likelihood for activation while listening to familiar music (Freitas et al., 2018). Furthermore, greater motor cortical activation for familiar music compared to unfamiliar music has been found in musicians (D'Ausilio et al., 2006). Thus, unfamiliarity with the musical stimuli in our study may have influenced the lack of difference in motor cortical excitability between musicians and non-musicians.

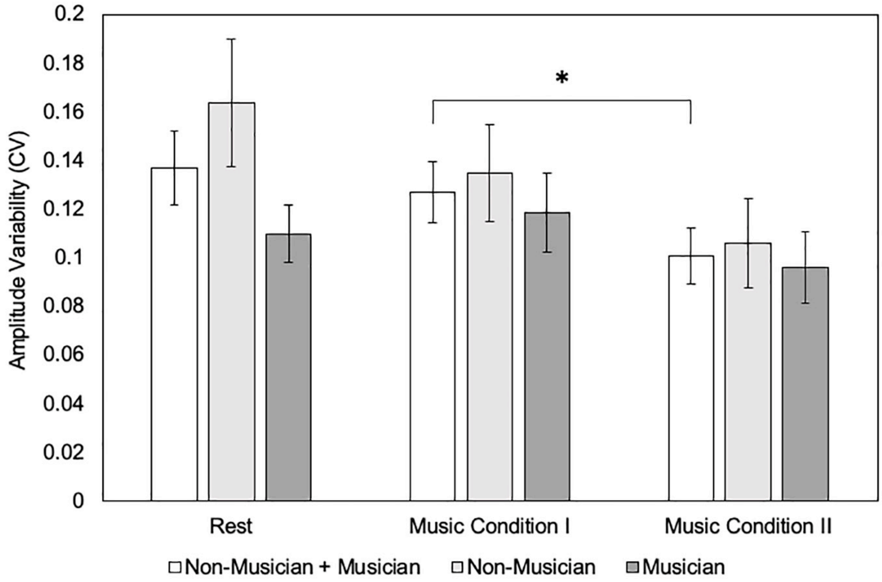



**FIGURE 4 |** Motor evoked potential (MEP) amplitude variability (coefficient of variation = CV) between conditions (*N* = 20) and groups (*N* = 10) using the single-pulse paradigm. Error bars reflect standard error. $^{*}p = 0.05$.

**TABLE 3 |** *Post hoc* tests using bonferroni correction for the interaction effect ($p < 0.005$).

| Comparison | Mean ± Standard Deviation | *P*-value |
|---|---|---|
| Music Condition I vs. Rest (Musicians Only) | 4.70 ± 0.41 vs. 4.62 ± 0.31 uV | 0.38 |
| Music Condition I vs. Music Condition II (Musicians Only) | 4.70 ± 0.41 vs. 4.66 ± 0.45 uV | 0.46 |
| Music Condition II vs. Rest (Musicians Only) | 4.66 ± 0.45 vs. 4.62 ± 0.31 uV | 0.57 |
| Music Condition I vs. Rest (Non-Musicians Only) | 4.76 ± 0.61 vs. 4.71 ± 0.46 uV | 0.83 |
| Music Condition I vs. Music Condition II (Non-Musicians Only) | 4.76 ± 0.61 vs. 5.18 ± 0.52 uV | 0.05 |
| Music Condition II vs. Rest (Non-Musicians Only) | 5.18 ± 0.52 vs. 4.71 ± 0.46 uV | 0.03 |
| Musicians vs. Non-Musicians (Rest) | 4.62 ± 0.31 vs. 4.71 ± 0.46 uV | 0.50 |
| Musicians vs. Non-Musicians (Music Condition I) | 4.70 ± 0.41 vs. 4.76 ± 0.61 uV | 0.79 |
| Musicians vs. Non-Musicians (Music Condition II) | 4.66 ± 0.45 vs. 5.18 ± 0.52 uV | 0.03 |

In short, our study is in keeping with previous literature on the neural basis of music familiarity. It may be that motor cortical differences are not dependent on musician/non-musician status but due to previous experience with a musical piece. This suggests that engagement with previously heard music may be beneficial for altering motor cortical activity. This has implications toward PD and music therapy, where people receiving music therapy are likely not musicians.

## Parkinson's Disease and Motor Cortical Activity

The findings from our study have important implications for using music therapy and music and medicine interventions in persons with PD. Differences in beta band oscillations in the motor cortex have been shown in previous literature in persons with PD (Brown, 2007; Stegemöller et al., 2016, 2017). This indicates that motor cortical activity in persons with PD is different than in healthy older adults. In studies of motor cortical activity using TMS, drug-naïve patients have been shown to have increased MEPs at rest (Derejko et al., 2013). Although, music listening and music therapy have been shown to improve motor performance in persons with PD (Sihvonen et al., 2017), results of this study suggest that increasing motor cortical activity using certain music conditions may not necessarily be beneficial. On the other hand, decreasing the variability in the motor system with music, as demonstrated in this study, may be beneficial for persons with PD. Thus, as research on the underlying mechanisms of music therapy continues to grow, a clear understanding of music impacts the motor system in neurological populations is needed. This study provides continued information in understanding the impact of music on motor cortical excitability.

## Limitations

A limitation of this study was that there was no survey for perception of music (i.e., whether the music was relaxing, activating, and/or emotionally stimulating). However, the music used for each condition was distinctly different and represents two contrasting forms of music regardless of the form (relaxing or activating) perceived. In addition, TMS was applied at rest and not during movement. Given the tempi of the music, movements would have been completed at either 70 or 140 beats per minute. Completing repetitive finger movements at these rates while applying TMS can be done, but also increases the potential error in obtaining MEPs due to underlying muscle activity. Thus, applying TMS at rest in the various conditions was the initial first step in understanding how music influences motor cortical activity.

## CONCLUSION

In conclusion, results revealed that unfamiliar music modulates motor cortical excitability, but is dependent upon the form of music and possibly music preference. In addition, the form of music has a differing effect on motor cortical variability. However, there are no differences in motor cortical excitability between musicians and non-musicians when listening to unfamiliar music. These results suggest that music could be used to influence excitatory activity in the primary motor cortex and potentially reduce variability of the motor system regardless if a person is a musician or non-musician. This has implications toward PD and music therapy, where people receiving music therapy are likely not musicians. An understanding of the basic mechanisms of how music affects motor cortical activity in healthy young adults is needed to provide the foundation for further examination of how music influences movement in healthy older adults and persons with PD. Future studies will involve a similar paradigm with healthy older adults and people with Parkinson's disease to further elucidate the influence of music on motor cortical activity in these populations.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Iowa State University Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

PI and ES made substantial contributions to the concept and design of the study, acquisition of the data, analysis and interpretation of the data, and drafting and revising the article. AZ made contributions to analysis and interpretation of the data. All authors gave final approval of the version to be submitted.

## FUNDING

This work was supported by the Research Enhancement Grant from the College of Human Sciences, Iowa State University.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00127/full#supplementary-material

## REFERENCES


Bangert, M., and Altenmüller, E. O., (2003). Mapping perception to action in piano practice: a longitudinal DC-EEG study. *BMC Neurosci.* 4:26. doi: 10.1186/1471-2202-4-26

Baumgartner, T., Willi, M., and Jäncke, L., (2007). Modulation of corticospinal activity by strong emotions evoked by pictures and classical music: a transcranial magnetic stimulation study. *Neuroreport* 18, 261–265. doi: 10.1097/wnr.0b013e328012272e

Bengtsson, S. L., Ullén, F., Ehrsson, H. H., Hashimoto, T., Kito, T., Naito, E., et al. (2009). Listening to rhythms activates motor and premotor cortices. *Cortex* 45, 62–71. doi: 10.1016/j.cortex.2008.07.002

Bernardi, N. F., Codrons, E., di Leo, R., Vandoni, M., Cavallaro, F., Vita, G., et al. (2017). Increase in synchronization of autonomic rhythms between individuals when listening to music. *Front. Physiol.* 8:785. doi: 10.3389/fphys.2017.00785

Brown, P. (2007). Abnormal oscillatory synchronisation in the motor system leads to impaired movement. *Curr. Opin. Neurobiol.* 17, 656–664. doi: 10.1016/j.conb.2007.12.001

Calabresi, P., Picconi, B., Tozzi, A., Di Filippo, M. (2007). Dopamine-mediated regulation of corticostriatal synaptic plasticity. *Trends Neurosci.* 30, 211–219. doi: 10.1016/j.tins.2007.03.001

Carson, R. G., Welsh, T. N., and Pamblanco-Valero, M. A. (2005). Visual feedback alters the variations in corticospinal excitability that arise from rhythmic movements of the opposite limb. *Exp. Brain Res.* 161, 325–334. doi: 10.1007/s00221-004-2076-x

Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cereb. Cortex* 18, 2844–2854. doi: 10.1093/cercor/bhn042

Clark, S., Tremblay, F., and Ste-Marie, D. (2004). Differential modulation of corticospinal excitability during observation, mental imagery and imitation of hand actions. *Neuropsychologia* 42, 105–112. doi: 10.1016/s0028-3932(03)00144-1

D'Ausilio, A., Altenmüller, E., Olivetti Belardinelli, M., and Lotze, M. (2006). Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *Eur. J. Neurosci.* 24, 955–958. doi: 10.1111/j.1460-9568.2006.04960.x

Darling, W. G., Wolf, S. L., and Butler, A. J. (2006). Variability of motor potentials evoked by transcranial magnetic stimulation depends on muscle activation. *Exp. Brain Res.* 174, 376–385. doi: 10.1007/s00221-006-0468-9

Derejko, M., Rakowicz, M., Antczak, J., Inglot, E., and Niewiadomska, M. (2013). Corticomotor excitability in drug-naive patients with Parkinson disease. *Neurol. Neurochir. Pol.* 47, 109–15.

Devanne, H., Lavoie, B. A., and Capaday, C. (1997). Input-output properties and gain changes in the human corticospinal pathway. *Exp. Brain Res.* 114, 329–338. doi: 10.1007/pl00005641

Foster, E. R., Golden, L., Duncan, R. P., and Earhart, G. M. (2013). Community-based argentine tango dance program is associated with increased activity participation among individuals with Parkinson's disease. *Arch. Phys. Med. Rehabil.* 94, 240–249. doi: 10.1016/j.apmr.2012.07.028

Freitas, C., Manzato, E., Burini, A., Taylor, M. J., Lerch, J. P., and Anagnostou, E. (2018). Neural correlates of familiarity in music listening: a systematic review and a neuroimaging meta-analysis. *Front. Neurosci.* 12:686. doi: 10.3389/fnins.2018.00686

Furubayashi, T., Ugawa, Y., Terao, Y., Hanajima, R., Sakai, K., Machii, K., et al. (2000). The human hand motor area is transiently suppressed by an unexpected auditory stimulus. *Clin. Neurophysiol.* 111, 178–183. doi: 10.1016/s1388-2457(99)00200-x

Furukawa, Y., Uehara, K., and Furuya, S., (2017). Expertise-dependent motor somatotopy of music perception. *Neurosci. Lett.* 650, 97–102. doi: 10.1016/j.neulet.2017.04.033

Hackney, M., and Earhart, G. (2010). Comparison of partnered and non-partnered dance movement. *Neurorehabil. Neural Repair* 24, 384–392. doi: 10.1177/1545968309353329

Hanna-Pladdy, B., and MacKay, A. (2011). The relation between instrumental musical activity and cognitive aging. *Neuropsychology* 25, 378–386. doi: 10.1037/a0021895

Haueisen, J., and Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *J. Cogn. Neurosci.* 13, 786–792. doi: 10.1162/08989290152541449

Houston, S., and McGill, A. (2013). A mixed-methods study into ballet for people living with Parkinson's. *Arts Health* 5, 103–119. doi: 10.1080/17533015.2012.745580

Janata, P., Tomic, S. T., and Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *J. Exp. Psychol. Gen.* 141, 54–75. doi: 10.1037/a0024208

Jenkinson, N., and Brown, P. (2011). New insights into the relationship between dopamine, beta oscillations and motor function. *Trends Neurosci.* 34, 611–618. doi: 10.1016/j.tins.2011.09.003

Kiers, L., Cros, D., Chiappa, K. H., and Fang, J. (1993). Variability of motor potentials evoked by transcranial magnetic stimulation. *Electroencephalogr. Clin. Neurophysiol.* 89, 415–423. doi: 10.1016/0168-5597(93)90115-6

Klein-Flügge, M. C., Nobbs, D., Pitcher, J. B., and Bestmann, S. (2013). Variability of human corticospinal excitability tracks the state of action preparation. *J. Neurosci.* 33, 5564–5572. doi: 10.1523/jneurosci.2448-12.2013

Koeneke, S., Lutz, K., Herwig, U., Ziemann, U., and Jäncke, L. (2006). Extensive training of elementary finger tapping movements changes the pattern of motor cortex excitability. *Exp. Brain Res.* 174, 199–209. doi: 10.1007/s00221-006-0440-8

Majid, D. S. A., Lewis, C., and Aron, A. R. (2015). Training voluntary motor suppression with real-time feedback of motor evoked potentials. *J. Neurophysiol.* 113, 3446–3452. doi: 10.1152/jn.00992.2014

Menon, V., and Levitin, D. J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *Neuroimage* 28, 175–184. doi: 10.1016/j.neuroimage.2005.05.053

Molina-Luna, K., Pekanovic, A., Röhrich, S., Hertler, B., Schubring-Giese, M., Rioult-Pedotti, M. S., et al. (2009). Dopamine in motor cortex is necessary for skill learning and synaptic plasticity. *PLoS One* 4:e7082 doi: 10.1371/journal.pone.0007082

Moraes, M. M., Rabelo, P. C. R., Pinto, V. A., Pires, W., Wanner, S. P., Szawka, R. E., et al. (2018). Auditory stimulation by exposure to melodic music increases dopamine and serotonin activities in rat forebrain areas linked to reward and motor control. *Neurosci. Lett.* 673, 73–78. doi: 10.1016/j.neulet.2018.02.058

Nielsen, J. F. (1996). Logarithmic distribution of amplitudes of compound muscle action potentials evoked by transcranial magnetic stimulation. *J. Clin. Neurophysiol.* 13, 423–434. doi: 10.1097/00004691-199609000-00005

Patel, A. D., Iversen, J. R., Bregman, M. R., and Schulz, I. (2009). Studying synchronization to a musical beat in nonhuman animals. *Ann. N. Y. Acad. Sci.* 1169, 459–469. doi: 10.1111/j.1749-6632.2009.04581.x

Popescu, M., Otsuka, A., and Ioannides, A. A. (2004). Dynamics of brain activity in motor and frontal cortical areas during music listening: a magnetoencephalographic study. *Neuroimage* 21, 1622–1638. doi: 10.1016/j.neuroimage.2003.11.002

Rosenkranz, K., Williamon, A., and Rothwell, J. C. (2007). Motorcortical excitability and synaptic plasticity is enhanced in professional musicians. *J. Neurosci.* 27, 5200–5206. doi: 10.1523/jneurosci.0836-07.2007

Rossi, S., De Capua, A., Pasqualetti, P., Ulivelli, M., Fadiga, L., Falzarano, V., et al. (2008). Distinct olfactory cross-modal effects on the human motor system. *PLoS One*, 3:e1702. doi: 10.1371/journal.pone.0001702

Sihvonen, A. J., Särkämö, T., Leo, V., Tervaniemi, M., Altenmüller, E., and Soinila, S. (2017). Music-based interventions in neurological rehabilitation. *Lancet Neurol.* 16, 648–660. doi: 10.1016/s1474-4422(17)30168-0

Stegemöller, E. L., Allen, D. P., Simuni, T., and MacKinnon, C. D. (2016). Motor cortical oscillations are abnormally suppressed during repetitive movement in patients with Parkinson's disease. *Clin. Neurophysiol.* 127, 664–674. doi: 10.1016/j.clinph.2015.05.014

Stegemöller, E. L., Allen, D. P., Simuni, T., and MacKinnon, C. D. (2017). Altered premotor cortical oscillations during repetitive movement in persons with Parkinson's disease. *Behav. Brain Res.* 317, 141–146. doi: 10.1016/j.bbr.2016.09.036

Stegemöller, E. L., Bates, B., Zaman, A., Tatz, J. R., Warnecke, A., and Hibbing, P. (2018a). Influence of music style and rate on repetitive finger tapping. *Motor Control* 22, 472–485. doi: 10.1123/mc.2017-0081

Stegemöller, E. L., Izbicki, P., and Hibbing, P. (2018b). The influence of moving with music on motor cortical activity. *Neurosci. Lett.* 683, 27–32. doi: 10.1016/j.neulet.2018.06.030

Strong, J. V., and Mast, B. T. (2018). The cognitive functioning of older adult instrumental musicians and non-musicians. *Aging Neuropsychol. Cogn.* 26, 367–386. doi: 10.1080/13825585.2018.1448356

Stupacher, J., Hove, M. J., Novembre, G., Schütz-Bosbach, S., and Keller, P. E. (2013). Musical groove modulates motor cortex excitability: a TMS investigation. *Brain Cogn.* 82, 127–136. doi: 10.1016/j.bandc.2013.03.003

Volpe, D., Signorini, M., Marchetto, A., Lynch, T., and Morris, M. E. (2013). A comparison of Irish set dancing and exercises for people with Parkinson's disease: a phase II feasibility study. *BMC Geriatr.* 13:54. doi: 10.1186/1471-2318-13-54

Weigmann, K. (2017). Feel the beat. *EMBO Rep.* 18, 359–362.

Witek, M. A. G., Clarke, E. F., Wallentin, M., Kringelbach, M. L., and Vuust, P., (2014). Syncopation, body-movement and pleasure in groove music. PLoS One 9:e94446. doi: 10.1371/journal.pone.0094446

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422. doi: 10.1038/nn1872

Ziemann, U., Tergau, F., Bruns, D., Baudewig, J., and Paulus, W. (1997). Changes in human motor cortex excitability induced by dopaminergic and anti-dopaminergic drugs. *Electroencephalogr. Clin. Neurophysiol.* 105, 430–437. doi: 10.1016/s0924-980x(97)00050-7

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 29 April 2020
doi: 10.3389/fnins.2020.00399

# The Acoustic Dimension of Reading: Does Musical Aptitude Affect Silent Reading Fluency?

*José Manuel Foncubierta[1,2]*, Francisco H. Machancoses[3], Kris Buyse[2,4] and M.C. Fonseca-Mora[5,6]*

[1] Education Department, Faculty of Education, Psychology and Sports Sciences, University of Huelva, Huelva, Spain, [2] Department of Linguistics, Faculty of Arts, KU Leuven, Leuven, Belgium, [3] Predepartamental Unit of Medicine, Science Health Faculty, Jaume I University, Castellón, Spain, [4] Department of Applied Languages, Faculty of Languages and Education, Nebrija University, Madrid, Spain, [5] English Studies Department, Faculty of Humanities, University of Huelva, Huelva, Spain, [6] Center of Contemporary Thinking and Innovation for Social Development (COIDESO), University of Huelva, Huelva, Spain

Fluent reading in a foreign language includes a complex coordination process of visual and auditory nature as the reading brain transforms written symbols into speaking auditory patterns through subvocalization (inner voice). The auditory information activated for reading involves the projection of speech prosody and allows, beyond letters and words decoding, the recognition of word boundaries and the construction of the melodic contours of the phrase. On the one hand, phonological awareness and auditory working memory have been identified in the literature as relevant factors in the reading process as skilled readers keep the acoustic information in their auditory working memory to predict the construction of larger lexical units. On the other hand, we observed that the inclusion of musical aptitude as an element belonging to the acoustic dimension of the silent reading aptitude of adults learning a foreign language remains understudied. Therefore, this study examines the silent reading fluency of 117 Italian adult students of Spanish as a foreign language. Our main aim was to find a model that could show if linguistic, cognitive and musical skills influence adults' silent reading fluency. We hypothesized that learners' contextual word recognition ability in L1 and FL in addition to, phonological awareness, auditory working memory and musical aptitude, elements related to the acoustic dimension of reading, would influence adults' silent reading fluency. Our structural modeling allows us to describe how these different variables interact to determine the silent reading fluency construct. In fact, the effect of musical aptitude on fluent silent reading in our model reveals to be stronger than phonological awareness or auditory working memory.

**Keywords: silent reading fluency, musical aptitude, foreign language, acoustic dimension, auditory working memory, phonological awareness, contextual word recognition, adult reader**

OPEN ACCESS

***Edited by:***
*Marta Olivetti,
Sapienza University of Rome, Italy*

***Reviewed by:***
*Mireille Besson,
Institut de Neurosciences Cognitives de la Méditerranée (INCM), France
Anja-Xiaoxing Cui,
University of British Columbia, Canada*

****Correspondence:***
*José Manuel Foncubierta
jmfoncu76@gmail.com*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *18 November 2019*
***Accepted:*** *31 March 2020*
***Published:*** *29 April 2020*

***Citation:***
*Foncubierta JM, Machancoses FH, Buyse K and Fonseca-Mora MC (2020) The Acoustic Dimension of Reading: Does Musical Aptitude Affect Silent Reading Fluency? Front. Neurosci. 14:399. doi: 10.3389/fnins.2020.00399*

## INTRODUCTION

### The Acoustic Dimension of Reading in a Foreign Language

Either in the mother tongue (L1) or in a foreign language (FL), the reading process implies the inter-relationship between written and spoken language. Ahissar et al. (2000), for instance, studied adults' reading abilities and concluded that auditory processing abilities accounted for more than 50% of the reading score variance, even in the group of adults who never had childhood histories of reading difficulties. Tichko and Skoe (2018) pointed out that "sensorineural auditory processing in central auditory structures is related to reading ability across the lifespan, beginning in the preliterate period and continuing into adulthood" (p.2), while Mankel and Bidelman (2018) stated that the brain's neural encoding and perception of sound differences is simply due to inherent auditory abilities that belong to the acoustic dimension. Therefore, an appropriate acquisition of oral skills eases the processes of triggering word recognition and fluency both necessary for reading comprehension (Dehaene, 2009). In alphabetic and shallow languages, such as Spanish and Italian, phonological awareness or the identification and manipulation of units in oral language is a reliable indicator of word recognition (McBride-Chang, 1995; Share, 1995, 2008): fluent reading is not possible without efficient contextual word recognition (Wang et al., 2005; Koda, 2007a; Macalister, 2010). In this sense, although letters to sounds conversion is a critical subskill for word recognition and reading fluency, the role of phonology appears to be more complex than simply support of word-by-word visual recognition. While reading silently or aloud, the identification of words is not enough, nor is it enough considering learners' ability of discriminating, remembering, and manipulating sounds at the sentence, word, syllable, and phoneme level, a lack of sensitivity toward the rhythmic and melodic properties of a given language also produces difficulties in accessing and comprehending a written text (D'Imperio et al., 2016). Thus, our study examines the acoustic dimension of reading. More concretely, the silent reading fluency of Italian adult students of Spanish as a foreign language in order to find a plausible model where the interaction between linguistic, cognitive and musical skills could explain adults' silent reading fluency. We hypothesized that learners' contextual word recognition abilities in L1 and FL in addition to phonological awareness, auditory working memory and musical aptitude, elements related to the acoustic dimension of reading, explain adults' silent reading fluency.

As regards phonological awareness, Ashby et al. (2013) showed in a longitudinal study the relationship between phonological awareness and silent reading fluency where results of phonemic tasks done by children studying Grade 2 accounted for nearly 42% of the variance in total time during silent reading in Grade 3. These data challenge the shift hypothesis and the accounts of reading development that claim that the role of phonology in reading is minimized as fluency develops and readers access word meanings directly from the orthographic form. They concluded that phonological processing continues to contribute to the efficiency of word recognition processes even in fluent readers. Macaruso and Shankweiler (2010, p. 464–465) carried out a study to identify a set of predictors that might be useful in distinguishing between less skilled and average college students readers. A discriminant analysis showed that the best predictors were a measure of phonological awareness (spoonerism) and a measure of verbal working memory (digit span). According to their results, phonological awareness and verbal working memory were more sensitive in identifying less skilled readers in the sample. Together these two variables predicted group membership correctly for 77% of the cases.

In foreign language reading, phonological awareness is considered as a precursor of the reading ability in different languages (Koda, 2007b). Kato (2009) studied Japanese students learning English as a second language and showed that phonological processes are required in foreign language silent reading at least until the learner becomes very proficient in the second language. The results of this research evidence that highly significant correlations are maintained between the sentence processing performance when reading silently and the reading comprehension score. For proficient readers, the involvement of the orthographic skills remained significant but phonological skills were still highly necessary for low proficient language learners.

Research on silent reading has shown that readers use their inner voice to project prosodic elements (intonation, tone, stress, and rhythm) on written symbols in order to disambiguate confusing sentences, create phonic chunks and predict lexical items (Kadota, 1987; Fodor, 2002; Ashby, 2016). According to the Prosodic Structure Hypothesis (Kadota, 1987), during FL silent reading the reader's inner voice or subvocalization follows speech rhythm patterns that support prediction of stressed syllables. This subvocalization plays an essential role when including words in syntactic and semantic relationships, allowing the reader to organize texts into lexical chunks. Even more, Ashby (2016) states that phonological decoding itself is a conscious process. The unconscious process of transforming visual information into their correlative sounds would only be possible when automatically activating the phonological word form before it is captured according to the prosodic information contained in the syllable, such as intensity, pitch and duration (phonological precoding stage). Therefore, the melodic and rhythmic structure of the text is built during contextual word recognition as well as during sentence integration, facilitating reading speed. As phonological precoding requires high-quality phonological representations of spoken words both during FL and L1 reading experience, research has been conducted into the influence of L1 orthographic and phonological coding on the FL reading ability (Sparks, 1995; Sparks et al., 2012). In this vein, transference from reading subskills like L1 phonological awareness into FL is well documented (Wang et al., 2005; Ziegler and Goswami, 2006; Bernhardt, 2010).

Unlike children, adult readers have more difficulties in distinguishing phonemic contrasts between L1 and FL (Kuhl et al., 2006). Apart from neurophysiological reasons such as the age of exposure to the foreign language (brain plasticity), in the case of FL reading fluency acquisition, the degree of phonological transfer may also be influenced by the proximity or similarity between the two languages (Ziegler and Goswami, 2006; Russak and Saiegh-Haddad, 2011; Yamashita, 2013) or by individual differences in working memory.

The second aspect of the acoustic dimension considered in our study is auditory working memory, another key concept of both reading and musical abilities (Kraus and Chandrasekaran, 2010). Baddeley et al. (1985) highlights the role that working memory plays as a component of fluent reading. Other works such as Strait et al. (2011) demonstrated the importance of auditory working memory for oral and silent reading fluency. In their study, higher auditory working memory correlated with better reading performance. Linguistic and musical information requires a temporary information storage system for their correct manipulation and integration, fundamental for reading prosody (Strait et al., 2011). To understand a phrase, the skilled reader needs to keep phonemic information in memory and integrate it in order to build lexemes and their semantic representation. In fact, reading with natural prosody facilitates sentence organization in memory and increases recall (Koriat et al., 2002). In the same way, processing melodic information requires tones to be kept in memory in order to integrate them in the melodic phrase representation. Pechmann and Mohr (1992) added the tonal loop, where prosodic and musical processing share resources of the auditory working memory.

Finally, musical aptitude, understood as a range of inherent abilities for music that an individual is born with and that are possibly shaped by informal exposition to music, has also been considered as a fundamental element of the acoustic dimension as it builds humans' auditory abilities (Patel, 2011; Slevc, 2012; Besson et al., 2017). In fact, music and speech prosody are communication sounding systems supported by the same acoustic parameters such as frequency, duration, intensity and timbre (Chobert and Besson, 2013). Slevc and Miyake (2006) considered that "being skilled at music means having a "good ear" for perceiving and analyzing foreign speech sounds" (p. 675) and showed that "individuals who are good at analyzing, discriminating, and remembering musical stimuli are better than other people at accurately perceiving and producing L2 sounds" (p. 679). Several studies have shown evidence of musical aptitude and pronunciation of a second language, both relying on cognitive processes of the auditory working memory, where tonal and verbal memory have a similar functional architecture (Tanaka and Nakamura, 2004; Koelsch et al., 2009; Jordan, 2018). This implies an overlap of neural structures from early ages on (Christiner and Reiterer, 2018). According to Jordan (2018: 177), "both musicians and non-musicians have an additional component, such as a tonal loop, which supports the retention of tone sequences". In other words, to some extent the brain processes speech as a kind of music (Koelsch, 2011). The effect of learners' musical aptitude has been mainly related to FL phonological perception and production (Milovanov et al., 2010), but less clear is its connection to FL reading skills. Studies about musical aptitude and "seemingly" visual reading skills such as silent reading fluency, remain to be scarce and inconclusive (Zeromskaite, 2014; Gordon et al., 2015), especially with adult readers who learn a language in a foreign context (Swaminathan et al., 2018). Gómez-Domínguez et al. (2019) provided insights into how music perception affects early reading skills in 63 Spanish children learning English. Their findings support a transfer of music perception abilities to L1 young learners' reading abilities that affect the alphabetic principle, the phonemic awareness and the word recognition skills in their FL early reading skills.

Studies focusing on the relationship between language perception, musical skills and reading abilities confirm the hypothesis that music and language rely on similar mechanisms of auditory temporal processing (Patel, 2011; Besson et al., 2017). Nevertheless, two issues are still debated: on the one hand, studies that argue that differences in reading abilities mediated by musical aptitude could be the result of genetic mediated differences (Schellenberg, 2015; Swaminathan and Schellenberg, 2017). On the other hand, empirical studies indicate that it is specific musical training that could exert a causal influence on the subjects' abilities to discriminate language sounds and to get better results in reading (Kraus and Chandrasekaran, 2010; Chobert and Besson, 2013; Besson et al., 2017). There are even longitudinal studies of educational intervention that show how musical training improves language perception and reading skills (Besson et al., 2007; Flaugnacco et al., 2015). However, Bigand and Poulin-Charronnat (2006) pointed out that musical aptitude could be acquired by "musically experienced listeners" only through exposure to music without explicit musical training. Thus, being a non-musician does not mean that one does not have musical aptitude. Individuals with extensive musical training do not always reach higher levels of musical competence than those without formal musical training (Law and Zentner, 2012).

In this study, the term musical aptitude represents the music abilities of individuals with or without musical training. Our hypothesis is that musical aptitude, as a capacity measured by the participant's Tuning, Melody, Accent and Tempo abilities, shapes the acoustic dimension of reading because fluent reading requires a sensibility toward the phonological, rhythmic and melodic properties of any language. Taking all this together, in our model we hypothesize that if "reading fluency involves every process and subskill involved in reading" (Wolf and Katzir-Cohen, 2001: 220), then silent reading fluency can be operationalized as a complex construct where different visual and oral components interact: phonological awareness, auditory working memory and L1/FL visual contextual word recognition.

Therefore, this study aims to uncover, through correlations and structural equation modeling (SEM), the acoustic dimension of silent reading fluency based on an analysis of factors such as L1 and FL contextual word segmentation, in addition to phonological awareness, auditory working memory and musical aptitude of 117 Italian university students of Spanish as a foreign language. Our research questions based on correlations are to confirm in our sample what previous research about phonological awareness, word identification and segmentation, auditory working memory and reading has already tested. Given that a lack of sensitivity toward the rhythmic and melodic properties of a given language could also produce difficulties in accessing and comprehending a written text (D'Imperio et al., 2016), our study is aimed at searching for a statistical-causal model between musical aptitude and silent reading fluency. Moreover, it is the first time to our knowledge that musical aptitude is correlated with L1 and FL word segmentation.

The study is structured around five research questions (see **Figure 1**), all of them related to the silent reading fluency of adult readers:

RQ1: Is there any relationship between L1 segmentation and FL segmentation?

RQ2: Is there any relationship between phonological awareness and FL segmentation?

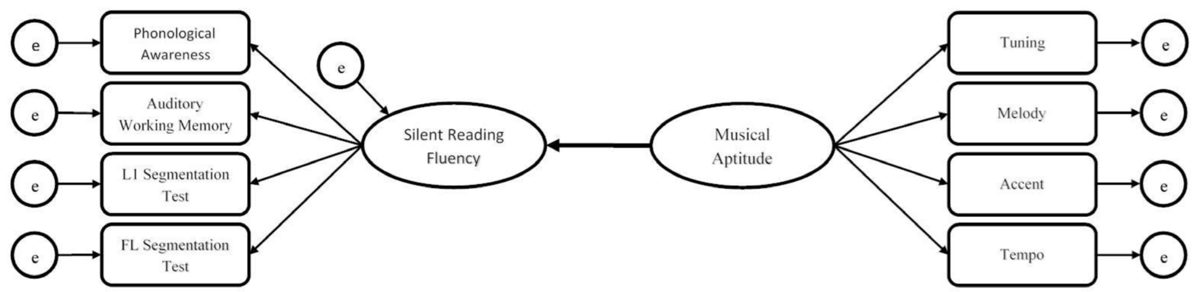



**FIGURE 1 |** Theoretical SRF model.

RQ3: Is there any relationship between auditory working memory and FL segmentation?

RQ4: Is there any relationship between musical aptitude subtests and L1/FL segmentation?

RQ4a: Is there any relationship between musical aptitude subtests and L1 segmentation?

RQ4b: Is there any relationship between musical aptitude subtests and FL segmentation?

RQ5: Can we establish a statistical-causal model for determining silent reading fluency on the basis of L1 and FL segmentation, phonological awareness, auditory working memory and musical aptitude?

The eight observed variables (phonological awareness, auditory working memory, L1 and FL contextual word recognition/segmentation, tuning, melody, accent and tempo) have been measured directly. From these measured variables, the latent variables (silent reading fluency and musical aptitude) are reflected if the model is true.

In order to find out how musical aptitude influences silent reading fluency as hypothesized in **Figure 1**, a SEM was carried out to understand if and how musical aptitude could influence silent reading fluency, and how the eight observed indicators would interact with each other and with the latent variables of this study in our sample. SEM provides a statistical method which "enables researchers to easily set up and reliably test hypothetical relationships among theoretical constructs as well as those between the constructs and their observed indicators" (Deng et al., 2018, p. 1).

These measurement components are shown in **Figure 1** by using thin lines. By convention, the direction of the arrows goes from the latent variables to the observed ones.

# MATERIALS AND METHODS

## Participants

Data was collected from 124 adult readers, all of them students of the University of Macerata, of whom only 117 answered all the tests. All participants were freshmen and passed a language level test called "Test di linguistic idoneitaÌ" that the university uses to classify them into a homogeneous pre-intermediate language level class. All participants belonged to the same class. Of the 117, 34.19% ($n$ = 40) were male, and 65.81% ($n$ = 77) were female students. Age ranged between 21 and 25 years, with an average of 21.72 (Sd = 0.771). All subjects were native speakers of Italian studying a Degree Program in Linguistic and Cultural Mediation in English and Spanish. They had never participated in any immersion program in Spain and acknowledged not suffering any kind of reading disability. Most of them had not received musical training (only 4.7% had received some training before).

## Measures

Students were administered five different tests: a contextual word recognition test in its Spanish version, a contextual word recognition test in its Italian version, a Spoonerism test to measure learners' phonological awareness, WAIS-IV to measure learner ̀s auditory working memory (Digit Span tests, Letters and Numbers Sequencing and Arithmetic) and the musical MiniProms Test in order to check their musical aptitude.

**Figure 2** includes our data collection protocol flowchart and in the following paragraphs each test is explained.

### L1 and FL Contextual Segmentation Tests

The Spanish and Italian contextual word recognition or in brief, L1 and FL segmentation tests, were adapted versions of the Test of Silent Contextual Reading Fluency (Hammill et al., 2006). These tests measure the participants' level of reading fluency in each language by counting the number of printed words that could be segmented within 3 min in a text without blank spaces. The participants were presented with the text of Human Rights in its Spanish and Italian version. Both versions were based on different articles of the Universal Declaration of Human Rights in order to avoid transfer of previous knowledge. Readability tests were performed with a view to control that the selected Spanish and Italian texts fit the college level (45.6 Spanish Flesch Reading Formula, 30.17 Italian Flesch-Vacca and 44 Italian GulpEase). GulpEase index was rated similar to the Italian Flesch-Vacca adaptation but better tailored to the Italian language (Forti et al., 2019, p.360).

Letters were all in lowercase because "the lowercase letters offer the reader a skyline of words" (Hiebert and Reutzel, 2014, p. 37). In order to measure speed and correctness of word recognition in the text, participants had 3 min to recognize as many words as possible using a ballpoint pen and making

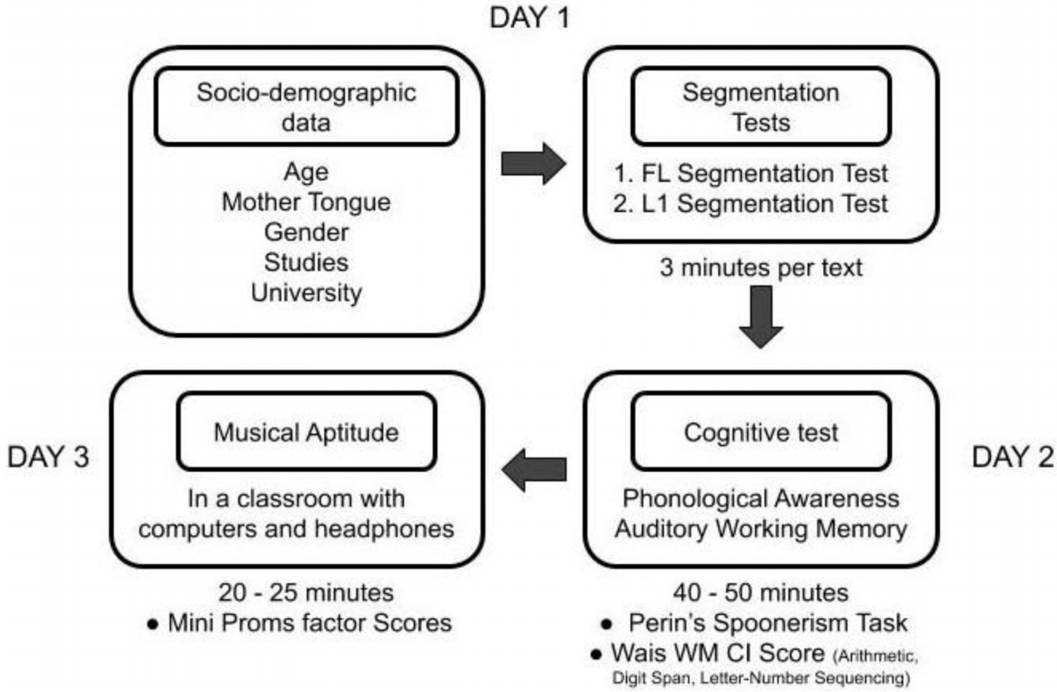



**FIGURE 2 |** Data Collection protocol flowchart.

separations with bars. First, they did the test with the FL text and subsequently the other text in L1. The results were obtained from the total number of correctly identified words within the fixed time period. Data collection time was 6 min.

### Phonological Awareness

The phonological awareness test is a Spanish adaptation of Perin (1983). In the original version of this task, American famous people's names were used; for example, "Chuck Berry." It was administered individually. Students had to listen to 18 pairs of first and last names of famous Spanish people (for example, Peneélope Cruz [peneélope krué$\theta$]), and were asked to change the initial consonant of the name by the initial consonant of the surname, producing Ceneélope Pruz [$\theta$eneélope prué$\theta$], in such a way that [t$\int$eneélope prué$\theta$] or [keneélope prué$\theta$] were considered non-valid. After hearing the name, they only had 4 s to respond. An Olympus Ws-650S tape recorder was used for data collection. The data collection time was 2 min per participant.

### Auditory Working Memory

Furthermore, participants scored individually their auditory working memory. Digit Span backward and forward, and Arithmetic of the WAIS-IV test (Wechsler et al., 2008) were administered, in addition to Letters and Number Sequencing. These subtests evaluate auditory working memory. Following the WAIS-IV test score indications, the AWM score was computed from the sum of Arithmetics, Digit Span and Letters and Numbers Sequencing, gathering the AWM Scalar Punctuation. Afterward, this score is transformed in CI scores using the scales offered by the WAIS-IV correction manual. Data collection time was approximately 30 to 35 min for each participant.

### Musical Aptitude

Mini-PROMS, the reduced version of the Proms test (Zentner and Strauss, 2017), was administered individually, each student with a computer and headphones. This reduced version was selected due to the high number of tests and the amount of class time needed. Mini-Proms consists of a battery of subtests that measure musical aptitude through the discrimination of different musical structures, namely Tuning, Melody, Accent and Tempo. The tuning subtest plays a C-chord whose tone E could be mistuned. Participants are asked to judge whether the tuning is the same in the reference and the probe stimulus. In the melody subtest participants hear a two-bar monophonic harpsichord melody twice, followed by the probe melody which can differ slightly by one or more tones. Accent assesses the capacity of detecting and retaining rhythmic patterns in a sequence of 5 to 12 beats. The tempo subtest comprises rhythmically and timbrally diverse stimuli which are the same between reference and probe stimulus, except, potentially, for their tempo. The data collection time was 20 to 25 min.

## Data Analysis

First, a descriptive analysis of the variables has been carried out (**Table 1**). The normality of these variables has been tested using the Kolmogorov-Smirnov (KS) normality test. Before starting the SEM analysis, we wanted to know if there were correlations in accordance with our research questions. As mentioned earlier, our correlational questions check if our results

**TABLE 1 |** Descriptive statistics.

| | Mean | SD | Median | Min | Max | Ks (p) |
|---|---|---|---|---|---|---|
| L1 segmentation test | 195.56 | 54.64 | 207 | 74 | 364 | 0.087 (0.043) |
| FL segmentation test | 120.67 | 55.18 | 119 | 29 | 288 | 0.085 (0.013) |
| Phonological awareness | 12.26 | 1.71 | 12 | 5 | 16 | 0.155 (0.000) |
| Auditory working memory | 87.66 | 8.53 | 87 | 72 | 118 | 0.203 (0.000) |
| Proms melody score | 7.22 | 3.17 | 7 | 1 | 15 | 0.115 (0.024) |
| Proms tuning score | 6.03 | 2.12 | 6 | 1 | 16 | 0.124 (0.000) |
| Proms accent score | 6.86 | 2.16 | 7 | 2 | 12 | 0.079 (0.048) |
| Proms tempo score | 7.39 | 2.48 | 7 | 2 | 15 | 0.120 (0.000) |

are consistent with the ones previously reported in literature although mainly for children and referring to L1. Phonological awareness and auditory working memory have already been consistently identified as predictors of early reading ability and we wanted to check the same type of correlations with our adult population. We think this gives more support to the SEM we carried out based on our working hypotheses.

To determine the statistical-causal model that interrelate all variables, we conducted a SEM analysis with the Multivariate Software program EQS 6.2 (Bentler, 2008). Although there is debate about the sample size needed for SEM, we considered our sample of 117 participants suitable to perform the proposed structural modeling because correlations were strong (Kenny, 2015). In order to describe how different variables interact in the silent reading fluency construct, SEM is a better-chosen analysis technique than the classical methods of regression because it assigns dependent and independent variables to cause and effect categories, including their order of appearance. SEM provides a statistical method for evaluating relationships among indicators and latent variables in a hypothesized model, and provides causal statistical fit indices of the hypothesized model. Our structural model integrates eight directly measured variables (L1 and FL contextual word segmentation, phonological awareness, auditory working memory, tuning, melody, accent and tempo) and two multi-factorial latent variables: silent reading fluency and musical aptitude (see **Figure 1**, where latent variables are represented by circles and observed variables by squares, with arrows showing the relations between these variables).

When the variables did not follow a normal distribution, the robust statistic of Satorra-Bentler (Satorra and Bentler, 1988; Satorra, 1990; Yuan and Bentler, 2007) was used. This robust statistical procedure allowed us to contrast hypotheses concerning relationships among latent variables and indicators, including the different interrelations between them, when the assumptions of normality and heteroscedasticity do not occur.

The EQS also offers the Lagrange Multiplier Test, a procedure designed to study the need for constraints on the model, both the equality constraints that may have been included, and the covariance not initially included and that should be counted as free parameters (Bentler, 2008). This test is analogous to the so-called LISREL Modification Indices, with the difference that the Lagrange Test operates multivariately in determining misspecified parameters in a model, while the LISREL Modification Indices operate univariately (Byrne, 2013, p. 84). As the Lagrange Test indicated the introduction of modifications, they were tested until we reached the fitted model.

## RESULTS

The main descriptive statistics of the variables under study, as well as the K-S test of normality, are presented in **Table 1**. In order to answer research questions 1 to 4, a correlational analysis using Spearman Rho ($\rho$) with a bilateral significance test was performed to test the relational hypothesis (**Table 2**), given the non-normality of the variables ($p < 0.05$).

The Spearman Rho ($\rho$) test reveals a highly significant relationship between L1 Segmentation and FL segmentation [RQ1] ($\rho = 0.750$), between FL segmentation and phonological awareness [RQ2] ($\rho = 0.645$), between auditory working memory and FL segmentation [RQ3] ($\rho = 0.609$), and between musical aptitude subtests and L1 segmentation [RQ4a] (Melody: $\rho = 0.692$; Tuning:0.656; Accent:0.705; Tempo:0.658). Also, there is a strong correlation between musical aptitude subtests and FL segmentation [RQ4b] (Melody: $\rho = 0.807$; Tuning:0.615; Accent:0.711; Tempo:0.523) (see **Table 2**). All these correlations have a significance $p < 0.01$.

In order to more comprehensively examine relationships among musical aptitude and silent reading fluency, we subjected these data to SEM in **Figure 1** [RQ5]. All covariances and saturations between the variables are represented in a path diagram with their fit indexes (**Figure 3**). According to MacCallum et al. (1996) and Schreiber et al. (2006), RMSEA values between 0.06 and 0.08, and other coefficients greater than or equal to 0.95 indicate an appropriate fit. Therefore, considering the results obtained, we can determine that our model fits appropriately.

High saturation of musical aptitude on silent reading fluency ($\beta$ = 0. 914) was observed. The latent construct musical aptitude is determined significantly by the four mentioned components: tuning ($\gamma = 0.564$), melody ($\gamma = 0.915$), accent ($\gamma = 0.862$) and tempo ($\gamma = 0.818$); and silent reading fluency is determined significantly by the indicators L1 segmentation ($\gamma = 0.832$) and FL segmentation ($\gamma = 0.920$), in addition to phonological awareness ($\gamma = 0.678$) and auditory working memory ($\gamma = 0.734$).

The inclusion of a series of covariances among the indicators, based on information provided by the Lagrange

**TABLE 2 |** Spearman's rho Correlations.

| | L1ST | FLST | PA | AWM | PMS | PTS | PAS | PTmS |
|---|---|---|---|---|---|---|---|---|
| L1 segmentation test | 1 | 0.750** | 0.645** | 0.609** | 0.692** | 0.656** | 0.705** | 0.658** |
| FL segmentation test | | 1 | 0.668** | 0.729** | 0.807** | 0.615** | 0.711** | 0.523** |
| Phonological awareness | | | 1 | 0.694** | 0.635** | 0.466** | 0.641** | 0.658** |
| Auditory working memory | | | | 1 | 0.781** | 0.541** | 0.543** | 0.587** |
| Proms melody score | | | | | 1 | 0.680** | 0.771** | 0.742** |
| Proms tuning score | | | | | | 1 | 0.554** | 0.527** |
| Proms accent score | | | | | | | 1 | 0.760** |
| Proms tempo score | | | | | | | | 1 |

***Correlation is significant at the 0.01 level (2-tailed).*

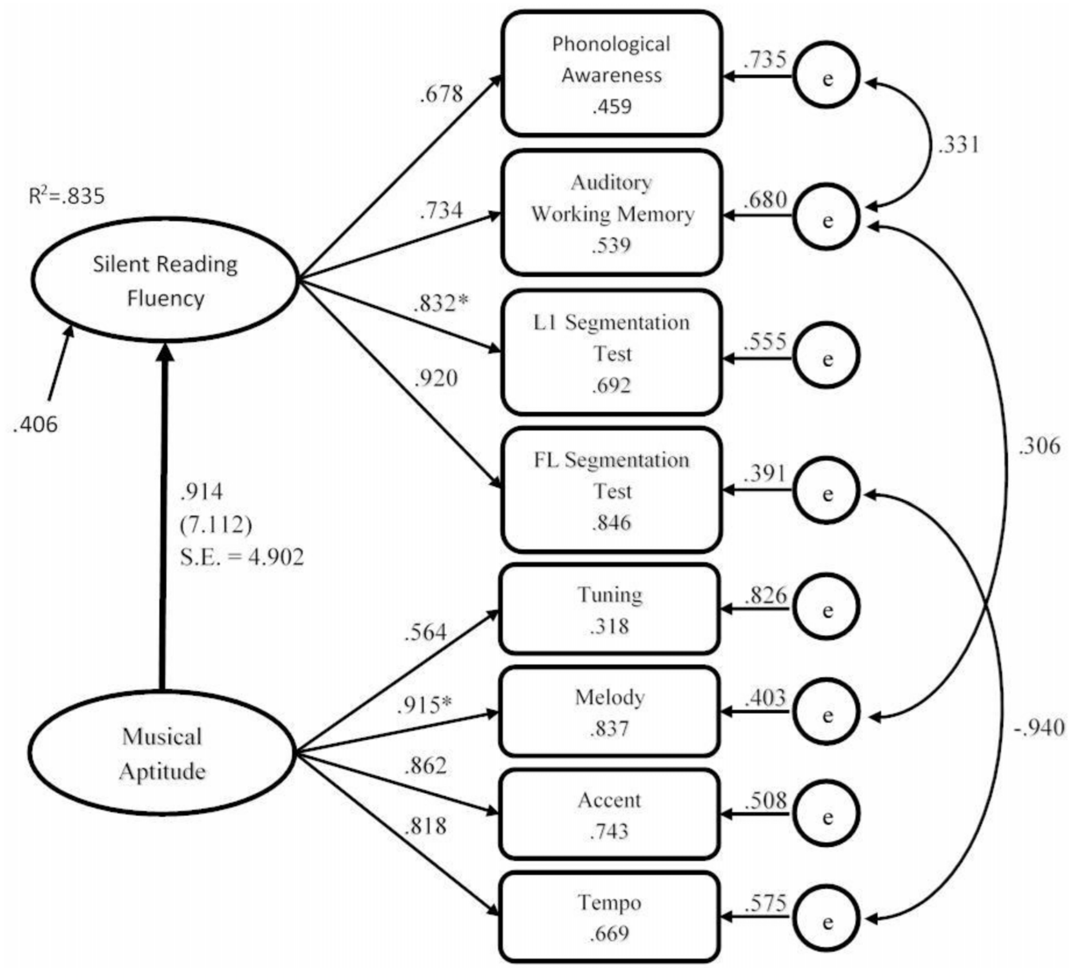



**FIGURE 3 |** Final SEM Model obtained in standardized values. All coefficients are significant. The fixed parameters were marked with "∗". Robust Independence Model χ2 = 619.753; Satorra-Bentler Scaled χ2 = 22.601 (p = 0.093): Non-Normed Fit Index 2816 = 0.976; Comparative Fit Index = 0.987; Root Mean Square Error of Approximation [90% CI] = 0.066 [0.000, 0.118]; e = error.

Test, helped to adjust the model. These covariances have been included through an iterative process, in which the fit of the model for each covariance introduced was tested. Especially relevant were covariances between phonological awareness and auditory working memory (φ = 0.331), and the one between auditory working memory and melody (φ = 0.306). Also, covariances between Tempo and FL segmentation (φ = −0.940) were found.

In order to observe the saturation between musical aptitude and silent reading fluency, a scatterplot analysis was carried out, showing a linear $R^2$ of 0.720 between the factorial scores in standardized values obtained for each subject (**Figure 4**).

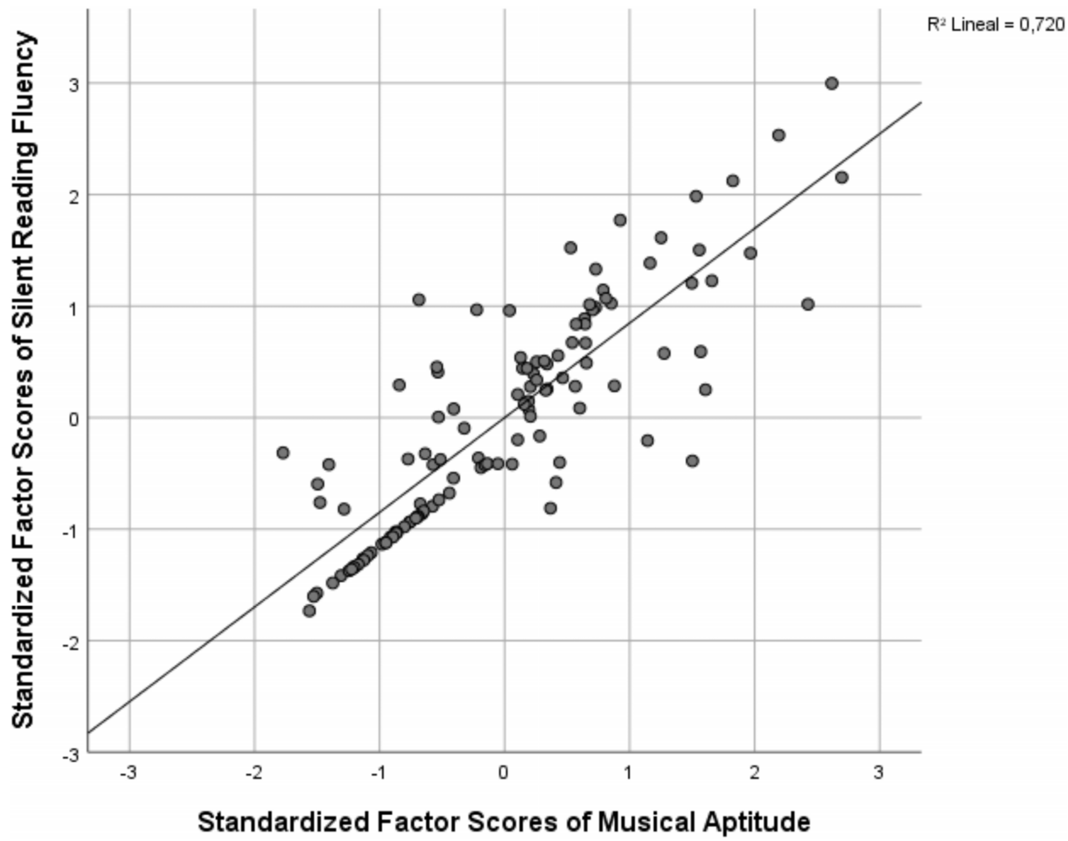



**FIGURE 4 |** Silent reading fluency positively correlates with musical aptitude (factor scores extracted from SEM).

# DISCUSSION

The objective of this study was to uncover the acoustic dimension of silent reading fluency based on an analysis of factors such as contextual word recognition in L1 and FL, in addition to phonological awareness, auditory working memory and musical aptitude among 117 Italian university students of Spanish as a FL. We expected that these variables could explain learners' individual differences in their silent reading fluency. More concretely, we wanted to know if musical aptitude affects silent reading fluency. The analysis provides us with the following answers to the different research questions.

## RQ1: Is There Any Relationship Between L1 Segmentation and FL Segmentation?

Regarding the first question, a strong correlation was found between the L1 segmentation and FL segmentation. As put forward by Sparks et al. (2012) in their *Linguistic Coding Differences Hypothesis*, skills acquired in the mother tongue, such as fluent reading, can be transferred to foreign language learning. This transfer, as well as its degree (Young-Kim et al., 2017) may also be due to the proximity or similarity between orthographic codes of the two languages (Wang et al., 2005; Ziegler and Goswami, 2006; Bernhardt, 2010). In fact, in transparent languages such as Spanish and Italian, with a consistent grapheme-phoneme relationship, fluent reading develops earlier than in alphabetic languages with a more complexly decodable spelling system such as English. Nevertheless, regardless of the typological and linguistic similarity of the two languages, the contextual word recognition ability in the foreign language scored lower. In this sense, the results are consistent with earlier studies where reading in a FL occurs slower than in L1 (Koda, 2007b; Bernhardt, 2010; Yamashita, 2013). According to previous literature, this deceleration could be due to the grade of familiarity between FL and L1 but also to the learners' accumulated reading experience in L1.

## RQ2: Is There Any Relationship Between Phonological Awareness and FL Segmentation?

The results of our analysis point out a strong correlation between both variables. Kato (2009) found out that phonological decoding plays an important role for low language proficient FL readers, at least, in two situations: while reading unfamiliar words, and when it is necessary to keep information in memory at the same time that processing complex structures. In our study, participants had a pre-intermediate level and we increased the difficulty of the silent reading fluency task by asking students to read a visually complex text, since words were not separated by blank spaces. In this way, being able to visually recognize letters, syllables and words requires to keep in memory the conversion of letters into sounds. As difficulties in phonological awareness are usually the hallmark of reading difficulty (Ziegler and Goswami, 2006; Perfetti, 2007; Russak and Saiegh-Haddad, 2011), we expected that the ability of retaining acoustic features in memory and to manipulate them was related to silent reading fluency. Phonological awareness is a construct composed of at least three components -general cognitive ability, verbal memory, and speech perception-, but a large part of phonological awareness is simple speech perception (McBride-Chang, 1995). As phonological awareness is a reliable indicator of visual word recognition in FL reading (Wang et al., 2005; Koda, 2007a), our results reveal that part of the individual differences in FL word recognition are due to the ability to perceive sounds and manipulate them in a non-native language. The proximity between the two languages also shows the strong correlation between phonological awareness and L1 word recognition ability.

## RQ3: Is There Any Relationship Between Auditory Working Memory and FL Segmentation?

The results of our study show a correlation between auditory working memory and FL segmentation, which is weaker in L1 than in FL, probably due to the learners' greater mastery and reader confidence in their L1 (Russak and Saiegh-Haddad, 2011). We expected that the use of a text without blank spaces between words or spelling signs would force readers to mentally pronounce the words they are discovering while reading (Kadota, 2002); for that, readers need to maintain acoustic information in mind to integrate sounds into larger units and build meaning. The orthographic information without phonological decoding is purely iconic and does not allow the grouping of sound blocks according to the melodic and rhythmic pattern of the language. In this sense, silent reading fluency implies the cooperation of sound information and its corresponding meaning beyond words. As silently reading a text without spaces requires to manage the letter-sound relationship in order to recognize words, and also to integrate this information into larger units, the theoretical construct of working memory presented by Baddeley (1986) plays an essential role in discussions on the mechanisms employed in L1/FL segmentation. Especially, the component called phonological loop allows readers to manipulate and store speech-based information and is further divided into a phonological short-term store and an articulatory control process. The former is in charge of temporarily maintaining phonological information, the latter of refreshing fading phonological information through subvocal rehearsal.

However, readers also need to process melodic information from syllables (intensity, pitch, duration), in order to predict the phonological form of words and their composition spelling (Koriat et al., 2002). According to Ashby (2016), this precoding occurs automatically and requires out of the syllabic information certain prosodic elements in order to complete a word, such as when we complete a song from the beginning of its melody. This process demands, therefore, a tonal loop so that tones are kept in memory and integrated in the melodic phrase representation (Pechmann and Mohr, 1992; Tanaka and Nakamura, 2004; Jordan, 2018).

## RQ4: Is There Any Relationship Between Musical Aptitude Subtests and L1/FL Segmentation?

Our results indicate that musical aptitude subtests correlate highly and positively with L1 segmentation (RQ4a) and FL segmentation (RQ4b). To our knowledge, there are no other studies on musical aptitude and L1/FL segmentation. Previous studies such as Slevc and Miyake (2006) or Milovanov et al. (2010) had already shown a relationship between musical aptitude and FL learning, especially at the phonological level and the acquisition of other oral skills. As for the relationship between sensitivity to different musical structures (tuning, melody, accent, and tempo) and visual word recognition, our data show that musical aptitude holds a high correlation with L1 segmentation as well as with FL segmentation. Zeromskaite (2014, p.85) in a literature review claims that "the theoretical basis behind the reading skills facilitation by music is less clear, but it may be best explained by increased listening sensitivity." In a meta-analysis by Gordon et al. (2015), only a weak trend was found toward significance of musical discrimination abilities on reading fluency. They hypothesize that music skills share more variance with phonological skills (due to their auditory bases) than with reading fluency skills (more visual skills), and thus music training may have larger effects on phonological awareness than on reading. Nevertheless, our results point out that likely adult readers' musical aptitude is affecting their contextual word recognition ability.

## RQ5: Can We Establish a Statistical Causal Model for Determining Silent Reading Fluency on the Basis of L1 and FL Segmentation, Phonological Awareness, Auditory Working Memory and Musical Aptitude?

In order to find out how musical aptitude affects silent reading fluency, a SEM was carried out (see **Figure 3**). The results allowed us to test our model proposed in **Figure 1**. We included three *post hoc* modifications. The Lagrange Test for computing parameters recommended us to add covariances between Auditory Working Memory and Phonological Awareness, Auditory Working Memory and Melody, and FL Segmentation Test and Tempo. All covariances and saturations between the variables are represented in a path diagram with their fit indexes (**Figure 3**).

The theoretical approach is highly relevant when trying to present a new model. So, when the test indicated these possible covariances between auditory working memory and melody, and in order to improve the fit of the model, we first checked whether they had a prior theoretical justification for adding them and we found the following support for the inclusion of these covariances. The use of covariance to fit the model is not conventional, but authors such as Byrne (2013, p. 184) point out that it is reasonable to use it when the theoretical basis supports it. Kline (2015, p. 380) states that "the capability to explicitly model the error covariance structure is an advantage of SEM over more traditional statistical techniques."

The covariance between the values of phonological awareness and auditory working memory shows that differences in silent reading fluency are also determined by the retention capacity of acoustic elements such as phonemes for word recognition, as pointed out in the Baddeley (1986) working-memory model that includes the phonological loop. Regarding the integration of information in the oral reconstruction of reading, the covariance between auditory working memory and melody may indicate that the ability to retain musical information, such as the succession of single tones, could be related to the reading intonation which is necessary to understand a text, as this intonation is also present in students' silent reading. This recognition of tonal frequencies points to the importance of tonal memory in the development of silent reading fluency (Pechmann and Mohr, 1992; Tanaka and Nakamura, 2004).

On the other hand, the results are consistent with previous studies that show how melody is the main musical feature affecting phonological awareness in adult readers (Posedel et al., 2012; Kempe et al., 2015). The covariance between phonological awareness and auditory working memory and between auditory working memory and melody, may reflect that tasks used for both variables (phonological awareness and melody), have in common the same cognitive processing demand which is the temporary information storage system required for their correct manipulation (Strait et al., 2011).

Similar studies carried out with children while reading aloud show that rhythm-related skills often predict phonological awareness (Tierney and Kraus, 2014). Nevertheless, as put forward by Swaminathan and Schellenberg (2017, p. 1930), among adult readers "the story is more complicated." Likely, adult readers are more experienced listeners than children.

As stated by Koelsch (2011), to some extent the brain processes speech as a kind of music, but when learning a foreign language some musical features of the mother tongue may remain. This seems to be the case of the negative covariance between tempo and FL segmentation. This covariance, known as negative transference (Melby-Lervåg and Lervåg, 2014), due to the proximity or similarity between the two languages, could indicate the influence of an individual characteristic of the L1 rhythmic pattern. As Italian is characterized by the elongation or duration of the accented vowels, this value may show that Italian learners of Spanish are using their Italian rhythmic patterns, which goes in line with the interference hypothesis of the L1 rhythmic pattern (Iversen et al., 2008). Their central idea is that depending on the L1 musical features, there is a certain influence on the perception of non-linguistic musical traits, hence that negative covariance influences FL and not L1. In theory, the Italians would perceive tempo differences better in Italian than in Spanish as it is a characteristic of their L1. In Italian, tonic vowels receive a greater emphasis on duration than Spanish tonic vowels. This would mean that duration is a relevant phonological aspect in Italian but not in Spanish, where the duration does not produce a change in meaning in the system; that is, it would only have a pragmatic value: when a speaker extends the duration of a vowel to add a connotative meaning. The negative covariance with (only) FL segmentation would be an example of negative transfer in FL reading: with less musical tempo, more FL silent reading fluency. The high value of the covariance between tempo and FL segmentation would not indicate that they are identical variables, but they may mean that for Italian language learners of Spanish the ease of recognizing a musical aspect such as tempo is inversely proportional to their ability of segmenting a text in Spanish.

Taken all these data together, it can be argued that the high saturation of musical aptitude on silent reading fluency confirms that the ability of perceiving the differences of tuning and tempo along with accent and melody may contribute more to the understanding of the individual differences in silent reading fluency than other factors.

## CONCLUSION

The general conclusions of this study allow us to consider that the musical aptitude of adult readers studying a foreign language gives shape to their reading skills. Other cognitive components involved in reading such as the auditory working memory appear to be fundamental to the integration of linguistic and musical information, playing a crucial role in explaining the individual differences in silent reading fluency. To some extent, we expected that the correlational study could yield positive results. Earlier studies had already reported the positive correlations between reading skills in L1 and second languages (Koda, 2007a; Gómez-Domínguez et al., 2019), or between phonological awareness and reading components such as fluency (van den Boer et al., 2014; Flaugnacco et al., 2015), but we decided to check it again to present our model. The SEM, as a statistical-causal method, allowed us to analyze how variables would behave after previously observed correlations, according to an *a priori* hypothesized model.

The many significant results may be also due to the nature of our research design where all tests represent a demand on participants' auditory working memory. The L1 and FL segmentation tests involve reading a complex text with no blanks, in which readers need to retain the sequences that they are recoding in their memory. The phonological awareness test requires keeping sounds in memory in order to manipulate them, and the musical aptitude test also calls for the retention of auditory information. Although further research is still needed, the level of significance found in our results may reveal the existence of common cognitive and neural mechanisms for language reading and musical skills, so that readers with better results in the musical aptitude, segmentation, and the phonological awareness tests are also demonstrating a better ability in the task of maintaining information in their auditory working memory.

Given the novelty of our vision on how musical aptitude explains adult readers' silent reading fluency, it still requires further study especially with other foreign languages and other adult populations. Our model based on the acoustic dimension of silent reading fluency offers an image about the interaction of visual and sound factors related to reading. In agreement with Grabe and Stoller (2011), readers are extraordinary word recognizers and, moreover, according to our data, good readers are excellent melody recognizers and this affects their silent reading fluency.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

MF-M contributed to the conceptualization, investigation, and funding acquisition. FM contributed to the methodology and formal analysis. JF wrote the original draft. JF, FM, KB and MF-M wrote, reviewed and edited the manuscript.

## FUNDING

This study has been supported by the R+D project "Musical aptitude, reading fluency and intercultural literacy of European university students" (FFI2016-75452-R, Spain, Ministerio de Economía, Industria y Competitividad).

## ACKNOWLEDGMENTS

Thanks to Dr. Marisa Martínez Pérsico (University of Macerata, Italy) for her assistance in the data collection and our gratitude to the Innsbruck University for their assistance in the Musical Aptitude Test data collection.

## REFERENCES

Ahissar, M., Protopapas, A., Reid, M., and Merzenich, M. M. (2000). Auditory processing parallels reading abilities in adults. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6832–6837. doi: 10.1073/pnas.97.12.6832

Ashby, J. (2016). "Why does prosody accompany fluency? Re-conceptualizing the role of phonology in reading," in *Reading Fluency. Literacy Studies: Current Insights from Neurocognitive Research and Intervention Studies*, Vol. 12, eds A. Khateb, and I. Bar-Kochva, (Basel: Springer International Publishing), 65–89. doi: 10.1007/978-3-319-30478-6_5

Ashby, J., Dix, H., Bontrager, M., Dey, R., and Archer, A. (2013). Phonemic awareness contributes to text reading fluency: evidence from eye movements. *School Psych. Rev.* 42, 157–170.

Baddeley, A. (1986). *Working Memory (Oxford Psychology Series, No. 11)*. New York, NY: Clarendon Press.

Baddeley, A., Logie, R., Nimmo-Smith, I., and Brereton, N. (1985). Components of fluent reading. *J. Mem. Lang.* 24, 119–131. doi: 10.1016/0749-596X(85)90019-1

Bentler, P. M. (2008). *EQS 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software.

Bernhardt, E. B. (2010). *Understanding Advanced Second-Language Reading*. New York, NY: Routledge.

Besson, M., Barbaroux, M., and Dittinger, E. (2017). "Music in the brain: music and language processing," in *Routledge Companion to Music Cognition*, eds R. Ashley, and R. Timmers, (New York, NY: Routledge).

Besson, M., Schön, D., Moreno, S., Santos, A., and Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restor. Neurol. Neurosci.* 25, 399–410.

Bigand, E., and Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition* 100, 100–130. doi: 10.1016/j.cognition.2005.11.007

Byrne, B. M. (2013). *Structural Equation Modeling with EQS: Basic Concepts, Applications, and Programming*. Abingdon: Routledge.

Chobert, J., and Besson, M. (2013). Musical expertise and second language learning. *Brain Sci.* 3, 923–940. doi: 10.3390/brainsci3020923

Christiner, M., and Reiterer, S. (2018). Early influence of musical abilities and working memory on speech imitation abilities: study with pre-school children. *Brain Sci.* 8:169. doi: 10.3390/brainsci8090169

Dehaene, S. (2009). *Reading in the Brain: The New Science of How We Read*. New York, NY: Penguin.

Deng, L., Yang, M., and Marcoulides, K. M. (2018). Structural equation modeling with many variables: a systematic review of issues and developments. *Front. Psychol.* 9:580. doi: 10.3389/fpsyg.2018.00580

D'Imperio, M., Dittinger, E., and Besson, M. (2016). "Prosodie et intonation : notions de base et données neuro-psycholinguistiques," in *Traité de Neurolinguistique*, eds S. Pinto, and M. Sato, (Paris: De Boeck), 133–145.

Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., and Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: a randomized control trial. *PLoS One* 10:e0138715. doi: 10.1371/journal.pone.0138715

Fodor, J. D. (2002). "Prosodic disambiguation in silent reading," in *Proceedings of the North East Linguistics Society*, Vol. 32, ed. M. Hirotani, (Amherst, MA: GSLA), 112–132.

Forti, L., Milani, A., Piersanti, L., Santarelli, F., Santucci, V., and Spina, S. (2019). "Measuring text complexity for Italian as a second language learning purposes," in *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, Florence, 360–368. doi: 10.18653/v1/w19-4438

Gómez-Domínguez, M., Fonseca-Mora, M. C., and Machancoses, F. H. (2019). First and foreign language early reading abilities: the influence of musical perception. *Psychol. Music* 47, 213–224. doi: 10.1177/0305735617746734

Gordon, R. L., Fehd, H. M., and McCandliss, B. D. (2015). Does music training enhance literacy skills? A meta-analysis. *Front. Psychol.* 6:1777. doi: 10.3389/fpsyg.2015.01777

Grabe, W., and Stoller, F. (2011). *Teaching and Researching Reading*, 2nd Edn. Harlow: Pearson Longman.

Hammill, D. D., Wiederholt, J. L., and Allen, E. A. (2006). *Test of Silent Contextual Reading Fluency*. Austin, TX: PRO-ED.

Hiebert, E., and Reutzel, D. R. (2014). *Revisiting Silent Reading: New Directions for Teachers and Researchers*, Ist Edn. Santa Cruz, CA: TextProject.

Iversen, J. R., Patel, A. D., and Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *J. Acoust. Soc. Am.* 124, 2263–2271. doi: 10.1121/1.2973189

Jordan, C. (2018). *Exploring a Possible Tonal Loop in Musicians and Non-Musicians and the Relationship between Musical Expertise and Cognitive Ageing*. Edinburgh: University of Edinburgh.

Kadota, S. (1987). The role of prosody in silent reading. *Lang. Sci.* 9, 185–206. doi: 10.1016/S0388-0001(87)80019-0

Kadota, S. (2002). *How Written and Spoken English are Related: the Cognitive Mechanism of Secondary Linguistic Comprehension*. Tokyo: Kuroshio Shuppan.

Kato, S. (2009). Suppressing inner speech in ESL reading: implications for developmental changes in second language word recognition processes. *Mod. Lang. J.* 93, 471–488. doi: 10.1111/j.1540-4781.2009.00926.x

Kempe, V., Bublitz, D., and Brooks, P. J. (2015). Musical ability and non-native speech-sound processing are linked through sensitivity to pitch and spectral information. *Br. J. Psychol.* 106, 349–366. doi: 10.1111/bjop.12092

Kenny, D. A. (2015). *Measuring Model Fit*. Available online at: http://www.davidakenny.net/cm/fit.htm (accessed December 10, 2019).

Kline, R. B. (2015). *Methodology in the social sciences. Principles and practice of structural equation modeling*, 4th edn. New York, NY: Guilford Press.

Koda, K. (2007a). *Insights into Second Language Reading: A Cross-Linguistic Approach. Reading in a Foreign Language*, Vol. 18. New York, NY: Cambridge University Press.

Koda, K. (2007b). Reading and language learning: crosslinguistic constraints on second language reading development. *Lang. Learn.* 57, 1–44. doi: 10.1111/0023-8333.101997010-i1

Koelsch, S. (2011). Toward a neural basis of music perception – a review and updated model. *Front. Psychol.* 2:110. doi: 10.3389/fpsyg.2011.00110

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., and Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an FMRI study. *Hum. Brain Mapp.* 30, 859–873. doi: 10.1002/hbm.20550

Koriat, A., Kreiner, H., and Greenberg, S. N. (2002). The extraction of structure during reading: evidence from reading prosody. *Mem. Cogn.* 30, 270–280. doi: 10.3758/BF03195288

Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605. doi: 10.1038/nrn2882

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev. Sci.* 9, F13–F21. doi: 10.1111/j.1467-7687.2006.00468.x

Law, L. N. C., and Zentner, M. (2012). Assessing musical abilities objectively: construction and validation of the profile of music perception skills. *PLoS One* 7:e52508. doi: 10.1371/journal.pone.0052508

Macalister, J. (2010). Speed reading courses and their effect on reading authentic texts?: a preliminary investigation. *Read. Foreign Lang.* 22, 104–116.

Macaruso, P., and Shankweiler, D. (2010). Expanding the simple view of reading in accounting for reading skills in community college students. *Read. Psychol.* 31, 454–471. doi: 10.1080/02702710903241363

MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychol. Methods* 1, 130–149. doi: 10.1037/1082-989X.1.2.130

Mankel, K., and Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proc. Natl. Acad. Sci. U.S.A.* 115, 13129–13134. doi: 10.1073/pnas.1811793115

McBride-Chang, C. (1995). What is phonological awareness? *J. Educ. Psychol.* 87, 179–192. doi: 10.1037/0022-0663.87.2.179

Melby-Lervåg, M., and Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: a meta-analysis of studies comparing first- and second-language learners. *Psychol. Bull.* 140, 409–433. doi: 10.1037/a0033890

Milovanov, R., Pietilä, P., Tervaniemi, M., and Esquef, P. A. A. (2010). Foreign language pronunciation skills and musical aptitude: a study of Finnish adults with higher education. *Learn. Individ. Differ.* 20, 56–60. doi: 10.1016/j.lindif.2009.11.003

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychol.* 2:142. doi: 10.3389/fpsyg.2011.00142

Pechmann, T., and Mohr, G. (1992). Interference in memory for tonal pitch: implications for a working-memory model. *Mem. Cogn.* 20, 314–320. doi: 10.3758/BF03199668

Perfetti, C. (2007). Reading ability: lexical quality to comprehension. *Sci. Stud. Read.* 11, 357–383. doi: 10.1080/10888430701530730

Perin, D. (1983). Phonemic segmentation and spelling. *Br. J. Psychol.* 74, 129–144. doi: 10.1111/j.2044-8295.1983.tb01849.x

Posedel, J., Emery, L., Souza, B., and Fountain, C. (2012). Pitch perception, working memory, and second-language phonological production. *Psychol. Music* 40, 508–517. doi: 10.1177/0305735611415145

Russak, S., and Saiegh-Haddad, E. (2011). Phonological awareness in Hebrew (L1) and English (L2) in normal and disabled readers. *Read. Writ.* 24, 427–442. doi: 10.1007/s11145-010-9235-1

Satorra, A. (1990). Robustness issues in structural equation modeling: a review of recent developments. *Qual. Quant.* 24, 367–386. doi: 10.1007/bf00152011

Satorra, A., and Bentler, P. M. (1988). "Scaling corrections for chi-square statistics in covariance structure analysis," in *Proceedings of the American Statistics Association*, Alexandria, 308–313.

Schellenberg, E. G. (2015). Music training and speech perception: a gene-environment interaction. *Ann. N. Y. Acad. Sci.* 1337, 170–177. doi: 10.1111/nyas.12627

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., and King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: a review. *J. Educ. Res.* 99, 323–338. doi: 10.3200/JOER.99.6.323-338

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition* 55, 151–218. doi: 10.1016/0010-0277(94)00645-2

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: the perils of overreliance on an "outlier" orthography. *Psychol. Bull.* 134, 584–615. doi: 10.1037/0033-2909.134.4.584

Slevc, L. R. (2012). Language and music: sound, structure, and meaning. *Wiley Interdiscip. Rev. Cogn. Sci.* 3, 483–492. doi: 10.1002/wcs.1186

Slevc, L. R., and Miyake, A. (2006). Individual differences in second-language proficiency. *Psychol. Sci.* 17, 675–681. doi: 10.1111/j.1467-9280.2006.01765.x

Sparks, R. L. (1995). Examining the linguistic coding differences hypothesis to explain individual differences in foreign language learning. *Ann. Dyslexia* 45, 187–214. doi: 10.1007/BF02648218

Sparks, R. L., Patton, J., Ganschow, L., and Humbach, N. (2012). Do L1 reading achievement and L1 print exposure contribute to the prediction of L2 proficiency? *Lang. Learn.* 62, 473–505. doi: 10.1111/j.1467-9922.2012.00694.x

Strait, D. L., Hornickel, J., and Kraus, N. (2011). Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behav. Brain Funct.* 7:44. doi: 10.1186/1744-9081-7-44

Swaminathan, S., and Schellenberg, E. G. (2017). Musical competence and phoneme perception in a foreign language. *Psychon. Bull. Rev.* 24, 1929–1934. doi: 10.3758/s13423-017-1244-5

Swaminathan, S., Schellenberg, E. G., and Venkatesan, K. (2018). Explaining the association between music training and reading in adults. *J. Exp. Psychol. Learn. Mem. Cogn.* 44, 992–999. doi: 10.1037/xlm0000493

Tanaka, A., and Nakamura, K. (2004). Auditory memory and proficiency of second language speaking: a latent variable analysis approach. *Psychol. Rep.* 95, 723–734. doi: 10.2466/pr0.95.3.723-734

Tichko, P., and Skoe, E. (2018). Musical experience, sensorineural auditory processing, and reading subskills in adults. *Brain Sci.* 8:77. doi: 10.3390/brainsci8050077

Tierney, A., and Kraus, N. (2014). Auditory-motor entrainment and phonological skills: precise auditory timing hypothesis (PATH). *Front. Hum. Neurosci.* 8:949. doi: 10.3389/fnhum.2014.00949

van den Boer, M., van Bergen, E., and de Jong, P. F. (2014). Underlying skills of oral and silent reading. *J. Exp. Child Psychol.* 128, 138–151. doi: 10.1016/j.jecp.2014.07.008

Wang, M., Perfetti, C. A., and Liu, Y. (2005). Chinese–English biliteracy acquisition: cross-language and writing system transfer. *Cognition* 97, 67–88. doi: 10.1016/j.cognition.2004.10.001

Wechsler, D., Coalson, D. L., and Raiford, S. E. (2008). *WAIS-IV Technical and Interpretive Manual*. San Antonio, TX: Pearson.

Wolf, M., and Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Sci. Stud. Read.* 5, 211–239. doi: 10.1207/S1532799XSSR0503_2

Yamashita, J. (2013). Word recognition subcomponents and passage level reading in a foreign language. *Read. Foreign Lang.* 25, 52–71.

Young-Kim, S., Liu, L., and Cao, F. (2017). Brain & language how does first language (L1) influence second language (L2) reading in the brain ? Evidence from Korean-English and Chinese-English bilinguals. *Brain Lang.* 171, 1–13. doi: 10.1016/j.bandl.2017.04.003

Yuan, K. H., and Bentler, P. M. (2007). "Robust procedures in structural equation modeling," in *Handbook of Computing and Statistics with Applications*, Vol. 1, ed. S. Y. Lee, (Amsterdam: Elsevier), 367–397. doi: 10.1016/S1871-0301(06)01017-1

Zentner, M., and Strauss, H. (2017). Assessing musical ability quickly and objectively: development and validation of the Short-PROMS and the Mini-PROMS. *Ann. N. Y. Acad. Sci.* 1400, 33–45. doi: 10.1111/nyas.13410

Zeromskaite, I. (2014). The potential role of music in second language learning: a review article. *J. Eur. Psychol. Stud.* 5, 78–88. doi: 10.5334/jeps.ci

Ziegler, J. C., and Goswami, U. (2006). Becoming literate in different languages: similar problems, different solutions. *Dev. Sci.* 9, 429–436. doi: 10.1111/j.1467-7687.2006.00509.x

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 16 June 2020
doi: 10.3389/fnins.2020.00619

# Timing Markers of Interaction Quality During Semi-Hocket Singing

*Alessandro Dell'Anna[1,2*†], Jeska Buhmann[1†], Joren Six[1], Pieter-Jan Maes[1] and Marc Leman[1]*

[1] *Department of Musicology – IPEM, Ghent University, Ghent, Belgium,* [2] *Department of Psychology, University of Turin, Turin, Italy*

Music is believed to work as a bio-social tool enabling groups of people to establish joint action and group bonding experiences. However, little is known about the quality of the group members' interaction needed to bring about these effects. To investigate the role of interaction quality, and its effect on joint action and bonding experience, we asked dyads (two singers) to perform music in medieval "hocket" style, in order to engage their co-regulatory activity. The music contained three relative inter-onset-interval (IOI) classes: quarter note, dotted quarter note and eight note, marking time intervals between successive onsets (generated by both singers). We hypothesized that singers co-regulated their activity by minimizing prediction errors in view of stable IOI-classes. Prediction errors were measured using a dynamic Bayesian inference approach that allows us to identify three different types of error called fluctuation (micro-timing errors measured in milliseconds), narration (omission errors or misattribution of an IOI to a wrong IOI class), and collapse errors (macro-timing errors that cause the breakdown of a performance). These three types of errors were correlated with the singers' estimated quality of the performance and the experienced sense of joint agency. We let the singers perform either while moving or standing still, under the hypothesis that the moving condition would have reduced timing errors and increased We-agency as opposed to Shared-agency (the former portraying a condition in which the performers blend into one another, the latter portraying a joint, but distinct, control of the performance). The results show that estimated quality correlates with fluctuation and narration errors, while agency correlates (to a lesser degree) with narration errors. Somewhat unexpectedly, there was a minor effect of movement, and it was beneficial only for good performers. Joint agency resulted in a "shared," rather than a "we," sense of joint agency. The methodology and findings open up promising avenues for future research on social embodied music interaction.

**Keywords: joint action, embodied interaction, expressive quality, timing, Bayesian inference**

OPEN ACCESS

***Edited by:***
*Marta Olivetti Belardinelli, Sapienza University of Rome, Italy*

***Reviewed by:***
*Evangelos Himonides, University College London, United Kingdom*
*Franco Delogu, Lawrence Technological University, United States*

***Correspondence:***
*Alessandro Dell'Anna*
*alessandro.dellanna@unito.it*

*[†]These authors share first authorship*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *16 December 2019*
***Accepted:*** *19 May 2020*
***Published:*** *16 June 2020*

***Citation:***
*Dell'Anna A, Buhmann J, Six J, Maes P-J and Leman M (2020) Timing Markers of Interaction Quality During Semi-Hocket Singing. Front. Neurosci. 14:619. doi: 10.3389/fnins.2020.00619*

## INTRODUCTION

Music is a rewarding and empowering activity (Chanda and Levitin, 2013; Fritz et al., 2013), having the capacity to connect people (Malloch and Trevarthen, 2009; Overy and Molnar-Szakacs, 2009), and increase their self-confidence, their feelings of wellbeing, for example after singing together (Kreutz, 2014), and their motivation, for example in treating neurological disorders

(Wan et al., 2010; MacDonald et al., 2012). While there exist other such facilitators of reward and empowerment, such as dance, ritual actions, sports, and other forms of joint actions (Sebanz et al., 2006), music is special in the sense that its social power is driven by auditory information, next to visual information. When you don't see what the other is doing, your action can still be perfectly synchronized. And when muscles and brains get synchronized, strong group bonding effects may occur (McNeill, 1995). Obviously, while the use of music may date back to the very beginning of human evolution (Honing et al., 2015), its power is still working in all kinds of human social activities, including academic meetings, banquets, funerals, rituals, football matches, concerts, festivals, and so on.

Music can drive joint actions, intended as a "form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment" (Sebanz et al., 2006: 70), and generate affects (e.g., of being in joint control and connected with others, see Keller et al., 2016 for a review). According to the minimal architecture hypothesis (Vesper et al., 2010), joint action can be investigated in terms of representations (the goal and tasks the subjects involved in it assign themselves), processes (prediction and monitoring of the various steps needed to accomplish it), and coordination smoothers (actions that simplify coordination). In this paper, we focus on timing prediction as a dynamic marker of the quality of a musical joint action in singing dyads and correlate it to subjective reports of that very quality and of the joint agency induced by it.

In recent studies, advantage has been taken of imposed musical tasks in order to understand how two or more subjects build representations of a joint action, employing both behavioral (Keller et al., 2007; Goebl and Palmer, 2009; Lesaffre et al., 2017), and neuroscientific (Arbib, 2013; Loehr et al., 2013; Keller et al., 2014) approaches to better understand music-based social empowerment (D'Ausilio et al., 2015). In a couple of works (Müller and Lindenberger, 2011; Müller et al., 2018), it has been shown that singing together (in a choir) implies also breathing and heart beating together, giving rise to a complex network of processes that, with the addition of body movements, imposes boundary conditions to its constituents (the singers), just like a "superordinate system." In other words, singing together consists of a "participatory sense-making" that spreads out in the dynamics of the interaction itself, back to the subjects who get individually affected by certain properties of the singing (De Jaegher and Di Paolo, 2007; Schiavio and De Jaegher, 2017). Thereby, interaction cannot be understood by analyzing one single subject at a time, but rather by analyzing the interaction itself (in the form of behavior relative to one another).

However, at this point the question arises as to how good a music interaction should be in order for the music's "bio-technological power" (Freeman, 2000) to become effective. A joint action such as singing can facilitate group-formation and generate feelings of connectedness, yet to what degree is this feeling depending on interaction qualities, that is, on the capacity to perform the rules as stated by the cultural context? In our opinion, few studies have addressed this question. Recent studies, indeed, use techniques that measure the bodily interaction of musicians mostly in view of timing, but quality as such is not addressed (Loehr et al., 2011; Eerola et al., 2018). To this effect, quality can be estimated through self-assessments by performers, or third persons. However, it is also of interest to consider the concept of joint agency. Pacherie (2012), in generalizing the concept of agency (Haggard and Eitam, 2015) from the individual to the group, distinguishes between a SHARED sense of joint agency and a WE sense of joint agency, pointing out that, if an action is a joint action, the resulting sense of agency should be a feeling of being in control of (at least) part of the joint action outcome. According to Pacherie's distinction, people may experience a SHARED sense of joint agency in small groups, with a certain degree of specialization among the different participants, but without hierarchies, while a WE-agency might be experienced in larger ensembles with less specialization among its members and (sometimes) directed by a leader. Fitting examples are a volleyball team for the former kind, and a volleyball stadium choreography for the latter and, from a musical point of view, a small combo and an orchestra, respectively. In both cases, Pacherie stresses the importance of predictability, but to a different degree: a SHARED sense of joint agency may draw upon a low predictability of the partners' action, whereas a WE-agency may draw upon a high predictability due to similarity among partners' actions (what she calls "coordination symmetry"). Pacherie's model has been successfully applied to dyads in studies comparing individual and shared control on a given joint action (Bolt et al., 2016; Bolt and Loehr, 2017). We extend this investigation so to include the WE-agency factor, trying to establish whether the (two, in our experiment) performers experience distinction from each other or blending into each other. The latter would imply a kind of boundary loss between agents.

To address the analysis of timing, we adopt a Bayesian inference framework to guide our methodological choices. Bayesian inference is the core approach behind the predictive coding theory (Vuust et al., 2009; Friston, 2010; Clark, 2016; Koelsch et al., 2019), a nowadays largely debated theory that sees the brain as an active generator of predictions, rather than a passive receptor of stimuli from the external world. Thanks to sensory feedback received in a continuous circular sensorimotor process, prediction errors are minimized for a given action the subject is about to accomplish. Brain networks thus formulate hypotheses about the possible state of the world (the prior) that are compared to the actual sensory information received (the error), in order to update the hypothesis (the posterior). Priors and posteriors are also called "beliefs," not in the sense of explicit propositions, but rather in the sense of probability distributions, hence, mainly latent variables. A continuous updating of such beliefs is allowed by acting on the world in ways that minimize the error, or sensory surprise. Such sensorimotor loops, then, work as "active inferences" (Adams et al., 2013).

In a social context of music interaction, feedback on timing is provided by the other interacting subjects' behavior. As stressed by Koelsch et al. (2019), music is a perfect case against which the predictive model may be tested, because the music's very syntactic structure implies rhythmic, melodic and harmonic expectancies, that is, prediction. We believe that accurate prediction of the partner's action is crucial in musical ensembles, even if it shows in different degrees, depending on musical genres, cultures, and kind of ensembles (Rohrmeier and Koelsch, 2012; Salimpoor et al., 2015). In this paper we focus on timing since this is one of the main features in which the quality of music is reflected, and probably the most tractable with our approach. The Bayesian inference framework is here used to develop a computational analysis that copes with prediction errors in conditions where timing can be unstable, thereby assuming that performers construct latent time-varying beliefs about their joint timing. Our quest for interaction quality is therefore also a quest for a proper methodology for estimating latent variables about joint timing. Importantly, such a timing marker has to be considered as a dynamic index of coordination insofar as it takes into account not only the timings of the two musicians separately and correlate them afterwards (for example, by means of windowed cross-correlation), but several inter-onset intervals constituted by the two interacting musicians' singing (see section "Materials and Methods"). It is worth stressing that, while timing errors may be due to a variety of factors (inability of reading the notes, general lack of ability to accurately follow to synchronize singing with beat, etc.), in this paper we are interested in the dynamics, rather than the causes, of timing quality.

Following the growing interest in embodied approaches to cognition (Thompson and Varela, 2001; Gallagher, 2005; Chemero, 2009), in particular music cognition (Iyer, 2002; Leman, 2007; Walton et al., 2015), the kinaesthetic dimension of musical performances has been widely explored in the last years, stressing the impact of body movements on both production and perception. The predictive coding framework, combined with these embodied approaches (Gallagher and Allen, 2016), can help explain how movement, along with other sensory modalities, could contribute to error minimization. Indeed, when body parts move in time with the music, their timing reflects the timing of the music and can help shape this timing during production (be it singing or playing an instrument, see Wanderley et al., 2005; Maes et al., 2014).

In the present study, we explore interaction quality in a singing dyad, taking advantage of the medieval "hocket" style, in which two (or more) musicians are required to build a melody together using strict alternation of notes. Our hypotheses are as follows:

1. The quality of music interaction is reflected in the timing of the joint action among performers. Performers can estimate their own interaction quality through continuous video annotation (video-stimulated recall) and they can assess the social effect of the interaction in terms of an assessment of their joint agency experienced during the interaction. As quality is reflected in timing, joint action timing can be measured as the performers' latent (or emerging) belief about joint timing. We predict that more accurate timing is correlated with higher quality, as reflected (i) in the performers' higher self-annotation of their own performed interaction quality, and (ii) in the performers' estimation of joint agency experiences.
2. Given the high similarity of the singers' music score (see **Figure 1**), a high quality in performing will correspond with a high sense of joint agency values, that is, by WE-agency.
3. Movement may help performers to make their timing more accurate. Indeed, since multiple senses take away uncertainty (according to the predictive coding theory) and movement is timing (according to embodiment theory), movement should affect quality.

# MATERIALS AND METHODS

## Ethics Statement

Participants were informed in advance about the task, the procedure and the technology used for measurement. They had the opportunity to ask questions and were informed that they could stop the experiment at any time. The ethics committee of the Faculty of Arts and Philosophy of Ghent University approved the study and the consent procedure.

## Participants

Fifteen couples of musicians were recruited (mean age $29.4 \pm 10.4$ years; 12 women), both participants being either men or women, so that their pitch range could match more easily. Only couples for which both participants knew one another were considered, in order to reduce performance stress for an intimate task such as singing. As musicians we considered people currently playing an instrument (or singing) with at least 5 years of regular (formal or informal) musical training (mean $10.1 \pm 9.7$), capable of singing a simple melody from sheet music.

## Task

In this experiment we let pairs of musicians sing "on stage" an interleaved melody provided on a score. They were told that their parts should never overlap, and that the combination of the parts would result in a melody consisting of an A- and a B-part. They were also instructed to try to keep going on with singing even if for some reason their interactive performance would break up. We asked the participants to sing the notes by producing the sound "ta" or "pa." The fact that these sounds start with a plosive facilitates automatic onset detection of sung notes, needed to extract inter-onset-intervals (IOIs). In hocket polyphonic style a single melody is broken down into two or more parts that never overlap, alternating almost regularly one tone after another. Here we use a semi-hocket technique for two singers, where alternation is somewhat less strict, meaning that sometimes a singer might sing two notes in a sequence (see the music score, **Figure 1**). We assume that the quality of the singing is reflected in the performers' timing of the sung notes, in particular also in the joint timing of the sung notes. In a good performance we expect the relative timing of the notes to correspond to the relative timing of the notes in the score, whereas a bad performance would contain note durations that do not correspond with those of the score. Due to a limited rehearsal time (5 min alone, 15 min together) the task was expected to be challenging, leading to different outcomes in performance quality. After the rehearsal, singers had to perform eight trials of two randomized conditions lasting two minutes each, either moving (four trials), or not

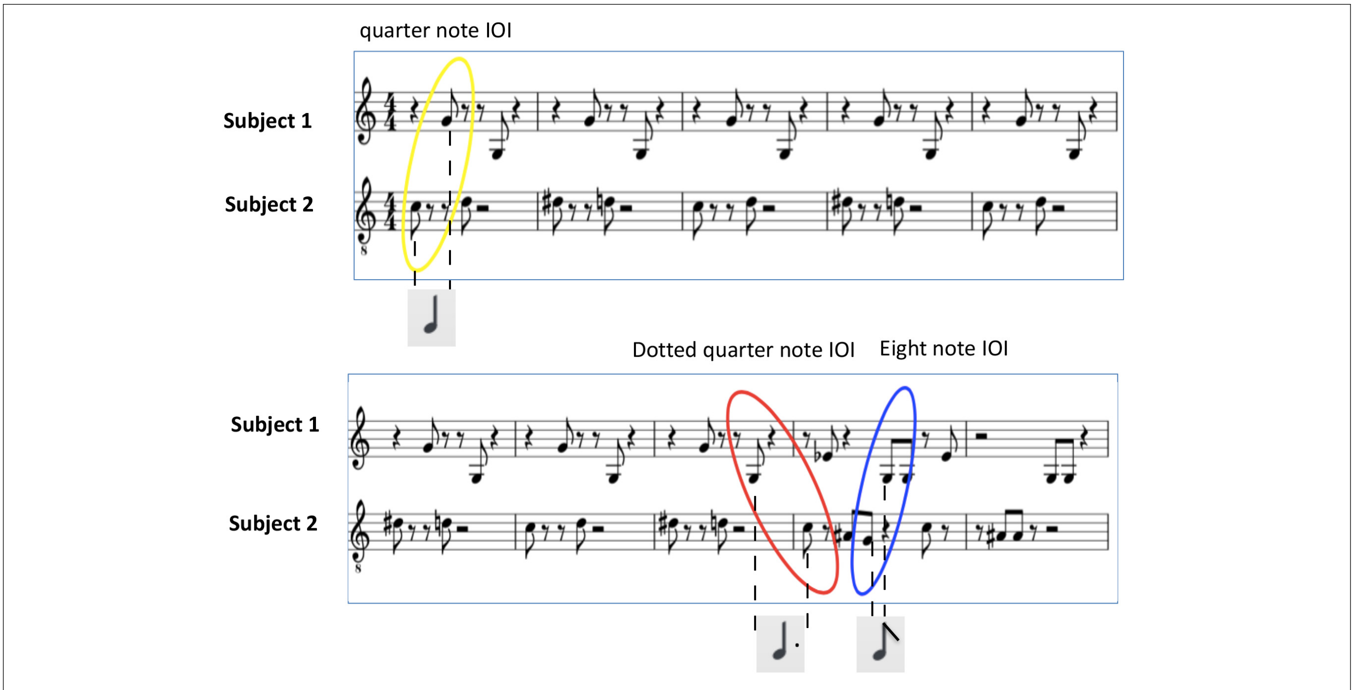



**FIGURE 1 |** Experimental stimulus. Part of the participants' scores. Together these scores form a semi-hocket, meaning that there are no simultaneous notes, and the combined scores merge into one melody. This melody is an adaptation of Michael Jackson's Billy Jean. The two parts contain an equal number of notes, displaying the same level of difficulty. In yellow, red and blue the three IOIs used in Bayesian regression (see below) are highlighted.

moving (four trials). In the non-movement trials participants were asked to stand as still as possible, while performing the singing task. In the movement trials participants were invited to move as they pleased while performing. This could result in simple hand- or foot-tapping, head-nodding, body-swaying, or even dance-like movements.

## Technical Setup

For each recording of the musical interaction task the two participants were standing on a force plate facing each other (see **Figure 2**). Both force plates have four weight sensors at the corners in order to register movement of the participants. The measured voltages are converted to MIDI control change (CC) messages by means of a Teensy 3.2 microcontroller. The MIDI stream was recorded in Ableton 9. The encoding of sensor signals into MIDI makes it straightforward to record audio and sensor data in sync using standard DAW software such as Ableton. A decoder script turns the MIDI into data fit for analysis. Participants were equipped with a headset containing a small microphone. The singing was thus captured and also recorded with Ableton. In addition, a video recording was made with a webcam (Logitech, c920). The webcam was modified to allow audio input. The audio input is connected to a SMPTE source to synchronize the video with the audio. These audio-visual recordings were used immediately after the recording session. The participants were requested to review their performances and annotate the quality level of their interaction. This was done via a script that synchronized and merged the audio and video recordings per trial. The scoring of the interaction happened on two separate computers via the mouse that could move a visual line up (better quality) and down (worse quality). The visual scores are stored as thousand samples of values between 0 and 127. The initial position of the cursor was set to value 64, a neutral starting point. All recording devices were connected with a master sync clock (Rosendahl Nanosyncs HD), preventing drift and enabling precise synchronization of audio movement, and video data.

## Procedure

Each couple was welcomed in our laboratory and, after filling in the informed consent, participants were explained that they had to build a melody together, combining their individual parts, stressing that these should never overlap, but almost always alternate one note after the other. Moreover, subjects were not allowed to read their partner's score. Then, they rehearsed their part in two separate rooms for 5 min, having the opportunity to listen to it once or twice in order to find the right pitch and learn the melody. Afterwards, they were gathered in the main lab, equipped with the headsets, and invited to get on "the stage," that is, on the two force plates facing each other at 1.5 m, to rehearse together for 15 min maximum, before the beginning of the real performance. After recording the eight trials, each participant individually executed a quality assessment task concerning the performance and the sense of joint agency (see below), without communicating with the partner. In total the experiment took between 1.5 and 2 h per couple.

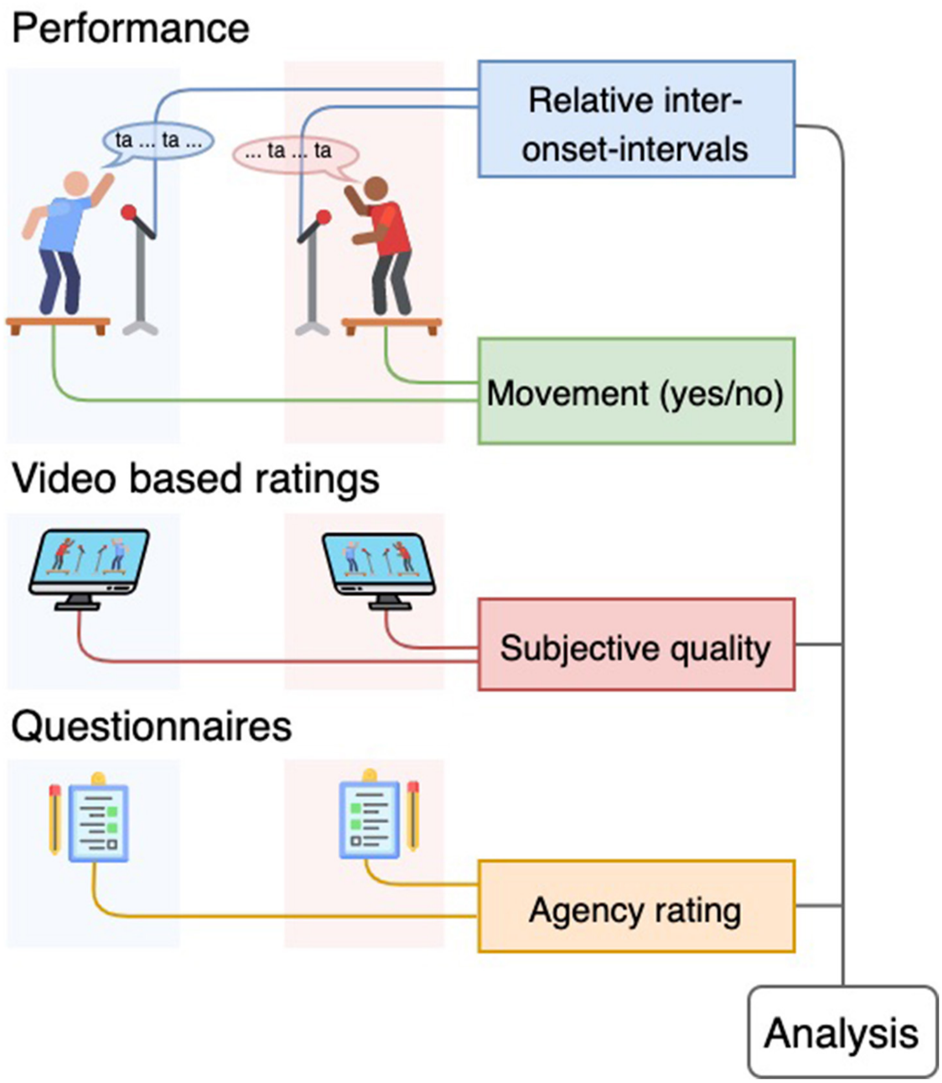



**FIGURE 2 |** Experimental procedure. Each experiment has three parts. The performance is followed by video based ratings and questionnaires. During performance the participants are allowed to move or not, which is checked by the movement outcome. Other measures are combined in the analysis.

## Data Pre-processing

### Audio Onset-Detection

As our approach has a focus on the timing of the singing, the audio recordings are reduced to the onsets of the singing, by doing an automatic onset detection in *Sonic Visualizer*, followed by a manual checking and correction step, that involved adding onsets in case they were not automatically added, and/or removing automatic onsets that did not accord with a sung note. The onsets are then converted into IOIs, that is durations that mark the time between two successive onsets (whoever sung them). The analysis is thus based on relative IOIs, that is, the IOIs formed by both performers. According to the score, a performance should result in three types of IOI durations, matching the durations of eighth notes, quarter notes, and dotted quarter notes (**Figure 1**). Depending on the tempo, a 2-min performance equals approximately singing the A- and B-parts four times. In theory this would result in 176 eighth notes, 96 quarter notes, and 47 dotted quarter notes.

## Subjective and Objective Markers of Interaction Quality

### Annotation and Questionnaires (Subjective)

The two participants of each couple were separately asked to assess the general quality of the interaction, that is, the performance as a whole, rather than the quality of their individual performance or the other participant's performance. This resulted in two time-series of quality values between 0 and 127 for each trial. Secondly, participants were asked to assess the joint sense of agency on a 7-point Likert scale. In particular, for each of the eight trials the subjects were asked to answer the question "When looking at the moments with the highest quality assessment, how was your feeling of control over the process on a scale between 0 (independent), 3 (shared), and 6 (complete unity with your partner)?" We explained this question by saying that the interaction could be either the product of two actions not really well coordinated between them (independent) or the product of two coordinated but distinct actions (shared), or the product of two actions that are not felt as different, but rather as the accomplishment of a single subject (unity or WE-agency).

### Third-Person Quality Assessment (Subjective)

As expected, we observed differences in the performance quality of the different couples. Given the fact that there was a large variation in performance quality, the authors of the paper agreed upon a subjective classification of the performances per duo into two groups, i.e., expert group and non-expert group. This was done by looking at the performance videos and evaluating the stability of the performance (could couples keep up their performances without too many break-ups) and how similar the performance was to what was written in the score. Six couples were assigned to the non-expert group and nine couples to the expert group. This subjective classification was done to validate our assumption that a good performance has less errors in the timing of the singing than a bad performance (**Figure 4**).

### Performance Timing Errors (Objective)

The score defines a musical norm for interactive performances, including rhythmic figures, tempo and an overall melodic narrative. However, due to the fact that the music emerges from the interaction, we assume that singers predict each other's performance in order to perform their own contribution correctly. As mentioned, not all performances may reach a high-quality level of interaction. Given the constraints of the musical rules, we consider three different types of prediction-errors, related to:

- *Fluctuation*: The fluctuation errors are defined as micro-timing (in milliseconds) prediction-errors that result from different sources such as timing-corrections due to small mistakes, due to active sampling, or even small onset measurement errors within the data pre-processing. Overall, fluctuation is a source of variance that can be considered necessary in order to maintain a stable performance state, even of high quality.
- *Narration*: The narration errors are defined as meso-timing (typically up to half a second, related to note durations) prediction-errors that may occur when a performer fails to follow the musical rule, for example, by forgetting a note, or making a mistake in note duration. Pitch is not taken into account, only timing. Overall, an error in the sequence (for example due to the omission of a note) may disturb the ongoing interaction. However, the dynamic system may be resilient enough to recover from such errors.
- *Collapse*: The collapse errors are defined as macro-timing (up to several seconds) prediction-errors that may occur when the performance, hence also the musical interaction, breaks down. The breakdown is catastrophic in the sense that both performers lose control of the expected musical narrative. This error is different from the narration errors that allow recovery due to resilience. To recover from such an interaction collapse, it may be necessary to start a new narrative from the beginning of the piece or the beginning of a section.

## Data Analysis

### Bayesian Inference Approach

As our data-analysis approach is based on the idea that performers try to reduce performance errors with respect to predictions, we consider performers as components of an interaction dynamics. We assume that each performer makes a prediction of the timing of the joint action (the interaction) based on a latent, or emergent, variable that estimates the timing of the relative IOIs in terms of milliseconds. As the piece contains only three different IOI classes, we assume that performers construct a latent variable for the estimated timing of each IOI-class. Obviously, the timings of the IOI classes are mutually constrained, thus contributing to a global latent variable, which is known as tempo. In our analysis we focus on how performed IOIs relate to the latent IOI-classes. Rather than inferring the prediction errors from an estimated global tempo (and proportional ratio of that tempo with respect to the IOI-classes) our method is tolerant to a systematic shortening or lengthening of IOI-classes according to performers' expressive timing preferences. The initial values of the variables that estimate the timing of the IOI-classes are set by a $k$-means clustering on all IOIs in three IOI-classes, using the first 15 s of a performance. Thereafter, a sequential Bayesian updating is performed for each of the IOI-classes separately, using a 15-s window of incoming IOI values (leading to the evidence distribution or the likelihood of measurement). Using Bayesian terminology, we interpret the prior as the mean of a distribution of old predicted durations of the IOI-class and the posterior as the mean of an updated distribution due to new evidence. This procedure is executed step by step (i.e., one IOI after another in the time series). It allows us to calculate the difference between the performed IOI and the predicted IOI, in milliseconds (**Figure 3**). For the entire performance, we calculate the root-mean-square error (RMSE) for each IOI-class, and take the average over all IOI-classes. This approach can deal with small changes in tempo and therefore, it accounts for the assumption of non-stationarity. In fact, for each IOI-class we use proportional timing errors by taking the log2 of the ratio of the measured IOI and the predicted IOI.

While the above approach may be working for fluctuation errors, we also have to consider the fact that IOIs may be wrongly classified due to narration errors. In order to account for these narration errors (which are restricted to duration errors), we keep track of the sung (duration) sequence and the expected corresponding IOI-class assignments. When expected IOIs get wrongly classified they are considered as narration errors, expressed in percentage of matching IOIs. Collapse errors are considered to be larger gaps in the performance (IOI durations that differ more than two standard deviations from the corresponding IOI-class prior), where normally onsets would have been expected. The collapse errors are expressed as

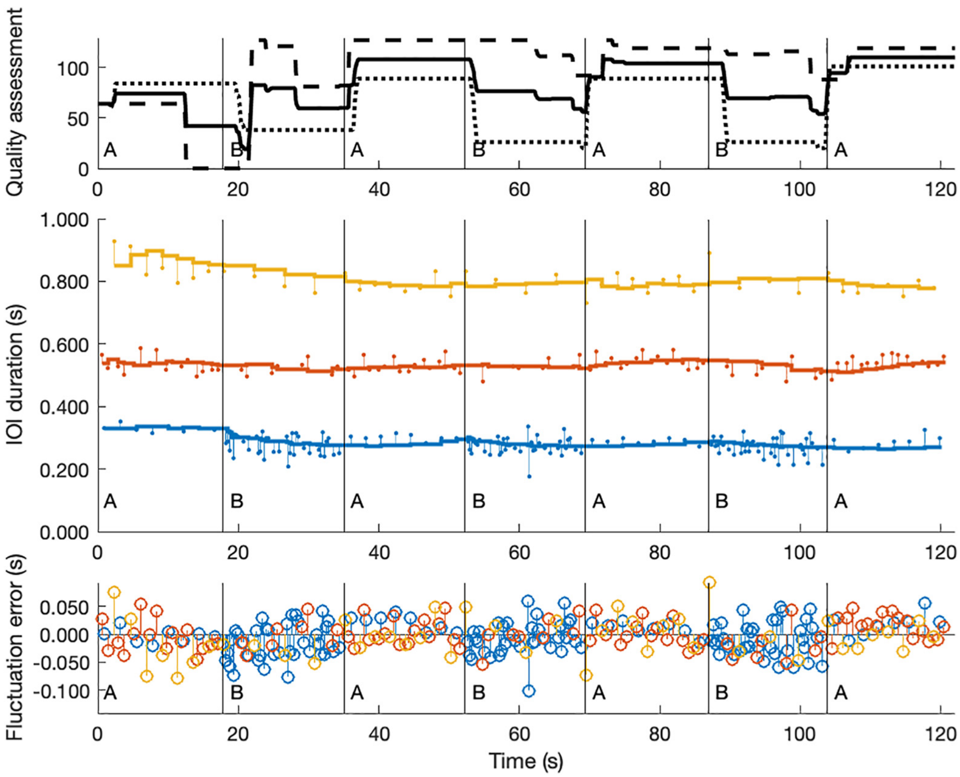



**FIGURE 3 |** Performance interaction measurements in time-series. In this figure, three time-series of the same performance (a 2-min trial) are visualized. The top plot shows the two quality assessments of the two participants (the dotted lines); the solid line represents the mean of the two assessments. The horizontal lines of the middle plot show how the priors of the three IOI classes evolve over time, according to our Bayesian sequential updating approach. The vertical lines with small dots indicate the deviations of the actual IOI durations with respect to those priors; in other words, they represent the errors in seconds. The plot at the bottom is a summary of the middle plot, where the zero-line represents the priors and the vertical lines are the errors (in seconds) for the three IOI classes (blue = short IOIs; orange = middle IOIs; and yellow = long IOIs).

a percentage of the number of collapses compared to the total number of IOIs in the performance.

### Correlations Between First-Person Viewpoints and Timing Errors

The correlation was calculated between the overall timing errors per trial and the average joint-agency scores that were indicated in the questionnaires. This was done for each performance-error type. Since the joint agency scores are not normally distributed and they contain a lot of identical values (7-point Likert scale), Kendall's Tau correlation was used. In a similar manner the correlation between timing errors and the quality assessment scores was calculated.

### The Effect of Movement and Expert Group on Performance Errors

In order to test our hypothesis that movement has an auxiliary function in error minimization during a joint singing task, for each type of performance error (fluctuation, narration, and collapse) we compare the average error value of the four movement trials with the average of the four non-movement trials. A 2 × 2 mixed ANOVA was performed with condition (movement/non-movement) as within-subject factor and expert-group (yes/no) as between-subjects factor. In a few cases the performance errors in a group were not normally distributed. When data distributions were not normal or the assumption of homogeneity of variance was violated, non-parametric tests were executed instead (Wilcoxon signed rank tests to compare movement with non-movement condition for each expert level).

### The Effect of Movement and Expert Group on Agency and Quality Assessment

To validate our hypothesis that movement and expert level have a positive impact (higher agency and quality scores for experts, while moving) on the subjective assessment of a performance interaction, a 2 × 2 mixed ANOVA was performed. Identical to the test on performance errors, condition (movement/non-movement) is the within-subject factor and expert-group (yes/no) the between-subjects factor.

### Movement Assessment

For each trial, continuous wavelet transforms were performed on the movement data of the two force plates, i.e., for each force plate the sensor that captured the highest amplitude. Only the wavelet information within the movement-relevant

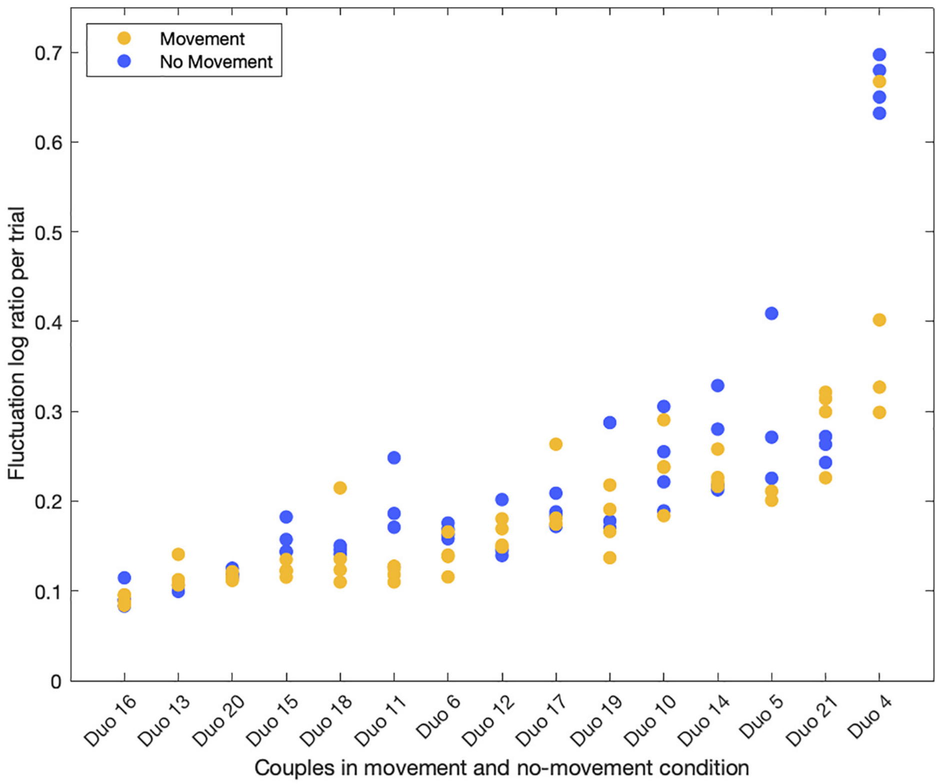



**FIGURE 4 |** Fluctuation errors per duo. The duos are ordered from smallest to largest average fluctuation error over the eight performance trials.

range of 0.25 to 5 Hz was considered. Within that range the frequency band with the highest average wavelet magnitude was selected. For each force plate this average magnitude was used to calculate the average movement magnitude for the couple. The right-skewed histogram of these values for all the non-movement trials covers a small range of magnitude values, with a maximum average magnitude value of 11. For the movement trials the histogram covers a much wider range of magnitude values, with a maximum of 88. In accordance with what was observed in the video recordings, a threshold of 25 was chosen as the cut-off for detected movement (above), or not (below).

# RESULTS

## Effect of Movement and Expert Group

### Performance Timing Errors

In total, nine out of the 120 performance trials (7.5%) were excluded from analysis. Two trials were excluded (the first and third trial of duo 5), because the participants made a lot of errors by singing (almost) simultaneously, resulting in IOIs that were too short to be valid eighth note durations. In other words, in these two trials the participants did not perform the singing task as described in the musical score. Seven more trails were excluded (duo 5, trial 2; duo 12, trial 1; duo 13, trial 2 and 4; duo 18, trial 3; duo 19, trial 5; and duo 21, trial 7), because too much movement was detected in the conditions where participants were instructed not to move, i.e., the continuous wavelet transforms of the sensor-data coming from the force plates revealed magnitude values that were higher than the defined threshold for non-movement.

Fluctuation errors are not significantly lower in the movement condition ($M$ = 0.185, $SE$ = 0.022) than in non-movement condition ($M$ = 0.217, $SE$ = 0.036), $F(1, 13) = 3.929$, $p = 0.069$, and $r = 0.48$. There was a significant effect of expert level, indicating that experts had lower error rates ($M$ = 0.133, $SE$ = 0.012 for movement; $M$ = 0.147, $SE$ = 0.013 for non-movement) than non-experts ($M$ = 0.245, $SE$ = 0.036 for movement; $M$ = 0.298, $SE$ = 0.064 for non-movement), $F(1, 13) = 7.938$, $p = 0.015$, and $r = 0.62$. No significant interaction effect was found between movement and expert level, $F(1, 13) = 1.436$, $p = 0.252$, and $r = 0.32$.

Narration matching in the movement condition ($M$ = 84.54, $SE$ = 3.43) is not significantly different from that in the non-movement condition ($M$ = 82.90, $SE$ = 3.69), $F(1, 13) = 1.437$, $p = 0.252$, and $r = 0.32$. There was a significant effect of expert level, indicating that experts had higher percentages of predictable IOI classes ($M$ = 94.92, $SE$ = 1.44 for movement; $M$ = 93.18, $SE$ = 1.98 for non-movement) than non-experts ($M$ = 72.67, $SE$ = 3.46 for movement; $M$ = 71.15, $SE$ = 4.45 for non-movement), $F(1, 13) = 31.913$, $p < 0.001$, and $r = 0.84$. No significant interaction effect was found between movement and expert level, $F(1, 13) = 0.007$, $p = 0.937$, and $r = 0.02$.

For collapse errors, a Wilcoxon signed ranks test revealed that for non-experts a significantly higher percentage of collapses occurred in the movement condition ($Mdn$ = 8.66) than in the non-movement condition ($Mdn$ = 6.18), $z = -2.366$, $p = 0.018$, and $r = -0.43$. However, for experts the percentage of collapses were not different for moving ($Mdn$ = 0.45), and not moving ($Mdn$ = 0.45), $z = -0.700$, $p = 0.484$, and $r = -0.13$.

### Joint Agency and Quality Assessments

With respect to agency, a Wilcoxon signed ranks test revealed that for non-experts, agency ratings were significantly higher when moving ($Mdn$ = 3.28) than when not moving ($Mdn$ = 2.63), $z = -2.043$, $p = 0.041$, and $r = -0.39$. However, for experts, agency ratings were not significantly different for moving ($Mdn$ = 3.88), and not moving ($Mdn$ = 3.13), $z = -0.762$, $p = 0.446$, and $r = -0.14$.

The quality assessment in the movement condition is not significantly different from that in the non-movement condition, $F(1, 13) = 1.880$, $p = 0.194$, and $r = 0.36$. There was a significant effect of expert level, indicating that experts gave higher annotation scores ($M$ = 89.70, $SE$ = 4.90 for movement; $M$ = 83.99, $SE$ = 5.90 for non-movement) than non-experts, ($M$ = 69.03, $SE$ = 5.56 for movement; $M$ = 63.70, $SE$ = 3.54 for non-movement) $F(1, 13) = 11.477$, $p = 0.005$, and $r = 0.68$. No significant interaction effect was found between movement and expert level, $F(1, 13) = 0.002$, $p = 0.962$, and $r = 0.01$.

## Correlations of Performance Timing Errors

All types of performance error are significantly correlated with one another. **Table 1** shows the correlation values and their corresponding significance values for the three types of performance errors.

## Correlations of Performance Timing Errors With Agency and Quality Assessments

Fluctuation errors are negatively correlated with agency assessments, although correlation values are low ($\tau = -0.15$). The lower the fluctuation error, the higher the agency assessment value. Narration is positively correlated with quality assessments. The higher the percentage of predictable IOI classes, the higher the agency assessment value. Collapse errors are negatively correlated with quality assessments. The lower the percentage of collapses, the higher the agency assessment value.

With respect to the quality assessments, higher correlations are found. Fluctuation is negatively correlated with quality assessment. The lower the fluctuation error, the higher the quality score. Narration is positively correlated with quality assessments. The higher the percentage of predictable IOI classes, the higher the quality assessment value. Collapse errors are negatively correlated with quality assessments. The lower the percentage of collapses, the higher the quality assessment value. **Table 2** shows all Kendall's tau correlation coefficients and the corresponding significance values.

# DISCUSSION

The present paper investigated whether the quality of interaction, while performing music, plays a role in the establishment of joint action and group bonding experiences. The hypothesis that interaction quality plays a role was tested with singing dyads. We thereby focused on timing markers. We achieved three main outcomes. Firstly, we found correlations, albeit weak, between the sense of joint agency and measured fluctuation, narration and collapse errors. Contrary to our prediction, the highest degrees of joint agency reached by the dyads point to a SHARED rather than a WE sense of agency, particularly in the movement condition. Secondly, we found correlations between the self-annotated performance quality and measured fluctuation, narration and collapse errors. Although movement as such did not produce overall improvement in the quality of the performances, we observed a tendency for participants to reduce fluctuation errors while moving. On the contrary, non-expert dyads showed more collapse errors in that condition. These results point toward a different kind of effect of movement on micro- and macro-timing: when movement might possibly reduce micro-timing errors in general (recall that the difference was close to significant, with a medium effect size), it disrupts the performance on a macro-timing level for less experienced performers. Finally, we contributed to a novel and effective methodology and framework to analyse the objective quality of the interaction from the point of view of timing.

An important limitation of our study concerns the fact that the assessment of the feeling of joint agency was done after watching the performance recording, while moments of joint agency are supposed to occur during the performance. The correlation results are promising, but they point toward the need for a more refined method for estimating agency. The idea of using a hocket composition in our study was also inspired by Bolt et al. (2016)'s discovery that sequences of (twelve) tones played at the piano by pairs of non-musicians, first by one subject and then by the second one, resulted in lower values of joint sense of agency compared to when the subjects alternated a tone after another. Moreover, these authors found that objective coordination between the subjects (measured by means of cross-correlation of the tones' onset series) impacted on joint agency, enhancing it when the coordination was strong. These findings are coherent with ours, though the data were obtained with different analytical tools and in a study that did not deal with expressive quality. In the present paper, we were also interested in the kind of joint agency such a performance could induce. Therefore we administered a questionnaire asking the subjects an assessment of their experienced sense of joint agency, stressing that the lowest values indicated an independent control, the medium values a shared control and the highest values a complete unity with the partner in controlling the musical joint action. Since the average collected values were in the medium range, our results point toward a SHARED rather than a WE sense of agency. This outcome complies, indeed, with Pacherie's definition of the two kinds of joint agency (Pacherie, 2012), in particular when she suggests that a SHARED agency would ensue from a small group joint action, in which roles can be easily distinguishable. At this

**TABLE 1 |** Performance error correlations.

| | Fluctuation vs. narration | | Fluctuation vs. nollapse | | Narration vs. nollapse | |
|---|---|---|---|---|---|---|
| ***Trials*** | **τ** | ***p*** | **τ** | ***p*** | **τ** | ***p*** |
| All | −0.67 | <0.001 | 0.46 | <0.001 | −0.52 | <0.001 |
| Movement | −0.60 | <0.001 | 0.50 | <0.001 | −0.57 | <0.001 |
| No movement | −0.75 | <0.001 | 0.46 | <0.001 | −0.50 | <0.001 |

**TABLE 2 |** Performance error correlations with subjective performance assessments.

| | | Fluctuation | | Narration | | Collapse | |
|---|---|---|---|---|---|---|---|
| | ***Trials*** | **τ** | ***p*** | **τ** | ***p*** | **τ** | ***p*** |
| Agency | All | −0.15 | 0.028 | 0.18 | 0.010 | −0.24 | 0.001 |
| | Movement | −0.10 | 0.311 | 0.17 | 0.083 | −0.31 | 0.002 |
| | No movement | −0.16 | 0.126 | 0.22 | 0.030 | −0.26 | 0.013 |
| Quality | All | −0.37 | < 0.001 | 0.40 | < 0.001 | −0.41 | < 0.001 |
| | Movement | −0.35 | < 0.001 | 0.35 | < 0.001 | −0.48 | < 0.001 |
| | No movement | −0.35 | < 0.001 | 0.43 | < 0.001 | −0.36 | < 0.001 |

moment, we can speculate that this feature overcame the high similarity we intentionally established between the two scores. Indeed, according to Pacherie, the high predictability of, and, as a consequence, low necessity to keep oneself distinguishable from, the partner could have caused a WE-agency, rather than a SHARED agency (see Fairhurst et al., 2013 for a similar idea and some neuro-scientific possible account of it). Sticking to this result, we may then conclude that, on average, our musical task did not induce any boundary loss between the subjects in the pair, but we cannot exclude that the difficulty of the task contributed to prevent it. All in all, this finding adds to the debate on joint agency not only in musicology, but also in the wider domain of cognitive science (van der Wel et al., 2012; Dewey et al., 2014; Bolt and Loehr, 2017).

Subjective self-annotations of the quality of the performance have to be treated carefully as well. The correlation results are promising and they seem to indicate that performance quality can be self-assessed in a proper way, although improvements to our slider approach in the video-stimulated recall protocol are still possible. Here, an important limitation of our study consisted in the latency between the recorded performance and the annotation the subjects did by means of the slider, meaning that the assessment cannot match perfectly the moments it refers to, but it is always a bit late. Furthermore, we asked the subjects to assess the quality of the performance as a whole, without focusing on timing, since we were interested also in other expressive features like pitch and tuning (whose analysis we are bracketing in the present study). Yet, given the crucial role of timing in music and its capacity to create social bonding in synchronization tasks (Hove and Risen, 2009; Wiltermuth and Heath, 2009; Kokal et al., 2011), we assumed timing was the main feature to be analyzed in our study. The good level of musicianship declared by our subjects, and visible in many of their performances, should bolster the validity of the correlation we found. Of course, not all couples reached the same quality levels, as it is manifest from both the objective and subjective measurements and from **Figure 4**, which shows the clustering of each couple's trials according to their fluctuation errors. Yet, we think that considering the relationships between those measurements gave us some hint about a proper treatment of the expressive quality in a singing dyad.

Also, the relatively large number of rejected data may induce some improvement of our paradigm. Indeed, most of the rejected trials were due to the fact that subjects did not comply with the experimental condition, either moving when they were supposed not to do so or singing completely different than what was in the score. Some kind of feedback, either a visual or an auditory feedback, could inform the subject about his/her passing a given movement threshold, thus allowing to adjust for it. After all, both visual and auditory bio-feedback systems may be conceived of in order to adjust the performance itself according to the amount of (mainly fluctuation and narration) errors collected in a given time interval. This is how we see a relevant application of our method aiming at enhancing musical learning processes (see Moens and Leman, 2015, for some applications of the same principle to running and walking to the music).

In this paper we developed a novel methodology to capture the interaction of a singing dyad. While the method was applied to the emergent timing of both singers, the method allows an analysis of each singer separately, despite possible changes in tempo. In accordance with the recent emphasis on the predictive coding framework (Friston, 2010; Clark, 2016), also in music studies (Vuust et al., 2009; Koelsch et al., 2019), we applied a Bayesian inference approach to dynamically analyse a semi-hocket interaction between two subjects. In fact, a singing dyad can be conceived of as a dynamical system whose components constrain each other's unfolding performance (Müller and Lindenberger, 2011; Konvalinka and Roepstorff, 2012; Müller et al., 2018), considering its variability and correcting for it, when needed. A sequential Bayesian process allowed for an analysis in the form of a continuous updating of timing-error minimisation. We focused on timing and identified fluctuation, narration and collapse errors as objective, third-person markers of the quality of a musical interaction, exploiting the idea that the "superordinate system," i.e., the dyad, rather than the single singer, constructed predictions of latent variables that keep track of the timing of each relative IOI. This approach has the advantage that we look finer in time than a method that would focus on the overall tempo. Obviously, it can be questioned whether this construct has any psychological plausibility, yet the emergence of latent variables is a known phenomenon, and in full agreement with the predictive coding approach. For example, the concept of latent variables that work as predictors for observable/measurable action can be compared with the two processes postulated to correct errors in a sensorimotor synchronization task at the individual level, phase correction and period correction, the former being an almost automatic process with which fluctuation errors can be equated, the latter requiring a conscious effort comparable to the one needed to overcome narration errors (Wing et al., 2010; Repp and Su, 2013). The distinction between fluctuation, narration and collapse errors was introduced in order to deal with typical performance errors. Fluctuation may be related to subconscious active sampling in order to be able to update the latent variable on timing. Further research is needed to refine its sources of variability. Narration relates to a symbol-based account of the performance and therefore, we assume that it has a cognitive origin related to memory and sequencing. While collapse errors induce a complete breakdown of the performance, the singers may still cope with narration errors (possibly with period correction), even if they surely threaten the quality of the performance. We believe that the Bayesian inference framework offers a useful method for assessing musical expression in high quality music performance. As our concept is based on relative IOIs, the method offers the perspective that it can be applied to groups comprising three and more singers and musicians.

Finally, movement did not improve the performance timing, but the fact that the worse couples made more collapse errors in the movement condition, along with the higher joint agency values reported in that condition and a tendency for all participants to reduce their fluctuation errors in that condition, suggests that above a certain level movement may impact on the overall quality of the performance. In particular, this result could imply that, while for bad couples movement constitutes an interference with their task, good couples may benefit from it at a micro-timing level. This hypothesis is compatible with a Bayesian approach insofar as bad couples, by definition, find it difficult to both coordinate their movements with the music and their singing with the partner's, that is, predicting the music and the partner at the same time. On the other hand, active inference may be enhanced by moving for those couples that are already fluent, but can take further advantage from moving at a micro-timing level. However, further research is surely needed to better disentangle the network of dynamic processes that is constituted by prediction, agency and movement in musical expressive moments (Leman, 2016).

As far as we know, this is the first study that applies principles of the predictive coding approach to a social musical interaction. And it does so by stressing the dynamic character of the interaction thanks to a parameter, the relative IOI, which treats two subjects as one, hence taking seriously the Gestalt concept that the whole is more than the sum of its parts. The same idea is implicit in the concept of participatory sense-making (De Jaegher and Di Paolo, 2007), which emphasizes that the sense of a joint action is not given in advance, but it is co-constituted by the interactive subjects. In a musical context, thereby, the musical object is not constituted either by the score or by the representations in the minds of each musician, not even by the auditory event in itself, but rather by the embodied interaction of the musicians on the fly (Schiavio and De Jaegher, 2017). The focus on the interaction, rather than on the single components of it, increases the complexity of studying an already complex phenomenon like music, although also in the domain of cognitive neurosciences several appeals have been recently made toward such a perspective change. For example, Schilbach et al. (2013) write that "After more than a decade of research, the neural mechanisms underlying social interaction have remained elusive and could – paradoxically – be seen as representing the 'dark matter' of social neuroscience" (ibidem: 394). Hyper-scanning, the simultaneous acquisition of cerebral data from two or more subjects, is a promising technique to approximate such an ambitious aim (Konvalinka and Roepstorff, 2012; Babiloni and Astolfi, 2014). Indeed, though not yet analyzed, not only did our experiment carry out a motion capture collection of data from the singing dyads, but it also planned the physiological recording of skin conductance by means of portable bracelets. Moreover, we are working exactly on the possibility to simultaneously electroencephalography (EEG) recording two interacting musicians, in search of the brain basis of social embodied music interaction. Such an empowered set-up would likely allow both to test the psychological plausibility of a dynamic marker of timing as the one we devised in the present paper and to identify possible dynamic neural markers of timing and other musical features and processes (see also Osaka et al., 2015; Nozaradan et al., 2016; Pan et al., 2018). Ultimately, such enterprise would probably require a thorough theoretical synthesis between embodied and predictive approaches to (music) cognition, of which the present work can be seen as a first empirical application.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Faculty of Arts and Philosophy of Ghent University. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

AD developed the experimental design, executed the experiment, and wrote the manuscript. JB executed the experiment, analyzed data, and wrote the manuscript. JS build the technical set-up for the experiment. P-JM contributed to the analysis and technical set-up building. ML conceived the analysis and contributed to write the manuscript.

## REFERENCES


Adams, R., Stewart, S., and Friston, K. (2013). Predictions not commands. Active inference in the motor system. *Brain Struct. Funct.* 218, 611–643. doi: 10.1007/s00429-012-0475-5

Arbib, M. (2013). *Language, Music, and the Brain: A Mysterious Relationship*. Cambridge, MA: MIT Press.

Babiloni, F., and Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: past, present and future. *Neurosci. Biobehav. Rev.* 44, 76–93 doi: 10.1016/j.neubiorev.2012.07.006

Bolt, N. K., Poncelet, E. M., Schultz, B. G., and Loehr, J. D. (2016). Mutual coordination strengthens the sense of joint agency in cooperative joint action. *Conscious. Cogn.* 46, 173–187. doi: 10.1016/j.concog.2016.10.001

Bolt, N., and Loehr, J., (2017). The predictability of a partner's actions modulates the sense of joint agency. Cognition 161, 60–65. doi: 10.1016/j.cognition.2017.01.004

Chanda, M. L., and Levitin, D. J. (2013). The neurochemistry of music. *Trends Cogn. Sci.* 17, 179–193. doi: 10.1016/j.tics.2013.02.007

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Clark, A. (2016). *Surfing Uncertainty*. Oxford: Oxford University Press.

D'Ausilio, A., Novembre, G., Fadiga, L., and Keller, P. E. (2015). What can music tell us about social interaction? *Trends Cogn. Sci.* 19, 111–114. doi: 10.1016/j.tics.2015.01.005

De Jaegher, H., and Di Paolo, E. (2007). Participatory sense-making: an enactive approach to social cognition. *Phenomenol. Cogn. Sci.* 6, 485–507. doi: 10.1007/s11097-007-9076-9

Dewey, J. A., Pacherie, E., and Knoblich, G. (2014). The phenomenology of controlling a moving object with another person. *Cognition* 132, 383–397. doi: 10.1016/j.cognition.2014.05.002

Eerola, T., Jakubowski, K., Moran, N., Keller, P., and Clayton, M. (2018). Shared periodic performer movements coordinate interactions in duo improvisations. *R. Soc. Open Sci.* 5:171520. doi: 10.1098/rsos.171520

Fairhurst, M. T., Janata, P., Keller, P. E. (2013). Being and feeling in sync with an adaptive virtual partner: brain mechanisms underlying dynamic cooperativity. *Cereb. Cortex* 23, 2592–2600.

Freeman, W. J. (2000). "A neurobiological role of music in social bonding," in *The Origins of Music*, eds WallinMerkerBrown (Cambridge, MA: The MIT Press), 411–424.

doi: 10.1038/nrn2787Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138.

Fritz, T., Hardikar, S., Demoucron, M., Niessen, M., Demey, M., Giot, O., et al. (2013). Musical agency reduces perceived exertion during strenuous physical performance. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17784–17789. doi: 10.1073/pnas.1217252110

Gallagher, S. (2005). *How the Body shapes the Mind*. Oxford: Oxford University Press.

Gallagher, S., and Allen, M. (2016). Active inference, enactivism and the hermeneutics of social cognition. *Synthese* 195, 2627–2648. doi: 10.1007/s11229-016-1269-8

Goebl, W., and Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Percept.* 26, 427–438. doi: 10.1525/mp.2009.26.5.427

Haggard, P., and Eitam, B. (2015). *The Sense of Agency*. Oxford: Oxford University Press.

Honing, H., Cate, C., Peretz, I., Trehub, S. E., (2015). Without it no music: cognition, biology and evolution of musicality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140088. doi: 10.1098/rstb.2014.0088

Hove, M. J., and Risen, J. L. (2009). It's all in the timing: interpersonal synchrony increases affiliation. *Soc. Cogn.* 27, 949–960. doi: 10.1521/soco.2009.27.6.949

Iyer, V. (2002). Embodied mind, situated cognition, and expressive micro-timing in African-American music. *Music Percept.* 19, 387–414. doi: 10.1525/mp.2002.19.3.387

Keller, P. E., Novembre, G., and Loehr, J. D. (2016). Musical ensemble performance: representing self, other, and joint action outcomes. in *Shared Representations: Sensorimotor Foundations of Social Life*, eds S. Obhi and E. Cross (Cambridge: Cambridge University Press), 280–311.

Keller, P. E., Knoblich, G., Repp, B. H. (2007). Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. *Conscious. Cogn.* 16, 102–111. doi: 10.1016/j.concog.2005.12.004

Keller, PE, Novembre, G, Hove, MJ. (2014). Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philos. Trans. R. Soc. B* 369:20130394. doi: 10.1098/rstb.2013.0394

Koelsch, S., Vuust, P., and Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends Cogn. Sci.* 23, 63–77. doi: 10.1016/j.tics.2018.10.006

Kokal, I., Engel, A., Kirschner, S., and Keysers, C. (2011). Synchronized drumming enhances activity in the caudate and facilitates prosocial commitment – if the rhythm comes easily. *PLoS One* 6:e27272. doi: 10.1371/journal.pone.0027272

Konvalinka, I., and Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Front. Hum. Neurosci.* 6:215. doi: 10.3389/fnhum.2012.00215

Kreutz, G. (2014). Does singing facilitate social bonding? *Music Med.* 6, 50–61.

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press.

Leman, M. (2016). *The Expressive Moment*. Cambridge, MA: MIT Press.

Lesaffre, M., Maes, P. J., Leman, M. (2017). *Routledge Companion to Embodied Music Interaction*. London: Routledge.

Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., and Günther, K. (2013). Monitoring individual and joint action outcomes in duet music performance. *J. Cogn. Neurosci.* 25, 1049–1061. doi: 10.1162/jocn_a_00388

Loehr, J. D., Large, E. W., and Palmer, C. (2011). Temporal coordination and adaptation to rate change in music performance. *J. Exp. Psychol.* 37, 1292–1309. doi: 10.1037/a0023102

MacDonald, R., Kreutz, G., and Mitchell, L. (2012). *Music, Health and Wellbeing*. Oxford: Oxford University Press.

Maes, P.-J.., Leman, M., Palmer, C., and Wanderley, M. M. (2014). Action-based effects on music perception. *Front. Psychol.* 4:1008. doi: 10.3389/fpsyg.2013.01008

Malloch, S., and Trevarthen, C. (2009). *Communicative Musicality*. Oxford: Oxford University Press.

McNeill, W. H., (1995). *Keeping Together in Time*. Cambridge, MA: Harvard University Press.

Moens, B., and Leman, M. (2015). Alignment strategies for the entrainment of music and movement rhythms. *Ann. N. Y. Acad. Sci.* 1337, 86–93. doi: 10.1111/nyas.12647

Müller, V., and Lindenberger, U. (2011). Cardiac and respiratory patterns synchronize between persons during choir singing. *PLoS One* **6**:e24893. doi: 10.1371/journal.pone.0024893

Müller, V., Delius, J. A. M., and Lindenberger, U. (2018). Complex networks emerging during choir singing. *Ann. N. Y. Acad. Sci.* 1431, 85–101. doi: 10.1111/nyas.13940

Nozaradan, S., Peretz, I., and Keller, P. (2016). Individual differences in rhythmic cortical entrainment are correlated with predictive behavior in sensorimotor synchronization. *Nat. Sci. Rep.* 6:20612.

Osaka, N., Minamoto, T., Yaoi, K., Azuma, M., Minamoto Shimada, Y., and Osaka, M. (2015). How two brains make one synchronized mind in the inferior frontal cortex: fNIRS-based hyper-scanning during cooperative singing. *Front. Psychol.* 6:1811 doi: 10.3389/fpsyg.2015.01811

Overy, K., and Molnar-Szakacs, I., (2009). Being together in time: musical experience and the mirror neuron system. *Music Percept.* 26, 489–504 doi: 10.1525/mp.2009.26.5.489

Pacherie, E., (2012). The phenomenology of joint action: self-agency vs. joint-agency. in *Joint Attention: New Developments*, ed. A. Seemann (Cambridge, MA: MIT Press), 343–389.

Pan, Y., Novembre, G., Song, B., Li, X., and Hu, Y. (2018). Interpersonal synchronization of inferior frontal cortices tracks social interactive learning of a song. *Neuroimage* 183, 280–290. doi: 10.1016/j.neuroimage.2018.08.005

Repp, B. H., and Su, Y. H. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). *Psychonom. Bull. Rev.* 20, 403–452. doi: 10.3758/s13423-012-0371-2

Rohrmeier, M. A., and Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *Int. J. Psychophysiol.* 83, 164–175. doi: 10.1016/j.ijpsycho.2011.12.010

Salimpoor, V. N., Zald, D., Zatorre, R., Dagher, A., and MacIntosh, A. R. (2015). Predictions and the brain: how musical sounds become rewarding. *Trends Cogn. Sci.* 19, 86–91. doi: 10.1016/j.tics.2014.12.001

Schiavio, A., and De Jaegher, H. (2017). "Participatory sense-making in joint musical practice," in *The Routledge Companion to Embodied Music Interaction*, eds M. Lesaffre, P. J. Maes, and M. Leman (Milton Park: Taylor and Francis), 31–39. doi: 10.4324/9781315621364-4

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/s0140525x12000660

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76. doi: 10.1016/j.tics.2005.12.009

Thompson, E., and Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends Cogn. Sci.* 5, 418–425. doi: 10.1016/s1364-6613(00)01750-2

van der Wel, R. P. R. D., Sebanz, N., and Knoblich, G. (2012). The sense of agency during skill learning in individuals and dyads. *Conscious. Cogn.* 21, 1267–1279. doi: 10.1016/j.concog.2012.04.001

Vesper, C., Butterfill, S., Knoblich, G., and Sebanz, N. (2010). A minimal architecture for joint action. *Neural Netw.* 23, 998–1003. doi: 10.1016/j.neunet.2010.06.002

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., and Roepstorff, A. (2009). Predictive coding of music- brain responses to rhythmic incongruity. *Cortex* 45, 80–92. doi: 10.1016/j.cortex.2008.05.014

Walton, A., Richardson, M. J., Langland-Hassan, P., and Chemero, A. (2015). Improvisation and the self- organization of multiple musical bodies. *Front. Psychol.* 6:313. doi: 10.3389/fpsyg.2015.00313

Wan, C., Rüber, T., Hohmann, A., and Schlaug, G. (2010). The therapeutic effects of singing in neurological disorders. *Music Percept.* 27, 287–295. doi: 10.1525/mp.2010.27.4.287

Wanderley, M., Vines, B., Middleton, N., McKay, C., and Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: an exploration of the field. *J. New Music Res.* 34, 97–113. doi: 10.1080/09298210500124208

Wiltermuth, S. S., and Heath, C. (2009). Synchrony and cooperation. *Psychol. Sci.* 20, 1–5 doi: 10.1111/j.1467-9280.2008.02253.x

Wing, A. M., Doumas, M., and Welchman, A. E. (2010). Combining multisensory temporal information for movement synchronisation. *Exp. Brain Res.* 200, 277–282. doi: 10.1007/s00221-009-2134-5

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 24 June 2020
doi: 10.3389/fnhum.2020.00210

# Tension Experience Induced By Nested Structures In Music

*Lijun Sun[1,2], Chen Feng[1,2] and Yufang Yang[1,2]**

[1]Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, [2]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

Tension experience is the basis for music emotion. In music, discrete elements are always organized into complex nested structures to convey emotion. However, the processing of music tension in the nested structure remains unknown. The present study investigated the tension experience induced by the nested structure and the underlying neural mechanisms, using a continuous tension rating task and electroencephalography (EEG) at the same time. Thirty musicians listened to music chorale sequences with non-nested, singly nested and doubly nested structures and were required to rate their real-time tension experience. Behavioral data indicated that the tension experience induced by the nested structure had more fluctuations than the non-nested structure, and the difference was mainly exhibited in the process of tension induction rather than tension resolution. However, the EEG data showed that larger late positive components (LPCs) were elicited by the ending chords in the nested structure compared with the non-nested structure, reflecting the difference in cognitive integration for long-distance structural dependence. The discrepancy between resolution experience and neural responses revealed the non-parallel relations between emotion and cognition. Furthermore, the LPC elicited by the doubly nested structure showed a smaller scalp distribution than the singly nested structure, indicating the more difficult processing of the doubly nested structure. These findings revealed the dynamic tension experience induced by the nested structure and the influence of nested type, shedding new light on the relationship between structure and tension in music.

**Keywords: tension, resolution, nested structure, LPC, integration**

OPEN ACCESS

***Edited by:***
*Franco Delogu,
Lawrence Technological University,
United States*

***Reviewed by:***
*Hamid Reza Marateb,
Universitat Politecnica de Catalunya,
Spain
José Manuel Reales,
National University of Distance
Education (UNED), Spain*

****Correspondence:***
*Yufang Yang
yangyf@psych.ac.cn*

***Specialty section:***
*This article was submitted to
Cognitive Neuroscience, a section of
the journal Frontiers in Human
Neuroscience*

***Received:*** *10 January 2020*
***Accepted:*** *08 May 2020*
***Published:*** *24 June 2020*

***Citation:***
*Sun L, Feng C and Yang Y
(2020) Tension Experience Induced
By Nested Structures In Music.
Front. Hum. Neurosci. 14:210.
doi: 10.3389/fnhum.2020.00210*

## INTRODUCTION

Musical tension is one of the core principles evoking musical emotions, playing an important role in musical listening (Lehne et al., 2013; Lehne and Koelsch, 2015). As the link between auditory stimuli and subjective experience, tension experience relies on the cognition of complex structures through the process of expectation build-up, violation, and fulfillment (Margulis, 2005; Huron, 2006; Rohrmeier and Koelsch, 2012). Indeed the relationship between tension and structure was depicted in the generative theory of tonal music (GTTM; Lerdahl and Jackendoff, 1983) and the tonal tension model (TTM; Lerdahl and Krumhansl, 2007). It was suggested that tension experience was highly hierarchical, based on the harmonic stability of each chord/note in the passage of tonal music. Depending on tonally hierarchical positions in Western tonal music, the patterns of tension and resolution were presented through a tree notation.

Through manipulating the tonal function of certain chords, previous studies have corroborated the GTTM and the TTM finding that unstable chords and structural breaches induced tension experience in short chord sequences (Bigand et al., 1996; Bigand and Parncutt, 1999; Steinbeis et al., 2006; Lerdahl and Krumhansl, 2007). However, there are far more complicated structures in real music and structure–tension relationships in music listening.

Discrete elements in music are organized into complex structure through finite state grammar (FSG) and phrase structure grammar (PSG; Rohrmeier et al., 2014; Ma et al., 2018b). It is the PSG, rather than the FSG, that organizes a set of finite elements into infinite sentences/phrases in the form of nested tree structures to express complicated and rich meaning (Chomsky, 1957, 1965; Fitch and Martins, 2014), constituting the core cognitive faculty of the human beings (Fitch and Hauser, 2004; Makuuchi et al., 2009; Dehaene et al., 2015). In terms of PSG, the discrete elements in music are always organized in a subordinate or dominant way (Lerdahl and Jackendoff, 1983; Longuet-Higgins, 1987; Rohrmeier, 2011; Prince and Schmuckler, 2014), such as the harmonic progression of A (the original key)—B (new key)—A (return to the original key), with a new key embedded in the original key at a higher level. Given the importance of PSG in music, it is essential for us to uncover the tension experience induced by the nested structure, which would shed new light on the relationship between structure and emotion in music.

As a core principle in the tension models (Lerdahl and Jackendoff, 1983; Lerdahl and Krumhansl, 2007), prolongational reduction assigns to pitches a hierarchical structure that expresses tension and relaxation. Thus, the type of hierarchical structures in music can influence the way of prolongational reduction and the tension–resolution pattern. Local and simple tension–resolution patterns are organized in a hierarchical fashion, forming a global and complex tension–resolution pattern. Indeed numerous tension arches are usually interweaved into large-scale tension arches in Western music (Koelsch, 2013). For example, in the case of the harmonic progression of C major—F major—C major, the occurrence of the tonal modulation of F major key induces tension experience because out-of-key chords violate the mental representation based on the original tonal context (Steinbeis et al., 2006; Lerdahl and Krumhansl, 2007). Meanwhile, listeners remember the beginning C major key and expect the subsequent unfolding musical events to modulate to the beginning tonality (Meyer, 1956). Thus, when the C major key returns, the listeners would integrate the nested F major key into the C major key context based on their knowledge of nested structure and acquire a resolution experience (Schenker, 1979; Krumhansl and Kessler, 1982). However, the tension experience induced by the nested structure remains unknown.

Although little research has revealed tension–resolution patterns induced by nested structure, the cognitive processing of nested structure and long-distance dependence in music has been explored by several studies. Behavioral studies found that listeners had difficulty perceiving a higher-level organization of musical structure, especially for the completeness and coherence of large-scale tonal relationship (Gotlieb and Konecni, 1985; Cook, 1987; Karno and Konecni, 1992; Deliege et al., 1996; Tillmann and Bigand, 1996, 2004; Tillmann et al., 1998). Until recently, Koelsch et al. (2013) first explored the neural responses to the processing of nested structure and found that structurally irregular endings elicited larger early right anterior negativity (ERAN) and N5 components than structurally regular endings, reflecting the structural integration for long-distance dependence. Similar components, such as N5 and late positive component (LPC), were also observed in music nested structure processing by Chinese listeners (Ma et al., 2018a,b; Zhou et al., 2019). These results demonstrated the integration of harmonic cadence into the originally tonal context and the cognitive processing of nested structure.

Considering the important contribution of structure to emotion in music, the present study examined the tension experience induced by the nested structure and its underlying neural mechanisms. We manipulated the structural type while keeping the cadence unchanged and created three conditions as follows: non-nested structure, singly nested structure, and doubly nested structure. Frequent key changes were included in the nested structure but not in the non-nested structure, leading to different ways of prolongational reduction. A real-time tension rating task was employed, that is, the tension value was continuously recorded during the unfolding of the whole pieces to reflect the dynamic and time-varying characteristics of tension experience (Fredrickson, 2000; Hackworth and Fredrickson, 2010; Schubert, 2010). Given that the pattern of tension experience was determined by prolongational reduction, we predicted that the nested structure would induce higher tension than the non-nested structure due to frequent key modulations. In particular, the more complex the structure was, the more fluctuations would occur in the pattern of tension and resolution experience. Furthermore, given that the key point representing the tonality return was the cadence in the nested structure, the event-related potential (ERP) responses were locked to the final chords. Based on the cognitive processing of nested structure reported in previous studies (Koelsch et al., 2013; Ma et al., 2018a,b; Zhou et al., 2019), we also predicted that larger N5 or LPC would be elicited by the nested structure compared to the non-nested structure, reflecting the cognitive processing of long-distance structural integration.

# MATERIALS AND METHODS

## Participants

*A priori* sample size was calculated using G*power (G*power version 3.1.9.4), and the result indicated that a sample of 27 was required in our study to reach 90% power and for detecting an effect size of $f = 0.30$, with $\alpha = 0.05$. The effect size was based on a previous study investigating the processing of nested structure in music (Ma et al., 2018a). Therefore, we recruited 30 subjects for our experiment. Given that the processing of doubly nested structure may be difficult for nonmusicians, we recruited musicians who had received more than 8 years of formal musical training and played at least one musical instrument. Then, we randomly selected 30 musicians ($M_{age}$ = 22.34 years, SD = 2.49, 20 females) to participate in the experiment. They were graduate students in music colleges and had received formal Western instrumental training, such as piano, violin, viola, and cello, for an average of 16 years (8–19 years). In music colleges, they learned many Western music theory curricula, including Western harmony, polyphony, orchestration, music form, history of Western music, etc. They were all right-handed and all of them had no history of neural impairment or psychiatric illness. The study was approved by the Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences, in accordance with the ethical principles of the Declaration of Helsinki. All the participants provided informed consent.

## Stimuli

Ten original chorale sequences including ten bars were composed in a 2/4 meter. The original sequences started with a tonic chord in C, A, or G major keys, and then developed around the key and ended with a harmonic cadence from dominant to tonic chords. The original sequences with non-nested structure unfolded in a single key and had no modulation in the middle of the sequences.

The nested sequences were obtained by modulating the key in the middle of the original sequences. For the singly nested structure, the chords were not changed until the first chord in measure 3, which was the featured chord in the dominant key. The featured chord included one pitch that was in-key in the present key but out-of-key in the previous key, signifying the modulation to the dominant key. Then, the sequences developed around the dominant key until the second chord in measure 8, which was the featured chord in the initial key. For the doubly nested structure, the chords were the same as the singly nested structure except those in measures 4 and 5. The first chord in measure 4 was the featured chord in the double dominant key, following which the sequences developed around the double dominant key in these two measures. In both the singly and doubly nested conditions, the endings of the sequences were harmonic progressions from dominant to tonic chord in the initial key, signifying the modulation return into the beginning key (see **Figure 1** for an example). Ten original sequences and 20 corresponding modified versions were all transposed to three other keys, yielding 120 sequences in total (10 excerpts × 4 keys × 3 structures). Using the Sibelius 7.5 software, we created the stimuli and adopted a Yamaha piano timbre with a velocity of 100 through the Cubase 5.1 software. All the stimuli files were played at a tempo of 100 beats per minute.

## Procedure

The sequences were presented in a pseudorandom order such that a given condition could not be repeated more than three times in succession and the same original sequence could not be consecutive sequences. The participants were required to judge the experienced tension continuously while listening to music, which was recorded by the Psychopy 1.0. software interface at a sampling rate of 20 Hz. They indicated the tension level by the position of a slider bar at the center of the screen, which was controlled by moving the mouse up or down. At the beginning of each trial, the slider was set to a quarter of the whole bar in order to prevent the participants' rating out of the bar scope over the whole music piece. Three practice trials were performed before the formal experiment to familiarize the participants with the stimuli and the procedure. The stimuli were presented binaurally through Audio Technica CKR30iS headphones.

## EEG Recording and Analysis

Electroencephalography (EEG) data were recorded by Brain Products with 64 Ag/AgCI electrodes in International 10-20 system scalp locations at the sampling rate of 500 Hz. FCz was used as an online reference electrode. The electrode between Fz and FPz served as the ground electrode, and the electrode placed below the right eye was used to track eye movements. We kept the impedance of all electrodes less than 5 kΩ during the whole experiment.

The raw EEG data were preprocessed with EEGLAB (Delorme and Makeig, 2004) in MATLAB. First of all, the data were referenced to the algebraic mean of the left and the right mastoid electrodes. Second, the data were filtered offline with the Basic FIR Filter function implemented in EEGLAB to remove linear trends. We set 0.1 Hz with a filter order of 13,750 points as the lower edge and 30 Hz with a filter order of 220 points as the higher edge of the frequency pass band. Then, the data were segmented into epochs of 1,400 ms, ranging from −200 to 1,200 ms relative to the final chord. Each trial was baseline-corrected using the 200-ms prestimulus interval, and ocular and muscle artifacts were corrected using an independent component analysis algorithm (Makeig et al., 1997; Delorme and Makeig, 2004) implemented in EEGLAB. Trials in any electrode exceeding ±75 μV were regarded as artifacts and rejected. The threshold of artifact rejection was consistent with previous studies (e.g., Ellis et al., 2015; Sun et al., 2018; Zhang et al., 2018). Finally, average ERPs were calculated for each participant at each electrode in each condition.

Based on previous studies (Besson and Faïta, 1995; Patel, 1998; Regnault et al., 2001; Zendel et al., 2015) and the visual inspection, we selected 650–900 ms as the time window of LPC, the mean amplitudes of which were entered into statistical analysis. The ERPs were analyzed statistically in four regions of interest: left anterior electrodes (F1, F3, F5, FC1, and FC3), right anterior electrodes (F2, F4, F6, FC2, and FC4), left posterior electrodes (P1, P3, P5, CP1, and CP3), and right posterior electrodes (P2, P4, P6, CP2, and CP4). Repeated-measures ANOVAs taking condition (non-nested vs. singly nested vs. doubly nested), laterality (left vs. right), and anteriority (anterior vs. posterior) as within-subject factors were conducted. We conducted Mauchly's test to test the assumption of sphericity in repeated-measures designs. If the assumption of sphericity was not met, the $p$-values corrected by the Greenhous–Geisser method were reported.

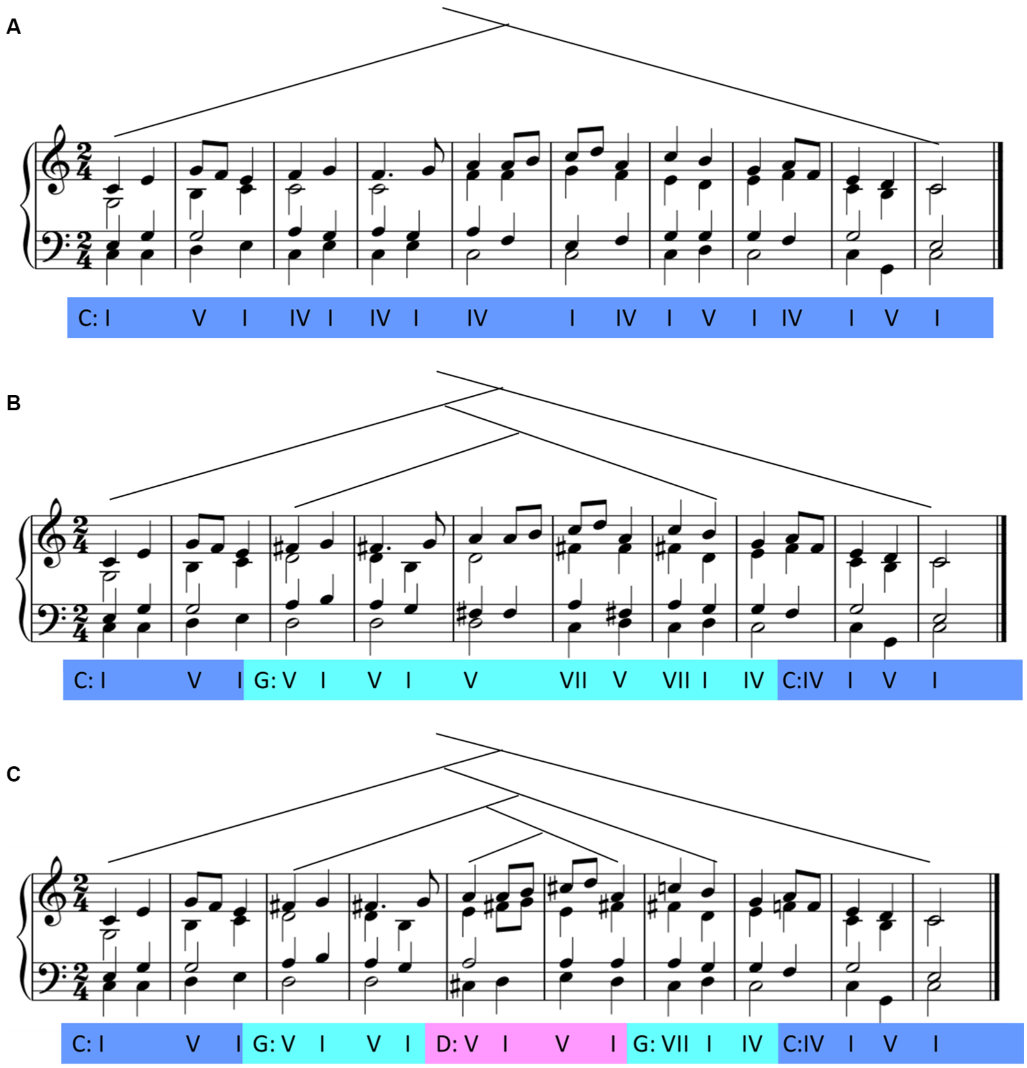



**FIGURE 1 |** Samples of the musical sequences used in the study. Sequences with non-nested structure **(A)**, singly nested structure **(B)**, and doubly nested structure **(C)**.

Simple effect tests and planned comparisons were conducted when there were any interactions with critical manipulations in ANOVAs. Bonferroni correction was applied to adjust the multiple comparisons.

# RESULTS

## Behavioral Results

At first, all data were normalized to Z-scores for each participant to minimize the differences across participants in terms of the slider ranges, which were also used by previous studies (e.g., Farbood, 2012; Lehne et al., 2013; Gingras et al., 2016). The tension values averaged across all participants are presented in **Figure 2**, showing dynamic changes in tension over the course of the whole musical sequences under the three conditions. **Figure 3** exhibits the median, the first and the third quartiles, and the highest and the lowest tension values under each condition.

Given the response delays that existed in the tension rating task, we did not choose a specific time window to calculate the average tension values. Instead we first calculated the range between the highest and the lowest tension values and conducted repeated-measures ANOVA (rmANOVA) analysis with the structural types as the main factor. The results found a significant main effect of structure ($F_{(1,29)} = 27.00$, $p < 0.001$, partial $\eta^2 = 0.48$). Further paired comparisons among the three conditions showed that the ranges in the doubly and the singly nested structures were larger than in the non-nested structure (doubly: $p < 0.001$; singly: $p = 0.001$), and the range in the doubly nested condition was wider than that in the singly nested condition ($p < 0.001$; non-nested: $M = 3.00 \pm 0.98$; singly nested: $M = 3.41 \pm 1.00$; doubly

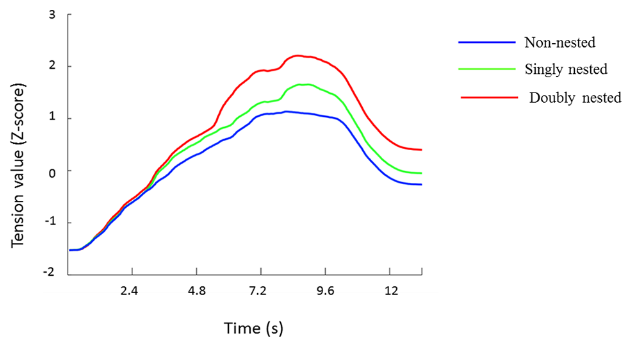



**FIGURE 2 |** Mean ratings (Z-score) of tension values for each stimulus type at each time point. The color scheme codes represent non-nested, singly nested and doubly nested structures, respectively.

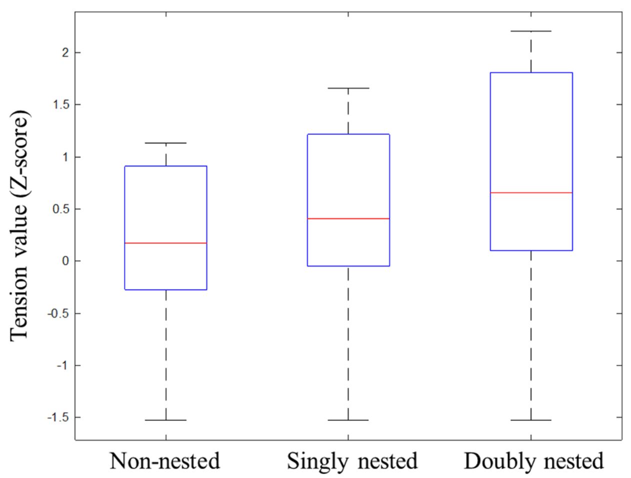



**FIGURE 3 |** Box plots of Z-scores of tension values for the tension ratings under non-nested, singly nested and doubly nested conditions. These box plots contain the extreme of the lower whisker, the lower hinge, the median, the upper hinge, and the extreme of the upper whisker. The two hinges are the first and third quartiles, and the whiskers extend to the most extreme data.

nested: $M$ = 4.02 ± 1.03). Second, we analyzed the tension peaks for each subject and conducted one-way rmANOVA analysis. The results showed a significant main effect of structure ($F_{(1,29)}$ = 21.24, $p < 0.001$, partial $\eta^2 = 0.42$), indicating that the tension peak in the doubly and the singly nested structures was larger than in the non-nested structure (doubly: $p < 0.001$; singly: $p = 0.005$) and larger in the doubly nested structure than in the singly nested structure ($p < 0.001$; non-nested: $M$ = 1.23 ± 0.97; singly nested: $M$ = 1.53 ± 1.01; doubly nested: $M$ = 2.08 ± 1.23).

In order to examine the process of tension induction and resolution, respectively, the difference between the original and the highest tension value (tension induction) and the difference between the highest and the final tension value (tension resolution) were calculated for each subject under each condition. In terms of tension induction, the one-way rmANOVA results showed a significant main effect of structure ($F_{(1,29)}$ = 16.70, $p < 0.001$, partial $\eta^2 = 0.37$), indicating that the tension difference in the doubly and the singly nested structures was larger in the non-nested structure (doubly: $p < 0.001$; singly: $p = 0.011$) and larger in the doubly nested condition than in the singly nested condition ($p = 0.005$; non-nested: $M$ = 2.76 ± 1.13; singly nested: $M$ = 3.07 ± 1.02; doubly nested: $M$ = 3.06 ± 1.02). In terms of tension resolution, the one-way rmANOVA results also revealed a significant main effect of structure ($F_{(1,29)}$ = 4.19, $p = 0.041$, partial $\eta^2 = 0.13$). However, the multiple-comparisons results showed no significant difference in any paired comparisons ($ps > 0.07$; non-nested: $M$ = 1.46 ± 1.01; singly nested: $M$ = 1.61 ± 1.53; doubly nested: $M$ = 1.72 ± 1.32).

## ERP Results

**Figure 4A** shows the brain electrical responses to non-nested, singly nested and doubly nested structures. **Figure 4B** shows the scalp distributions of the singly nested structure minus the non-nested structure and the doubly nested structure minus the non-nested structure difference waves. In the time window of 650–900 ms, the final chords in both the singly nested and the doubly nested structures elicited a larger positivity compared to the non-nested structure. However, the LPC effect elicited by the singly nested structure was distributed in the whole scalp, while the effect elicited by the doubly nested structure was only distributed in the posterior scalp.

For the time window of 650–900 ms, the one-way rmANOVA revealed an effect of structure ($F_{(1,29)}$ = 5.94, $p = 0.004$, partial $\eta^2 = 0.17$). Moreover, there was an interaction between structure and regions ($F_{(1,26)}$ = 4.09, $p = 0.022$, partial $\eta^2 = 0.12$). A further simple-effect analysis revealed that the final chords in the singly nested condition elicited a larger positivity in both the anterior ($p = 0.048$) and the posterior regions ($p = 0.001$; anterior: $M$ = 0.65 ± 0.27; posterior: $M$ = 0.94 ± 0.23). However, the final chords in the doubly nested condition elicited a larger positivity in the posterior regions ($p = 0.002$) than in the anterior regions ($p = 1.00$; anterior: $M$ = 0.20 ± 0.29; posterior: $M$ = 0.79 ± 0.20). No other significant main effect or interaction was found (all $ps > 0.09$).

# DISCUSSION

The present study investigated musical tension induced by music sequences with nested structure and the underlying neural mechanisms, using tension experience ratings in real time and EEG recordings simultaneously. We found that the tension experience induced by the nested structure had more fluctuations than by the non-nested structure, and the difference was mainly exhibited in tension induction rather than in tension resolution. However, it was shown that a larger LPC was induced by the ending chord in nested structure compared with that in non-nested structure, and the LPC for singly nested structure had a broader scalp distribution than that for doubly nested condition, indicating that the processing of doubly nested structure was more difficult for the listeners. Following is the discussion of our main findings.

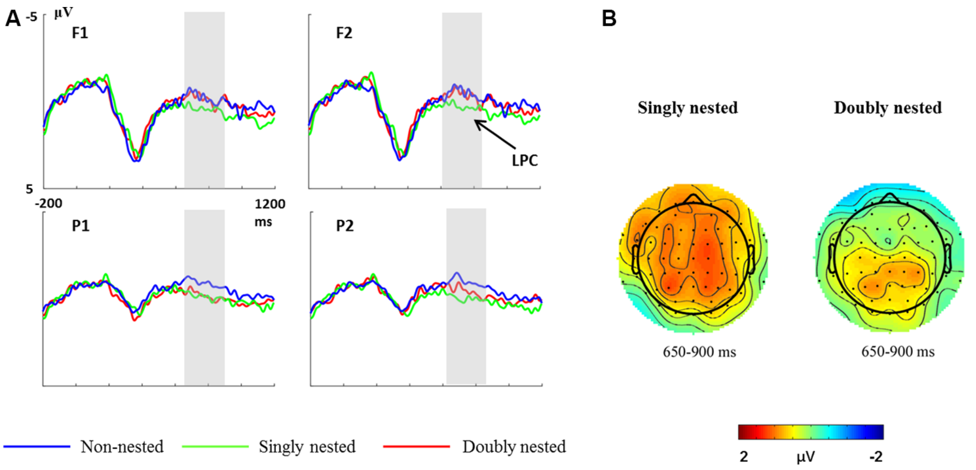



**FIGURE 4 | (A)** Grand mean event-related potential (ERP) waveforms elicited by final chords in non-nested, singly nested and doubly nested structures at four electrode sites. The gray-shaded areas indicate the time window used for statistical analysis. **(B)** Scalp distributions of the singly nested minus non-nested structure difference waves and the doubly nested minus non-nested structure difference waves for the 650–900 ms latency range.

## The Whole Dynamic Tension Curves Induced by Nested Structures

The tension curves showed different tension patterns induced by nested and non-nested conditions. The tension variation range was wider for the nested conditions than for the non-nested condition because of their higher tension rising speed in the tension induction processes.

Previous studies using short chord sequences have provided evidence that tonal breaches can induce tension experience because of their violation of the established mental representations of tonal context and the prediction for the upcoming notes (Meyer, 1956; Bigand et al., 1996; Margulis, 2005; Steinbeis et al., 2006). In our study, more out-of-key chords and tonal modulations were included in the nested conditions than in the non-nested conditions, whereas rhythmic patterns and melodic contours were controlled to be consistent. Therefore, the tension increases could be more likely attributed to the frequent key modulations in the nested structures. The overlap of each tension arch associated with key modulation led to the tension increases in the sequences with nested structures. Unfortunately, we could not obtain evidence from the EEG data. In order to ensure the same final chord in each condition, the acoustic elements changed in the middle of the sequences, which hindered us from locking any specific chords to analyze ERPs in the tension induction process.

In contrast to the tension induction process, the tension curves in the resolution process were almost parallel to each other, and no significant difference was found between the tension reduction values, defined as the difference between the highest and the final tension values of each curve. Based on the assumption of prolongational reduction, all tension arches should be closed at the end of the sequences and the maximum amount of resolution should be reached (Koelsch, 2014). However, our results suggested that the listeners' tension experience was not resolved by each key returning in a hierarchical way in the nested conditions. This could be attributed to the difficulty in perceiving and memorizing harmonic relationships in multiple nested hierarchical structures. In our study, the nested structures were shaped in the short chorale sequences with frequent key modulations. Thus, the ambiguous expectations for the tonal returning in the nested conditions might bring about very subtle variations in emotional experience and weaken the resolution experience.

From a dynamic perspective, tension resolution was slower than the process of tension induction and difficult to be resolved completely at the ending of music pieces. The reason might be that tension induction elicited by out-of-key chords was related to local violations, whereas tension resolution would be shaped by global integration. Several studies have demonstrated the difficulty of global processing in music using scrambled music pieces at different time scales and found that bar-level, but not phrase-level scrambling influenced the perception of tonal structure (Tillmann and Bigand, 1998; Granot and Jacoby, 2011). It has also been confirmed by an ERP study that the information in the local context has an earlier influence than in the global context, as reflected by an early ERAN component for local violation rather than for global violation (Zhang et al., 2018). Further study is needed to examine the difference of tension experience elicited by the processing of local and global structures.

## The Discrepancy Between Cognitive and Emotional Responses to the Final Chord

In our study, the final chords in both the singly nested and the doubly nested structures elicited larger LPCs compared with the non-nested structure. LPC is an ERP correlate of syntactic processing in language (Friederici et al., 1993; Kaan et al., 2000; Hahne, 2001; Mueller et al., 2005; Phillips et al., 2005) and music (Patel, 1998; Neuhaus, 2013;

Sun et al., 2018), reflecting the integrative process and the cognitive resources allocation. Evidence of LPCs for the processing of non-adjacent tonal integration is also given by previous research on musical syntax violation (Koelsch et al., 2013; Ma et al., 2018a,b; Zhou et al., 2019). The LPC effect observed in our study may be ascribed to the more cognitive resources required by combining local information into higher global hierarchical units for the nested structure than the non-nested structure. According to Meyer (1956) and Lerdahl and Jackendoff (1983), when the music returns to the beginning tonality, the listeners would generate the feeling of harmonic completeness. This view was supported by our ERP results. Generally speaking, our results suggested that the listeners were able to process the long-distance harmonic dependency in the complex structures.

However, the behavioral data showed no significant difference in the resolution process, although the tension value induced by the last chord seems to be different between the nested and the non-nested conditions. The tension declines with a similar slope from the tension peak to the ending of the whole chorale sequences so that the difference in ending values should be ascribed to tension accumulation. Interestingly, our unpublished data (under review) also found the phenomenon that the tension experience elicited by structural violations was not resolved entirely and immediately at the ending of each phrase but accumulated during subsequent music pieces. Taken together, our studies suggested that the dynamic temporal mode in which musical tension experienced instantly was influenced by previous time windows and produced an additionally increased tension experience (Farbood, 2012).

Combining the behavioral and the EEG data together, it seems that the cognitive processing of the distant tonal relationships in the nested structure did not bring about a resolution experience. It may be explained in terms of the relationship between cognition and emotion. It has been acknowledged that activations of both the autonomic nervous system and cognitive evaluation are prerequisites for the emotional experience (Schachter, 1959, 1964). In music, cognitive evaluation is also one of the mechanisms underlying emotional induction (Juslin and Västfjäll, 2008; Juslin, 2013). Although the experience of tension and resolution relies heavily on cognitive processing, insufficient physiological activation cannot definitely elicit the experience. Our study supported the emotional theory by demonstrating the divergence of cognitive and emotional implementations.

## The Influence of Nested Complexity on Tension Experience

In our study, two types of nested structures induced different tension experiences. The tension experience was more dramatic in the doubly nested condition than in the singly nested condition, with a wider tension range and a higher tension peak in the doubly nested condition. Furthermore, the tension curves indicated an acceleration of tension rise in the doubly nested condition than in the singly nested condition, which might be attributed to the number of subcomponents inserted into the main phrase, as the occurrence of each subcomponent increased the tension experience. To our knowledge, this is the first study to reveal the difference in tension experience induced by the singly nested structure vs. the doubly nested structure in music.

Compared with the non-nested structure, the LPC elicited by the singly nested structure is distributed in the whole brain, whereas the LPC elicited by the doubly nested structure is only found in the posterior brain area. In addition, our study also suggested that the cognitive processing of the doubly nested structure was more difficult than that of the singly nested structure, given that the distribution and the amplitude of the LPC were modulated by task difficulty (Gunseli et al., 2014; Bertoli and Bodmer, 2016; Timmer et al., 2017). Consistent with our results, one previous study also found more difficult processing for the doubly nested structure than the singly nested structure while using atonal music and artificial grammars of interval and melodic lines to construct the nested structures (Cheung et al., 2018). Language materials with more nested structures required longer reading time (Babyonyshev and Gibson, 1999; Nakatani and Gibson, 2010) and activated more activities of the left pars opercularis in the case of controlling working memory load (Makuuchi et al., 2009) compared with the fewer nested structures.

In conclusion, the tension experience elicited by the nested structure was higher and had more fluctuations than that by the non-nested structure. Furthermore, the difference was mainly exhibited in tension induction rather in resolution experience. Although the explicit resolution experience was unaffected by the nested structure, larger LPCs were elicited by the ending chords in the nested condition than in the non-nested condition, reflecting the divergence between cognitive integration and the resolution experience. Given that the LPC effect elicited by the doubly nested structure has a smaller scalp distribution than the singly nested structure, we speculated that it was more difficult for listeners to integrate the final chords into such a complex musical context. Our study demonstrated the influence of nested structure on tension experience and revealed dynamic and different processes for tension induction and resolution for the first time.

Given that the processing of a doubly nested structure may be difficult for nonmusicans, only highly proficient musicians were included in our study. Although previous studies have found that both Western and Chinese nonmusicians exhibited specific neural responses to integrate tonally long-distance dependency, the musical sequences with doubly nested structure were barely used in their studies (e.g., Koelsch et al., 2013; Ma et al., 2018a,b). Future studies should investigate whether nonmusicans can process musical tension induced by complex structures and the influence of musical training on such processing. Moreover, despite the fact that tension is the basis for emotion induction in music, we know little about how musical tension contributed to the emotion experience. Thus, more attention should also be paid to the relationship between musical tension and emotion induction in real music pieces. Investigations focusing on the above issues will shed new light on the mechanisms of musical emotion processing.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institute of Psychology, Chinese Academy of Sciences. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

LS and YY proposed the study and designed the experiment. LS and CF conducted the tests. All the authors contributed to data analysis, drafting and revising the article.

## FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31470985 and 31971034) and the China Postdoctoral Science Foundation (Y9BH032001).

## REFERENCES

Babyonyshev, M., and Gibson, E. (1999). The complexity of nested structures in Japanese. *Language* 75, 423–450.

Bertoli, S., and Bodmer, D. (2016). Effects of age and task difficulty on ERP responses to novel sounds presented during a speech-perception-in-noise test. *Clin. Neurophysiol.* 127, 360–368. doi: 10.1016/j.clinph.2015.02.055

Besson, M., and Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: comparison of musicians with nonmusicians. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 1278–1296. doi: 10.1037/0096-1523.21.6.1278

Bigand, E., and Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychol. Res.* 62, 237–254. doi: 10.1007/s004260050053

Bigand, E., Parncutt, R., and Lerdahl, F. (1996). Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Percept. Psychophys.* 58, 125–141. doi: 10.3758/bf03205482

Cheung, V. K., Meyer, L., Friederici, A. D., and Koelsch, S. (2018). The right inferior frontal gyrus processes nested non-local dependencies in music. *Sci. Rep.* 8:3822. doi: 10.1038/s41598-018-22144-9

Chomsky, N. (1957). *Syntactic Structures.* The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Cook, N. (1987). The perception of large-scale tonal closure. *Music Percept.* 5, 197–206. doi: 10.2307/40285392

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* 88, 2–19. doi: 10.1016/j.neuron.2015.09.019

Deliege, I., Mélen, M., Stammers, D., and Cross, I. (1996). Musical schemata in real-time listening to a piece of music. *Music Percept.* 14, 117–159. doi: 10.2307/40285715

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Ellis, C., Kuipers, J. R., Thierry, G., Lovett, V., Turnbull, O., and Jones, M. W. (2015). Language and culture modulate online semantic processing. *Soc. Cogn. Affect. Neurosci.* 10, 1392–1396. doi: 10.1093/scan/nsv028

Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Percept.* 29, 387–428. doi: 10.1525/mp.2012.29.4.387

Fitch, W. T., and Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380. doi: 10.1126/science.1089401

Fitch, W. T., and Martins, M. D. (2014). Hierarchical processing in music, language and action: lashley revisited. *Ann. N Y Acad. Sci.* 1316, 87–104. doi: 10.1111/nyas.12406

Fredrickson, W. E. (2000). Perception of tension in musicians versus nonmusicians. *J. Music Ther.* 37, 40–50. doi: 10.1093/jmt/37.1.40

Friederici, A. D., Erdmut, P., and Anja, H. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cogn. Brain Res.* 1, 183–192. doi: 10.1016/0926-6410(93)90026-2

Gingras, B., Pearce, M. T., Goodchild, M., Dean, R. T., Wiggins, G., and Mcadams, S. (2016). Linking melodic expectation to expressive performance timing and perceived musical tension. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 594–609. doi: 10.1037/xhp0000141

Gotlieb, H., and Konecni, V. J. (1985). The effects of instrumentation, playing style, and structure in the Goldberg Variations by Johann Sebastian Bach. *Music Percept. Interdiscip. J.* 3, 87–102. doi: 10.2307/40285323

Granot, R. Y., and Jacoby, N. (2011). Musically puzzling I: sensitivity to overall structure in the sonata form? *Music Sci.* 15, 365–386. doi: 10.1177/1029864911409508

Gunseli, E., Meeter, M., and Olivers, C. N. (2014). Is a search template an ordinary working memory? Comparing electrophysiological markers of working memory maintenance for visual search and recognition. *Neuropsychologia* 60, 29–38. doi: 10.1016/j.neuropsychologia.2014.05.012

Hackworth, R. S., and Fredrickson, W. E. (2010). Effect of text translation on perceived musical tension in Debussy's Nöeldes enfants qui n'ont plus de maisons. *J. Res. Music Educ.* 58, 184–195. doi: 10.2307/40666242

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *J. Psycholinguist. Res.* 30, 251–266. doi: 10.1023/a:1010490917575

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation.* Cambridge, MA: MIT Press.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Phys. Life Rev.* 10, 235–266. doi: 10.1023/a:1010490917575

Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 559–575. doi: 10.1017/S0140525X08005293

Kaan, E., Harris, A., Gibson, E., and Holcomb, P. (2000). The p600 as an index of syntactic integration difficulty. *Lang. Cogn. Process.* 15, 159–201. doi: 10.1080/016909600386084

Karno, M., and Konecni, V. J. (1992). The effects of structural intervention in the first movement of Mozart's Symphony in G-Minor, K. 550 on Aesthetic Preference. *Music Percept. Interdiscip. J.* 10, 63–72.

Koelsch, S. (2013). *Brain and Music.* Oxford: John Wiley & Sons.

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666

Koelsch, S., Rohrmeier, M., Torrecuso, R., and Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proc. Natl. Acad. Sci. U S A* 110, 15443–15448. doi: 10.1073/pnas.1300272110

Krumhansl, C. L., and Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychol. Rev.* 89, 334–368. doi: 10.1037/0033-295x.89.4.334

Lehne, M., and Koelsch, S. (2015). Toward a general psychological model of tension and suspense. *Front. Psychol.* 6:79. doi: 10.3389/fpsyg.2015.00079

Lehne, M., Rohrmeier, M., Gollmann, D., and Koelsch, S. (2013). The influence of different structural features on felt musical tension in two piano pieces by Mozart and Mendelssohn. *Music Percept.* 31, 171–185. doi: 10.1525/mp.2013.31.2.171

Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music.* Cambridge, MA: MIT Press.

Lerdahl, F., and Krumhansl, C. L. (2007). Modeling tonal tension. *Music Percept.* 24, 329–366. doi: 10.1525/mp.2007.24.4.329

Longuet-Higgins, H. C. (1987). "The grammar of music," in *Mental Processes: Studies in Cognitive Science*, eds H. C. Longuet-Higgins (Cambrige, MA: MIT Press), 130–147.

Ma, X., Ding, N., Tao, Y., and Yang, Y. F. (2018a). Differences in processing mechanisms underlying the processing of center-embedded and non-embedded musical structures. *Front. Hum. Neurosci.* 12:425. doi: 10.3389/fnhum.2018.00425

Ma, X., Ding, N., Tao, Y., and Yang, Y. F. (2018b). Syntactic complexity and musical proficiency modulate neural processing of non-native music. *Neuropsychologia* 121, 164–174. doi: 10.1016/j.neuropsychologia.2018.10.005

Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., and Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci. U S A* 94, 10979–10984. doi: 10.1073/pnas.94.20.10979

Makuuchi, M., Bahlmann, J., Anwander, A., and Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proc. Natl. Acad. Sci. U S A* 106, 8362–8367. doi: 10.1073/pnas.0810928106

Margulis, E. H. (2005). A model of melodic expectation. *Music Percept.* 22, 663–714. doi: 10.1525/mp.2005.22.4.663

Meyer, L. B. (1956). *Emotion and Meaning in Music.* Chicago, IL: University of Chicago Press.

Mueller, J. L., Hahne, A., Fujii, Y., and Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *J. Cogn. Neurosci.* 17, 1229–1244. doi: 10.1162/0898929055002463

Nakatani, K., and Gibson, E. (2010). An on-line study of Japanese nesting complexity. *Cogn. Sci.* 34, 94–112. doi: 10.1111/j.1551-6709. 2009.01067.x

Neuhaus, C. (2013). Processing musical form: behavioural and neurocognitive approaches. *Music Sci.* 17, 109–127. doi: 10.1177/10298649124S68998

Patel, A. D. (1998). Syntactic processing in language and music: different cognitive operations, similar neural resources? *Music Percept.* 16, 27–42. doi: 10.2307/40285775

Phillips, C., Kazanina, N., and Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cogn. Brain Res.* 22, 407–428. doi: 10.1016/j.cogbrainres.2004.09.012

Prince, J. B., and Schmuckler, M. A. (2014). The tonal-metric hierarchy: a corpus analysis. *Music Percept.* 31, 254–270. doi: 10.1525/mp.2014.31.3.254

Regnault, P., Bigand, E., and Besson, M. (2001). Different brain mechanisms mediate sensitivity to sensory consonance and harmonic context: evidence from auditory event-related brain potentials. *J. Cogn. Neurosci.* 13, 241–255. doi: 10.1162/089892901564298

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *J. Math. Music* 5, 35–53. doi: 10.1080/17459737.2011.573676

Rohrmeier, M., Dienes, Z., Guo, X., and Fu, Q. (2014). *Implicit Learning and Recursion.* New York, NY: Springer.

Rohrmeier, M., and Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *Int. J. Psychophysiol.* 83, 164–175. doi: 10.1016/j.ijpsycho.2011.12.010

Schachter, S. (1959). The psychology of affiliation: experimental studies of the sources of gregariousness. *Q. Rev. Biol.* 17, 15–25.

Schachter, S. (1964). "The interaction of cognitive and physiological determinants of emotional state," in *Advances in Experimental Social Psychology* (Vol. 1), ed. L. Berkowitz (New York, NY: Academic Press), 49–80.

Schenker, H. (1979). *Free Composition (E. Oster, Trans).* New York, NY: Longman.

Schubert, E. (2010). "Continuous self-report methods," in *Handbook of Music and Emotion —Theory, Research, Applications*, eds P. Juslin and J. A. Sloboda (Oxford: Oxford University Press), 223–253.

Steinbeis, N., Koelsch, S., and Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological, and neural responses. *J. Cogn. Neurosci.* 18, 1380–1393. doi: 10.1162/jocn.2006.18.8.1380

Sun, L., Liu, F., Zhou, L., and Jiang, C. (2018). Musical training modulates the early but not the late stage of rhythmic syntactic processing. *Psychophysiology* 55:e12983. doi: 10.1111/psyp.12983

Tillmann, B., and Bigand, E. (1996). Does formal musical structure affect perception of musical expressiveness? *Psychol. Music* 24, 3–17. doi: 10.1177/0305735696241002

Tillmann, B., and Bigand, E. (1998). Influence of global structure on musical target detection and recognition. *Int. J. Psychol.* 33, 107–122.

Tillmann, B., and Bigand, E. (2004). The relative importance of local and global structures in music perception. *J. Aesthet. Art Crit.* 66, 211–222. doi: 10.1111/j.1540-594x.2004.00153.x

Tillmann, B., Bigand, E., and Madurell, F. (1998). Local versus global processing of harmonic cadences in the solution of musical puzzles. *Psychol. Res.* 61, 157–174.

Timmer, K., Grundy, J. G., and Bialystok, E. (2017). Earlier and more distributed neural networks for bilinguals than monolinguals during switching. *Neuropsychologia* 106, 245–260. doi: 10.1016/j.neuropsychologia.2017.09.017

Zendel, B. R., Lagrois, M. É., Robitaille, N., and Peretz, I. (2015). Attending to pitch information inhibits processing of pitch information: the curious case of amusia. *J. Neurosci.* 35, 3815–3824. doi: 10.1523/jneurosci.3766-14.2015

Zhang, J., Zhou, X., Chang, R., and Yang, Y. (2018). Effects of global and local contexts on chord processing: an ERP study. *Neuropsychologia* 109, 149–154. doi: 10.1016/j.neuropsychologia.2017.12.016

Zhou, L., Liu, F., Jiang, J., Jiang, H., and Jiang, C. (2019). Abnormal neural responses to harmonic syntactic structures in congenital amusia. *Psychophysiology* 56:e13394. doi: 10.1111/psyp.13394

**Conflict of Interest**: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 25 June 2020
doi: 10.3389/fnhum.2020.00216

# Positive and Detached Reappraisal of Threatening Music in Younger and Older Adults

*Sandrine Vieillard[1]*, Charlotte Pinabiaux[1] and Emmanuel Bigand[2]*

*[1] Unité DysCo, Fonctionnement et Dysfonctionnement Cognitifs: Les âges de la vie, Université Paris Nanterre, Nanterre, France, [2] Unité LEAD, Laboratoire d'Etude de l'Apprentissages et du Développement, UMR CNRS 5022, Université Bourgogne-Franche-Comté, Dijon, France*

Past empirical studies have suggested that older adults preferentially use gaze-based mood regulation to lessen their negative experiences while watching an emotional scene. This preference for a low cognitively demanding regulatory strategy leaves open the question of whether the effortful processing of a more cognitively demanding reappraisal task is really spared from the general age-related decline. Because it does not allow perceptual attention to be redirected away from the emotional source, music provides an ideal way to address this question. The goal of our study was to examine the affective, behavioral, physiological, and cognitive outcomes of positive and detached reappraisal in response to negative musical emotion in younger and older adults. Participants first simply listened to a series of threatening musical excerpts and were then instructed to either positively reappraise or to detach themselves from the emotion elicited by music. Findings showed that, when instructed to simply listen to threatening music, older adults reported a more positive feeling associated with a smaller SCL in comparison with their younger counterparts. When implementing positive and detached reappraisal, participants showed more positive and more aroused emotional experiences, whatever the age group. We also found that the instruction to intentionally reappraise negative emotions results in a lesser cognitive cost for older adults in comparison with younger adults. Taken together, these data suggest that, compared to younger adults, older adults engage in spontaneous downregulation of negative affect and successfully implement downregulation instructions. This extends previous findings and brings compelling evidence that, even when auditory attention cannot be redirected away from the emotional source, older adults are still more effective at regulating emotions. Taking into account the age-associated decline in executive functioning, our results suggest that the working memory task could have distracted older adults from the reminiscences of the threat-evoking music, thus resulting in an emotional downregulation. Hence, even if they were instructed to implement reappraisal strategies, older adults might prefer distraction over engagement in reappraisal. This is congruent with the idea that, although getting older, people are more likely to be distracted from a negative source of emotion to maintain their well-being.

**Keywords: age-related effects, positive reappraisal, detached reappraisal, affective outcomes, behavioral responses, physiological measures, cognitive cost**

OPEN ACCESS

***Edited by:***
*Cunmei Jiang, Shanghai Normal University, China*

***Reviewed by:***
*Junichi Chikazoe, National Institute for Physiological Sciences (NIPS), Japan*
*Elena Solesio-Jofre, Autonomous University of Madrid, Spain*

****Correspondence:***
*Sandrine Vieillard*
*sandrine.vieillard@parisnanterre.fr*

***Specialty section:***
*This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience*

***Received:*** *15 October 2019*
***Accepted:*** *13 May 2020*
***Published:*** *25 June 2020*

***Citation:***
*Vieillard S, Pinabiaux C and Bigand E (2020) Positive and Detached Reappraisal of Threatening Music in Younger and Older Adults. Front. Hum. Neurosci. 14:216. doi: 10.3389/fnhum.2020.00216*

## INTRODUCTION

The most widely known model of emotion regulation (Gross, 1998) proposes several emotion-regulation modalities falling along a continuum from less (i.e., situation selection, situation modification, attentional deployment) to more cognitively demanding (i.e., cognitive reappraisal, behavioral suppression) emotion-regulation modalities. Among them, cognitive reappraisal, defined as the aim to change an emotional response by reinterpreting the meaning of the emotional event, has received increasing attention over the past decade in the field of psychological aging. Nevertheless, the studies on age-related changes in the self-reported use of cognitive reappraisal yield mixed outcomes. Although a majority reveal no significant effect of aging (Hess et al., 2010; Li et al., 2011; Tucker et al., 2012; Brummer et al., 2014), some other studies report greater preference for cognitive reappraisal over other emotion-regulation modalities for older adults compared with their younger counterparts (John and Gross, 2004; Urry and Gross, 2010; Gerolimatos and Edelstein, 2012).

Researchers operationalize cognitive reappraisal in varied ways mainly with the objective to study age-related effects on the efficiency of emotional regulation. Some make a clear distinction between two main categories of positive and detached reappraisal, which were, respectively, associated with instructions consisting of *thinking about positive aspects* or *adopting a detached and unemotional attitude* while seeing an emotional scene in order to feel less negative emotion (Shiota and Levenson, 2009; Lohani and Isaacowitz, 2014; Liang et al., 2017; Livingstone and Isaacowitz, 2018). In some other studies, the authors directly liken the notion of cognitive reappraisal to detached emotion regulation by asking participants to distance themselves from the emotional event (Winecoff et al., 2011; Tucker et al., 2012; Pedder et al., 2016). Still others do not make any distinction between positive and detached reappraisal and are based on an experimental design in which participants are just asked to *decrease* the (negative) emotion they felt (Opitz et al., 2012; Allard and Kensinger, 2017). At last, positive and detached reappraisal instructions are sometimes employed as two interchangeable options of a same and unique emotion-regulation modality (Allard and Kensinger, 2014).

Not surprisingly, these considerable variations in the operationalization of cognitive reappraisal lead to equivocal findings. The measures recorded in these previous studies (i.e., affective outcomes, facial expression, physiological reactions, and—much more rarely—eye-gaze deployment) indeed show that older adults are sometimes less (Winecoff et al., 2011; Opitz et al., 2012; Tucker et al., 2012; Allard and Kensinger, 2014; Scheibe et al., 2015; Pedder et al., 2016), sometimes equally (Allard and Kensinger, 2017; Livingstone and Isaacowitz, 2018), or sometimes better (Lohani and Isaacowitz, 2014) able to successfully implement cognitive reappraisal in comparison with their younger counterparts. Authors who show that older adults are better at regulating their emotions explain their findings within the dominant framework of socio-emotional selectivity theory (SST; Carstensen et al., 1999). This theoretical model posits that the perceived time remaining in life is a critical determinant of motivational processes in life span. While getting older, the consciousness of the limited remaining time in life leads older adults to be more motivated to maintain or enhance social and emotional well-being. Consequently, older adults preferentially process positive information over negative, resulting in a positivity effect that has been widely documented in literature on aging and emotion. Accumulated findings indicate that, when cognitive resources are experimentally distracted, the positivity effect is no longer observed, suggesting that this effect is cognitively demanding processing (e.g., Knight et al., 2007). Basically, the SST model postulates that the positivity effect operates as an emotion-regulation goal (but see Isaacowitz and Blanchard-Fields, 2012, for a critical view of this hypothetical link between the positivity effect and the outcome of positive affective experience). According to this view, intentional emotion regulation is, by essence, a cognitively demanding task. Alternative but most confidential theoretical frameworks of the dynamic integration theory (DIT; Labouvie-Vief, 2008) attempt to offer a different explanation to interpret the improvement in emotional regulation with aging. As the SST model, DIT assumes that emotion regulation is cognitively costly. Focusing on the effect of cognitive decline on emotional processing, the DIT model postulates that older adults may still be able to regulate their emotion only when facing emotional stimuli of low intensity. When facing highly negative emotions, older adults tend to automatically distance from them (e.g., Brady et al., 2018), then abolishing the cognitive cost elicited by the elaborate processing of negative events considered as cognitively much more complex (Labouvie-Vief, 2008). SST assumes that the more cognitive resources older adults have, the better they are able to reappraise the emotional significance of the event, whereas DIT argues that the fewer cognitive resources older adults have, the better they are able to downregulate negative emotion, not through an emotional reappraisal *per se* (which would be too resource demanding), but through a mechanism allowing them to disengage from the source of emotion. According to the DIT model, such a mechanism would take place in an automated way precisely because the older adults' cognitive resources are limited.

Regarding the different types of cognitive reappraisal modalities, several studies suggest that positive and detached reappraisals are not equally affected by age. For instance, Shiota and Levenson (2009) demonstrate that, even if older adults report greater success than younger adults at implementing both types of reappraisal when viewing sad and disgusting films, the age-related effect on affective outcomes and physiological reactivity varies as a function of strategy. Older adults use positive reappraisal more successfully than younger adults, whereas younger adults use detached reappraisal more successfully. Corroborating these previous findings, Lohani and Isaacowitz (2014) show that older adults are more successful than younger adults at implementing positive reappraisal in response to sadness-eliciting film clips. This is shown by the fact that, relative to the no-regulation condition, older adults report a decreased negative mood with no heightened physiological activity, and younger adults do not experience a decrease in negative mood, but an increase in physiological response. Going one step further, Liang et al. (2017) conducted a study on a single sample of older adults providing evidence that, in comparison with positive reappraisal, detached reappraisal relies more heavily on cognitive control, especially on mental set shifting used to capture mental flexibility. These results are interpreted as corroborating the idea that, given that positive reappraisal necessitates maintaining attention on emotion rather than inhibiting it, it is finally not surprising that such emotion regulation is less cognitively demanding compared to detached reappraisal. These data are also in line with the hypothesis that positive reappraisal is effective in everyday life (Guiliani and Gross, 2009; McRae et al., 2012) notably because of its benefits on mental health in the context of physical illness or stress (Nowlan et al., 2014, 2016). Although the studies above suggest the existence of a differential impact of positive and detached reappraisal depending on age, other recent findings indicate that both age groups show mood improvement when using cognitive appraisal whatever the positive or detached modality (Allard and Kensinger, 2017; Livingstone and Isaacowitz, 2018). A recent meta-analysis on age-related differences in ability to implement emotion-regulation instruction, including previously cited studies and also unpublished sources of data, concludes that both young and older adults better regulate the behavioral indicators of emotion when using detached reappraisal relative to positive reappraisal (Brady et al., 2018). In view of such discrepant findings, further investigative efforts are needed.

An important aspect of the previously cited works is that they were conducted using emotional visual scenes (film clips or pictures). Most of them often neglect the possibility that less-demanding modalities of emotion regulation, such as attention deployment using gaze direction, might help to achieve an emotion-control goal. This is a question of importance because it has been shown that older adults deploy their visual attention away from negative stimuli (Isaacowitz et al., 2006). Moreover, when controlling gaze direction, older people are less successful than younger adults at regulating unpleasant emotion elicited by a visual scene (Opitz et al., 2012). More recent findings indicate that, when getting older, distraction (avoidance by gaze redirection) is less cognitively effortful than reinterpreting negative information through a positive reappraisal instruction (Martins et al., 2018). It is then reasonable to think that the potential preferential use of attentional deployment could partly explain successful emotion regulation in older adults. This assumption is consistent with empirical evidence that, when alternative modalities are easily reachable and just as efficient, older adults prefer to use less cognitively demanding modalities of emotion regulation (Scheibe et al., 2015). This raises the question of whether, with no possibility of recourse to attentional deployment, the effortful processing of some cognitive reappraisal contexts is really spared from the general age-related decline. One way to address this issue is to use emotional stimuli that do not allow participants to reallocate perceptual attention away from the emotional source. In this respect, music represents an ideal mean. Music has also been shown to be sensitive to the age-related positivity effect (Vieillard and Gilet, 2013; Vieillard and Bigand, 2014). Using musical material ensures that participants remain engaged in the auditory processing of emotional material, but this is not to say that, once the emotion is processed, the participant could not implement an emotion-regulation strategy.

Another important aspect of the works described above is that they mainly focus on affective, behavioral, and physiological consequences of reappraisal, failing to address the impact of positive and detached reappraisal on cognitive processing. Previous studies addressing the cognitive cost of reappraisal in younger adults have yielded mixed conclusions. Some authors suggest that the use of cognitive reappraisal, in its detached version, may lead to lesser cognitive costs compared with other emotion-regulation modalities, such as expressive suppression (Richards and Gross, 2000; Richards et al., 2003; Gross and Thompson, 2007). For instance, Richards and Gross (2000) find that participants who receive an instruction to detach from their feelings before watching pictures of injured people show better memory for the picture details compared to those who are told to use expressive suppression or no regulation. Richards et al. (2003) show that the memory for conversion utterances is increased when people are told to positively reappraise their emotional inner states during a naturalistic conflictual discussion compared to when they are told to suppress the expression of their feelings. Some researchers question such less cognitive cost, arguing that the successful use of cognitive reappraisal, whether positive or detached, requires active reinterpretation of the meaning and significance of emotional stimuli and, thus, involves demanding processes, such as working memory, flexibility, or inhibition, especially in emotionally intense situations (Hofmann et al., 2012; Ortner et al., 2016). Hence, it has been shown that detached reappraisal is associated with decreased performance on reaction-time tasks and decreased self-control resources when used in high-intensity negative situations (Sheppes and Meiran, 2008; Ortner et al., 2016). Detached reappraisal of positive pictures is also shown to cause a decrease in subsequent memory recognition (Ortner and de Koning, 2013). On a subjective level, positive reappraisal seems to be more difficult to implement than acceptance as indicated by the greater perceived cognitive cost of positive reappraisal reported by young adults (Troy et al., 2018). To date, cognitive consequences of detached and positive reappraisals has never been compared in a within-subject design experiment, leaving open the question as to whether one of them could be less costly than the other in younger adults. Regarding the effect of age, to our knowledge, only one study has attempted to test whether cognitive reappraisal was synonymous with higher cognitive cost in advancing age. Scheibe and Blanchard-Fields (2009) investigate the age-related effect on cognitive consequences of intentional downregulation of disgust induced by film clips. To this end, they asked younger and older participants to implement a positive reappraisal consisting of turning negative feelings potentially elicited by the emotional stimulus into positive ones. Measuring the working memory performances as a cognitive load index of the emotion regulation activity with a *N*-back task, the authors find that in comparison with their younger counterparts, older adults show better working memory performances when implementing positive reappraisal after emotional induction. Given the cognitive control *a priori* required by the execution of positive reappraisal, such findings may sound counterintuitive in particular with regard to the age-related decline in executive functioning. In an attempt to provide a plausible account, Scheibe and Blanchard-Fields (2009) reason in terms of long-term practice in regulating emotion while getting older, speculating that such practice might render the emotion regulation activity less costly. Another explanation in terms of distraction is advanced by authors with the idea that the working memory task itself could be operated in an emotion-regulation way, distracting older adults from their memories of the negative emotion elicited by film clips. Nevertheless, authors do not consider the possibility that the emotion regulation has been less cognitively costly in older adults precisely because they preferentially avert their eyes from disgusting film clips. To determine which of these two hypotheses best accounts for Scheibe and Blanchard-Fields (2009) results, we need to replicate and extend their findings in a context in which participants cannot reallocate perceptual attention away from the emotional source. We conduct such a study to investigate the effect of age on both positive and detached reappraisal with the hope to provide a thinner understanding on emotion-regulation consequences in aging. We also addressed this issue with an attempt to determine which of the theoretical frameworks, SST or DIT, offers a better account for empirical findings.

In the current study, we examine the affective, behavioral, physiological, and cognitive outcomes of positive and detached reappraisal in response to negative musical emotion in both younger and older adults. Participants were first asked to simply listen to a series of threatening musical excerpts and then were given the instruction to either positively reappraise or to detach themselves from the emotion elicited by music. The participants' affective ratings (Likert scale), facial expressions (facial EMG), and physiological state (SCL, HR) were recorded as a spontaneous control condition during the simply listening phase as well as under the instruction of emotion regulation. The simply listening condition allowed us to examine how the affective state of older adults spontaneously changes when listening to threatening music. In line with previous findings (Vieillard and Gilet, 2013; Vieillard and Bigand, 2014), we expected that the affective ratings of the threatening musical excerpts would be less negative and intense for older adults than for younger ones. This may be associated with age-related changes in facial expressions and physiological state reflecting the tendency of older adults to reduce the processing of negative emotion. Regarding the cognitive reappraisal, the scarce and contradictory results do not facilitate the formulation of specific hypotheses on affective outcomes, expressive responses, and physiological reactions. However, if we rely on the dominant framework of the SST, it would be expected that, whatever the kind of cognitive reappraisal, older adults would report more positive affective outcomes and expressive responses than their younger counterparts. Because the SST postulates that reappraisal is cognitively costly, it is reasonable to postulate that this cost would be reflected in more physiological reaction. The DIT framework predicts similar findings on the condition that threatening musical excerpts are experienced as having relatively low intensity. If threatening musical excerpts are experienced as having high intensity, the DIT conjectures that, compared to younger adults, older adults would be less effective at implementing emotion-regulation instructions, whatever the kind of reappraisal.

We also wanted to test the robustness of what appears to be better emotional control with age. To extend previous findings to a different cognitive task, we use a memory span task adapted from Schmeichel (2007) work. Our goal was to measure cognitive performances of participants just after implementing the simply listening instruction as well as reappraisal (positive, detached) instructions. As postulated by the SST, if we consider that the spontaneous modulation of negative affects operated by older adults is resource demanding, it should lessen their subsequent working memory performance more than for younger adults, whatever the spontaneous or intentional condition. On the other hand, as stated by the DIT, if older adults spontaneously downregulate their negative feelings in an automated way when faced with not too emotionally intense stimuli, their subsequent working memory performance should not be affected by their spontaneous emotion regulation or by their intentional emotion regulation, contrary to younger adults. Moreover, based on Scheibe and Blanchard-Fields (2009) findings showing that reappraisal is less effortful as people grow older, we hypothesize that reappraisal in our study would be less cognitively demanding in older adults in comparison with younger adults. A plausible explanation is that focusing on the working memory task may distract participants from the emotions elicited during the emotion regulation (Scheibe and Blanchard-Fields, 2009). Because distraction has been demonstrated to be an efficient emotion-relation modality in aging (Martins et al., 2018), we can sketch out a general prediction that, whatever its positive or detached type, the cognitive reappraisal would be less costly in older than in younger adults. Given that, in our experiment, participants cannot avoid listening to musical excerpts, a replication of Scheibe and Blanchard-Fields (2009) results would mean that, in line with previous works (e.g., Vieillard and Bigand, 2014), older adults have higher propensity to disengage from negative stimuli, in particular when they are helped by a source of distraction (working memory task). This would be in line with DIT (Labouvie-Vief, 2008) and agree with the hypothesis that older adults become more efficient at implementing intentional emotion regulation (not based on real reappraisal but on distraction). In the case in which we do not replicate Scheibe and Blanchard-Fields (2009) results, an alternative explanation, in line with the SST, would be that the implementation of a cognitive reappraisal strategy, whether positive or detached, may importantly lessen the working memory performances of older adults in comparison with their younger counterparts.

## MATERIALS AND METHODS

### Participants

Forty-six non-musician younger adults from the University of Franche-Comté, France, and 37 non-musician older adults recruited through senior social programs in Besançon, France, participated in the experiment. We ensured that they had no neurological or psychiatric antecedent and reported normal or corrected visual acuity. Due to the affective aspect of the experiment, we excluded 11 younger and three older adults reporting high depressive symptoms (BDI-II score higher than 30) and/or high levels of anxiety (state and trait STAI-Y standard scores higher than 55) from the analyses. The final sample consists of 35 younger adults ranging from 18 to 27 years (*M* = 21, *SD* = 2.23; 60% females) and 34 older adults ranging from 60 to 79 years (*M* = 66, *SD* = 4.80; 62% females).

As illustrated in **Table 1**, older adults did not report statistically different levels of education or self-reported health, but younger adults reported spending more time on music listening per week (*M* = 10.97 h, *SD* = 10.99) than older adults (*M* = 7.55 h, *SD* = 11.34; $p = 0.036$). For each participant, potential hearing loss was examined using a professional audiometer. As expected, the hearing level (dB) varied as a function of age group. The examination of cognitive functioning showed that younger adults have higher performance on the Victoria Stroop index (Bayard et al., 2011) than older adults, but no age-related effect was found on the letter–digit sequencing test (WAIS-III, Wechsler, 2000). The investigation of affective functioning indicated that older and younger adults did not significantly differ on depression (BDI-II, Beck et al., 1996), state anxiety (STAI- Y; Spielberger et al., 1983), trait anxiety (STAI-Y; Spielberger et al., 1983), PANAS positive affects (Watson et al., 1988), and reappraisal subscale of ERQ (Christophe et al., 2009) measures. However, there was a statistically significant main effect of age on PANAS negative affects (Watson et al., 1988). Finally, the examination of personality traits (measured by the French validation of NEO-P-IR by Plaisant et al., 2010) showed age-related differences

**TABLE 1** | Sample characteristics.

| | Younger adults (*n* = 35) | Older adults (*n* = 34) | Shapiro–Wilk Test | Levene's homogeneity test | Age group difference |
|---|---|---|---|---|---|
| | **Mean** | **Mean** | ***p*-value** | ***p*-value** | ***Corrected p* value[a,b]** |
| **Demographic characteristics** | | | | | |
| Age (year) | 21 (2.22) | 66 (4.81) | – | – | 0.000[(b)] |
| Education (year)[(c)] | 13.14 (1.55) | 13.88 (2.01) | 0.00 | 0.13 | 0.205[(b)] |
| Sex (% female) | 60 | 62 | – | – | – |
| Music listening per week (hours) | 10.97 (10.99) | 7.55 (11.34) | 0.00 | 0.47 | 0.036[(b)] |
| Self reported health (max. 5) | 4.43 (0.65) | 4.32 (0.58) | 0.00 | 0.28 | 0.446[(b)] |
| Hearing level (dB) | | | | | |
| 500 Hz | 8.43 (6.10) | 15.51 (5.53) | 0.00 | 0.62 | 0.000[(b)] |
| 1000 Hz | 6.86 (7.73) | 15.59 (7.39) | 0.00 | 0.48 | 0.000[(b)] |
| 2000 Hz | 3.07 (6.81) | 18.90 (12.93) | 0.00 | 0.00 | 0.000[(b)] |
| 4000 Hz | 0.57 (7.45) | 30.29 (18.43) | 0.00 | 0.00 | 0.000[(b)] |
| 8000 Hz | 6.86 (10.73) | 47.43 (21.13) | 0.00 | 0.00 | 0.000[(b)] |
| **Cognitive Scores** | | | | | |
| Inhibition: Victoria Stroop (IF) | 1.77 (0.36) | 2.04 (0.32) | 0.59 | 0.31 | 0.002[(a)] |
| Working Memory: Digit Span (max. 30) | 11.14 (2.85) | 10.15 (1.94) | 0.00 | 0.08 | 0.171[(b)] |
| MMSE | – | 29.76 (0.50) | – | – | – |
| **Affectives Scores** | | | | | |
| PANAS positive affects (max. 50) | 32.83 (5.88) | 34.68 (4.72) | 0.02 | 0.44 | 0.296[(b)] |
| PANAS negative affects (max. 50) | 18.97 (6.93) | 16.15 (5.58) | 0.00 | 0.79 | 0.032[(b)] |
| BDI-II (max. 63) | 7.69 (5.26) | 6.09 (5.36) | 0.00 | 0.72 | 0.181[(b)] |
| STAI-Y trait (max. 80) | 38.14 (6.54) | 35.09 (6.51) | 0.28 | 0.72 | 0.283[(a)] |
| STAI-Y state (max. 80) | 29.23 (5.17) | 27.79 (5.84) | 0.06 | 0.64 | 0.086[(a)] |
| **Emotion Regulation Questionnaire Scores** | | | | | |
| ERQ Suppression Score (max. 28) | 15.83 (5.43) | 13.68 (4.87) | 0.15 | 0.66 | 0.127[(a)] |
| ERQ Regulation Score (max. 42) | 27.74 (5.83) | 29.53 (6.06) | 0.10 | 0.63 | 0.248[(a)] |
| **NEO PIR Scores** | | | | | |
| Neuroticism (max. 192) | 89.83 (22.36) | 83.12 (16.23) | 0.03 | 0.04 | 0.000[(b)] |
| Extraversion (max. 192) | 114.20 (16.94) | 103.30 (12.13) | 0.53 | 0.07 | 0.006[(a)] |
| Openess to experience (max. 192) | 123.20 (18.45) | 117.53 (12.41) | 0.81 | 0.03 | 0.000[(b)] |
| Consciousness (max. 192) | 115.09 (21.45) | 121.41 (13.02) | 0.04 | 0.02 | 0.000[(b)] |
| Agreeableness (max. 192) | 121.54 (24.26) | 131.03 (14.40) | 0.00 | 0.01 | 0.000[(b)] |

*Standard deviations are listed in parentheses. (a) Student's t test, p values are corrected with the Hochberg procedure for controlling the family-wise error rate; (b) Mann–Whitney U test, p values are corrected with the Hochberg procedure for controlling the family-wise error rate; (c) Number of years of education has been calculated from 6 years old (age at which school is compulsory in France).*

on the mean scores of extraversion. No other significant difference was found.

## Material

Forty threatening musical excerpts taken from a Platel et al. (unpublished) database of film soundtracks and three peaceful musical excerpts taken from the database of Bigand et al. (2005) were used in this experiment (**Supplementary Material**). Among the 40 threatening excerpts, two of them were devoted to a training phase, six were allocated to a condition in which the instruction was to spontaneously respond to music, and six were allocated to a condition in which the instruction was to implement an emotion-regulation modality of reappraisal. Two peaceful musical excerpts were added. One was used as a familiarization phase with rating scales, and the other was used as a debriefing phase. All these musical stimuli were extracted from the classic (for peaceful music) and modern repertoires (for threatening music), excluding songs. Additional musical material was used as a baseline condition and was especially created for this study in order to test whether affective, behavioral, and physiological responses to such stimuli varied as a function of age groups. These control musical stimuli consisted of four auditory stimuli, including a tuning orchestra or playing scales in cello or piano. All the musical stimuli have a duration of 20 s.

## Procedure

The experiment was divided into two sessions separated by an interval of about 2 weeks at the university of Franche-Comté. In the first session, participants were instructed to sign a consent form according to the declaration of Helsinki. They filled out a demographic questionnaire, including information about their age, education level, self-reported health, visual acuity, and medical history and were then presented with a set of cognitive and affective tests. The first session lasted about 1 h.

In the second session, participants were tested individually in a quiet room at stable ambient temperature (**Figure 1A**). Once the participants were installed with the physiological system, they were asked to listen to two threatening musical excerpts, one by one, in a set of training trials. These later were used to allow each participant to adjust the sound loudness so that it was judged to be as comfortable as possible. Immediately after, one peaceful excerpt was presented with the aim to familiarize each participant with rating scales designed to evaluate the intensity of the emotional experience ("The emotion I feel is" from 0 "weak" to 9 "strong") and the hedonic valence of the emotional experience ("The emotion I feel is" from 0 "negative" to 9 "positive"). The order of presentation of the rating scales was counterbalanced across participants.

For each participant, the emotion-regulation task always began with a baseline condition followed by the simply listening condition, which was, in turn, followed by the reappraisal condition. Under the reappraisal condition, half of the participants were randomly allocated to a detached reappraisal condition, and the other half were randomly allocated to a positive reappraisal condition. As illustrated in the appendix, the assignment of the 12 threatening excerpts was controlled so that, when the first half was allocated to a simply listening condition, the other half was allocated to a reappraisal condition (either detached or positive reappraisal) and vice versa. In each emotion-regulation condition, the order of the presentation of the musical excerpts was randomized. The experiment ended with a debriefing block of two peaceful musical excerpts for which participants were asked to apply a simply listening instruction.

In the baseline condition, the four auditory excerpts were presented with the following instruction: "You will listen to auditory excerpts. Be careful because, after each excerpt, you will be asked to evaluate what you thought and felt during the listening." The simply listening condition, including a block of six threatening musical excerpts, was presented with the following instruction: "You will listen to musical excerpts, which can elicit in you some feelings. We ask you to listen to them carefully and to feel your emotion as you want. After each musical excerpt, you will have to evaluate what you have felt during listening." The reappraisal emotion-regulation condition was divided into two distinct instructions: a positive and a detached reappraisal. The positive reappraisal instruction was: "You will listen to musical excerpts conveying a negative feeling. While listening to the music, we ask you to reconsider what the music conveys in such a way that you will focus on its positive aspect. To this end, please try to think that this music has been composed to comic purposes in order to feel the least negative emotion as possible." The detached reappraisal instruction was: "You will listen to musical excerpts conveying a negative feeling. While listening to the music, we ask you to detach yourself from what the music conveys. To this end, please try to think about other things than what you are listening to in order to feel the least negative emotion as possible." After presenting the two threatening musical excerpts devoted to the training phase for the reappraisal instruction implementation and before presenting the six threatening stimuli devoted to the testing part (i.e., simply listening and cognitive reappraisal), a set of instructions related to a working memory span task (Schmeichel, 2007) was displayed. In this task, participants had to judge the correctness of math equations while encoding target words for a subsequent recall. For instance, participants saw ("3 + 4 = 5") and had to indicate "Yes" or "No" as to whether the given answer was correct. Then the participants read a target word aloud (e.g., nail) for later recall. One target word was presented after each equation. The participants saw three, four, or five equation–word pairs before being prompted to recall the target words in the set. They did not know in advance how many words a set would include. The working memory span task included 12 sets totaling 48 equation–word pairs in all, counterbalanced across participants. The experiment was designed such that, after each of the six simply listening trials and the six reappraisal trials, the participants were presented with one trial of a memory span task. Equation–word pairs were displayed on a computer screen and participants controlled their display with their responses.

In this experiment, one trial always began with the emotion-regulation instructions after the participants indicated they were ready (**Figure 1B**). A 10-s blank screen was displayed while the physiological baseline was recorded, followed by a fixation cross of 1 s. Immediately after, the musical excerpt was delivered for 20 s. At the end of the musical excerpt's presentation,

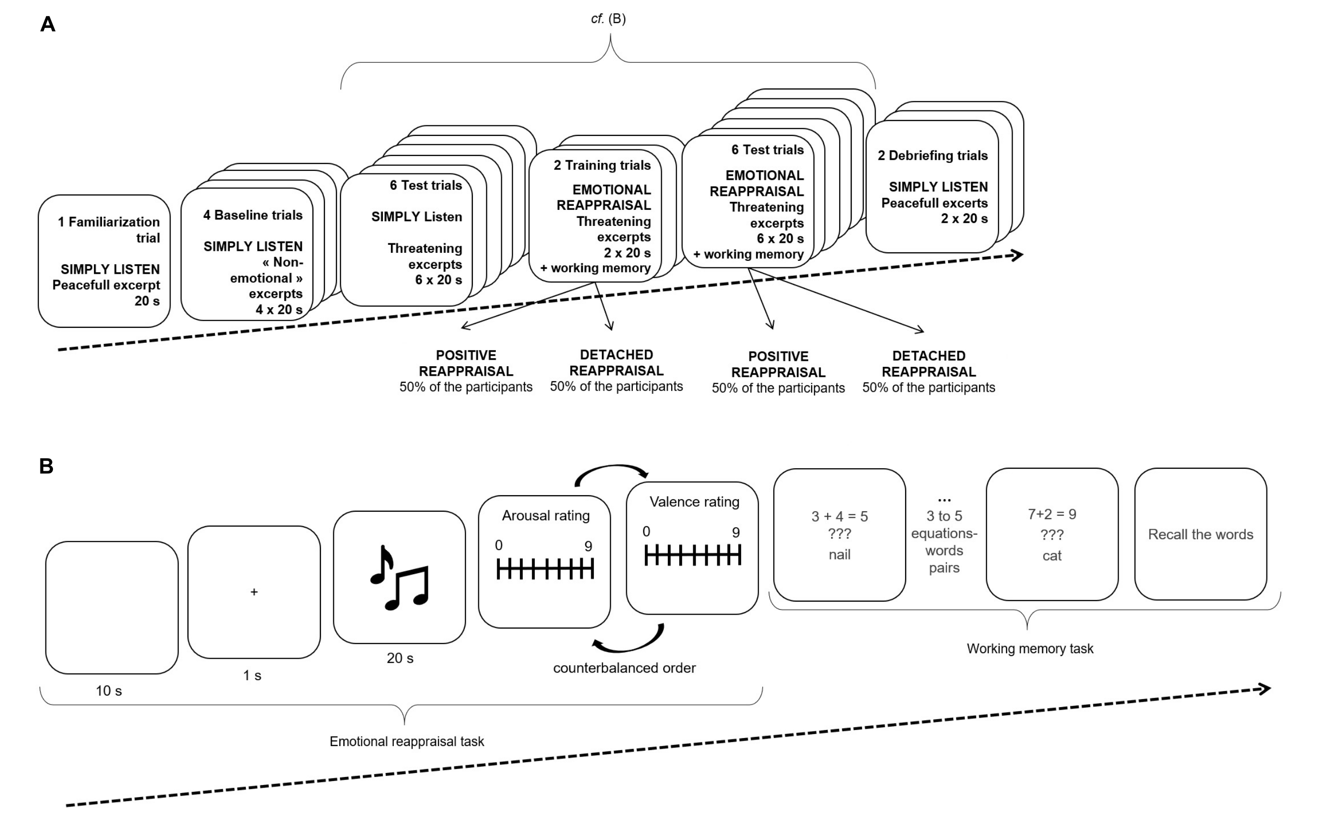



**FIGURE 1 |** General procedure **(A)** and trial shematic **(B)** for emotion regulation task.

the participants evaluated it on rating scales and performed one trial from the memory span task. For each of this set of events, the experimenter carefully examined the level of cutaneous conductance and did not start the next trial until it had stabilized, ensuring that this measure returned to baseline. Once the tasks were completed, the physiological system was removed, and the participant was then debriefed. The second session lasted about 1 h.

## Data Acquisition and Transformation

Physiological responses were monitored throughout the experiment using an MP150 Biopac system (Biopac Systems, Inc., Goleta, CA) at a sampling rate of 500 Hz and were processed using AcqKnowledge software. The facial expressivity of the participants was assessed through their zygomaticus electromyographical activity (facial EMG, μ volts) with two 4-mm shielded electrodes located on the left zygomaticus muscle (see recommendations of Tassinary et al., 2007). The signal was rectified using the root mean square function of the software and then smoothed using a 50-Hz band stop filter. The EMG score was calculated by subtracting the EMG signal recorded for the 1-s duration before the onset of the musical excerpt from the EMG signal (area under the curve per second) recorded for the 20 s of musical excerpt presentation. Skin conductance level (SCL, μ Siemens) was recorded on the left index and little fingers with two electrodes filled with isotonic gel. The reported SCL score was calculated by subtracting the SCL signal recorded at the onset of the musical excerpt from the SCL recorded from the third second to the end of the musical excerpt presentation. Electrocardiogram (heart rate in beats per minute – HR in bpm) activity was recorded with three 8-mm electrodes located at the left wrist (+), right wrist (-), and left ankle (ground). The signal was smoothed using a 50-Hz band stop filter. The reported HR score was calculated by subtracting the HR signal recorded during the 1-s before the onset of the musical excerpt from the HR signal recorded during the 20-s of musical excerpt presentation. The distribution of each variable was examined to identify possible remaining outliers (mean ± 3 *SD*). Based on this criterion, about 3.4% of all measurements were excluded from the analyses.

## Data Analyses

The normality and homogeneity of variances were tested using Shapiro Wilk and Levene's tests before the statistical analyses were applied. Because these conditions were not met, the different variables were analyzed with non-parametric statistics using Mann–Whitney *U* tests for between-groups comparisons. For each affective (arousal and valence judgment), expressive (facial EMG), physiological (SCL and HR), and cognitive (mean ratio at the working memory task) measure, we conducted the analyses on the efficiency of the reappraisal strategies (positive vs. detached) with a differential score obtained by subtracting the measure obtained for the cognitive reappraisal condition from the measure obtained in the simply listening condition. Positive scores corresponded to relatively higher value of the measure under the reappraisal instructions compared to the simply listening condition. On the contrary, negative scores corresponded to higher values of the measure in the simply listening condition compared to the reappraisal instruction. The efficiency of both types of reappraisal was then tested for each age group separately (one-sample *t* test, comparison of the differential score with $\mu = 0$) and compared between age groups (Mann–Whitney *U* tests). In the end, the efficiency of the two types of reappraisal (positive vs. detached) was compared within each age group (Mann–Whitney *U* tests). To control for the equivalence between each age group (younger vs. older) through reappraisal conditions (positive vs. detached) two sets of Mann–Whitney *U* tests were conducted. Because multiple comparisons are conducted, we applied the Hochberg procedure for controlling the family-wise error rate to correct the *p* values. To investigate the consistency between subjective ratings and cognitive cost of emotion-regulation instructions, correlation analyses were calculated for each experimental condition separately for younger and older adults. We also examined how the affective and the cognitive functioning relate to emotion regulation across age groups. Every time that age-related differences were found, we computed a set of correlation analyses between sample characteristics and the differential scores of the affective, behavioral, physiological, and cognitive measures for each experimental condition, separately for younger and older adults.

# RESULTS

## Age Equivalence in Baseline Reactivity Levels (Control Musical Excerpts)

As shown in **Table 2**, during the baseline condition, the older adults reported a more negative emotional experience than the younger adults ($p = 0.036$) together with an increased HR compared to their younger counterparts ($p = 0.045$). However, both age groups reported a globally negative (i.e., <4,5/9) emotional experience when listening to the so-called control musical excerpts. No other statistically significant difference between groups was found.

## Age-Related Effects on Spontaneous Responses to Threatening Musical Excerpts

As shown in **Table 3**, our results indicate that, in comparison with their younger counterparts, the older adults has a more positive emotional experience ($p = 0.038$) and showed a smaller SCL ($p = 0.007$). No other statistically significant difference between age groups was observed.

## Age-Related Effects on Cognitive Reappraisal

### Positive Reappraisal

As shown in **Table 4**, the set of one-sample *t* tests ($\mu = 0$) indicates that, when they were asked to positively reappraise threatening musical excerpts, both younger ($p = 0.007$) and older ($p = 0.033$) adults reported a significantly more positive emotional experience than when they were asked to simply listen to it. No statistically significant effect of positive reappraisal instruction was found on the reported intensity of the emotional experience, whatever the age group. Similarly, neither expressive nor physiological responses (SCL, HR) varied as a function of the instruction to positively reappraise musical excerpts, whether in younger or older adults. Concerning the cognitive cost of positive reappraisal, when they were asked to positively reappraise the musical excerpts, the older adults showed a significant gain in their working memory performances in comparison with the simply listening condition ($p = 0.007$). This is not the case in the younger adults, who showed no beneficial effect when implementing positive reappraisal. The age group comparisons confirm that the gain in working memory performances observed in the positive reappraisal condition is specific of older adults

**TABLE 2 |** Age equivalence in baseline condition (control musical stimuli).

| | Younger adults ($n = 35$) | Older adults ($n = 34$) | Age comparisons |
|---|---|---|---|
| | Mean | Mean | Corrected *p*-value |
| **Affective responses** | | | |
| Arousal rating (max. 9) | 3.250 (1.463) | 3.956 (2.341) | 0.478 |
| Valence rating (max. 9) | 3.979 (1.085) | 3.125 (1.790) | 0.036 |
| **Expressive responses** | | | |
| Facial EMG Zygomaticus Major (Area under the curve, mV*sec) | 0.002 (0.008) | 0.0002 (0.0008) | 0.593 |
| Facial EMG Corrugator Supercilii (Area under the curve, mV*sec) | 0.0009 (0.001) | 0.0007 (0.0011) | 0.474 |
| **Physiological responses** | | | |
| SCL's magnitude (μS) | 0.0250 (0.0630) | 0.0120 (0.0330) | 0.108 |
| HR (bpm) | −0.0493 (3.586) | 0.6786 (3.715) | 0.045 |

*Standard deviations are listed in parentheses. Age comparisons were assessed by MannWhitney U test. p values are corrected with the Hochberg procedure for controlling the family-wise error rate.*

**TABLE 3 |** Spontaneous responses to threatening musical excerpts in younger and older adults (simply listen condition).

| | Younger adults ($n$ = 35) | Older adults ($n$ = 34) | Age comparisons |
|---|---|---|---|
| | **Mean** | **Mean** | **Corrected $p$-value** |
| **Affective responses** | | | |
| Arousal rating (max. 9) | 5.295 (1.280) | 5.476 (1.646) | 0.671 |
| Valence rating (max. 9) | 4.024 (1.176) | 4.843 (1.651) | 0.038 |
| **Expressive responses** | | | |
| Facial EMG Zygomaticus Major (Area under the curve, mV*sec) | 0.0016 (0.0052) | 0.0003 (0.0006) | 0.673 |
| Facial EMG Corrugator Supercilii (Area under the curve, mV*sec) | 0.0006 (0.0012) | 0.0005 (0.0006) | 0.947 |
| **Physiological responses** | | | |
| SCL's magnitude (μS) | 0.1663 (0.1342) | 0.0815 (0.1187) | 0.007 |
| HR (bpm) | −0.493 (3.586) | 0.371 (3.913) | 0.671 |
| **Working memory** | | | |
| Mean ratio | 0.7400 (0.1500) | 0.7500 (0.1100) | 0.671 |

*Standard deviations are listed in parentheses. Age comparisons were assessed by Mann–Whitney U test. p values are corrected with the Hochberg procedure for controlling the family-wise error rate.*

**TABLE 4 |** Positive reappraisal's efficiency (differential score: reappraisal condition minus simple listening condition) in younger and older adults.

| | Younger adults ($n$ = 18) | | Older adults ($n$ = 17) | | Age comparisons between differential scores |
|---|---|---|---|---|---|
| | **Mean** | **Statistical difference from 0 (corrected $p$-value)** | **Mean** | **Statistical difference from 0 (corrected $p$-value)** | **Corrected $p$-value** |
| **Affective responses** | | | | | |
| Arousal rating (max. 9) | −0.3796 (1.2895) | 0.534 | −0.2059 (1.5916) | 0.701 | 0.792 |
| Valence rating (max. 9) | 1.2130 (1.3368) | 0.007 | 0.8039 (1.1935) | 0.033 | 0.510 |
| **Expressive responses** | | | | | |
| Facial EMG Zygomaticus Major[a] | −0.0007 (0.0015) | 0.228 | 0.0001 (0.0015) | 0.793 | 0.273 |
| Facial EMG Corrugator Supercilii[a] | 0.0000 (0.0012) | 0.883 | −0.0001 (0.0005) | 0.430 | 0.711 |
| **Physiological responses** | | | | | |
| SCL's magnitude (μS) | 0.0077 (0.0647) | 0.725 | −0.0154 (0.0231) | 0.033 | 0.299 |
| HR (bpm) | 0.8733 (4.7497) | 0.725 | −1.0350 (3.5762) | 0.430 | 0.273 |
| **Working memory** | | | | | |
| Mean ratio | 0.0107 (0.0879) | 0.725 | 0.0869 (0.0917) | 0.007 | 0.042 |

*Standard deviations are listed in parentheses. Age comparisons between differential scores were assessed by Mann–Whitney U test, whereas statistical differences from 0 were assessed by Student's t test. [a]Area under the curve, mV*sec. p values are corrected with the Hochberg procedure for controlling the family-wise error rate.*

($p$ = *0.042*). No other statistically significant difference between age groups was found.

## Detached Reappraisal

As illustrated in **Table 5**, the set of one-sample $t$ tests ($\mu$ = 0) indicates that, when they were asked to detach from the emotion elicited by threatening musical excerpts, older ($p$ = 0.038) but not younger ($p$ = 0.133) adults reported a significantly less negatively valenced experience than when they were in the simply listening condition. No difference was observed for arousing ratings. Neither the expressive nor the physiological responses (SCL, HR) varied as a function of the instruction to use detached reappraisal, whether in younger or older adults. Once again, our findings indicate that, when they were told to detach from the emotion elicited by the threatening musical excerpts, the older adults showed a statistically significant gain in their working memory performances in comparison with the simply listening condition ($p$ = 0.038). This is not the case in the younger adults. However, this is not associated with a significant age group comparison effect ($p$ = 0.493). No other statistically significant difference between age groups was found.

## Positive and Detached Reappraisal Together

When analyzed together, the impact of both positive and detached reappraisal yields supplementary results as shown in **Table 6**. First, the set of one-sample $t$ tests ($\mu$ = 0) indicates that the instruction to reappraise the musical excerpts led to a less arousing emotional experience for younger ($p$ = 0.032) but not older ($p$ = 0.067) adults. Younger adults also reported a more positive emotional experience when they were asked to reappraise than when they were not ($p$ = 0.007). Concerning the cognitive cost of reappraisal, we show again that older adults benefit

**TABLE 5 |** Detached reappraisal's efficiency (differential score: reappraisal condition minus simple listening condition) in younger and older adults.

| | Younger Adults ($n$ = 17) | | Older Adults ($n$ = 17) | | Age comparisons between differential scores |
|---|---|---|---|---|---|
| | Mean | Statistical difference from 0 (corrected *p*-value) | Mean | Statistical difference from 0 (corrected *p*-value) | Corrected *p*-value |
| **Affective responses** | | | | | |
| Arousal rating (max. 9) | −1.1476 (1.8199) | 0.133 | −0.1.3235 (1.8648) | 0.038 | 1.00 |
| Valence rating (max. 9) | 0.3529 (1.0815) | 0.345 | −0.7843 (1.3081) | 0.058 | 0.154 |
| **Expressive responses** | | | | | |
| Facial EMG Zygomaticus Major[a] | 0.0020 (0.0069) | 0.346 | −0.0000 (0.0006) | 0.930 | 0.369 |
| Facial EMG Corrugator Supercilii[a] | −0.0002 (0.0023) | 0.772 | 0.0002 (0.0005) | 0.296 | 0.883 |
| **Physiological responses** | | | | | |
| SCL's magnitude (μS) | −0.0216 (0.0474) | 0.184 | −0.0005 (0.0239) | 0.930 | 0.369 |
| HR (bpm) | −2.3197 (4.9224) | 0.184 | −0.1307 (4.5720) | 0.930 | 0.399 |
| **Working memory** | | | | | |
| Mean ratio | 0.0241 (0.1143) | 0.466 | 0.0600 (0.0865) | 0.038 | 0.493 |

*Standard deviations are listed in parentheses. Age comparisons between differential scores were assessed by Mann–Whitney U test, whereas statistical differences from 0 were assessed by Student's t test. [a]Area under the curve, mV*sec. p values are corrected with the Hochberg procedure for controlling the family-wise error rate.*

**TABLE 6 |** Cognitive (positive and detached) reappraisal's efficiency (differential score: reappraisal condition minus simple listening condition) in younger and older adults.

| | Younger adults ($n$ = 35) | | Older adults ($n$ = 34) | | Age comparisons between differential scores |
|---|---|---|---|---|---|
| | Mean | Statistical difference from 0 (corrected *p*-value) | Mean | Statistical difference from 0 (corrected *p*-value) | Corrected *p*-value |
| **Affective responses** | | | | | |
| Arousal rating (max. 9) | −0.7524 (1.5942) | 0.032 | −0.7647 (1.7986) | 0.067 | 0.952 |
| Valence rating (max. 9) | 0.7952 (1.2783) | 0.007 | 0.0098 (1.4731) | 0.969 | 0.140 |
| **Expressive responses** | | | | | |
| Facial EMG Zygomaticus Major[a] | 0.0006 (0.0050) | 0.595 | 0.0000 (0.0012) | 0.969 | 0.952 |
| Facial EMG Corrugator Supercilii[a] | −0.0001 (0.0017) | 0.852 | 0.0000 (0.0005) | 0.969 | 0.952 |
| **Physiological responses** | | | | | |
| SCL's Magnitude (μS) | −0.0065 (0.0580) | 0.595 | −0.0080 (0.0243) | 0.152 | 0.952 |
| HR (bpm) | −0.6776 (5.0303) | 0.595 | −0.5828 (4.0677) | 0.717 | 0.952 |
| **Working memory** | | | | | |
| Mean ratio | 0.0172 (0.1003) | 0.595 | 0.0735 (0.0888) | 0.001 | 0.091 |

*Standard deviations are listed in parentheses. Age comparisons between differential scores were assessed by Mann–Whitney U test, whereas statistical differences from 0 were assessed by Student's t test. [a]Area under the curve, mV*sec. p values are corrected with the Hochberg procedure for controlling the family-wise error rate.*

from the instruction to reappraise their emotional experience in a way that their working memory performances improve in comparison with the simply listen condition ($p = 0.001$). No effect of the instructions on the behavioral or physiologic measures was found. We found no significant age group comparisons.

**Table 7** shows the comparison between the positive and the detached reappraisal conditions for affective, expressive, physiological, and cognitive outcomes in the younger and older adults, respectively. No statistically significant difference between the two groups of younger adults, respectively, assigned to positive and detached reappraisal was found regardless of the measure considered. Regarding the older adults, one statistically significant difference between the two groups, respectively, assigned to positive and detached reappraisal was found for valence rating ($p = 0.035$): The older adults who were asked to positively reappraise the musical excerpts reported significantly more positive ratings of their emotional experience than those who were asked to implement the detached modality.

## Additional Control Analyses

We conducted non-parametric comparisons (Mann–Whitney) in each age group separately to verify whether the participants assigned to the detached and positive reappraisal conditions differed in their demographical, cognitive, and affective characteristics. Those revealed no statistically significant difference (all corrected $ps > 0.05$, cf. **Supplementary Material** and **Tables 2**, **3**). In addition, we searched for correlations between the affective, expressive, physiological, and cognitive

**TABLE 7 |** Comparison between positive and detached reappraisal's efficiency in younger and older adults.

| | Younger adults ($n$ = 35) | | | Older adults ($n$ = 34) | | |
|---|---|---|---|---|---|---|
| | Positive reappraisal ($n$ = 18) | Detached reappraisal ($n$ = 17) | Group comparisons (corrected $p$-value) | Positive reappraisal ($n$ = 17) | Detached reappraisal ($n$ = 17) | Group comparisons (corrected $p$-value) |
| **Affective responses** | | | | | | |
| Arousal rating (max. 9) | −0.3796 (1.2895) | −1.1476 (1.8199) | 0.250 | −0.2059 (1.5916) | −0.1.3235 (1.8648) | 0.198 |
| Valence rating (max. 9) | 1.2130 (1.3368) | 0.3529 (1.0815) | 0.144 | 0.8039 (1.1935) | −0.7843 (1.3081) | 0.035 |
| **Expressive responses** | | | | | | |
| Facial EMG Zygomaticus Major[a] | −0.0007 (0.0015) | 0.0020 (0.0069) | 0.060 | 0.0001 (0.0015) | −0.0000 (0.0006) | 0.945 |
| Facial EMG Corrugator Supercilii[a] | 0.0000 (0.0012) | −0.0002 (0.0023) | 0.766 | −0.0001 (0.0005) | 0.0002 (0.0005) | 0.198 |
| **Physiological responses** | | | | | | |
| SCL's magnitude (μS) | 0.0077 (0.0647) | −0.0216 (0.0474) | 0.766 | −0.0154 (0.0231) | −0.0005 (0.0239) | 0.198 |
| HR (bpm) | 0.8733 (4.7497) | −2.3197 (4.9224) | 0.060 | −1.0350 (3.5762) | −0.1307 (4.5720) | 0.651 |
| **Working memory** | | | | | | |
| Mean ratio | 0.0107 (1.3368) | 0.0241 (0.1143) | 0.766 | 0.0869 (0.0917) | 0.0600 (0.0865) | 0.221 |

*Standard deviations are listed in parentheses. Group comparisons between differential scores were assessed by Mann–Whitney U test. [a]Area under the curve, mV*sec. p values are corrected with the Hochberg procedure for controlling the family-wise error rate.*

measurements and all the other demographical, affective, and cognitive variables for which we found statistically significant differences between age groups (i.e., music listening, hearing levels, Stroop Victoria IF, PANAS negative affect, and NEO-PI-R scores; cf. **Table 1**). These correlational analyses are aimed to identify potential confounding factors that may explain our results in another way than by age-related or type of reappraisal effects. For each set of correlations, a Bonferroni correction was used to account for the increased chance of a type I error associated with conducting multiple correlations. To this end, we adjusted the $\alpha$ level from 0.05 to 0.00006. No statistically significant correlations were found. For each age group, we searched for correlations between the affective ratings (i.e., valence and arousal), behavioral responses (i.e., EMG), physiological reactions (i.e., skin conductance level and heart beat), and cognitive costs (i.e., mean working span ratio) for the positive and detached reappraisal conditions, respectively (all the tables of correlations are presented as **Supplementary Material** and **Tables 4–7**). We also sought correlations between cognitive inhibition (i.e., Stroop interference score) and cognitive costs (i.e., mean working span ratio) for all participants whatever their age group. After correcting the $\alpha$ level from 0.05 to 0.005 using Bonferroni adjustment, no statistically significant correlation was found ($p = 0.081$).

## DISCUSSION

In the current study, we investigated age-related differences in the use of cognitive reappraisal to regulate emotional responses to threatening musical excerpts. Wanting to reduce the effect of visual attentional deployment (a strategy that may be preferentially used by older adults to disengage from unpleasant events) as a confounding factor on cognitive reappraisal abilities, we used a musical source of emotion. This choice was motivated by the fact that auditory stimuli should prevent participants from reallocating perceptual attention away from the emotional source (except if they plugged their own ears). In a mixed-design framework, participants were instructed to simply listen to negative content of music (spontaneous response) and then to implement cognitive reappraisal to reduce the negative emotion elicited by the music. In the cognitive reappraisal condition, one half of the participants were asked to focus on positive aspects of music (positive reappraisal), and the other half was instructed to detach from what the music conveyed to them (detached reappraisal). Aiming for a full investigation of younger and older adults' emotional responses, we scrutinized the effects of these instructions on affective ratings (Likert scales), facial expressions (facial EMG), physiological state (SCL, HR), and cognitive performances (working memory task) of the participants.

First, our findings give empirical evidence for an age-related effect in spontaneous response to threatening musical excerpts. In line with previous findings (Mather et al., 2004; Vieillard and Bigand, 2014), we found that, when getting older, a reduction of emotional processing for negative stimuli occurs. When instructed to simply listen to threatening music, older adults reported a more positive feeling associated with a smaller SCL in comparison with their younger counterparts. In accordance with previous findings, it appears that, even in the absence of instructed emotional regulation, older adults engage in spontaneous downregulation of their negative emotions (Mather et al., 2004). It is worth noting that younger and older adults judged threatening musical excerpts as moderately arousing. In that sense, our results extend a previous fMRI study providing evidence that, when presented with low arousing negative pictures, compared with younger adults, older adults show an increased spontaneous activity in the prefrontal areas that have been interpreted as suggesting that the regulation networks of older adults are chronically activated in response to low arousing negative stimuli (Dolcos et al., 2014). In the same vain, Samanez-Larkin and Carstensen (2011) postulate that emotion regulation in aging might be more automatic and less cognitively effortful due to chronically activated goals. Nevertheless, the notion of chronically activated goals remains somewhat unclear in the literature. Such anotion logically evokes the idea that the emotion regulation would be based on more automatic processes with aging. Intriguingly, the motivational perspective of SST pleads for the idea that the goal of voluntarily maintaining emotional well-being is chronically activated among older adults and postulates, at the same time, that the preference for positivity and well-being in older adults reflects controlled cognition. This is ambiguous about whether spontaneous emotion regulation is resource demanding or not. As outlined by Isaacowitz and Blanchard-Fields (2012), to date, the logical link postulated by the SST between the cognitive resources required by the positivity effect and the emotion-regulation activity has never received empirical support. This leaves open the question whether spontaneous emotion regulation in aging must be conceived as an automatic or more controlled process. In our study, older adults showed no significant gain in their working memory performances in the simply listening condition in comparison with their younger counterparts. Such a finding does not fully corroborate the hypothesis that the older adults' spontaneous tendency to downregulate negative emotion depends on more automatic processes than in younger adults. This is not to say that spontaneous emotion regulation is synonymous with effortful activity. Previous findings indicating that working memory performances were disrupted by prior efforts at self-regulating some behaviors (i.e., controlling the focus of visual attention, inhibiting predominant writing tendencies, or exaggerating emotional expressions) has been interpreted as supporting the idea that prior effortful tasks may determine the following operation of executive processes (Schmeichel, 2007). Based on such results, the fact that, in our study, the experience of fear *per se* (i.e., simply listening condition) does not affect performances on the working memory task, whatever the age group, suggests that spontaneous emotion regulation is not cognitively demanding in younger or older adults. The fact remains that, even at an equivalent cognitive cost, affective and physiological outcomes show that older adults are more effective at spontaneously regulating their emotions than their younger counterparts.

When the participants were asked to implement a cognitive reappraisal of the threatening musical excerpts that consists of either positively reevaluating the music or detaching from it, no effects of emotion regulation were found on facial expressions (facial EMG) or physiological state (SCL, HR) regardless of the age group. Such findings indicate that the participants did not use a strategy of expressive suppression or enhancement to implement the emotion-regulation instruction. This suggests that the attempt to regulate emotions was not based on a control of behavioral expression, allowing us to think that the participants conformed to the instructions. The fact that neither the positive nor the detached reappraisal instruction elicited a significant increase of physiological arousal is also congruent with previous findings indicating that this kind of emotion regulation has a positive impact on the affective sphere because it is associated with a decreased negative emotion experience without any increase in physiological activation (Cutuli, 2014). In line with this, we observe that both younger and older adults report a reduced negative emotion (valence rating) in the case of positive reappraisal and a reduced arousal (arousal rating) in the case of detached reappraisal. At first glance, these results are consistent with previous results showing equal success in implementing positive and detached reappraisal for younger and older adults (Allard and Kensinger, 2017; Livingstone and Isaacowitz, 2018). However, this view is a little bit challenged by additional findings indicating that, when asked to apply detached reappraisal, the older adults of our sample reported a more negative emotional experience (valence rating) than when they were asked to simply listen to it although this is not the case in their younger counterparts. Even if they are congruent with the phenomenon of emotional dedifferentiation with age (Grühn and Scheibe, 2008; Vieillard et al., 2012), conflicting results observed in older adults are hardly fully explainable in the context of emotion-regulation skills. One could explain them speculating that explicit instructions of detached reappraisal invite adopting a distant/indifferent attitude toward the emotion stimuli directly affecting older adults'—who may show greater compliance with instructions—overt valence ratings (e.g., Henkel, 2014). With respect to the age-related effect on emotion-regulation abilities, current results cannot conclusively determine whether or not older adults use detached reappraisal less successfully than younger adults. This verifies discrepant findings in the literature and asks for further investigations.

As in the previous study conducted by Scheibe and Blanchard-Fields (2009), our results seem to indicate that consequences of cognitive reappraisal for cognitive functioning varies as a function of age group. But it also varies as a function of the type of cognitive reappraisal. When instructed to positively reappraise the threatening musical excerpts, older adults show a significant gain in their working memory performances in comparison with the simply listening condition as well as in comparison with their younger counterparts. A similar pattern of results has been observed for detached reappraisal except that no age-related effect was found. Such findings are in line with the general idea that older adults may be more effective at regulating emotion (e.g., Charles and Carstensen, 2010). Importantly, the current results demonstrate that implementing cognitive reappraisal (positive and detached) does not deplete the limited cognitive resources in older adults. This is not consistent with the claim that older adults would use cognitive resources to regulate their emotions (e.g., Kryla-Lighthall and Mather, 2009). Because our experiment was designed to prevent participants from reallocating auditory attention away from the musical emotions, current results provide further support for the distraction hypothesis of Scheibe and Blanchard-Fields (2009), according to which older adults would be more easily diverted by the working memory task, thus leading to a downregulation of negative emotion. Even if we do not find a statistically significant correlation between the cognitive effect of the implementation of the emotion-regulation activity and the inhibitory capacities of participants, the distraction hypothesis is consistent with the fact that, compared to their younger counterparts, older adults do display lower cognitive performances. Literature on aging revealed that, due to their limited cognitive resources, older adults tend to be more distracted by irrelevant information when they are instructed to perform a cognitive task. In the context of our experiment, the working memory task might have been considered as irrelevant information while implementing reappraisal instructions, at least at the first stage of processing, but might also quickly have worked as a *relevant* mean to divert older adults from their negative emotions. In other words, we suggest that, although instructed to intentionally reappraise negative emotions, older adults may prefer diverting from them the source of emotion rather than implementing a reappraisal strategy. This is in line with the observation that age is associated with an increased preference to choose distraction over reappraisal (Scheibe et al., 2015). Such a hypothesis also corroborates previous findings showing that, compared with those younger, older adults are better at reducing their attention to unpleasant musical stimuli (Vieillard and Bigand, 2014). In that respect, the current pattern of findings is in accordance with the DIT model (Labouvie-Vief, 2008) postulating that older adults become more efficient at downregulating emotion precisely because their cognitive resources are limited. The fact that, in our experiment, older adults judged threatening music as moderately intense is consistent with the idea that musical stimuli elicited enough emotional intensity to lead older adults to divert from it. Does this mean that the current findings do not match the SST model? If, as we suspect, older adults do not really implement a reappraisal strategy *per se* but rather prefer to decrease negative emotions through a distraction strategy, there is no strong empirical evidence to claim that positive and detached reappraisal are less cognitively costly in older than in younger adults. Thus, current findings do not strictly provide evidence against SST model.

Regarding the spontaneous responses to musical excerpts designed for our baseline condition, an unexpected age-related difference in affective and physiological consequences was found. When faced with musical excerpts, such as a tuning orchestra, the older adults reported feeling a more negative emotion and showed greater HR than their younger counterparts, suggesting that age groups were not equivalent in their emotional reaction to music. Although it remains difficult to explain why older adults were more reactive to orchestral sounds in a negative way, our above findings demonstrating that older adults spontaneously downregulate their negative emotion suggest that the older adults' negative appraisal of orchestral sounds (i.e., tuning) did not correspond to a general age-related emotional bias likely to impact the measures of our study. However, the orchestral sounds and threatening musical excerpts were extracted from the classical and the modern repertoire, respectively, which could also explain the different results obtained. More generally, findings outline the difficulty of choosing musical excerpts fit for a baseline condition because musical stimuli of neutral emotional value do not really exist in the natural environment. The present study carries another limitation. Even if we choose to operationalize the cognitive reappraisal with two different instructions conveying either positive or detached reappraisal, our results seem to indicate that older participants may still prefer to use another way, such as distraction, to downregulate their negative emotions. This raises the question of whether and how participants follow verbal instructions in emotion-regulation studies. Further research is needed to shed a light on this key question.

Overall, the current study provides empirical evidence that older adults are better at downregulating their negative emotions. Our data might suggest that such ability is the result of older adults' inclination to use a subsequent working memory task as a distractor, thus facilitating the disengagement from negative music. It may be argued that, because cognitive reappraisal is not required during the working memory task *per se*, cognitive resources required by cognitive reappraisal way not directly affect those in working memory task. As a consequence, current results could be explained in a different manner. For instance, it would be argued that the improved working memory performances in older adults reflect the more sustained mood states in this age group. Based on Dolcos et al. (2006) findings showing that the arousal level during the working memory task may explain the change in working memory performances, one could state that the improvement of working memory performances in older adults is the result of a higher level of arousal able to sustain their mood state. However, we found that the arousal ratings were lower in all participants whatever the age group across reappraisal conditions. This suggests that the improved working memory performances in older adults cannot be explained by a difference in the way the mood state is sustained for the duration of the working memory task. Regarding the question of whether the cognitive resources required by cognitive reappraisal do or do not directly affect those in working memory tasks, previous findings have persuasively demonstrated that prior efforts at executive control do have a significant effect on subsequent operations requiring executive processes (Schmeichel, 2007). Such results do not corroborate the idea that cognitive resources required in the emotional reappraisal task are completely independent of those involved in the working memory task. Based on the pattern of findings observed for older adults, no evidence was found to advocate that positive and detached reappraisal become less cognitively costly while getting old. However, our results indicate that attention processes play an important role in emotion-regulation success in the elderly. If we assume that, in older adults, a low resource–demanding regulation is the best guarantee of success, then the distraction strategy appears as a good candidate for them. This is consistent with the hypothesis, regularly demonstrated in the field of visual attention (e.g., Isaacowitz et al., 2006; Scheibe et al., 2015; Livingstone and Isaacowitz, 2018) that older adults would have pervasive preference to turn away from negative stimuli. It could be argued that the lack of a baseline measure of participant's working memory capacity does not allow us to extract clear conclusions regarding our previous findings because we do not exactly know to what extent the working memory task may reduce cognitive gains or costs depending on the age group. However, the assessment of cognition, including working memory tasks (i.e., reverse counting and spelling, MMSE), provided a guarantee that older adults in our sample have a good level of executive functioning. The aim of the current experimental design was to investigate to what extent initial efforts at executive control required by a emotion-regulation task may affect subsequent efforts at implementing a working memory task. Previous findings observed in younger adults have shown that prior efforts at executive control do have a significant effect on subsequent operations requiring executive processes in this age group (Schmeichel, 2007). Regarding the literature on age-related effect on executive control, it is very unlikely that the magnitude of this effect would be reduced in older adults in comparison with their younger counterparts. At most, if we suppose that older adults in our sample are high functioning, they could have equivalent executive abilities as their younger counterparts. Yet our results indicated that the instruction to intentionally reappraise negative emotions results in a lesser cognitive cost for older adults in comparison with younger adults. This suggests that the potential variation (*a priori* down in older adults) in executive functioning performances across age groups could be an explaining factor for current results as follows: Older adults used the working memory task as a distractor to facilitate their disengagement from negative music. Even if they need further investigations, current findings are in line with the previous work of Scheibe and Blanchard-Fields (2009). Taken together with these previous findings, available empirical data indicate that the distraction strategy used by older adults to reduce their negative feelings applies at least to two types of negative emotions: disgust and fear. Further investigation is required to investigate other categories of feeling. Additional research efforts are also required to determine how situational factors encourage older adults to preferentially apply a specific emotion regulation over another.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ANR, the Institutional Review Board that has funded the research program (ANR-11-EMCO-0003). The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

SV designed the research and conducted the statistical analyzes. SV and EB validated the musical stimuli. SV, CP, and EB wrote the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by a research grant EMCO program (ANR 11 EMCO 003 01) to SV.

## ACKNOWLEDGMENTS

We are grateful to Uliana Borzova for her help in testing the participants.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00216/full#supplementary-material

## REFERENCES


Allard, E. S., and Kensinger, E. A. (2014). Age-related differences in neural recruitment during the use of cognitive reappraisal and selective attention as emotion regulation strategies. *Front. Psychol.* 5:296. doi: 10.3389/fpsyg.2014.00296

Allard, E. S., and Kensinger, E. A. (2017). Cognitive emotion regulation in adulthood and old age: positive gaze preferences across two strategies. *Neuropsychol. Dev. Cogn. Sect B Aging Neuropsychol. Cogn.* 25, 213–230. doi: 10.1080/13825585.2017.1279265

Bayard, S., Erkes, J., and Moroni, C. (2011). Victoria stroop test: normative data in a sample group of older people and the study of their clinical applications in the assessment of inhibition in Alzheimer's Disease. *Archiv. Clin. Neuropsychol.* 26, 653–661. doi: 10.1093/arclin/acr053

Beck, A. T., Steer, R. A., Ball, R., and Ranieri, W. (1996). Comparison of beck depression inventories -IA and -II in psychiatric outpatients. *J. Pers. Assess.* 67, 588–597. doi: 10.1207/s15327752jpa6703_13

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., and Alice, D. (2005). Multidimensional scaling of emotional responses to music: the effect of musical expertise and of the duration of the excerpts. *Cogn. Emot.* 19, 1113–1139. doi: 10.1080/02699930500204250

Brady, B., Kneebone, I. I., Denson, N., and Bailey, P. E. (2018). Systematic review and meta-analysis of age-related differences in instructed emotion regulation success. *PeerJ* 6:e6051. doi: 10.7717/peerj.6051

Brummer, L., Stopa, L., and Bucks, R. (2014). The influence of age on emotion regulation strategies and psychological distress. *Behav. Cogn. Psychother.* 42, 668–681. doi: 10.1017/S1352465813000453

Carstensen, L. L., Isaacowitz, D. M., and Charles, S. T. (1999). Taking time seriously. a theory of socioemotional selectivity. *Am. Psychol.* 54, 165–181. doi: 10.1037/0003-066x.54.3.165

Charles, S. T., and Carstensen, L. L. (2010). Social and emotional aging. *Annu. Rev. Psychol.* 61, 383–409. doi: 10.1146/annurev.psych.093008.100448

Christophe, V., Antoine, P., Leroy, T., and Delelis, G. (2009). Évaluation de deux stratégies de régulation émotionnelle: la suppression expressive et la réévaluation cognitive. *Eur. Rev. Appl. Psychol.* 59, 59–67. doi: 10.1016/j.erap.2008.07.001

Cutuli, D. (2014). Cognitive reappraisal and expressive suppression strategies role in the emotion regulation: an overview on their modulatory effects and neural correlates. *Front. Syst. Neurosci.* 8:175. doi: 10.3389/fnsys.2014.00175

Dolcos, F., LaBar K. S., and Cabeza, R (2006). "The memory-enhancing effect of emotion: functional neuroimaging evidence, in *Memory and Emotion: Interdisciplinary Perspectives*, eds B. Uttl, N. Ohta, and A. L. Siegenthaler (Oxford: Blackwell).

Dolcos, S., Katsumi, Y., and Dixon, R. A. (2014). The role of arousal in the spontaneous regulation of emotions in healthy aging: a fMRI investigation. *Front. Psychol.* 5:681. doi: 10.3389/fpsyg.2014.00681

Gerolimatos, L. A., and Edelstein, B. A. (2012). Predictors of health anxiety among older and young adults. *Intern. Psychogeriatr.* 24, 1998–2008. doi: 10.1017/S1041610212001329

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: divergent consequences for experience, expression, and physiology. *J. Pers. Soc. Psychol.* 74, 224–237. doi: 10.1037/0022-3514.74.1.224

Gross, J. J., and Thompson, R. A. (2007). "Emotion regulation: conceptual foundations," in *Handbook of Emotion Regulation*, ed. J. J. Gross (New York, NY: The Guilford Press), 3–24.

Grühn, D., and Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the international affective picture system (IAPS): do ratings become more extreme with age? *Behav. Res. Methods* 40, 512–521. doi: 10.3758/BRM.40.2.512

Guiliani, N. R., and Gross, J. J. (2009). "Reappraisal," in *Oxford Companion to the Affective Sciences*, eds D. Sander and K. R. Sherer (New York, NY: Oxford University Press).

Henkel, L. A. (2014). Do older adults change their eyewitness reports when re-questioned? *J. Gerontol. Ser B Psychol. Sci. Soc. Sci.* 69, 356–365. doi: 10.1093/geronb/gbt071

Hess, T. M., Beale, K. S., and Miles, A. (2010). The impact of experienced emotion on evaluative judgments: the effects of age and emotion regulation style. *Neuropsychol. Dev. Cogn. Sect. B Aging Neuropsychol. Cogn.* 17, 648–672. doi: 10.1080/13825585.2010.493207

Hofmann, W., Schmeichel, B. J., and Baddeley, A. D. (2012). Executive functions and self-regulation. *Trends Cogn. Sci.* 16, 174–180. doi: 10.1016/j.tics.2012.01.006

Isaacowitz, D. M., and Blanchard-Fields, F. (2012). Linking process and outcome in the study of emotion and aging. *Perspect. Psychol. Sci. J. Assoc. Psychol. Sci.* 7, 3–17. doi: 10.1177/1745691611424750

Isaacowitz, D. M., Wadlinger, H. A., Goren, D., and Wilson, H. R. (2006). Is there an age-related positivity effect in visual attention? A comparison of two methodologies. *Emotion* 6, 511–516. doi: 10.1037/1528-3542.6.3.511

John, O. P., and Gross, J. J. (2004). Healthy and unhealthy emotion regulation: personality processes, individual differences, and life span development. *J. Pers.* 72, 1301–1333. doi: 10.1111/j.1467-6494.2004.00298.x

Knight, M., Seymour, T. L., Gaunt, J. T., Baker, C., Nesmith, K., and Mather, M. (2007). Aging and goal-directed emotional attention: distraction reverses emotional biases. *Emotion* 7, 705–714. doi: 10.1037/1528-3542.7.4.705

Kryla-Lighthall, N., and Mather, M. (2009). "The role of cognitive control in older adults' emotional well-being," in *Handbook of Theories Of Aging*, 2nd Edn, eds V. L. Bengston, D. Gans, N. M. Pulney, and M. Silverstein (New York, NY: Springer Publishing Co), 323–344.

Labouvie-Vief, G. (2008). When differentiation and negative affect lead to integration and growth. *Am. Psychol.* 63, 564–565. doi: 10.1037/0003-066X.63.6.564

Li, T., Fung, H. H., and Isaacowitz, D. M. (2011). The role of dispositional reappraisal in the age-related positivity effect. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 66, 56–60. doi: 10.1093/geronb/gbq074

Liang, Y., Huo, M., Kennison, R., and Zhou, R. (2017). The role of cognitive control in older adult cognitive reappraisal: detached and positive reappraisal. *Front. Behav. Neurosci.* 11:27. doi: 10.3389/fnbeh.2017.00027

Livingstone, K. M., and Isaacowitz, D. M. (2018). The roles of age and attention in general emotion regulation, reappraisal, and expressive suppression. *Psychol. Aging* 33, 373–383. doi: 10.1037/pag0000240

Lohani, M., and Isaacowitz, D. M. (2014). Age differences in managing response to sadness elicitors using attentional deployment, positive reappraisal and suppression. *Cogn. Emot.* 28, 678–697. doi: 10.1080/02699931.2013.853648

Martins, B., Florjanczyk, J., Jackson, N. J., Gatz, M., and Mather, M. (2018). Age differences in emotion regulation effort: pupil response distinguishes reappraisal and distraction for older but not younger adults. *Psychol. Aging* 33, 338–349. doi: 10.1037/pag0000227

Mather, M., Canli, T., English, T., Whitfield, S., Wais, P., Ochsner, K., et al. (2004). Amygdala responses to emotionally valenced stimuli in older and younger adults. *Psychol. Sci.* 15, 259–263. doi: 10.1111/j.0956-7976.2004.00662.x

McRae, K., Gross, J. J., Weber, J., Robertson, E. R., Sokol-Hessner, P., Ray, R. D., et al. (2012). The development of emotion regulation: an fMRI study of cognitive reappraisal in children, adolescents and young adults. *Soc. Cogn. Affect. Neurosci.* 7, 11–22. doi: 10.1093/scan/nsr093

Nowlan, J. S., Wuthrich, V. M., and Rapee, R. M. (2014). Positive reappraisal in older adults: a systematic literature review. *Aging Ment. Health* 19, 475–484. doi: 10.1080/13607863.2014.954528

Nowlan, J. S., Wuthrich, V. M., and Rapee, R. M. (2016). The impact of positive reappraisal on positive (and negative) emotion among older adults. *Intern. Psychoger.* 28, 681–693. doi: 10.1017/S1041610215002057

Opitz, P. C., Rauch, L. C., Terry, D. P., and Urry, H. L. (2012). Prefrontal mediation of age differences in cognitive reappraisal. *Neurobiol. Aging* 33, 645–655. doi: 10.1016/j.neurobiolaging.2010.06.004

Ortner, C. N. M., and de Koning, M. (2013). Effects of regulating positive emotions through reappraisal and suppression on verbal and non-verbal recognition memory. *PLoS One* 8:e62750. doi: 10.1371/journal.pone.0062750

Ortner, C. N. M., Ste Marie, M., and Corno, D. (2016). Cognitive costs of reappraisal depend on both emotional stimulus intensity and individual differences in habitual reappraisal. *PLoS One* 11:e0167253. doi: 10.1371/journal.pone.0167253

Pedder, D. J., Terrett, G., Bailey, P. E., Henry, J. D., Ruffman, T., and Rendell, P. G. (2016). Reduced facial reactivity as a contributor to preserved emotion regulation in older adults. *Psychol. Aging* 31, 114–125. doi: 10.1037/a0039985

Plaisant, O., Courtois, R., Réveillère, C., Mendelsohn, G. A., and John, O. P. (2010). Validation par analyse factorielle du big five inventory français (BFI-Fr). Analyse convergente avec le NEO-PI-R. *Ann. Méd. Psychol. Revue Psychiatr.* 168, 97–106. doi: 10.1016/j.amp.2009.09.003

Richards, J. M., Butler, E. A., and Gross, J. J. (2003). Emotion regulation in romantic relationships: the cognitive consequences of concealing feelings. *J. Soc. Pers. Relat.* 20, 599–620. doi: 10.1177/02654075030205002

Richards, J. M., and Gross, J. J. (2000). Emotion regulation and memory: the cognitive costs of keeping one's cool. *J. Pers. Soc. Psychol.* 79, 410–424. doi: 10.1037/0022-3514.79.3.410

Samanez-Larkin, G. R., and Carstensen, L. L. (2011). "Socioemotional functioning and the aging brain," in *Oxford Library Of Psychology. The Oxford Handbook Of Social Neuroscience*, eds J. Decety and J. T. Cacioppo (Oxford: Oxford University Press), 507–521.

Scheibe, S., and Blanchard-Fields, F. (2009). Effects of regulating emotions on cognitive performance: what is costly for young adults is not so costly for older adults. *Psychol. Aging* 24, 217–223. doi: 10.1037/a0013807

Scheibe, S., Sheppes, G., and Staudinger, U. M. (2015). Distract or reappraise? Age-related differences in emotion-regulation choice. *Emotion* 15, 677–681. doi: 10.1037/a0039246

Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *J. Exper. Psychol. Gen.* 136, 241–255. doi: 10.1037/0096-3445.136.2.241

Sheppes, G., and Meiran, N. (2008). Divergent cognitive costs for online forms of reappraisal and distraction. *Emotion* 8, 870–874. doi: 10.1037/a0013711

Shiota, M. N., and Levenson, R. W. (2009). Effects of aging on experimentally instructed detached reappraisal, positive reappraisal, and emotional behavior suppression. *Psychol. Aging* 24, 890–900. doi: 10.1037/a0017896

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., and Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Tassinary, L. G., Cacioppo, J. T., and Vanman, E. J. (2007). "The skeletomotor system: surface electromyography," in *Handbook of Psychophysiology*, 3rd Edn, eds J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (Berlin: Springer), 267–299. doi: 10.1017/CBO9780511546396.012

Troy, A. S., Shallcross, A. J., Brunner, A., Friedman, R., and Jones, M. C. (2018). Cognitive reappraisal and acceptance: effects on emotion, physiology, and perceived cognitive costs. *Emotion* 18, 58–74. doi: 10.1037/emo0000371

Tucker, A. M., Feuerstein, R., Mende-Siedlecki, P., Ochsner, K. N., and Stern, Y. (2012). Double dissociation: circadian off-peak times increase emotional reactivity; aging impairs emotion regulation via reappraisal. *Emotion* 12, 869–874. doi: 10.1037/a0028207

Urry, H. L., and Gross, J. J. (2010). Emotion regulation in older age. *Curr. Direct. Psychol. Sci.* 19, 352–357. doi: 10.1177/0963721410388395

Vieillard, S., and Bigand, E. (2014). Distinct effects of positive and negative music on older adults' auditory target identification performances. *Q. J. Exper. Psychol.* 67, 2225–2238. doi: 10.1080/17470218.2014.914548

Vieillard, S., Didierjean, A., and Maquestiaux, F. (2012). Changes in the perception and the psychological structure of musical emotions with advancing age. *Exper. Aging Res.* 38, 422–441. doi: 10.1080/0361073X.2012.699371

Vieillard, S., and Gilet, A.-L. (2013). Age-related differences in affective responses to and memory for emotions conveyed by music: a cross-sectional study. *Front. Psychol.* 4:711. doi: 10.3389/fpsyg.2013.00711

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037//0022-3514.54.6.1063

Wechsler, D. (2000). *Wechsler Adult Intelligence Scale—3rd Edition (WAIS-3)*. San Antonio, TX: Harcour Assessment.

Winecoff, A., Labar, K. S., Madden, D. J., Cabeza, R., and Huettel, S. A. (2011). Cognitive and neural contributors to emotion regulation in aging. *Soc. Cogn. Affect. Neurosci.* 6, 165–176. doi: 10.1093/scan/nsq030

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 07 July 2020
doi: 10.3389/fnins.2020.00618

# Beneficial Effects of Musicality on the Development of Productive Phonology Skills in Second Language Acquisition

*Franco Delogu*[1]* *and Yi Zheng*[2]

[1] Department of Humanities, Social Sciences and Communication, Lawrence Technological University, Southfield, MI, United States, [2] Department of Psychology, Stony Brook University, Stony Brook, NY, United States

OPEN ACCESS

**Edited by:**
Gavin M. Bidelman,
University of Memphis, United States

**Reviewed by:**
L. Robert Slevc,
University of Maryland, College Park, United States
Mireille Besson,
Institut de Neurosciences Cognitives de la Méditerranée (INCM), France

***Correspondence:**
Franco Delogu
fdelogu@ltu.edu

**Specialty section:**
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

**Received:** 03 December 2019
**Accepted:** 19 May 2020
**Published:** 07 July 2020

**Citation:**
Delogu F and Zheng Y (2020) Beneficial Effects of Musicality on the Development of Productive Phonology Skills in Second Language Acquisition. Front. Neurosci. 14:618. doi: 10.3389/fnins.2020.00618

Previous studies show beneficial effects of musicality on the acquisition of a second language (L2). While most research focused on perceptual aspects, only few studies investigated the effects of musicality on productive phonology. The present study tested if musicality can predict productive phonological skills in L2 acquisition. Sixty-three students with no previous exposure to Arabic were asked to repeatedly listen to and immediately reproduce short sentences in standard Arabic. Before the sentence reproduction task, they completed an auditory discrimination task in three different between-subjects condition: attentive, in which participants were asked to discriminate phonological variations in the same Arabic sentence that they were asked to reproduce later; non-attentive, in which participants were asked to detect beeps in the same Arabic sentences without paying attention to their phonological content; and no-exposure, in which participants performed the discrimination task in another language (Serbian). The first, third and seventh reproductions of each participant were rated for intelligibility, accent, and syllabic errors by two independent evaluators, both native speakers of Arabic. Primary results showed that the intelligibility of the reproduced sentences was higher in participants with high musicality scores in the Advanced Measures of Music Audiation. Moreover, the intelligibility of sentences produced by highly musical participants improved more over time than the intelligibility of participants with lower musicality scores. Previous exposure to the Arabic sentence was beneficial in both the attentive and non-attentive conditions. Our results support the idea that musicality can have effects on productive skills even in the very first stages of L2 acquisition.

**Keywords: Arabic, auditory working memory, musicality in language, AMMA, productive phonology, second language processing**

## INTRODUCTION

This study focuses on the analysis of the relationships between musicality and speech production skills in L2 acquisition in adults. More specifically, we focus on the association between musicality and the acquisition of very early phonological and prosodic abilities in a second language.

L2 proficiency requires perceptual and productive competencies. Perceptual-receptive competencies include the understanding of lexical and grammatical structures and the detection of phonological contrasts and prosodic cues in the new language. Productive-motor competencies include the ability to compose and produce correct words and sentences in the L2, to produce the correct phonemes, and to produce the correct prosody. Until recently, receptive abilities attracted the most interest of psycholinguistic and cognitive neuroscience research, whereas productive abilities were less investigated. The limited attention of research on the productive aspects of L2 learning proficiency is probably due to the fact that the receptive and productive abilities are believed to be strongly associated. However, several studies found no correlation between the participants' ability to produce and to perceive a given contrast, a segment, or a consonant sequence (Golestani and Pallier, 2007; Kabak and Idsardi, 2007; Park, 2011; Kalathottukaren et al., 2017). Moreover, the evidence of possible different degrees of competence in receptive (passive) and productive (active) bilingualism is consistent with the existence of a gap between the understanding and the production of a second language (e.g., Umbel et al., 1992; Wei et al., 1992).

A limitation of most of the previous studies on L2 productive phonology is the rather heterogeneous competencies of participants, which varies for native language background and the amount of exposure and experience with the target L2. Because the amount of L2 exposure is very hard to control, even in those studies that select the sample from a homogenous population with the same number of years of immersion in the linguistic context, in this study we tested participants on a language in which they had no previous experience and provided a controlled exposure of the target L2 language. With such an approach, we sought to measure the association of factors such as musicality, working memory capacity, and passive and active exposure to the target L2 with the early acquisition of productive phonological and prosodic skills in an unknown L2.

What is the role of musical expertise and musicality in the acquisition of productive phonological skills in an L2? The understanding of the nature of the associations between these inherently human cognitive domains helps shed light on the evolution of the mechanisms of human communication in general. According to Patel's OPERA theory, musical training facilitates language processing due to the overlaps of brain regions that process language and music, and to the great demand of *precision*, *emotion*, *repetition*, and *attention* during music training (Patel, 2010, 2012). Studies have shown that musicians outperformed non-musicians in extracting prosodic information from speech in a familiar language, as well as in a foreign language (Thompson et al., 2003, 2004). However, in contrast with perceptual studies, in which the association between musical skills and second language acquisition has been amply demonstrated (e.g., Delogu et al., 2006, 2010; Magne et al., 2006; Slevc and Miyake, 2006; Marie et al., 2011), the evidence of a positive association between musical skills and L2 productive skills is still debated.

A quantifiable index of productive proficiency is foreign accent (FA), which is often defined as the pronunciation of a language that shows deviations from native standards. Recurrently identified factors influencing L2 phonological skills in general and FA in particular (see Piske et al., 2001 for a review) are the length of residence in an L2-speaking country, gender, formal instruction, motivation, language learning aptitude, amount of native language use, and L1 background. The association between musicality and FA is still a matter of debate. On the one side, there are studies in which musicality failed to correlate with FA (Tahta et al., 1981; Flege et al., 1995). Similarly, Posedel et al. (2012) showed that the number of years of musical training does not predict the pronunciation quality in L2 Spanish. On the other side, Slevc and Miyake (2006) showed a beneficial effect of musical ability on phonological proficiency in L2 English. Pastuszek-Lipinska (2008) showed that musicians outperform non-musicians in shadowing foreign-language sequences in six different languages. Similarly, Stegemööller et al. (2008) suggested that music training facilitates vocal production of speech and song. Also, Milovanov et al. (2010) have shown that university students with higher musical aptitude have a better L2 pronunciation than students with less musical aptitude. More recently, results consistent with an association between musicality and L2 phonological proficiency were obtained using the accent faking paradigm (Coumel et al., 2019).

These contrasting results are probably associated to the variability of methods used across studies to operationalize both the influential factor (musicality) and the dependent variable (productive skills). In some cases, musicality is operationalized as years of musical training (e.g., in Posedel et al., 2012), whereas in other cases as the result of a musicality test (e.g., in Slevc and Miyake, 2006). Likewise, productive skills are also differently measured, whether as verbal shadowing of foreign languages (e.g., Pastuszek-Lipinska, 2008), as amount of FA (e.g., Flege et al., 1995), as the ability to imitate FAs (Christiner and Reiterer, 2013; Reiterer et al., 2013; Coumel et al., 2019), or as the reproduction accuracy of L2 words and sentences (Slevc and Miyake, 2006). Another factor that can cause variability of results is the amount of expertise in (or exposure to) the targeted L2. The amount of previous experience in the language, in fact, is often considered as one of the most important influential factors in phonological proficiency (Flege and Liu, 2001). In this study, as anticipated earlier, we control the influence of previous exposure to the target language by using participants with no previous experience with the target language, and we measure the effect of a controlled exposure on a productive phonology task.

This study aimed at testing several factors able to predict the acquisition of productive skills in second language acquisition. We focused in particular on the influence of musicality. We also took into account the influence of gender, short-term memory capacity, and presence and quality of previous exposure to the target L2. Operatively, we measured the ability of our participants to reproduce a set of sentences in a foreign language after exposing them (or not) to the target language in a discrimination task that preceded the productive task and that required them to pay attention (or not) to the phonological features of the target language.

# METHODS

## Overview

The study included two linguistic tasks: a discrimination task in which participants had to decide if two consecutive sentences in a foreign language were identical or not and a productive task in which participants were asked to listen and reproduce sentences in a foreign language. The experiment also included the Advanced Measures of Music Audiation (AMMA; Gordon, 1989), which is a musicality test that assesses melodic and rhythmic skills, and a digit span memory test, which provided a measure of short-term memory capacity. We decided to use a measure of musicality such as AMMA, instead of a measure of music training (e.g., number of years of formal musical training) for different reasons. First, while music training does not necessarily reflect the actual and current state of musical abilities of participants, a musicality measure such as AMMA provides a quantitative score of musicality on the very day that the L2 task is performed. Second, AMMA allows the inclusion of participants regardless of their musical training and daily hours of practice, allowing a wider generalizability of the results.

## Participants

A total of 63 students from Lawrence Technological University (mean age = 23.60 years, SD = 11.13 years; 32 were female) participated in the study. All participants were native speakers of English with no self-reported auditory impairment conditions and no previous competence or extended exposure to Arabic or Serbian languages. None of them were bilingual. They were randomly assigned to three groups, each of which received different conditions in a same–different linguistic discrimination task. They all received either monetary compensation or course credits for their participation.

## Materials and Tasks

A total of 96 sentences, 48 in Arabic and 48 in Serbian, were recorded by native speakers of Arabic and Serbian. For each language, the recorded sentences were evenly divided between male and female speakers. The variant of Arabic used in this study is modern standard Arabic, which is used in many countries in educational contexts and in the media. Sentences were specifically constructed for this experiment with the following constraints: they were all between 9 and 11 syllables long and have an average duration of ∼2 s. Specifically, the Arabic sentences had an average duration of 2.18 s (SD = 0.29 s), and the Serbian sentences had an average duration of 2.13 s (SD = 0.33 s). The average duration of the sentences in the two languages was not significant [$F_{(1, 46)} = 0.17$, $p = 0.68$]. Most of the sentences in both languages had the same syntactic subject–verb–object or subject–verb–adverb structure. An example of an Arabic sentence is the following: "Alsafar Moreh Bealtaya Rah" (al-sa-far mo-reh be-al-ta-ya rah) [English translation: Traveling by plane is comfortable]. As the meaning of the sentence was not relevant for our experimental goals, the sentences in Arabic and Serbian were not translated from the same English sentences. The complete list of the sentences can be found in **Appendix 1** in **Supplementary Material**. After recording, alterations were applied to single syllables of each of the 96 sentences, so that for each sentence there existed an original version, a pitch-alteration version, a time-alteration version, and a pure tone–added version, for a total of 384 stimuli. Pitch- and time-altered versions were included to operationalize, in a same–different recognition task, two main sources of prosodic variations (pitch and time) that are also crucial in music. The pure-tone version was added to create a condition in which the detection task does not require the participant to put specific attention in the phonological and prosodic dimensions of the sentences. More details about the recognition task will be provided in the description of the procedure. An example of an original sentence, pitch-altered version, time-altered version, and pure tone–added version are available as additional media materials. In order to avoid participants focusing only on specific parts of a sentence in the same–different recognition task, the syllable in which the alteration was applied varies in different sentences. A total of 80 sentences (including original, pitch-, time-, and pure tone–altered sentences) were used as experimental materials, whereas 16 sentences were used in the training sessions.

In the pitch condition, the pitch contour of a single syllable (F0) was altered by transposing the syllable up two or three semitones in order to create weak (two semitones) and strong (three semitones) alterations, respectively (see Schön et al., 2004, for a similar paradigm). The amount of pitch alteration was tested in a pilot study with 10 non-musicians in which the strong alteration was correctly detected 85% of the time and the weak alteration 72% of the time. In the time condition, the duration of a single syllable was altered by stretching the syllable by either 75 or 100% in order to create weak and strong time alterations, respectively. The detectability of time alterations was tested in a pilot study with 10 non-musicians in which the strong alteration was correctly detected 80% of the time and the weak alteration 74% of the time. A third alteration condition was built by inserting a 100-ms-long pure tone in a syllable of each sentence. The pitch of the tone was always 6% (approximately one semitone) higher than the F0 of the syllable's vowel to which the tone was overlapped. Two conditions of detectability were created: an easy condition, in which the loudness of the tone was 10% softer than the perceived loudness of the vowel, and a difficult condition, in which the loudness of the tone was 20% softer than the perceived loudness of the vowel.

## Procedure

### Same-Different Recognition Task

Participants listened to two consecutive sentences separated by 1 s of silence. Their task was to indicate if the two sentences were identical or different by pressing either a "same" key or a "different" key on a keyboard. The same keys were used in all three between-subjects conditions to indicate that the two sentences were identical ("same") or to indicate that a variation occurred ("different"). The same–different recognition task was intended to provide participants with a controlled amount of exposure to the target language (Arabic) before they perform the productive task. The same–different discrimination was performed in three conditions, defined as *attentive* (provides an attentive exposure to Arabic), as *non-attentive* (provides a non-attentive exposure to Arabic), and as *control* (does not provide any exposure to Arabic). The conditions varied according to a between-subject design, where each group of subjects performed only one of the conditions of the experiment. In the *attentive* condition, the second sentence could differ from the first in the pitch or duration of a single syllable. In the non-attentive condition, the second sentence could differ from the first because of a brief beep contained in one of the syllables. We must clarify that by labeling the condition as "non-attentive," we specifically refer to the lack of attention toward the prosodic and phonological information. Still, the auditory stimuli in general must be attentively attended in order to detect the presence of brief beeps during the same–different discrimination task. In the control condition, the task is identical to the one in the attentive condition, but it is performed in a different language (Serbian) than the one (Arabic) to be reproduced in the productive task. The goal of the three conditions in the discrimination tasks was to manipulate the presence and the type of exposure to the language to be reproduced. First, in the *attentive* condition, participants were requested to attentively listen to the sentences in order to detect syllabic variations. Furthermore, because the same sentences they later reproduced in the productive task were used, the participants had the occasion to practice with the sounds and the sentences of the to-be-reproduced language while performing the discrimination task. In fact, while listening, they had to focus on the phonological and prosodic aspects of the sentences in order to detect possible variations between the first and the second sentence. By contrast, in the non-attentive beep condition, participant listened to the same sentences as in the attentive condition, but they did not have to focus on phonological and prosodic content in order to determine variations. In fact, as they simply have to detect the presence of a beep in the second sentence, their task was limited to the detection of a target signal (a pure tone) embedded in an unknown language. In the third, control condition, the attentive task was performed in a different language (Serbian) than the one they have to reproduce, with the consequence that there was no exposure to the language they will reproduce in the productive task.

### Productive Task

For each sentence, participants listened to the recording twice and then reproduced it twice. In order to reduce as much as possible the F0 distance between models and reproductions, male participants listened and reproduced the male version of the native-speaker recordings, whereas female participants listened and reproduced the female version of the native-speaker recordings. In the first reproduction, participants repeated the sentence along with the native-speaker recording, whereas in the second reproduction they produced the sentence aloud without the recording. The second reproduction was recorded via a microphone. For each sentence, this process was sequentially repeated seven times. After seven recorded reproductions of a sentence, the participant started listening to and reproducing a new sentence until each of 10 different sentences was reproduced seven times for a total of 70 recordings per each participant. The order of presentation of the 10 sentences was randomized across participants.

### Digit Span

A computerized version of the digit span forward task was used to determine participants' digit span. Sequences of single digit spoken numbers from one to nine were presented, and participants were required to repeat the sequence aloud. The sequence length began with four numbers, and every two sequences increased in length by one. The sequence presentation ended when the participant made two consecutives mistakes. The longest sequence repeated without making two consecutive mistakes determined the participants' span.

### Advanced Measures of Music Audiation

AMMA (Gordon, 1989) measures the ability to detect tonal and rhythmic variations in a pair of musical fragments. The test includes 30 experimental trials plus three practice trials. For each trial, participants chose between tonal variation, rhythmic variation, and no variation.

## Analysis

The testing phase produced 4,410 audio recordings of sentences spoken by participants (63 participants * 10 sentences * 7 reproductions). Two independent evaluators were recruited to perform a perceptual analysis of the distance between the spoken reproductions and the models. As there is evidence that judges perform better when they are bilingual themselves (Scovel, 1977; Beardsmore, 1980; Flege and Hillenbrand, 1984), we selected two native speakers of Arabic that were also fluent speakers of English. The evaluators were two senior students at Lawrence Technological University: one, a 20-year-old female student of psychology born and raised in Jeddah, Saudi Arabia, and the other, a 21-year-old student of Mechanical Engineering born and raised in Damascus, Syria. Both evaluators used standard modern Arabic daily in school contexts. The evaluators assessed three different aspects of sentence reproduction proficiency: (a) intelligibility, that is, how much of the sentence was comprehensible. This dimension is measured as the percentage of the message that is understood by the evaluator. (b) Accuracy, that is, number of errors. This dimension is measured by the ratio of number of incorrectly reproduced syllables over the total number of reproduced syllables. (c) Foreign accent: evaluators were asked to rate the strength of FA from 0 to 10, where 0 indicates indistinguishable from a native speaker, and 10 indicates an extremely strong FA.

Productions 1, 4, and 7 of each sentence from each participant were evaluated by the two evaluators for intelligibility, correctness, and accent. The interrater agreement was calculated with intraclass correlation coefficient (ICC). Intraclass correlation coefficient values are reported in **Table 1**.

According to widely accepted guidelines (Cicchetti, 1994), the ICC between our two raters expressed an excellent interrater agreement for intelligibility, a good agreement for accent, and a fair agreement for correctness. We ran an analysis of variance (ANOVA) with gender, previous exposure, and AMMA as between-subjects variables and learning as a within-subject

**TABLE 1 |** Intraclass correlation coefficients (ICCs) between the ratings of the two evaluators.

| | Intelligibility | Phonological errors | Foreign accent |
|---|---|---|---|
| ICC 1st reproductions | 0.82 | 0.56 | 0.65 |
| ICC 4th reproductions | 0.88 | 0.53 | 0.63 |
| ICC 7th reproductions | 0.87 | 0.52 | 0.74 |

**TABLE 2 |** Correlation coefficient (Pearson *R*) and relative *p*-values between the digit span scores and each measure of the productive task.

| | Intelligibility | *p* | Syllabic errors | *p* | Accent | *p* |
|---|---|---|---|---|---|---|
| First reproduction | −0.07 | 0.57 | 0.16 | 0.21 | −0.04 | 0.74 |
| Fourth reproduction | −0.04 | 0.65 | 0.07 | 0.61 | −0.05 | 0.78 |
| Seventh reproduction | −0.05 | 0.68 | 0.03 | 0.83 | −0.07 | 0.62 |

**TABLE 3 |** Summary of averages, standard errors and group comparison of intelligibility, phonological errors, and accent scores in the two AMMA groups.

| | | 1st Reproduction | 4th Reproduction | 7th Reproduction |
|---|---|---|---|---|
| Intellegibility | Low AMMA | 0.77 (0.15) | 1.27 (0.20) | 1.6 (0.23) |
| | High AMMA | 1.32 (0.14) | 2.35 (0.18) | 2.91 (0.21) |
| | Significance | *p = 0.036* | *p < 0.0001* | *p < 0.0001* |
| Accent | Low AMMA | 5.05 (0.18) | 5.63 (0.17) | 5.98 (0.18) |
| | High AMMA | 5.33 (0.17) | 6.03 (0.16) | 6.44 (0.16) |
| | Significance | *p > 0.05* | *p = 0.025* | *p = 0.013* |
| Phonological errors | Low AMMA | 4.80 (0.26) | 4.66 (0.28) | 4.55 (0.32) |
| | High AMMA | 5.57 (0.23) | 4.58 (0.25) | 4.23 (0.29) |
| | Significance | *p > 0.05* | *p > 0.05* | *p > 0.05* |

*Italics indicate significant p values.*

variable. As described above, previous exposure has three levels [attentive, non-attentive, control (Serbian)].

## RESULTS

In a preliminary analysis, the digit span scores showed no correlation with any of the three dependent variables at any of the learning stages of the productive task (**Table 2**).

Considering the absence of correlation with any of the dependent variables of the main design and in order to keep low the number of factors, we decided to exclude digit span from the list of factors of the main analysis.

Participants were classified in two different groups according to their score in the AMMA test. Participants who scored equal to or lower than the median value of the distribution of AMMA scores (median = 14) were included in the low AMMA group, whereas participants scoring higher than the median value (≥15) were included in the high AMMA group.

The main results are summarized in **Table 3**.

Concerning intelligibility, results showed that the AMMA was significant, $F_{(1, 51)} = 13.41$, $p = 0.0006$, $\eta^2 = 0.21$, indicating that the high AMMA group (mean = 2.24, SE = 0.19) outperformed the low AMMA group (mean = 1.20, SE = 0.20) in the intelligibility scores. Previous exposure was also significant, $F_{(2, 51)} = 5.12$, $p = 0.009$, $\eta^2 = 0.16$. *Post-hoc* analysis [Fisher least significant difference (LSD)] showed that the active Arabic discrimination group (mean = 2.22, SE = 0.24) had a greater intelligibility score than the Serbian discrimination group (mean = 1.19, SE = 0.28, $p = 0.003$), but did not differ from the passive Arabic group (mean = 1.84, SE = 0.22, $p = 0.26$). The passive Arabic group had significantly higher intelligibility scores than the Serbian group ($p = 0.049$). Gender was also significant, $F_{(1, 51)} = 5.67$, $p = 0.02$, $\eta^2 = 0.10$, indicating that female participants (mean = 2.05, SE = 0.20) scored higher in intelligibility that the male ones (mean = 1.39, SE = 0.19). Learning was significant, $F_{(2, 102)} = 131.92$, $p < 0.0001$, $\eta^2 = 0.72$. *Post-hoc* analysis (Fisher LSD) showed that the seventh reproduction of the sentence (mean = 2.31, SE = 0.17) was more intelligible than the fourth reproduction (mean = 1.86, SE = 0.15, $p < 0.001$), as well as the first reproduction (mean = 1.08, SE = 0.12, $p < 0.001$). The first reproduction was scored as less intelligible than the fourth reproduction ($p < 0.001$).

The interaction between learning and AMMA (**Figure 1**) was significant, $F_{(2, 102)} = 12.733$, $p < 0.0001$, $\eta^2 = 0.199$. *Post-hoc* analysis (Fisher LSD) showed that the *p*-values progressively decreased when comparing low and high AMMA in the first ($p = 0.034$), the fourth ($p = 0.00015$), and the seventh reproduction ($p < 0.00001$), indicating that the difference in intelligibility in the low and high AMMA groups progressively increased with the repetitions.

The interaction between learning and gender was significant, $F_{(2, 102)} = 4.73$, $p = 0.011$. *Post-hoc* analysis indicated that while the intelligibility scores of the first reproduction of female participants were not different from the ones of the male participants ($p = 0.09$), the difference became significant in reproduction 4 ($p = 0.005$) and reproduction 7 ($p = 0.0005$). The interaction between learning and previous exposure was not significant, $F_{(4, 102)} = 1.64$, $p = 0.169$. Also, the interaction between gender and AMMA was not significant, $F_{(1, 51)} = 0.237$, $p = 0.628$, as well as the interaction between gender and previous exposure, $F_{(2, 51)} = 0.03$, $p = 0.97$. The interaction between previous exposure and AMMA was not significant, $F_{(2, 51)} = 0.391$, $p = 0.68$. All of the three-way interactions and four-way interactions were not significant.

Concerning accent, *learning* was significant, $F_{(2, 102)} = 146.2$, $p < 0.00001$, $\eta^2 = 0.74$. *Post-hoc* analysis (Fisher LSD) showed that FA in the seventh reproduction (mean = 6.27, SE = 0.12) was significantly weaker ($p < 0.001$) than in the fourth (mean = 5.88, SE = 0.11) and in the first production (mean = 5.24, SE = 0.11, $p < 0.001$). Additionally, the first reproduction had more accent than the fourth reproduction ($p < 0.001$). Gender was not significant, $F_{(1, 51)} = 2.14$, $p = 0.14$. AMMA was not significant,

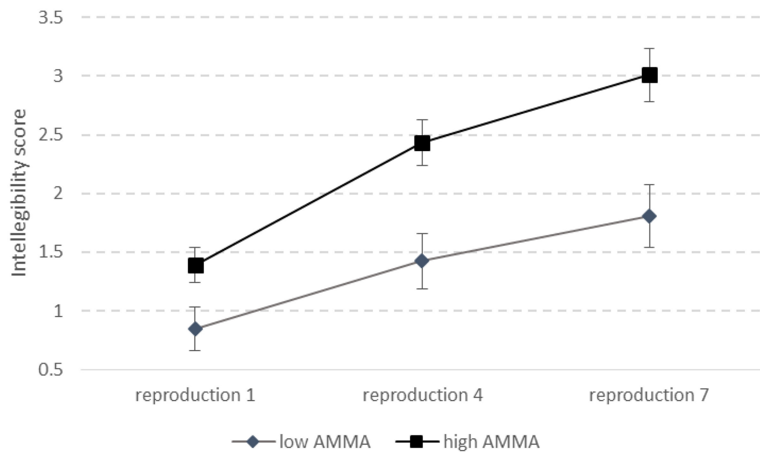



**FIGURE 1 |** Rates of intelligibility as a function of musicality and reproduction attempts.

$F_{(1, 51)} = 2.87$, $p = 0.096$, indicating that participants with low musicality scores (mean = 5.46, SE = 0.17) and participants with high musicality scores (mean = 5.79, SE = 0.14) did not significantly differ in their accent performance. Previous exposure was also not significant, $F_{(2, 51)} = 2.81$, $p = 0.069$, indicating that the active Arabic group (mean = 5.79, SE = 0.19), passive Arabic group (mean = 5.69, SE = 0.17), and Serbian group (mean = 5.39, SE = 0.21) did not vary significantly in their accent scores. The interaction between previous exposure and gender was significant, $F_{(1, 51)} = 3.99$, $p = 0.024$. *Post-hoc* analysis showed that female participants outperformed male participants in accent only in the condition where they were not exposed to Arabic, but to Serbian ($p = 0.0018$), whereas there were no gender differences in the quality of accent in the active ($p = 0.31$) or passive ($p = 0.66$) conditions. All of the other interactions were not significant.

Concerning accuracy (syllabic errors), learning was significant, $F_{(2, 96)} = 32.081$, $p < 0.0001$, $\eta^2 = 0.400$. *Post-hoc* analysis (Fisher LSD) showed that in the seventh reproduction of the sentence (mean = 4.23, SE = 0.22) the number of phonological errors committed was significantly less ($p = 0.0098$) than in the fourth (mean = 4.59, SE = 0.20) as well as in the first ($p < 0.0001$). Additionally, in the first reproduction (mean = 5.49, SE = 0.17) participants committed more errors than the fourth reproduction ($p < 0.0001$). AMMA was not significant, $F_{(1, 48)} = 0.274$, $p = 0.603$, $\eta^2 = 0.0057$, with the low musicality group (mean = 4.86, SE = 0.27) scoring with equivalent levels of accuracy as the high musicality group (mean = 4.68, SE = 0.22). Perceptual type was not significant as well, $F_{(2, 48)} = 0.699$, $p = 0.502$, $\eta^2 = 0.028$, as the active Arabic group (mean = 4.54, SE = 0.30), the passive Arabic group (mean = 4.69, SE = 0.26), and the Serbian group (mean = 5.07, SE = 0.34) did not significantly differ in number of errors. Regarding gender, female participants (mean = 4.62, SE = 0.24) and male participants (mean = 4.92, SE = 0.26) did not differ significantly in their amount of errors, $F_{(1, 48)} = 0.757$, $p = 0.389$, $\eta^2 = 0.016$.

The interaction between AMMA and learning was significant, $F_{(2, 96)} = 4.01$, $p = 0.02$, $\eta^2 = 0.08$. *Post-hoc* analysis (LSD Fisher) showed that among members of the high AMMA group, there was a progressive decrease in errors in successive attempts at reproducing the sentence. $p$ <0.001, 0.0287, and < 0.001 indicate that the differences between the first and fourth, fourth and seventh, and first and seventh reproductions, respectively, were all significant. In contrast, in the low AMMA group, the difference between the fourth and seventh reproductions was not significant ($p = 0.147$). A significant difference appeared only between the first and fourth reproductions and between the first and seventh reproductions, with the following $p$-values, respectively: $p = 0.007$ and $p = 0.00006$. All of the other interactions were not significant.

As we assessed the association between musicality and L2 phonological production by splitting the sample in two groups according to their performance in the AMMA test, we could potentially incur in reliability issues. In fact, as pointed out by MacCallum et al. (2002), an artificial dichotomization of a continuous variable such as AMMA could open the field to a number of issues, including loss of power and effect size and the risk of overlooking non-linear effects. In order to reduce such risks, we calculated the correlation between the continuous scores of AMMA and the intelligibility, phonological errors, and accent scores in their three longitudinal measures (first, fourth, and seventh reproduction). As shown in **Table 4**, the results of the correlation analysis provide analogous results as the ANOVA. It is worth noting that the correlation coefficients for intelligibility and accent increase progressively during the reproduction task.

## DISCUSSION

In this study, we investigated the influence of several factors on L2 speech phonological and prosodic competencies. Specifically, we tested whether musicality, gender, and previous exposure to the linguistic materials influence the intelligibility, accuracy, and level of FA of speech reproductions of Arabic sentences.

Concerning musicality, results showed that musicality is associated with a greater accuracy in sentence reproduction in an unknown language. In fact, the sentences produced by highly

**TABLE 4 |** Correlation between AMMA scores and production measures in the 1st, 4th, and 7th reproductions.

| | Intel. 1 | Intel. 4 | Intel. 7 | Errors 1 | Errors 4 | Errors 7 | Accent 1 | Accent 4 | Accent 7 |
|---|---|---|---|---|---|---|---|---|---|
| AMMA | 0.3139 | 0.4307 | 0.4670 | 0.2289 | 0.0442 | −0.0308 | 0.3378 | 0.4189 | 0.4646 |
| $p$ | $p = 0.012$ | $p = 0.000$ | $p = 0.000$ | $p = 0.071$ | $p = 0.731$ | $p = 0.811$ | $p = 0.007$ | $p = 0.001$ | $p = 0.000$ |

musical participants were significantly more intelligible than the sentences produced by participants with lower musicality scores. It is also notable that highly musical participants tended to improve more with a short amount of training. This result provides new evidence of music–language domain interactions. As far as we know, no previous study tested whether musicality can be associated with the accuracy of the reproduction of spoken sentences in an unknown language. If we consider the association between music and language in general, this result is consistent with previous evidence of music-to-language transfer effects and in particular on L2 acquisition (Slevc and Miyake, 2006; Delogu et al., 2010; Marie et al., 2011).

The role of musicality on the amount of FA is not completely clear in our results. In fact, while the main factor AMMA is not significant, there is a trend for less-accented reproduction by highly musical group in the fourth and the seventh reproductions. Such trend is confirmed by the progressively stronger correlation between the continuous measure of AMMA and the accent scores in the first, fourth, and seventh reproductions (**Table 4**). Such contrasting findings, possibly influenced by the numerosity of the sample, are in line with conflicting evidence in the literature. Specifically, whereas some previous studies did not find an association between musicality or musical expertise and levels of FA (Tahta et al., 1981; Flege et al., 1995; Posedel et al., 2012), other studies found evidence of the influence of musicality on prosodic aspects in second language acquisition (Slevc and Miyake, 2006) and of an association between musical aptitude and better foreign language pronunciation skills in adults (Milovanov et al., 2010). The lack of an effect is perhaps linked to a combination of methodological factors such as the reduced sample size and the not-so-high interrater agreement for the accent measure.

Additionally, the level of musicality of the participants did not influence correctness, measured as the number of syllabic errors. The lack of evidence of an influence is probably due to the difficulty of reporting the number of syllabic errors from the two evaluators, who were not formally trained in linguistics. Such difficulty is expressed by their interrater agreement for syllabic errors estimation, which is the lowest of the three measures.

Regarding prior exposure to the sentences, participants who performed an active or passive perceptual test on the same Arabic sentences produced more intelligible sentences than participants who performed an active perceptual test in another language (Serbian). Interestingly, the groups who performed the active Arabic task (phonological discrimination) and the groups who performed a passive Arabic task (beep detection) did not differ in the level of intelligibility of their sentence reproduction. These results could indicate that previous exposure to the language is a beneficial factor for the production of more intelligible sentences not only when a pre-reproduction exposure is focused on the phonological and prosodic features of the sentence to be reproduced, but also when the listener's attention is diverted toward the detection of near-threshold intrusive signals embedded within the sentence. An alternative interpretation that we currently cannot exclude is that the processing of the Serbian sentences in the same–different recognition task could have created an interference effect that made it more difficult for participants to later reproduce sentence in Arabic. Previous exposure to the sentences did not have an influence on the amount of accent or on the amount of syllabic errors committed.

Gender was important for accent. Consistent with previous evidence (Asher and García, 1969; Tahta et al., 1981; Thompson, 1991), women showed less FA than men. Interestingly, women with low musicality displayed a lower level of FA when compared with men in the same low-musicality group. However, in the high musicality group, the two genders did not differ. This result suggests that the advantage of women over men in accent/prosody/speech (Hyde and Linn, 1988; Buckner et al., 1995; Soleman et al., 2013) can be compensated by musicality. Female and male participants' reproductions did not differ in intelligibility or number of syllabic errors committed.

Performance at the digit span task did not show any significant association with any of the three language production measures. As many studies demonstrate the influence of working memory in language acquisition (Ellis, 1996; Mackey and Sachs, 2012; Williams, 2015), this finding was unexpected. A possible reason is that the digit span forward test we used does not provide an estimate of working memory processing, but only of its capacity. Our productive task could be particularly demanding in working memory processing and not very challenging in terms of capacity. However, it is worth reporting that many findings indicate that short-term memory measures do predict SLA. For example, Ellis (1996) illustrates experiments that use digit span (as used in a previous version of the ITPA) as a successful predictor SLA. Additionally, the PSTM used by Mackey and Sachs (2012) involves a similar procedure to the digit span as a measure of phonological short-term memory.

Intelligibility, accent and syllabic accuracy of the reproduced sentences markedly improved during the course of the productive task. It is of particular importance to note that the amount of improvement observed in both intelligibility and syllabic accuracy was significantly modulated by the musicality of the subjects. In particular, in their first attempt at speech reproduction, participants, regardless of their AMMA score, performed similarly in terms of intelligibility. Crucially, in the fourth and seventh reproductions, although all participants improved, those with high musicality showed greater improvement than those with low musicality.

In conclusion, the results of this study support the hypothesis that musicality can be a beneficial factor in the acquisition of productive skills in a second language even in the very first stages of L2 acquisition, when the imitation of the phonological and prosodic contents of the unfamiliar language is a challenging task for any learner. Not only do people with higher musicality produce more intelligible sentences, but also the comprehensibility of their sentences improves significantly more over time when compared with the sentences produced by people with lower musicality. It should be clarified that our study does not test L2 acquisition *per se*, but simply investigates the ability to reproduce foreign language sounds, which is only a limited aspect of L2 acquisition. Also, the evidence of an association between musicality and sentence reproduction accuracy in an unknown language does not in itself provide evidence of a transfer effect from musicality to language acquisition, as there is still the possibility of underlying unknown factors causing musicality scores and the quality of sentence reproductions to covary. Nevertheless, such evidence of an association between musicality and the accuracy of sentence reproduction supports the idea of the beneficial effects of integrating musical education in second language acquisition pedagogy.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IRB of Lawrence Technological University. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

FD wrote the manuscript and conducted the experimental study. FD and YZ performed the statistical analysis and reviewed the paper. All authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

We thank Amal Alattas and Bilal Sakbani for the precious help in Arabic sentence recording and phonological evaluation, Radovan Miucic and his mother for Serbian sentence recording, and Christopher Yono and Philip McMurray for their help in testing participants.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2020.00618/full#supplementary-material

## REFERENCES


Asher, J. J., and García, R. (1969). The optimal age to learn a foreign language. *Mod. Lang. J.* 53, 334–341. doi: 10.1111/j.1540-4781.1969.tb04603.x

Beardsmore, H. B. (1980). Bilingualism in Belgium. *J. Multiling. Multicult. Dev.* 1, 145–154. doi: 10.1080/01434632.1980.9994007

Buckner, R. L., Raichle, M. E., and Petersen, S. E. (1995). Dissociation of human prefrontal cortical areas across different speech production tasks and gender groups. *J. Neurophysiol.* 74, 2163–2173. doi: 10.1152/jn.1995.74.5.2163

Christiner, M., and Reiterer, S. M. (2013). Song and speech: examining the link between singing talent and speech imitation ability. *Front. Psychol.* 4:874. doi: 10.3389/fpsyg.2013.00874

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* 6:284. doi: 10.1037/1040-3590.6.4.284

Coumel, M., Christiner, M., and Reiterer, S. M. (2019). Second language accent faking ability depends on musical abilities, not on working memory. *Front. Psychol.* 10:257. doi: 10.3389/fpsyg.2019.00257

Delogu, F., Lampis, G., and Belardinelli, M. O. (2006). Music-to-language transfer effect: May melodic ability improve learning of tonal languages by native nontonal speakers? *Cogn. Process.* 7, 203–207. doi: 10.1007/s10339-006-0146-7

Delogu, F., Lampis, G., and Belardinelli, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *Eur. J. Cogn. Psychol.* 22, 46–61. doi: 10.1080/09541440802708136

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Stud. Second Lang. Acquis.* 18, 91–126. doi: 10.1017/S0272263100014698

Flege, J. E., and Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *J. Acoust. Soc. Am.* 76, 708–721. doi: 10.1121/1.391257

Flege, J. E., and Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Stud. Second Lang. Acquis.* 23, 527–552. doi: 10.1017/S0272263101004041

Flege, J. E., Munro, M. J., and MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *J. Acoust. Soc. Am.* 97, 3125–3134. doi: 10.1121/1.413041

Golestani, N., and Pallier, C. (2007). Anatomical correlates of foreign speech sound production. *Cereb. Cortex* 17, 929–934. doi: 10.1093/cercor/bhl003

Gordon, E. E. (1989). *Advanced Measures of Music Audiation*. Chicago, IL: GIA.

Hyde, J. S., and Linn, M. C. (1988). Gender differences in verbal ability: a meta-analysis. *Psychol. Bull.* 104, 53–69. doi: 10.1037/0033-2909.104.1.53

Kabak, B., and Idsardi, W. J. (2007). Perceptual distortions in the adaptation of english consonant clusters: syllable structure or consonantal contact constraints. *Lang. Speech* 50, 23–52. doi: 10.1177/00238309070500010201

Kalathottukaren, R. T., Purdy, S. C., and Ballard, E. (2017). Prosody perception and production in children with hearing loss and age-and gender-matched controls. *J. Am. Acad. Audiol.* 28, 283–294. doi: 10.3766/jaaa.16001

MacCallum, R. C., Zhang, S., Preacher, K. J., and Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychol. Methods* 7:19. doi: 10.1037/1082-989X.7.1.19

Mackey, A., and Sachs, R. (2012). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Lang. Learn.* 62, 704–740. doi: 10.1111/j.1467-9922.2011.00649.x

Magne, C., Schön, D., and Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *J. Cogn. Neurosci.* 18, 199–211. doi: 10.1162/089892906775783660

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., and Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *J. Cogn. Neurosci.* 23, 2701–2715. doi: 10.1162/jocn.2010.21585

Milovanov, R., Pietilä, P., Tervaniemi, M., and Esquef, P. A. (2010). Foreign language pronunciation skills and musical aptitude: a study of Finnish adults with higher education. *Learn. Individ. Differ.* 20, 56–60. doi: 10.1016/j.lindif.2009.11.003

Park, J. S. Y. (2011). The promise of English: Linguistic capital and the neoliberal worker in the South Korean job market. *Int. J. Biling. Educ. Biling.* 14, 443–455. doi: 10.1080/13670050.2011.573067

Pastuszek-Lipinska, B. (2008). "Musicians outperform nonmusicians in speech imitation," in *Computer Music Modeling and Retrieval. Sense of Sounds. CMMR 2007. Lecture Notes in Computer Science, Vol. 4969,* eds R. Kronland-Martinet, S. Ystad, and K. Jensen (Berlin; Heidelberg: Springer), 56–73. doi: 10.1007/978-3-540-85035-9_4

Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford, UK: Oxford University Press.

Patel, A. D. (2012). The OPERA hypothesis: assumptions and clarifications. *Ann. N. Y. Acad. Sci.* 1252, 124–128. doi: 10.1111/j.1749-6632.2011.06426.x

Piske, T., MacKay, I. R., and Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *J. Phon.* 29, 191–215. doi: 10.1006/jpho.2001.0134

Posedel, J., Emery, L., Souza, B., and Fountain, C. (2012). Pitch perception, working memory, and second-language phonological production. *Psychol. Music* 40, 508–517. doi: 10.1177/0305735611415145

Reiterer, S. M., Hu, X., Sumathi, T. A., and Singh, N. C. (2013). Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Front. Psychol.* 4:782. doi: 10.3389/fpsyg.2013.00782

Schön, D., Magne, C., and Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349. doi: 10.1111/1469-8986.00172.x

Scovel, T. (1977). *The Ontogeny of the Ability to Recognize Foreign Accents.*

Slevc, L. R., and Miyake, A. (2006). Individual differences in second-language proficiency: does musical ability matter? *Psychol. Sci.* 17, 675–681. doi: 10.1111/j.1467-9280.2006.01765.x

Soleman, R. S., Schagen, S. E., Veltman, D. J., Kreukels, B. P., Cohen-Kettenis, P. T., Lambalk, C. B., et al. (2013). Sex differences in verbal fluency during adolescence: a functional magnetic resonance imaging study in gender dysphoric and control boys and girls. *J. Sex. Med.* 10, 1969–1977. doi: 10.1111/jsm.12083

Stegemööller, E. L., Skoe, E., Nicol, T., Warrier, C. M., and Kraus, N. (2008). Music training and vocal production of speech and song. *Music Percept.* 25, 419–428. doi: 10.1525/mp.2008.25.5.419

Tahta, S., Wood, M., and Loewenthal, K. (1981). Age changes in the ability to replicate foreign pronunciation and intonation. *Lang. Speech* 24, 363–372. doi: 10.1177/002383098102400405

Thompson, I. (1991). Foreign accents revisited: the English pronunciation of Russian immigrants. *Lang. Learn.* 41, 177–204. doi: 10.1111/j.1467-1770.1991.tb00683.x

Thompson, W. F., Schellenberg, E. G., and Husain, G. (2003). Perceiving prosody in speech: effect of music lessons. *Ann. N. Y. Acad. Sci.* 999, 530–532. doi: 10.1196/annals.1284.067

Thompson, W. F., Schellenberg, E. G., and Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emotion* 4:46. doi: 10.1037/1528-3542.4.1.46

Umbel, V. M., Pearson, B. Z., Fernández, M. C., and Oller, D.K. (1992). Measuring bilingual children's receptive vocabularies. *Child Dev.* 4, 101 2–1020.

Wei, L., Milroy, L., and Ching, P. S. (1992). A two?step sociolinguistic analysis of code?switching and language choice: the example of a bilingual Chinese community in Britain. *Int. J. Appl. Linguist.* 2, 63–86. doi: 10.1111/j.1473-4192.1992.tb00024.x

Williams, J. (2015). 17 Working memory in sla research: challenges and prospects. *Work. Mem. Second Lang. Acquis. Process.* 87:301. doi: 10.21832/9781783093595-021

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 18 August 2020
doi: 10.3389/fpsyg.2020.01927

# Musicians Show Improved Speech Segregation in Competitive, Multi-Talker Cocktail Party Scenarios

*Gavin M. Bidelman[1,2,3]* and Jessica Yoo[2]*

*[1] Institute for Intelligent Systems, University of Memphis, Memphis, TN, United States, [2] School of Communication Sciences and Disorders, University of Memphis, Memphis, TN, United States, [3] Department of Anatomy and Neurobiology, University of Tennessee Health Sciences Center, Memphis, TN, United States*

Studies suggest that long-term music experience enhances the brain's ability to segregate speech from noise. Musicians' "speech-in-noise (SIN) benefit" is based largely on perception from simple figure-ground tasks rather than competitive, multi-talker scenarios that offer realistic spatial cues for segregation and engage binaural processing. We aimed to investigate whether musicians show perceptual advantages in cocktail party speech segregation in a competitive, multi-talker environment. We used the coordinate response measure (CRM) paradigm to measure speech recognition and localization performance in musicians vs. non-musicians in a simulated 3D cocktail party environment conducted in an anechoic chamber. Speech was delivered through a 16-channel speaker array distributed around the horizontal soundfield surrounding the listener. Participants recalled the color, number, and perceived location of target callsign sentences. We manipulated task difficulty by varying the number of additional maskers presented at other spatial locations in the horizontal soundfield (0–1–2–3–4–6–8 multi-talkers). Musicians obtained faster and better speech recognition amidst up to around eight simultaneous talkers and showed less noise-related decline in performance with increasing interferers than their non-musician peers. Correlations revealed associations between listeners' years of musical training and CRM recognition and working memory. However, better working memory correlated with better speech streaming. Basic (QuickSIN) but not more complex (speech streaming) SIN processing was still predicted by music training after controlling for working memory. Our findings confirm a relationship between musicianship and naturalistic cocktail party speech streaming but also suggest that cognitive factors at least partially drive musicians' SIN advantage.

**Keywords: acoustic scene analysis, stream segregation, experience-dependent plasticity, musical training, speech-in-noise perception**

OPEN ACCESS

***Edited by:***
*Cunmei Jiang,*
*Shanghai Normal University, China*

***Reviewed by:***
*Outi Tuomainen,*
*University of Potsdam, Germany*
*Mireille Besson,*
*Institut de Neurosciences Cognitives de la Méditerranée (INCM), France*

****Correspondence:***
*Gavin M. Bidelman*
*gmbdlman@memphis.edu*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology*

***Received:*** *10 April 2020*
***Accepted:*** *13 July 2020*
***Published:*** *18 August 2020*

***Citation:***
*Bidelman GM and Yoo J (2020) Musicians Show Improved Speech Segregation in Competitive, Multi-Talker Cocktail Party Scenarios. Front. Psychol. 11:1927. doi: 10.3389/fpsyg.2020.01927*

## INTRODUCTION

In naturalistic sound environments, the auditory system must extract target speech and simultaneously filter out extraneous sounds for effective communication – the classic "cocktail-party problem" (Cherry, 1953; Bregman, 1978; Yost, 1997). Auditory stream segregation refers to the ability to identify and localize important auditory objects (cf. sources) in the soundscape. The ability to stream is highly relevant to both speech and music perception, e.g., communicating in a noisy restaurant or following a symphonic melody (Fujioka et al., 2005; Shamma et al., 2011). Successful streaming depends on Gestalt-like processing (Holmes and Griffiths, 2019) but also hearing out important acoustic cues including harmonic structure, spatial location, and onset asynchrony, all of which can promote or deny perceptual segregation (Carlyon, 2004; Alain, 2007).

Several experiential (e.g., language expertise and musical training) and cognitive factors [e.g., attention and working memory (WM)] have been shown to influence auditory stream segregation (Broadbent, 1958; Fujioka et al., 2005; Singh et al., 2008; Zendel and Alain, 2009; Ruggles and Shinn-Cunningham, 2011; Bidelman and Dexter, 2015; Zendel et al., 2015). Musicianship, in particular, has been associated with widespread perceptual–cognitive enhancements that help the brain resolve the cocktail party problem. Indeed, musically savvy individuals are highly sensitive to changes in auditory space (Munte et al., 2001) and tracking voice pitch (Bidelman et al., 2011) and are better than their non-musician peers at detecting inharmonicity in sound mixtures (Zendel and Alain, 2009). These features are prominent cues that signal the presence of multiple acoustic sources (Popham et al., 2018), and musicians excel at these skills.

A widely reported yet controversial benefit of music engagement is the so-called "musician advantage" in speech-in-noise (SIN) processing (for review, see Coffey et al., 2017). Several studies demonstrate that musicians outperform non-musicians in figure-ground perception, as measured in a variety of degraded speech recognition tasks (Bidelman and Krishnan, 2010; Parbery-Clark et al., 2011; Swaminathan et al., 2015; Anaya et al., 2016; Clayton et al., 2016; Brown et al., 2017; Deroche et al., 2017; Du and Zatorre, 2017; Mankel and Bidelman, 2018; Torppa et al., 2018; Yoo and Bidelman, 2019). Amateur musicians (~10 years training) are better at identifying and discriminating target speech amidst acoustic interferences including reverberation (Bidelman and Krishnan, 2010) and noise babble (Parbery-Clark et al., 2009a). In standardized (audiological) measures of SIN perception [e.g., Hearing in Noise Test (HINT) and QuickSIN test] (Nilsson et al., 1994; Killion et al., 2004), musicians also tolerate ~1 dB more noise than their non-musician peers during degraded speech recognition (Parbery-Clark et al., 2009b; Zendel and Alain, 2012; Mankel and Bidelman, 2018; Yoo and Bidelman, 2019). Similar results transfer to non-speech sounds (Fuller et al., 2014; Başkent et al., 2018). Still, not all studies report a positive effect, and some fail to find a musician advantage even on identical SIN tasks (e.g., QuickSIN and HINT) (Ruggles et al., 2014; Boebinger et al., 2015; Madsen et al., 2017; Yeend et al., 2017; Escobar et al., 2020). The failure to replicate could be due to the small nature of this effect and/or, as we have previously suggested, unmeasured differences in music aptitude even among self-reported musicians that confer perceptual gains in SIN processing (Bidelman and Mankel, 2019). Musicians' SIN benefits are also more apparent in older adults (Zendel and Alain, 2012), so the predominance of studies on young adults may not be representative of music-related SIN benefits. Nevertheless, a handful of studies suggest (albeit equivocally) that music training might improve the ability to segregate multiple sound streams in relation to cocktail party listening.

To date, prior studies on the effects of long-term music experience and SIN processing have focused on simple headphone-based figure-ground tasks rather than stream segregation or true cocktail party listening, *per se* (but see Madsen et al., 2019). We know that musicians are less affected by informational masking (Oxenham and Shera, 2003; Swaminathan et al., 2015; Yoo and Bidelman, 2019) and that their SIN advantages are stronger when targets and maskers are both speech (Swaminathan et al., 2015; Yoo and Bidelman, 2019). For example, using a task decomposition strategy (e.g., Coffey et al., 2017), we recently examined musicians' performance in a number of speech (and non-speech) masking tasks in order to identify conditions under which musicians show listening benefits in adverse acoustic conditions (Yoo and Bidelman, 2019). We found that musicians excelled in SIN perception but most notably for speech-on-speech masking conditions, i.e., those containing substantial linguistic interference and higher degrees of information masking (see also Swaminathan et al., 2015). Thus, the "musician SIN benefit" depends largely on task structure (Yoo and Bidelman, 2019). Moreover, cocktail party listening draws upon general cognitive faculties (e.g., memory and attention), and musicians are known to differ from non-musicians in the domains of WM (Bugos et al., 2007; Bidelman et al., 2013; Yoo and Bidelman, 2019), attention (Strait et al., 2010; Strait and Kraus, 2011; Sares et al., 2018; Medina and Barraza, 2019; Yoo and Bidelman, 2019), and executive functioning (Bugos et al., 2007; Bialystok and Depape, 2009; Moreno et al., 2011; for review, see Moreno and Bidelman, 2014; Zuk et al., 2014; Lerousseau et al., 2020).

While we and others have shown musicians are unusually good at parsing simultaneous speech (at least diotically/monaurally), it remains unclear if these benefits translate to more naturalistic acoustic environments that offer spatial cues for segregation and engage binaural processing. Spatialization is an important acoustic cue listeners exploit to parse multiple talkers and aid speech recognition in normal cocktail party scenarios (Nelson et al., 1998). This realistic component of normal scene analysis is not testable using conventional SIN tests, limiting ecological validity of previous work. Moreover, given evidence that musicianship might engender enhanced cognitive functioning (Schellenberg, 2005; Moreno and Bidelman, 2014), we were interested to test the degree to which musicians' cocktail party benefits might be explained by domain general skills.

In light of the equivocal nature of musicians' SIN benefit(s), our aim was to assess whether they show perceptual advantages in speech segregation in a competitive, multi-talker environment, thereby confirming their putative SIN benefits but extending them to more ecological "cocktail party" scenarios. To this end, we measured speech streaming abilities in musicians and non-musicians using a realistic, 3-D cocktail party environment. The study was conducted in the unique setting of an anechoic chamber with surround sound stimulus presentation. We hypothesized musicians would show more accurate performance than non-musicians in cocktail party speech recognition and localization tasks, extending prior results from laboratory-based SIN tasks. We further expected to find associations among cognitive factors such as attention and WM with stream-segregation performance (e.g., Yoo and Bidelman, 2019). This would suggest a role of cognitive factors in partially driving musicians' cocktail party advantages.

## MATERIALS AND METHODS

### Participants

Young ($N$ = 28, age range: 19–33 years), normal-hearing adults were recruited for the study. The sample was divided into two groups based on self-reported music experience. Fourteen musicians (M; nine females and five males) had at least 9 years of continuous training (15.07 $\pm$ 4.14 years) on a musical instrument starting before age 10 (7.2 $\pm$ 2.49 years). Fourteen non-musicians (NM; 10 females and 4 males) were those with $\leq$4 years (0.89 $\pm$ 1.23 years) of lifetime music training on any combination of instruments. Instruments included piano (2), percussion (3), oboe (1), tuba (1), voice (1), saxophone (1), trumpet (1), French horn (2), guitar/bass (1), and clarinet (1). All were currently active in playing their instrument in an ensemble or private setting. All showed normal-hearing sensitivity (puretone audiometric thresholds $\leq$ 25 dB HL; 250 to 8,000 Hz) and had no previous history of brain injury or psychiatric problems. Non-native speakers perform worse on SIN tasks than their native-speaking peers. Thus, all participants were required to be native speakers of English (Rogers et al., 2006; Bidelman and Dexter, 2015). The two groups were otherwise matched in age ($t_{26}$ = −0.43, $p$ = 0.67), right-handedness as measured by the Edinburgh Handedness inventory (Oldfield, 1971; $t_{26}$ = 1.84, $p$ = 0.08), gender (Fisher's exact test: $p$ = 1.0), formal education ($t_{26}$ = 0.51, $p$ = 0.62), and socioeconomic status ($t_{26}$ = 0.48, $p$ = 0.64), scored based on the highest level of parental education: 1 (high school without diploma or GED)–6 (doctoral degree) (Norton et al., 2005; Mankel and Bidelman, 2018). Each gave written informed consent in accordance with a protocol approved by the University of Memphis Institutional Review Board.

### Stimuli and Task Paradigms

We measured naturalistic cocktail party listening skills via a sentence-on-sentence speech recognition task (Bolia et al., 2000) conducted in a 3D spatial field (described below). As a comparison to normed SIN measures, we also measured QuickSIN scores (Killion et al., 2004), which have previously revealed musician advantages in SIN perception (Zendel and Alain, 2012; Mankel and Bidelman, 2018; Yoo and Bidelman, 2019). Domain general cognitive skills [i.e., fluid intelligence (IQ), WM, and sustained attention] were evaluated using Raven's progressive matrices (Raven et al., 1998), backwards digit span (Wechsler et al., 2008), and the Sustained Attention to Response Task (SART) (Robertson et al., 1997), respectively.

#### Speech Streaming Task

We measured speech recognition and localization performance in a simulated multi-talker cocktail party environment within the University of Memphis Anechoic Chamber (**Figure 1A**)[1]. A 16-channel circular speaker array was positioned vertically 130 cm above the mesh floor of the anechoic chamber (approximately ear height). Each speaker had a radial distance of 160 cm to the center of the head. Speaker-to-speaker distance was ∼20°. Stimuli were presented at 70 dB SPL (z-weighted, free field), calibrated using a Larson–Davis sound level meter (Model LxT).

We used coordinate response measure (CRM) sentences (Bolia et al., 2000) to measure speech recognition in a multi-talker sounds mixture. CRM sentences contain a different target callsign (Charlie, Ringo, Laker, Hopper, Arrow, Tiger, Eagle, and Baron), color (blue, red, white, and green), and number (1–8) combination embedded in a carrier phrase (e.g., "Ready *Charlie*, go to *blue three* now"). The corpus contained all possible permutations of these callsign–color–number combinations spoken by eight different talkers (male and female). We used CRM sentences as they are sufficiently novel to listeners to avoid familiarity effects that might confound SIN recognition (Johnsrude et al., 2013; Holmes et al., 2018). They are also natural productions that offer a level of control (e.g., similar length and same sentence structure). Participants were cued to the target callsign before each block and were instructed to recall its color–number combination via a sequential button press on the keyboard as fast and accurately as possible (e.g., "b2" = blue–two and "r6" = red–six). We logged both recognition accuracy and reaction times (RTs). RTs were clocked from the end of the stimulus presentation. There were a total of 32 trials per block, repeated twice (i.e., 64 trials per masker condition).

On each trial, listeners heard a mixture of sentences, one of which contained the target callsign and additional CRM sentence(s) that functioned as multi-talker masker(s). Three additional constraints were imposed on sentence selection to avoid unnecessary task confusion: (1) targets were always from the same talker and callsign (within a block); (2) maskers were absent of any callsign, color, and number used in the target phrase (i.e., the callsign's information was unique among the speech mixture); and (3) target and masker(s) were presented from unique spatial locations (i.e., different speakers). The target speaker/callsign was allowed to vary between blocks but was fixed within block. Males and females were selected randomly. Thus, on average, targets and maskers were 50% male and 50% female. Presentation order and spatial location of the sentences in the 360° soundfield were otherwise selected randomly (**Figure 1B**).

In separate blocks, we manipulated task difficulty by parametrically varying the number of additional maskers (0, 1, 2, 3, 4, 6, and 8) presented at other spatial locations in the speaker array. We required participants to identify *both* the call color and number of the target callsign phrase to be considered a correct response (chance level = 3.13% = 1/32). It is possible

[1]The University of Memphis facility is a room-within-a room design featuring a 24 ft × 24 ft × 24 ft IAC anechoic chamber with floor/wall/ceiling Metadyne® acoustic wedge coverage. The noise lock provides an STC 61 noise rating (low cutoff frequency = 100 Hz). A 36-channel Renkus-Heinz speaker array surrounds the seating location (16 were used in the experiment). Multichannel audio control is achieved by a TDT RX8 Multi-I/O Processor (Tucker Davis Technologies). Six Focusrite and Ashley Ne8250 amplifiers drive the speakers via a RedNet Dante MADI interface.

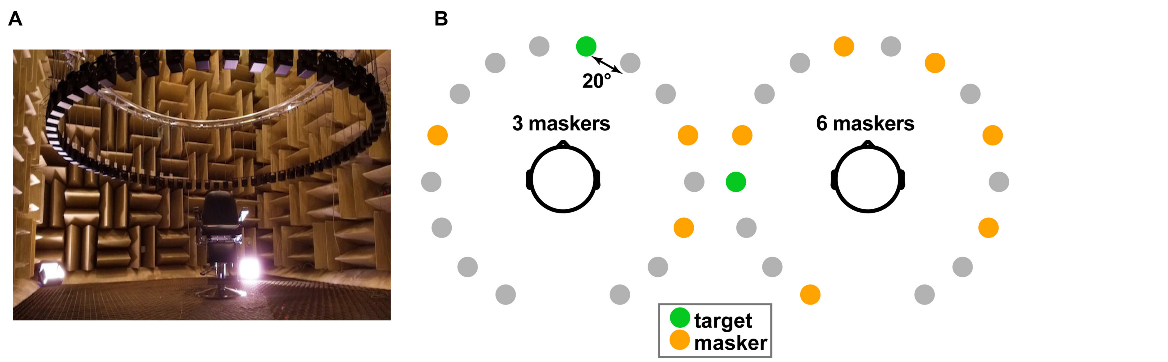



**FIGURE 1 |** Cocktail party streaming task. **(A)** Participants were seated in the center of a 16-channel speaker array within an anechoic chamber. Speaker heights were positioned at ear level (~130 cm) during the task with a radial distance of 160 cm to the center of the head and speaker-to-speaker distance of ~20°. **(B)** Example stimulus presentation (three- and six-talker conditions). Participants were asked to recall the color, number, and perceived location of target callsign sentences from the coordinate response measure (CRM) corpus (Bolia et al., 2000). Target location was varied randomly from trial to trial and occurred simultaneous with between zero and eight concurrent masking talkers.

for listeners to localize sound sources even if they cannot identify them (Rakerd et al., 1999). Consequently, after recognition, we had participants indicate the perceived location (azimuth) of the target by clicking on a visual analog of the speaker array displayed on the screen (cf. **Figure 1B**).

### QuickSIN

The QuickSIN provided a normed test of SIN reception thresholds. Participants heard six sentences embedded in four-talker noise babble, each containing five keywords. Sentences were presented at 70 dB HL. The signal-to-noise ratio (SNR) decreased parametrically in 5 dB steps from 25 to 0 dB SNR. At each SNR, participants were instructed to repeat the sentence, and correctly recalled keywords were logged. We computed their SNR loss by subtracting the number of recalled target words from 25.5 (i.e., SNR loss = 25.5 - total correct). The QuickSIN was presented binaurally via Sennheiser HD 280 circumaural headphones. Two lists were run, and the second was used in subsequent analysis to avoid familiarization effects (Yoo and Bidelman, 2019).

### SART

Attention was assessed using the SART (Robertson et al., 1997) implemented in PsychoPy2 (Peirce et al., 2019). Participants rapidly pressed a button for digits (1–9) presented on the computer screen but withheld their response for the digit 3 (i.e., Go/No-Go paradigm). Both correct and incorrect responses were logged, allowing for analysis of omission and commission errors (Van Schie et al., 2012).

### Digit Span

Backwards digit span was used to assess WM ability. The test consisted of seven questions (each repeated twice). A series of digits was verbally presented to listeners (~1/s), which varied in sequence length. The length started with two digits (e.g., 2 and 4) and progressively increased to eight digits (e.g., 7, 2, 8, 1, 9, 6, 5, and 3). Participants had to recall the sequence in reverse order. Participants were given 1 point for each correct response. The total score (out of 14) was taken as the individual's WM capacity.

### Raven's Matrices

Raven's (1998) progressive matrices was used to evaluate non-verbal fluid IQ. Each question contained a 3 × 3 matrix of different abstract patterns and shapes, and participants were instructed to select the missing pattern from one of eight options. Questions became progressively more difficult, which required greater reasoning ability and intellectual capacity. One of two test versions was randomly chosen. They were given 10 min to complete 29 questions. Percent correct scores were recorded.

## Statistical Analysis

Group differences were evaluated for each auditory/cognitive task using independent-samples *t*-tests. Tukey–Kramer adjustments corrected for multiple comparisons. We conducted two-way, mixed-model ANOVAs (group × masker count; subjects = random effect) on speech streaming measures (% accuracy, RTs, and localization error). The control (zero masker) condition was excluded from the ANOVA, though we note that the results were qualitatively similar with or without its inclusion. Dependent measures were log(.) transformed to satisfy homogeneity of variance assumptions necessary for parametric ANOVAs. Pearson correlations assessed (i) the relation between performance on the different speech and cognitive tasks and (ii) whether individuals' years of music training predicted their perceptual–cognitive skills. Multiple regressions were corrected using the false discovery rate (FDR) (Benjamini and Hochberg, 1995). Effect sizes are reported as $\eta_p^2$ for ANOVAs and Cohen's *d* for *t*-tests.

# RESULTS

Group speech streaming performance (i.e.,% accuracy, RTs, and localization error) is shown in **Figure 2**. Speech recognition expectedly declined from ceiling (M = 98%; NM = 99%) to

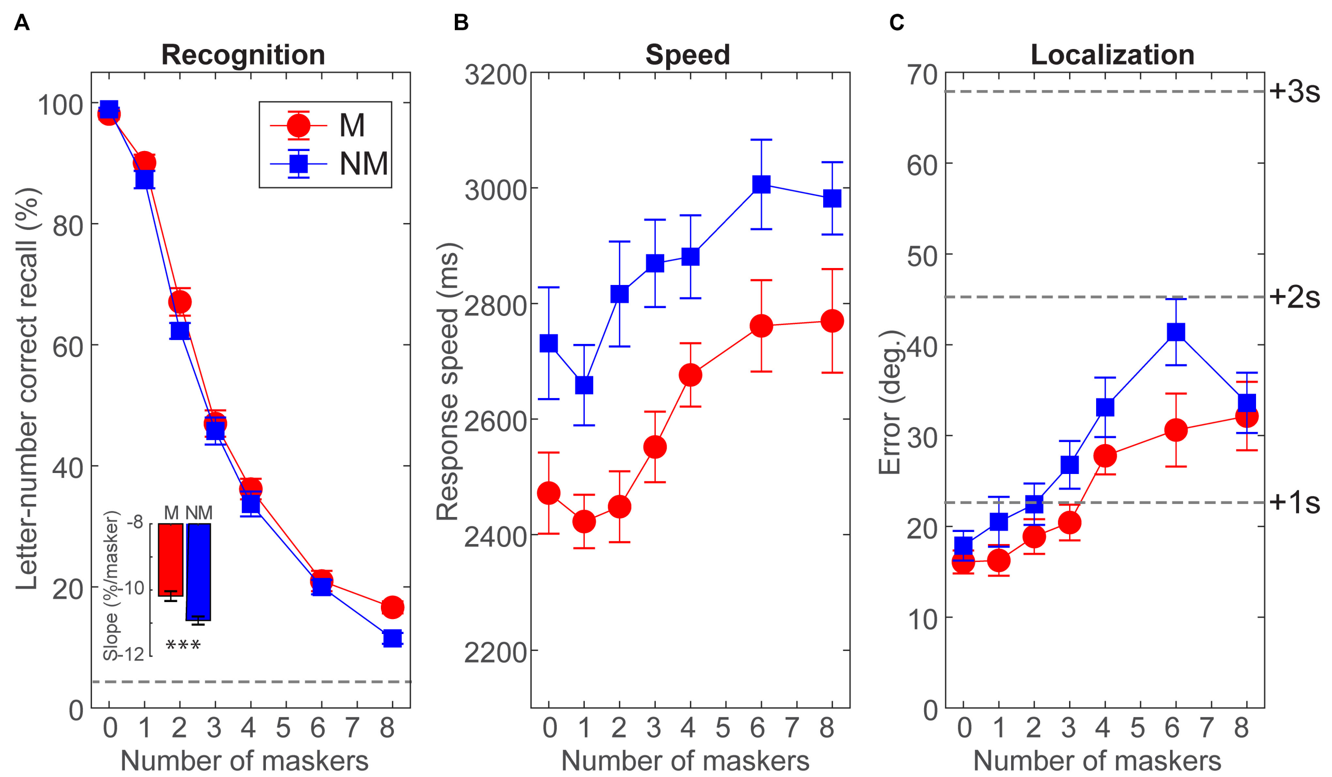



**FIGURE 2 |** Cocktail party listening is superior in musicians. **(A)** Speech recognition declines with increasing masker counts in both groups, but musicians show less performance decrement up to eight interfering talkers (*inset*). Dotted line = chance performance. **(B)** Musicians show faster (∼200–400 ms) speech recognition speeds than non-musicians. **(C)** Both groups localized correctly identified targets within two speakers (<40° error) with better localization in musicians. Error bars = ± 1 s.e.m.

near-floor (M = 17%; NM = 12%) performance with increasing masker counts from zero (unmasked) to eight multi-talkers. Both groups showed the single largest decrement with two talkers, consistent with prior auditory stream segregation studies (Rosen et al., 2013). Still, both groups showed above-chance recognition even amid eight maskers (all *ps* $< 0.0001$; *t*-test against 0). Notably, we found a group × masker interaction on target speech recognition accuracy [$F_{(5,\ 130)} = 4.48$, $p = 0.0008$, $\eta_p^2 = 0.15$; **Figure 2A**]. This interaction was attributable to the change in performance from zero to eight talkers being shallower in musicians compared to non-musicians (**Figure 2A**, *inset*; $t_{26} = 3.84$, $p = 0.0007$, $d = 1.45$). This suggests that musicians were less challenged by cocktail party speech recognition with an increasing number of interfering talkers.

For speed, we found main effects of group [$F_{(1,\ 26)} = 9.73$, $p = 0.0044$, $\eta_p^2 = 0.18$] and masker count [$F_{(5,\ 130)} = 28.20$, $p < 0.0001$, $\eta_p^2 = 0.52$] on speech recognition RTs (**Figure 2B**). These data reveal that while decision speeds were predictably slower in more challenging multi-talker scenarios, musicians were faster at streaming target speech across the board.

Localization errors are shown in **Figure 2C**. Both groups localized targets (correct trials) within about two speakers (<40° error). Localization varied with masker count [$F_{(5,\ 130)} = 21.61$, $p < 0.0001$, $\eta_p^2 = 0.45$], suggesting that target speech segregation worsened with additional talkers. However, musicians showed better localization than non-musicians overall [$F_{(1,\ 26)} = 4.32$, $p = 0.0478$, $\eta_p^2 = 0.14$].

Group differences in cognitive performance are shown in **Figure 3**. Replicating prior studies (e.g., Yoo and Bidelman, 2019), we found musician-related advantages in fluid IQ ($t_{26} = 1.72$, $p = 0.0491$, $d = 0.65$; **Figure 3A**) and backwards WM score ($t_{26} = 5.72$, $p < 0.0001$, $d = 2.16$; **Figure 3B**). Musicians also outperformed non-musicians by ∼1–2 dB on the QuickSIN test ($t_{26} = -1.71$, $p = 0.049$, $d = 0.65$; **Figure 3C**), consistent with their superior performance on the speech streaming task (present study) and prior work showing musician benefits in basic SIN perception (Zendel and Alain, 2012; Mankel and Bidelman, 2018; Yoo and Bidelman, 2019). Sustained attention, as measured via the SART, did not differ between groups for either commission (Ms: 35.5% vs. NMs: 29.8%; $p = 0.41$) or omission (Ms: 1.71% vs. NMs: 8.0%; $p = 0.38$) error rates (data not shown). Collectively, these results demonstrate that musicians have better performance than non-musicians in both SIN listening and some general cognitive abilities including IQ and WM.

We used pairwise correlations to evaluate relations between perceptual and cognitive measures as well as links between

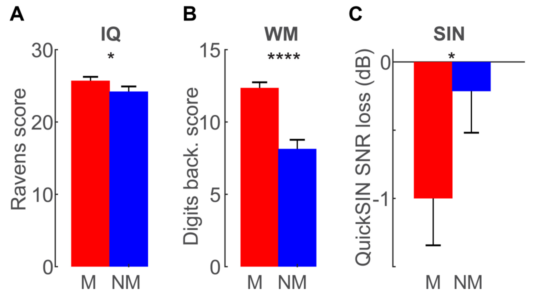



**FIGURE 3 |** Cognitive skills are superior in musicians. **(A)** Raven's fluid IQ and **(B)** auditory working memory are enhanced in musicians. **(C)** Musicians also obtain ~1 dB lower reception thresholds on the QuickSIN test, consistent with the notion of a musician advantage in speech-in-noise (SIN) perception. No group differences were observed in sustained attention (data not shown). Error bars = ± 1 s.e.m. $^*p < 0.05$, $^{****}p < 0.0001$.

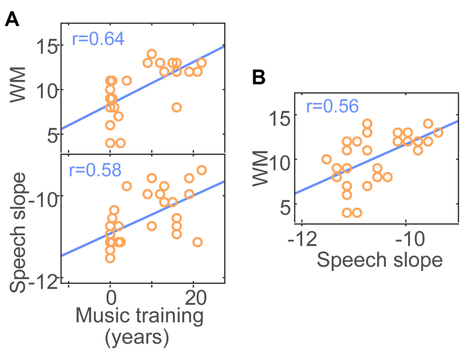



**FIGURE 4 |** Correlation results. **(A)** Formal music training predicts musicians' perceptual–cognitive advantages in working memory (WM) and speech streaming at the cocktail party. More extensive music training is associated with better auditory WM and shallower masker-related declines in speech streaming (see **Figure 2A**, inset). **(B)** Speech streaming is also related to WM; higher WM capacity predicts better cocktail party performance.

musical training and task performance (e.g., Yoo and Bidelman, 2019). Among the family of correlations we assessed (see **Supplementary Figure S1** for all 64 bivariate correlations; $p < 0.05$, uncorrected), three survived FDR correction for multiple comparisons: music training was associated with WM and speech streaming performance (**Figure 4A**). That is, listeners' years of formal music training predicted better auditory WM scores ($r = 0.64$, $p_{FDR} = 0.0069$) and shallower masker-related declines in speech streaming ($r = 0.58$, $p_{FDR} = 0.0189$). However, we also found that speech streaming correlated with WM such that higher WM capacity predicted better performance at the cocktail party ($r = 0.56$, $p_{FDR} = 0.0189$; **Figure 4B**). These data suggest that while musicianship is positively associated with improved speech streaming, successful cocktail party listening is at least partially related to cognitive abilities[2]. The association between musical training and SIN processing survived after controlling for WM for the QuickSIN ($r_{partial} = -0.38$, $p = 0.045$) but not speech streaming ($r_{partial} = -0.34$, $p = 0.08$) (cf. Yoo and Bidelman, 2019; Escobar et al., 2020).

[2]The QuickSIN uses four-talker babble and thus might be better related to streaming performance in the more comparable four-talker CRM condition. However, QuickSIN and $CRM_{4-talker}$ performances were not correlated ($r = -0.10$; $p = 0.60$).

## DISCUSSION

By measuring speech recognition in a multi-talker soundscape, we show that trained musicians are superior to their non-musician peers in deciphering speech within a naturalistic cocktail party environment. We found that musicians had faster and better target speech recognition amidst up to almost eight simultaneous talkers and enjoyed less noise-related decline in performance with increasing masker counts relative to musically naïve listeners. These SIN benefits were paralleled in normative measures of figure-ground perception (i.e., QuickSIN test). Our findings confirm and extend prior studies by demonstrating a relationship between musicianship and cocktail party listening skills (stream segregation) but also suggest that cognitive factors may at least partially account for music-related advantages in auditory scene analysis.

Regardless of music background, all listeners showed reduced ability to recognize target speech with increasing talker interferences. Poorer speech recognition with additional talkers is consistent with a reduction in spatial release from masking as more concurrent streams reduce the separability of the target in the soundfield (Pastore and Yost, 2017). More limited performance at higher masker counts is consistent with previous behavioral studies which show that spatial release from masking is effectively limited to fewer than six sound sources (Yost, 2017). Nevertheless, group differences revealed musicians showed smaller masker-related changes in recognition accuracy; trained listeners experienced a 10% decrease in accuracy for additional talkers vs. the 11–12% observed for the untrained group. This small but measurable boost in performance was paralleled in measures of conventional figure-ground SIN perception. We found that musicians had 1–2 dB better speech reception thresholds on the QuickSIN test. While modest, a 1–2 dB benefit in SNR can equate to improvements in speech recognition by as much as 10–15% (Middelweerd et al., 1990), which is comparable to the benefit we find in our cocktail party task. Our findings replicate and extend prior work on the so-called musician advantage for SIN perception (Parbery-Clark et al., 2009b; Zendel and Alain, 2012; Coffey et al., 2017; Mankel and Bidelman, 2018; Yoo and Bidelman, 2019) by demonstrating improved performance in challenging cocktail party speech streaming.

Our findings converge with prior behavioral studies using similar sample sizes ($N = 20$–$30$) that suggest a musician advantage in spatial release from masking as measured by the improvement in perception with spatially separated vs. co-located speech (Swaminathan et al., 2015; Clayton et al., 2016). However, using a similar paradigm as Swaminathan et al. (2015), but in an anechoic soundfield, Madsen et al. (2019) did not find a musician advantage in streaming performance in their sample of $N = 64$ listeners. However, we note that the evaluation of "cocktail party" listening in all three studies was limited to only a centrally located target (presented in front of the listener) concurrent with two flanking maskers (±15°). In contrast, our design used highly complex multi-talker mixtures (up to eight concurrent talkers) and roved the spatial relation(s) between target and masker(s) in the entire 360° soundfield. Furthermore, our listeners were able to stream using their individualized (natural) head-related transfer functions (HRTFs) rather than simulations as in headphone studies – which limits localization and externalization (cf. Swaminathan et al., 2015). Our data show that musicians outperform non-musicians in these highly ecological cocktail party scenarios at medium to large effect sizes. Collectively, we infer that musician benefits in cocktail party speech perception are not blanket effects. Rather, they seem to manifest only under the most challenging and ecological listening scenarios in tasks that tap linguistic and cognitive processing (e.g., Swaminathan et al., 2015; Yoo and Bidelman, 2019).

Group differences in localization were smaller. Both cohorts localized targets (correct trials) within one to two speakers (i.e., <20–40 degrees), with slightly better performance in musicians (**Figure 2C**). One explanation for this more muted effect is that the localization task was delayed compared to recognition. There is evidence listeners can localize sound sources even if they cannot identify them (Rakerd et al., 1999). Determining *where* a signal is emitted in the soundscape has clear biological advantage over identifying *what* it is. It is also conceivable that musicians who play in an orchestra might have higher-level localization performance than those who play in a smaller ensemble. We did collect information on the *size* of musicians' ensemble experience(s) to evaluate this possibility. However, supporting this notion, spatial tuning, and therefore localization abilities, does vary even among musicians depending on their relative position within an ensemble (e.g., conductor vs. player; Munte et al., 2001).

Musicians' SIN benefits could result from both auditory and cognitive enhancements. From an auditory standpoint, musicians are more sensitive to basic perceptual attributes of sound including pitch, spectrotemporal features, and temporal fine structure, all critically important for normal and degraded speech perception (Kishon-Rabin et al., 2001; Micheyl et al., 2006; Bidelman and Krishnan, 2010; Bidelman et al., 2014a; Mishra et al., 2015; Madsen et al., 2017, 2019; Tarnowska et al., 2019). Moreover, physiological studies indicate that musicianship may enhance cochlear gain control via the olivocochlear efferent system (Bidelman et al., 2017), a pathway thought to provide an "antimasking" function to the inner ear (Guinan, 2006) that enhances signal in noise detection (Micheyl and Collet, 1996; Bidelman and Bhagat, 2015). However, we also found evidence for enhanced cognitive faculties in musicians (i.e., IQ and WM). IQ, WM, and attention presumably play a large role in SIN processing. Indeed, we found that WM was associated with better speech streaming and reduced target localization error at the cocktail party. Thus, musicians' cocktail party benefits could reflect enhancements in domain-general cognitive abilities. Our findings parallel Schellenberg (2011) who found that musicianship was associated with IQ and Digit Span (WM and attention). They also converge with studies demonstrating relations between cognition (e.g., WM and auditory attention) and SIN performance in musical individuals (Strait and Kraus, 2011; Sares et al., 2018; Yoo and Bidelman, 2019; but see Escobar et al., 2020). Thus, musicians' cocktail party benefits observed here might result from a refinement in both auditory-perceptual and cognitive abilities, both of which could aid degraded speech-listening skills. They might also result from musicians' improved neural encoding of speech apparent at both brainstem and cortical levels (e.g., Parbery-Clark et al., 2009a; Bidelman et al., 2014b; Zendel et al., 2015; Mankel and Bidelman, 2018). Future electrophysiological studies are needed to evaluate the neural mechanisms underlying musicians' improved cocktail party listening observed here.

Alternatively, musicians could have lower levels of internal noise, which would tend to aid cocktail party listening (Lufti et al., 2017). Given that our listeners were young, normal-hearing individuals, the locus of this noise would probably stem from group differences in central factors (e.g., lesser lapses in attention and higher WM), which can be considered their own form of internal noise. This interpretation is at least qualitatively supported by the superior WM we find in the music group (present study; Bugos et al., 2007; Bidelman et al., 2013; Yoo and Bidelman, 2019).

Links between listeners' years of music training and (i) cocktail party recognition and (ii) cognitive measures (WM) suggest that musicians' SIN benefits scale with experience. Interestingly, we found that listeners' degree of music training predicted their QuickSIN performance even after controlling for WM. This suggests that musicianship might provide an additional boost to basic figure-ground speech perception beyond cognitive factors alone (e.g., Mankel and Bidelman, 2018; Yoo and Bidelman, 2019; but see Escobar et al., 2020). However, in contrast to the QuickSIN, the relation between musical training and *speech streaming* did not survive after controlling for WM. These results imply that while musicianship accounts for independent variance in simpler measures of SIN processing (i.e., QuickSIN), more complex SIN processing (i.e., cocktail party streaming) is driven more heavily by WM capacity. The degree to which listeners show successful speech/SIN processing likely represents a layering of inherent auditory listening skills (Mankel and Bidelman, 2018; Mankel et al., 2020), experience (Mankel and Bidelman, 2018), and cognitive factors including WM and attention (present study; Füllgrabe and Rosen, 2016; Oberfeld and Klöckner-Nowotny, 2016; Yoo and Bidelman, 2019). Our results are correlational in nature. Nevertheless, longitudinal (Torppa et al., 2018) and both quasi- and randomized-training studies in both younger and older adults (e.g., Kraus et al., 2014; Slater et al., 2015; Tierney et al., 2015; Zendel et al., 2019; Lo et al., 2020) provide converging evidence that musicianship causes gains in SIN processing in an experience-dependent manner.

Somewhat surprisingly, we did not find group differences in sustained attention, as measured via the SART, nor did attention correlate with cocktail party performance. These findings contrast with studies reporting attentional benefits in musicians (Strait et al., 2010; Thompson et al., 2017; Yoo and Bidelman, 2019) and work suggesting correlations between selective attention and individual differences in cocktail party listening (Oberfeld and Klöckner-Nowotny, 2016). Presumably, differences in results might be attributed to how attention is assessed. For example, selective attention, as measured via auditory backward masking (Strait et al., 2010; Yoo and Bidelman, 2019) and voice tracking (Madsen et al., 2019) paradigms, is superior in musicians. In contrast, we do not find group differences in *sustained* attention, as measured via the SART. Selective attention (Oberfeld and Klöckner-Nowotny, 2016), but not sustained attention (present study), correlates with cocktail party speech perception (but see Thompson et al., 2017). These studies suggest that the relation between attention and cocktail party listening varies with the specific (sub)construct of attention: selectively attending to a talker is arguably more relevant to parsing multi-talker mixtures than sustained, vigilance processes. Although not at ceiling performance, the relatively low error rates in the SART tasks (<30%) implies the lack of group effect might be due to the ease of the task. Moreover, the SART is a visual task. While there is some evidence that musicianship enhances visual processing (e.g., WM and multisensory binding) (George and Coch, 2011; Bidelman et al., 2013; Bidelman, 2016), visual attention may not differ between musicians and non-musicians (Strait et al., 2010). Nevertheless, in the cognitive domain of WM, we find a consistent musician boost in auditory mental capacity and strong links to SIN performance (e.g., Parbery-Clark et al., 2009b, 2011; Grassi et al., 2017; Yoo and Bidelman, 2019).

In conclusion, our findings confirm a relationship between musicianship and naturalistic cocktail party listening skills (stream segregation) but also suggest that cognitive factors may at least partially account for musicians' SIN advantage. Nevertheless, the degree to which music experience causally improves cocktail party speech processing (e.g., see Kraus et al., 2014; Slater et al., 2015; Tierney et al., 2015; Zendel et al., 2019) or is governed by preexisting factors unrelated to formal music training (e.g., inherent auditory aptitude; Mankel and Bidelman, 2018; Bidelman and Mankel, 2019; Mankel et al., 2020) awaits empirical confirmation with the present cross-sectional data.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher. Requests for data and materials should be directed to GB.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Memphis Institutional Review Board. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

GB and JY designed the study, analyzed the data, and wrote the manuscript. JY collected the data. Both authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under award number R01DC016267 (GB).

## ACKNOWLEDGMENTS

We thank Dr. Francisco Caballero for his assistance in piloting these experiments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01927/full#supplementary-material

## REFERENCES

Alain, C. (2007). Breaking the wave: effects of attention and learning on concurrent sound perception. *Hear. Res.* 229, 225–236. doi: 10.1016/j.heares.2007.01.011

Anaya, E. M., Pisoni, D. P., and Kronenberger, W. G. (2016). Long-term musical experience and auditory and visual perceptual abilities under adverse conditions. *J. Acoust. Soc. Am.* 140, 2074–2081. doi: 10.1121/1.4962628

Başkent, D., Fuller, C. D., Galvin, J. J., Schepel, L., Gaudrain, E., and Free, R. H. (2018). Musician effect on perception of spectro-temporally degraded speech, vocal emotion, and music in young adolescents. *J. Acoust. Soc. Am.* 143, EL311–EL316.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bialystok, E., and Depape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *J. Exper. Psychol. Hum. Percept. Perform.* 35, 565–574. doi: 10.1037/a0012735

Bidelman, G. M. (2016). Musicians have enhanced audiovisual multisensory binding: experience-dependent effects in the double-flash illusion. *Exper. Brain Res.* 234, 3037–3047. doi: 10.1007/s00221-016-4705-6

Bidelman, G. M., and Bhagat, S. P. (2015). Right ear advantage drives the link between olivocochlear efferent "antimasking" and speech-in-noise listening benefits. *Neuroreport* 26, 483–487. doi: 10.1097/wnr.0000000000000376

Bidelman, G. M., and Dexter, L. (2015). Bilinguals at the "cocktail party": dissociable neural activity in auditory-linguistic brain regions reveals neurobiological basis for nonnative listeners' speech-in-noise recognition deficits. *Brain Lang.* 143, 32–41. doi: 10.1016/j.bandl.2015.02.002

Bidelman, G. M., Gandour, J. T., and Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* 23, 425–434. doi: 10.1162/jocn.2009.21362

Bidelman, G. M., Hutka, S., and Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PLoS One* 8:e60676. doi: 10.1371/journal.pone.0060676

Bidelman, G. M., and Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res.* 1355, 112–125. doi: 10.1016/j.brainres.2010.07.100

Bidelman, G. M., and Mankel, K. (2019). Reply to Schellenberg: is there more to auditory plasticity than meets the ear? *Proc. Natl. Acad. Sci. U.S.A.* 116, 2785–2786. doi: 10.1073/pnas.1900068116

Bidelman, G. M., Schneider, A. D., Heitzmann, V. R., and Bhagat, S. P. (2017). Musicianship enhances ipsilateral and contralateral efferent gain control to the cochlea. *Hear. Res.* 344, 275–283. doi: 10.1016/j.heares.2016.12.001

Bidelman, G. M., Schug, J. M., Jennings, S. G., and Bhagat, S. P. (2014a). Psychophysical auditory filter estimates reveal sharper cochlear tuning in musicians. *J. Acoust. Soc. Am.* 136, EL33–EL39.

Bidelman, G. M., Weiss, M. W., Moreno, S., and Alain, C. (2014b). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* 40, 2662–2673. doi: 10.1111/ejn.12627

Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., and Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *J. Acoust. Soc. Am.* 137, 378–387. doi: 10.1121/1.4904537

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). A speech corpus for multitalker communications research. *J. Acoust. Soc. Am.* 107, 1065–1066. doi: 10.1121/1.428288

Bregman, A. S. (1978). Auditory streaming: competition among alternative organizations. *Percept. Psychophys.* 23, 391–398. doi: 10.3758/bf03204141

Broadbent, D. (1958). *Perception And Communication*. Elmsford, NY: Pergamon Press.

Brown, C. J., Jeon, E.-K., Driscoll, V., Mussoi, B., Deshpande, S. B., Gfeller, K., et al. (2017). Effects of long-term musical training on cortical auditory evoked potentials. *Ear. Hear.* 38, e74–e84. doi: 10.1097/aud.0000000000000375

Bugos, J. A., Perlstein, W. M., Mccrae, C. S., Brophy, T. S., and Bedenbaugh, P. H. (2007). Individualized piano instruction enhances executive functioning and working memory in older adults. *Aging Ment. Health* 11, 464–471. doi: 10.1080/13607860601086504

Carlyon, R. P. (2004). How the brain separates sounds. *Trends Cogn. Sci.* 8, 465–471. doi: 10.1016/j.tics.2004.08.008

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229

Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., and Patel, A. D. (2016). Executive function, visual attention and the cocktail party problem in musicians and non-musicians. *PLoS One* 11:e0157638. doi: 10.1371/journal.pone.00157638

Coffey, E. B. J., Mogilever, N. B., and Zatorre, R. J. (2017). Speech-in-noise perception in musicians: a review. *Hear. Res.* 352, 49–69. doi: 10.1016/j.heares.2017.02.006

Deroche, M. L., Limb, C. J., Chatterjee, M., and Gracco, V. L. (2017). Similar abilities of musicians and non-musicians to segregate voices by fundamental frequency. *J. Acoust. Soc. Am.* 142, 1739–1755. doi: 10.1121/1.5005496

Du, Y., and Zatorre, R. J. (2017). Musical training sharpens and bonds ears and tongue to hear speech better. *Proc. Natl. Acad. Sci. U.S.A.* 114, 13579–13584. doi: 10.1073/pnas.1712223114

Escobar, J., Mussoi, B. S., and Silberer, A. B. (2020). The effect of musical training and working memory in adverse listening situations. *Ear. Hear.* 41, 278–288. doi: 10.1097/aud.0000000000000754

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., and Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J. Cogn. Neurosci.* 17, 1578–1592. doi: 10.1162/089892905774597263

Fuller, C. D., Galvin, J. J. III, Maat, B., Free, R. H., and Baskent, D. (2014). The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? *Front. Neurosci.* 8:179. doi: 10.3389/fnins.2014.00179

Füllgrabe, C., and Rosen, S. (2016). On the (un)importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. *Front. Psychol.* 7:1268. doi: 10.3389/fpsyg.2016.01268

George, E. M., and Coch, D. (2011). Music training and working memory: an ERP study. *Neuropsychologia* 49, 1083–1094. doi: 10.1016/j.neuropsychologia.2011.02.001

Grassi, M., Meneghetti, C., Toffalini, E., and Borella, E. (2017). Auditory and cognitive performance in elderly musicians and nonmusicians. *PLoS One* 12:e0187881. doi: 10.1371/journal.pone.0187881

Guinan, J. J. Jr. (2006). Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. *Ear. Hear.* 27, 589–607. doi: 10.1097/01.aud.0000240507.83072.e7

Holmes, E., Domingo, Y., and Johnsrude, I. S. (2018). Familiar voices are more intelligible, even if they are not recognized as familiar. *Psychol. Sci.* 29, 1575–1583. doi: 10.1177/0956797618779083

Holmes, E., and Griffiths, T. D. (2019). 'Normal' hearing thresholds and fundamental auditory grouping processes predict difficulties with speech-in-noise perception. *Sci. Rep.* 9:16771.

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., and Carlyon, R. P. (2013). Swinging at a cocktail party: voice familiarity aids speech perception in the presence of a competing voice. *Psychol. Sci.* 24, 1995–2004. doi: 10.1177/0956797613482467

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 116, 2395–2405. doi: 10.1121/1.1784440

Kishon-Rabin, L., Amir, O., Vexler, Y., and Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *J. Basic Clin. Physiol. Pharmacol.* 12, 125–143.

Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., et al. (2014). Music enrichment programs improve the neural encoding of speech in at-risk children. *J. Neurosci.* 34, 11913–11918. doi: 10.1523/jneurosci.1881-14.2014

Lerousseau, J. P., Hidalgo, C., and Schön, D. (2020). Musical training for auditory rehabilitation in hearing loss. *J. Clin. Med.* 9, 1–15.

Lo, C. Y., Looi, V., Thompson, W. F., and Mcmahon, C. M. (2020). Music training for children with sensorineural hearing loss improves speech-in-noise perception. *J. Speech Lang. Hear. Res.* 63, 1990–2015. doi: 10.1044/2020_jslhr-19-00391

Lufti, R., Tan, A., and Lee, J. (2017). Individual differences in cocktail party listening: the relative role of decision weights and internal noise. *Proc. Meet. Acoust.* 30:050010.

Madsen, S. M. K., Marschall, M., Dau, T., and Oxenham, A. J. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Sci. Rep.* 9:10404.

Madsen, S. M. K., Whiteford, K. L., and Oxenham, A. J. (2017). Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. *Sci. Rep.* 7:12624.

Mankel, K., Barber, J., and Bidelman, G. M. (2020). Auditory categorical processing for speech is modulated by inherent musical listening skills. *Neuroreport* 31, 162–166. doi: 10.1097/wnr.0000000000001369

Mankel, K., and Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proc. Natl. Acad. Sci. U.S.A.* 115, 13129–13134. doi: 10.1073/pnas.1811793115

Medina, D., and Barraza, P. (2019). Efficiency of attentional networks in musicians and non-musicians. *Heliyon* 5:e01315. doi: 10.1016/j.heliyon.2019.e01315

Micheyl, C., and Collet, L. (1996). Involvement of the olivocochlear bundle in the detection of tones in noise. *J. Acoust. Soc. Am.* 99, 1604–1610. doi: 10.1121/1.414734

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hear. Res.* 219, 36–47. doi: 10.1016/j.heares.2006.05.004

Middelweerd, M. J., Festen, J. M., and Plomp, R. (1990). Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. *Audiology* 29, 1–7. doi: 10.3109/00206099009081640

Mishra, S. K., Panda, M. R., and Raj, S. (2015). Influence of musical training on sensitivity to temporal fine structure. *Int. J. Audiol.* 54, 220–226. doi: 10.3109/14992027.2014.969411

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., and Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychol. Sci.* 22, 1425–1433. doi: 10.1177/0956797611416999

Moreno, S., and Bidelman, G. M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hear. Res.* 308, 84–97. doi: 10.1016/j.heares.2013.09.012

Munte, T. F., Kohlmetz, C., Nager, W., and Altenmuller, E. (2001). Superior auditory spatial tuning in conductors. *Nature* 409:580. doi: 10.1038/35054668

Nelson, W. T., Bolia, R. S., Ericson, M. A., and Mckinley, R. L. (1998). *Monitoring the Simultaneous Presentation Of Spatialized Speech Signals In A Virtual Acoustic Environment. Air Force Research Lab Wright-Patterson AFB Oh Human Effectiveness Directorate*. Available online at: https://apps.dtic.mil/sti/citations/ADA430284 (accessed March 10, 2020).

Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.* 95, 1085–1099. doi: 10.1121/1.408469

Norton, A., Winner, E., Cronin, K., Overy, K., Lee, D. J., and Schlaug, G. (2005). Are there pre-existing neural, cognitive, or motoric markers for musical ability? *Brain and Cogn*. 59, 124–134. doi: 10.1016/j.bandc.2005.05.009

Oberfeld, D., and Klöckner-Nowotny, F. (2016). Individual differences in selective attention predict speech identification at a cocktail party. *eLife* 5:e16747.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Oxenham, A. J., and Shera, C. A. (2003). Estimates of human cochlear tuning at low levels using forward and simultaneous masking. *J. Assoc. Res. Otolaryngol.* 4, 541–554. doi: 10.1007/s10162-002-3058-y

Parbery-Clark, A., Skoe, E., and Kraus, N. (2009a). Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* 29, 14100–14107. doi: 10.1523/jneurosci.3256-09.2009

Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009b). Musician enhancement for speech-in-noise. *Ear. Hear.* 30, 653–661. doi: 10.1097/aud.0b013e3181b412e9

Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., and Kraus, N. (2011). Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PLoS One* 6:e18082. doi: 10.1371/journal.pone.018082

Pastore, M. T., and Yost, W. A. (2017). Spatial release from masking with a moving target. *Front. Psychol.* 8:2238. doi: 10.3389/fnins.2014.002238

Peirce, J., Gray, J. R., Simpson, S., Macaskill, M., Höchenberger, R., Sogo, H., et al. (2019). PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* 51, 195–203. doi: 10.3758/s13428-018-01193-y

Popham, S., Boebinger, D., Ellis, D. P. W., Kawahara, H., and Mcdermott, J. H. (2018). Inharmonic speech reveals the role of harmonicity in the cocktail party problem. *Nat. Commun.* 9:2122.

Rakerd, B., Hartmann, W. M., and Mccaskey, T. L. (1999). Identification and localization of sound sources in the median sagittal plane. *J. Acoust. Soc. Am.* 106, 2812–2820. doi: 10.1121/1.428129

Raven, J. (1998). *Raven Manual Section 4: Advanced Progressive Matrices*. Oxford: Oxford Psychologists Press.

Raven, J., Raven, J. C., and Court, J. H. (1998). *Advanced Progressive Matrices*. San Antonio, TX: Harcourt Assessment.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., and Yiend, J. (1997). 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia* 35, 747–758. doi: 10.1016/s0028-3932(97)00015-8

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Appl. Psycholinguist.* 27, 465–485. doi: 10.1017/s014271640606036x

Rosen, S., Souza, P., Ekelund, C., and Majeed, A. A. (2013). Listening to speech in a background of other talkers: effects of talker number and noise vocoding. *J. Acoust. Soc. Am.* 133, 2431–2443. doi: 10.1121/1.4794379

Ruggles, D., and Shinn-Cunningham, B. (2011). Spatial selective auditory attention in the presence of reverberant energy: individual differences in normal-hearing listeners. *J. Assoc. Res. Otolaryngol.* 12, 395–405. doi: 10.1007/s10162-010-0254-z

Ruggles, D. R., Freyman, R. L., and Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PLoS One* 9:e86980. doi: 10.1371/journal.pone.086980

Sares, A. G., Foster, N. E. V., Allen, K., and Hyde, K. L. (2018). Pitch and time processing in speech and tones: the effects of musical training and attention. *J. Speech Lang. Hear. Res.* 61, 496–509. doi: 10.1044/2017_jslhr-s-17-0207

Schellenberg, E. G. (2005). Music and cognitive abilties. *Curr. Direct. Psychol. Sci.* 14, 317–320.

Schellenberg, E. G. (2011). Music lessons, emotional intelligence, and IQ. *Music Percept. Interdiscipl. J.* 29, 185–194. doi: 10.1525/mp.2011.29.2.185

Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* 34, 114–123. doi: 10.1016/j.tins.2010.11.002

Singh, G., Pichora-Fuller, M. K., and Schneider, B. A. (2008). The effect of age on auditory spatial attention in conditions of real and simulated spatial separation. *J. Acoust. Soc. Am.* 124, 1294–1305. doi: 10.1121/1.2949399

Slater, J., Skoe, E., Strait, D., O'connell, S., Thompson, E., and Kraus, N. (2015). Music training improves speech-in-noise perception: longitudinal evidence from a community-based music program. *Behav. Brain Res.* 291, 244–252. doi: 10.1016/j.bbr.2015.05.026

Strait, D. L., and Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Front. Psychol.* 2:113. doi: 10.3389/fnins.2014.00113

Strait, D. L., Kraus, N., Parbery-Clark, A., and Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear. Res.* 261, 22–29. doi: 10.1016/j.heares.2009.12.021

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., and Kidd, G. Jr. (2015). Musical training, individual differences and the cocktail party problem. *Sci. Rep.* 5:11628.

Tarnowska, E., Wicher, A., and Moore, B. C. J. (2019). The effect of musicianship, contralateral noise, and ear of presentation on the detection of changes in temporal fine structure. *J. Acoust. Soc. Am.* 146, 1–10. doi: 10.1121/1.5114820

Thompson, E. C., Woodruff Carr, K., White-Schwoch, T., Otto-Meyer, S., and Kraus, N. (2017). Individual differences in speech-in-noise perception parallel neural speech processing and attention in preschoolers. *Hear. Res.* 344, 148–157. doi: 10.1016/j.heares.2016.11.007

Tierney, A. T., Krizman, J., and Kraus, N. (2015). Music training alters the course of adolescent auditory development. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10062–10067. doi: 10.1073/pnas.1505114112

Torppa, R., Faulkner, A., Kujala, T., Huotilainen, M., and Lipsanen, J. (2018). Developmental links between speech perception in noise, singing, and cortical processing of music in children with cochlear implants. *Music Percept.* 36, 156–174. doi: 10.1525/mp.2018.36.2.156

Van Schie, M. K., Thijs, R. D., Fronczek, R., Middelkoop, H. A., Lammers, G. J., and Van Dijk, J. G. (2012). Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness. *J. Sleep Res.* 21, 390–395. doi: 10.1111/j.1365-2869.2011.00979.x

Wechsler, D., Coalson, D. L., and Raiford, S. E. (2008). *WAIS-IV: Wechsler Adult Intelligence Scale*. San Antonio, TX: Pearson.

Yeend, I., Beach, E. F., Sharma, M., and Dillon, H. (2017). The effects of noise exposure and musical training on suprathreshold auditory processing and speech perception in noise. *Hear. Res.* 353, 224–236. doi: 10.1016/j.heares.2017.07.006

Yoo, J., and Bidelman, G. M. (2019). Linguistic, perceptual, and cognitive factors underlying musicians' benefits in noise-degraded speech perception. *Hear. Res.* 377, 189–195. doi: 10.1016/j.heares.2019.03.021

Yost, W. A. (1997). "The cocktail party problem: forty years later," in *Binaural And Spatial Hearing In Real And Virtual Environments*, eds R. H. Gilkey and T. R. Anderson (New Jersey: Lawrence Erlbaum Associates Inc), 329–347.

Yost, W. A. (2017). Spatial release from masking based on binaural processing for up to six maskers. *J. Acoust. Soc. Am.* 141, 2093–2093.

Zendel, B. R., and Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *J. Cogn. Neurosci.* 21, 1488–1498. doi: 10.1162/jocn.2009.21140

Zendel, B. R., and Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychol. Aging* 27, 410–417. doi: 10.1037/a0024816

Zendel, B. R., Tremblay, C.-D., Belleville, S., and Peretz, I. (2015). The impact of musicianship on the cortical mechanisms related to separating speech from background noise. *J. Cogn. Neurosci.* 27, 1044–1059. doi: 10.1162/jocn_a_00758

Zendel, B. R., West, G. L., Belleville, S., and Peretz, I. (2019). Musical training improves the ability to understand speech-in-noise in older adults. *Neurobiol. Aging* 81, 102–115. doi: 10.1016/j.neurobiolaging.2019.05.015

Zuk, J., Benjamin, C., Kenyon, A., and Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLoS One* 9:e99868. doi: 10.1371/journal.pone.099868

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 21 August 2020
doi: 10.3389/fnhum.2020.00329

# Long-Term Musical Training Alters Auditory Cortical Activity to the Frequency Change

*Jihyun Lee[1], Ji-Hye Han[1] and Hyo-Jeong Lee[1,2]**

[1]Laboratory of Brain & Cognitive Sciences for Convergence Medicine, Hallym University College of Medicine, Anyang, South Korea, [2]Department of Otorhinolaryngology, College of Medicine, Hallym University, Anyang, South Korea

**Objective**: The ability to detect frequency variation is a fundamental skill necessary for speech perception. It is known that musical expertise is associated with a range of auditory perceptual skills, including discriminating frequency change, which suggests the neural encoding of spectral features can be enhanced by musical training. In this study, we measured auditory cortical responses to frequency change in musicians to examine the relationships between N1/P2 responses and behavioral performance/musical training.

**Methods**: Behavioral and electrophysiological data were obtained from professional musicians and age-matched non-musician participants. Behavioral data included frequency discrimination detection thresholds for no threshold-equalizing noise (TEN), +5, 0, and −5 signal-to-noise ratio settings. Auditory-evoked responses were measured using a 64-channel electroencephalogram (EEG) system in response to frequency changes in ongoing pure tones consisting of 250 and 4,000 Hz, and the magnitudes of frequency change were 10%, 25% or 50% from the base frequencies. N1 and P2 amplitudes and latencies as well as dipole source activation in the left and right hemispheres were measured for each condition.

**Results**: Compared to the non-musician group, behavioral thresholds in the musician group were lower for frequency discrimination in quiet conditions only. The scalp-recorded N1 amplitudes were modulated as a function of frequency change. P2 amplitudes in the musician group were larger than in the non-musician group. Dipole source analysis showed that P2 dipole activity to frequency changes was lateralized to the right hemisphere, with greater activity in the musician group regardless of the hemisphere side. Additionally, N1 amplitudes to frequency changes were positively related to behavioral thresholds for frequency discrimination while enhanced P2 amplitudes were associated with a longer duration of musical training.

**Conclusions**: Our results demonstrate that auditory cortical potentials evoked by frequency change are related to behavioral thresholds for frequency discrimination in musicians. Larger P2 amplitudes in musicians compared to non-musicians reflects musical training-induced neural plasticity.

**Keywords: frequency change, spectral processing, musical training, N1/P2 auditory evoked potential, hemispheric asymmetry**

OPEN ACCESS

***Edited by:***
*Cunmei Jiang,
Shanghai Normal University, China*

***Reviewed by:***
*Gavin M. Bidelman,
University of Memphis, United States
Fawen Zhang,
University of Cincinnati, United States*

***Correspondence:**
*Hyo-Jeong Lee
hyojlee@hallym.ac.kr*

***Specialty section:***
*This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience*

***Received:** 29 April 2020*
***Accepted:** 24 July 2020*
***Published:** 21 August 2020*

***Citation:***
*Lee J, Han J-H and Lee H-J (2020) Long-Term Musical Training Alters Auditory Cortical Activity to the Frequency Change. Front. Hum. Neurosci. 14:329. doi: 10.3389/fnhum.2020.00329*

## INTRODUCTION

Understanding speech and other everyday sounds require the processing of the temporal and spectral information in sounds. Psychoacoustically, pitch perception is the ability to extract the frequency information of a complex stimulus. It relies on spectral cues because it requires the mapping of frequencies onto meaningful speech or music (Stangor and Walinga, 2014). In both speech and music, pitch provides spectral information to facilitate the perception of musical structure and the acquisition of speech understanding inferred from the pitch contour and prosody information (Moore, 2008; Oxenham, 2012). Pitch processing is even more crucial for understanding sounds under adverse listening conditions such as background noise (Fu et al., 1998; Won et al., 2011). Difficulties with listening in noise have been attributed to the reduced ability to segregate the spectral cues and noise (Gaudrain et al., 2007).

Attempts to demonstrate a relationship between frequency coding and music perception have been made to investigate the neural activities underlying the auditory function of people who have undergone musical training. There is a large body of literature on assessing whether long-term musical training affects perceptual changes in frequency coding (Shahin et al., 2003; Micheyl et al., 2006; Deguchi et al., 2012; Liang et al., 2016) since examining the neural processing of sounds in musicians can provide a conceptual model of auditory training. Several studies have reported a strong relationship between musical training and speech perception. For example, the acquisition of a foreign language can be facilitated by musical training due to the enhanced neural encoding for speech relevant cues such as formant frequencies of speech (Intartaglia et al., 2017). Musical training in early life produces an even greater influence on both neural and behavioral speech processing than in adult life. It has been revealed that young children engaged in piano training have enhanced cortical responses to pitch changes, and the neural changes are associated with their behavioral performances in word discrimination (Nan et al., 2018). These improvements suggest a link between musical training and functional and structural changes in the human brain. Neuroimaging studies have provided converging evidence that the volume of the brain regions related to speech processing is larger in musicians compared to non-musicians, which indicates neurophysiological changes occurred by training-induced brain plasticity (Schneider et al., 2002; Gaser and Schlaug, 2003; Bermudez et al., 2009; Hyde et al., 2009). These findings indicate that music and speech processing rely on partially overlapping neural and cognitive resources.

Perceiving music requires listeners to integrate various sources of sound information, including pitch, timbre, and rhythm, and these musical features are linked to cognitive/perceptual processing at the cortical level. Previous studies have found that musical training can change various auditory functions, including sound discrimination (Zuk et al., 2013), listening to a foreign language (Marques et al., 2007), as well as auditory attention (Seppänen et al., 2012). These studies have also demonstrated that the neural changes underlying the perceptual and cognitive changes as a function of musical training can be reflected in cortical responses to complex stimuli (Pantev et al., 1998; Shahin et al., 2003, 2007). In general, musicians show greater N1/P2 and late positive responses to musical and tone stimuli compared to non-musicians. The improved cortical activities in musicians exhibit experience-driven neural changes (Shahin et al., 2003; Marques et al., 2007; Seppänen et al., 2012). However, some studies have suggested that the neuroplasticity evidence in musicians may be an innate property in people whose auditory function is superior to others rather than music experience-driven factors (Schellenberg, 2015, 2019; Mankel and Bidelman, 2018).

Cortical N1/P2 responses can be elicited by changes in various types of sound: speech (Han et al., 2016), tonal (Martin and Boothroyd, 2000), and noise (Bidelman et al., 2018). Using frequency-modulated tonal stimuli, it has been found that the N1/P2 responses vary depending on the rate and the magnitude of frequency (Dimitrijevic et al., 2008; Pratt et al., 2009; Vonck et al., 2019). Furthermore, N1/P2 responses to frequency change are enhanced by long-term auditory training. For instance, a recent study by Liang et al. (2016) found that N1/P2 amplitudes are enhanced with an increase in the magnitude of frequency changes, and this was more evident in musicians compared to non-musicians. However, in their study, no relationship was found between behavioral performance for frequency change detection and the N1/P2 response measures.

Although there has been extensive research on the subject (Koelsch et al., 1999; Schneider et al., 2002, 2005; Bidelman et al., 2014; Hutka et al., 2015) which indicates that musical training can change neural processing to spectral change, the idea that altered neural responses are induced by long-term musical training or by other environmental/congenital properties is still controversial. Given that tracking the frequency pattern of tone and extracting the pitch of musical sounds both rely on spectral processing in the auditory cortex, an attempt to assess the neural sensitivity of musicians to subtle frequency changes can provide knowledge about the underlying auditory processing for music and speech. Therefore, we examined how the central auditory system of musicians encodes frequency information differently and related this to their behavioral perception ability. It is important to determine whether relationships exist between behavioral performance and objective cortical activities since it would indicate that cortical responses evoked by frequency change can be used as a marker for behavioral frequency discrimination. To answer the research question, we applied base frequencies of 250 and 4,000 Hz, because a previous study reported that cortical responses elicited by stimuli with the frequencies had a strong relationship with psychoacoustical thresholds of frequency discrimination (Dimitrijevic et al., 2008). We hypothesized that the cortical activity to frequency change is more enhanced in musicians compared to non-musicians. We further predicted that behavioral thresholds and the duration of musical training relate to the measures of cortical activity. Also, we measured the behavioral frequency discrimination both in quiet and noise conditions to associate with cortical responses, because musical training has improved sound in noise perception ability

(Parbery-Clark et al., 2009b; Yoo and Bidelman, 2019). We assumed that musicians reveal better noise perception than non-musicians, indicating the musician's advantage on the sound in noise perception.

# MATERIALS AND METHODS

## The Participants

A total of 13 (six male) musicians [mean age ± standard deviation (SD) = 27.1 ± 5.0 years, all right-handed] and 11 (six male) age-matched non-musicians (mean age ± SD = 26.8 ± 5.31 years, all right-handed) participated in this study. We ran a two-sample *t*-test to examine whether our findings were driven by age. Results showed no significant difference in the age of musician and non-musician groups ($p = 0.904$), suggesting that age is not a contributing factor. All musicians reported that they had been receiving professional musical training for over 10 years regularly and received musical training at least three times a week during the training period. The types of musical training were vocal, piano, drum, haegeum, guitar, and violin. Details on musical training of the musicians are provided in **Table 1**. All participants were recruited through online advertising and were compensated for their participation. Both groups had normal pure-tone thresholds below 20 dB hearing loss (HL) at octave test frequencies from 250 to 8,000 Hz, and they had no history of neurological or hearing disorders. The study protocol was approved by the Institutional Review Board of Hallym University Sacred Hospital, Gangwon-Do, South Korea (File No. 2018-02-019-001), and written informed consent were obtained from each participant.

## Psychoacoustics: Frequency Discrimination Test

Frequency discrimination was applied as a standard adaptive, three-interval, three-alternative, forced-choice, two-down, one-up procedure to detect the threshold for each subject. The base frequencies used for the frequency discrimination test were 250 and 4,000 Hz. During each trial, two of three intervals contained base frequency pure tones, while the remaining one had a pure tone with a higher frequency change. Individual tones were 300 ms in duration and separated by an inter-stimulus interval of 500 ms. The three intervals were presented randomly. The initial differences in frequency between the base and the change were 50 and 100 Hz for 250 and 4,000 Hz tones, respectively. The step size from the second trial was 12 Hz for 250 Hz and 50 Hz for 4,000 Hz. The difference was decreased or increased for the subsequent trial depending on whether there were two consecutive correct responses or a single incorrect response, respectively. When the current step size was larger than the difference, it was varied to half the difference. Each condition was ended after a maximum of 60 trials or 12 reversals. The averaged thresholds were measured with the last eight reversals to compute each subject's difference limen (DL). Subjects were instructed to find the frequency change among the three interval choices by clicking a mouse on a computer screen. For a noise condition, four types of background threshold-equalizing noise (TEN) were used; no TEN and +5, 0, and −5 dB signal-to-noise ratio (SNR) TEN. These noise conditions for each base tone were randomly presented to avoid the carryover effect (Logue et al., 2009; Dochtermann, 2010; Bell, 2013). During the testing, the subjects were seated in a sound-attenuated booth, and sound stimuli were presented through 2 channel speakers at a level of 70 dB HL.

Outliers were determined based on the interquartile range (IQR) method (Kokoska and Zwillinger, 2000). The outliers were identified by defining limits on the sample values that are a factor $k$ of the IQR below the 25th percentile or above the 75th percentile. We used 3 for $k$ to identify values that are extreme outliers. None of the musicians were defined as an outlier while two non-musicians were rejected.

## Electroencephalogram (EEG) Acquisition and Analysis

### Stimuli and Experimental Procedure

Stimuli for the frequency change and experimental procedures were based on a previous frequency change experiment (Dimitrijevic et al., 2008). Auditory stimuli were generated in MATLAB (MathWorks, Inc., Natick, MA, USA), and they were sampled at a rate of 48,828 Hz. Frequency change stimuli were constructed using two continuous base tones, 250 and 4,000 Hz, each with upward frequency changes of 10%, 25%, or 50% for 400 ms. The order of the frequency changes was

**TABLE 1** | Characteristics of musicians.

| Subjects | First instrument | Age began musical training | Secondary instrument | Age began musical training | Years of musical training |
|---|---|---|---|---|---|
| **Mus 1** | Vocal (Korean tradition) | 14 | | | 20 |
| **Mus 2** | Piano | 9 | | | 12 |
| **Mus 3** | Piano | 9 | | | 11 |
| **Mus 4** | Piano | 6 | Vocal | 17 | 20 |
| **Mus 5** | Bass guitar | 20 | | | 16 |
| **Mus 6** | Piano | 6 | Cello | 25 | 19 |
| **Mus 7** | Piano | 5 | Vocal | 23 | 25 |
| **Mus 8** | Haegeum (Korean tradition) | 16 | Vocal | 24 | 10 |
| **Mus 9** | Piano | 8 | Clarinet | 16 | 20 |
| **Mus 10** | Drum | 17 | | 17 | 12 |
| **Mus 11** | Guitar | 10 | | | 10 |
| **Mus 12** | Violin | 10 | Piano | 10 | 13 |
| **Mus 13** | Guitar | 13 | | | 13 |

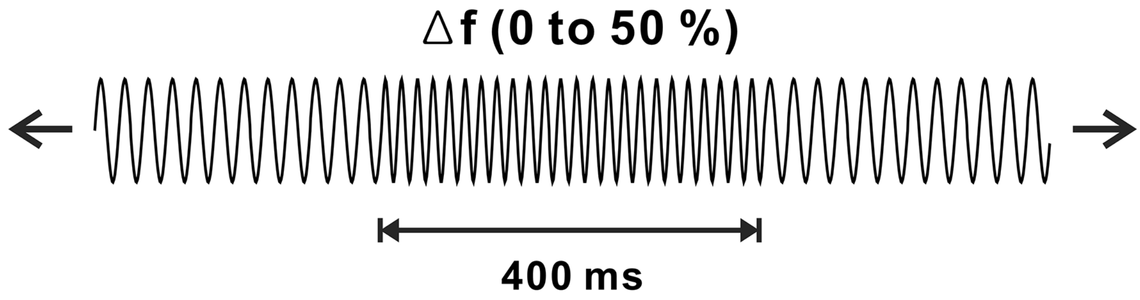



**FIGURE 1** | Schematic representation of frequency change stimulus. Continuous tones with base frequencies of 250 or 4,000 Hz are presented with occasional changes in frequency change of 10%, 25%, or 50% lasting 400 ms.

randomly determined. The intensity of the frequency change was equated to equal loudness concerning the base frequency. The ongoing stimuli consisted of frequency change stimuli followed by base frequency tones varied from 1.6 to 2.2 s to prevent anticipating the point where the frequency change occurred. To avoid a transient click, which was produced when changing the stimuli, we manipulated the stimuli to occur at the zero phase. **Figure 1** shows a schematic of the frequency changes of the stimulus. A minimum of 100 trials for each frequency change was presented in two blocks. The total electroencephalogram (EEG) recording time for each subject was approximately 30 min, during which the subjects were seated in a comfortable reclining chair and watched a close-captioned movie of their choice while the frequency change stimuli were presented through 2 channel speakers located 1.0 m away from the subject.

### EEG Acquisition and Data Processing

Multi-channel EEG data were acquired using the actiCHamp Brain Products recording system (Brain Products GmbH, Germany). Scalp potentials were recorded at 64 equidistant electrode sites, all electrodes were referenced to the reference electrode, electrical impedances were reduced below 10 kΩ, and EEG signals were amplified and digitized at 1,000 Hz. During the EEG recording, continuous data were band-pass-filtered from 0.1 to 120 Hz and a notch filter for 60 Hz noise was applied.

### EEG Data Analysis

All EEG data were preprocessed offline using Brain Vision Analyzer 2.2 (Brain Products GmbH, Germany). Continuous eye blink and horizontal movement artifacts were rejected using the independent component analysis (ICA) algorithm. After the ICA correction, the data were further analyzed in MATLAB. Continuous EEG data were down-sampled to 250 Hz and band-pass-filtered from 0.1 to 40 Hz. The data were segmented from −100 to 400 ms with 0 ms at the onset of frequency change. Segmented data were baseline-corrected from −100 to 0 ms and re-referenced to an average reference. Separate averages for individual frequency changes were also performed. Peak detection was performed for N1/P2 on the frontal central electrodes located at the near vertex. N1 peaks were determined as the first negative potentials between 70 and 150 ms after stimulus onset, while the most positive potentials between 120 and 230 ms were defined as P2 peaks.

### Dipole Source Analysis

This was performed using BESA Research 7.0 (Brain Electrical Source Analysis, GmbH, Germany), as described previously (Han et al., 2016). The source analysis was performed on individual averaged waveforms with band-pass filtering (0.5–40 Hz, 12 dB/octave, zero-phase). In the first step, two symmetric regional dipole sources were inserted near the auditory cortical regions. For N1 and P2 dipole fitting, the mean area over a 20 ms window around the N1 and P2 peaks on the global field power was used for further analysis. The dipole source activities were allowed to vary in location, orientation, and strength, and the maximum tangential sources were fitted on the N1 and P2 peaks. The residual variance was examined for each 20 ms window, for which all subjects obtained 5% or less variance. Statistical differences in the grand mean source waveforms were assessed across the different conditions and subject groups.

## Statistical Analysis

For the behavioral thresholds, the main effect of the subject groups (musician vs. non-musician), the noise (+5 SNR, 0 SNR, and −5 SNR; for noise condition only) and base frequency (250 and 4,000 Hz) settings were examined using repeated-measures analysis of variance (rmANOVA) for quiet and noise condition, separately. rmANOVA was used to assess the main effects of frequency change (10%, 25%, and 50%), the base frequency for within-subject comparison on the cortical measures (the frequency change and the base frequency were set as continuous variables). For between-subject factors, musician and non-musician groups were included. We performed this analysis using the fitrm and ranova functions in MATLAB. *Post hoc* testing was applied using Tukey's honestly significant difference tests, and paired *t*-tests were conducted for group comparisons. Pearson's product-moment correlation coefficient was applied to assess relationships among the behavioral measures and demographic factors with the electrophysiological measures. Multiple pairwise comparisons were adjusted with the false discovery rate (FDR). All data

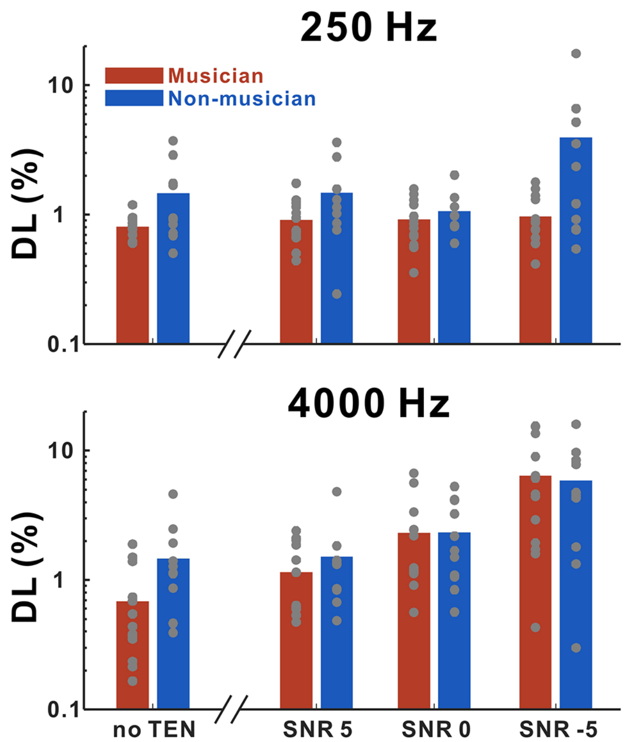



**FIGURE 2** | Mean frequency discrimination thresholds for 250 and 4,000 Hz in musicians and non-musicians as a function of listening conditions including no threshold-equalizing noise (TEN), SNR +5 dB, SNR 0 dB, and SNR −5 dB. Note that gray dots indicate each subject. Musicians show decreased thresholds compared to non-musicians for both 250 and 4,000 Hz in no TEN condition.

are expressed as the mean ± standard error (SE) unless otherwise stated.

## RESULTS

### Behavioral Frequency Discrimination

**Figure 2** shows frequency discrimination thresholds for 250 and 4,000 Hz as a function of listening conditions. We performed rmANOVA to examine the main effects of group, base frequency, and noise level (for the noise condition only) for quiet and noise conditions, separately. In the quiet condition, the results revealed significant main effects of the groups ($F_{(1,20)} = 5.18$; $p = 0.034$) in that the thresholds in the musicians were lower than those in the non-musicians. In the noise condition, a significant interaction between noise level and base frequency ($F_{(2,34)} = 64.75$; $p < 0.0001$) was found. Tukey's HSD (honestly significant difference) test showed that the thresholds at 250 Hz base frequency were significantly lower than those at the 4,000 Hz base frequency for −5 SNR ($p < 0.0001$) and 0 SNR ($p = 0.0086$) conditions. In addition, significant differences between +5 SNR and 0 SNR ($p = 0.0168$), +5 SNR and −5 SNR ($p < 0.0001$), and 0 SNR and −5 SNR ($p < 0.0001$) were found for the 4,000 Hz base frequency.

## Electrophysiology

### N1/P2 Cortical Responses

Grand mean waveforms as a function of frequency change for non-musicians and musicians are given in **Figure 3**. In general, the N1/P2 cortical responses were modulated by frequency changes, and the modulations were more evident at 4,000 Hz and the musician group, compared with 250 Hz and the non-musician group.

**Figure 4** shows N1 and P2 amplitudes as a function of frequency change starting at 250 and 4,000 Hz for the musician and non-musician groups. rmANOVA to examine the effect of frequency change on N1 response revealed a significant frequency change × base frequency interaction for N1 amplitude ($F_{(1,22)} = 24.32$; $p < 0.0001$) and latency ($F_{(1,22)} = 54.43$; $p < 0.0001$). The *post hoc* analysis confirmed that the N1 amplitude to the 50% change was larger than those to the 10% ($p = 0.005$), and 25% ($p = 0.024$) changes at the 250 Hz base frequency. The *post hoc* analysis also showed that the N1 amplitude to the 25% change was larger than those to the 10% ($p = 0.038$) and 50% ($p < 0.0001$) changes at the 4,000 Hz base frequency. In addition, the N1 amplitude for 4,000 Hz was significantly larger compared to 250 Hz with 10% ($p < 0.0001$) and 25% ($p < 0.0001$) frequency changes. For the N1 latency, the response at 4,000 Hz was significantly shorter than that at 250 Hz for all frequency changes ($p < 0.0001$). No significant group differences were found for N1.

Significant base frequency × frequency change interactions were found for P2 amplitude ($F_{(1,22)} = 10.97$; $p = 0.003$) and latency ($F_{(1,22)} = 14.64$; $p < 0.0001$). The *post hoc* results show that P2 amplitude to 25% change was greater than that to the 10% change for 250 Hz only ($p = 0.004$). For the P2 latency, the 25% frequency change elicited significantly shorter responses compared to the 10% change for 4,000 Hz ($p = 0.019$). Compared to 4,000 Hz, the P2 responses for 250 Hz significantly decreased in amplitude for 10% ($p = 0.012$), 25% ($p = 0.022$), and 50% ($p = 0.015$) frequency changes, while the latency increased for 10% ($p < 0.0001$), 25% ($p < 0.0001$), and 50% ($p = 0.015$). Significant differences between the musician and non-musician groups were found in the P2 amplitudes: that of the P2 in musician group was significantly larger than that of the non-musician ($F_{(1,22)} = 6.58$; $p = 0.018$). In addition, an interaction between the groups and frequency change was revealed ($F_{(1,22)} = 4.67$; $p = 0.042$). The *post hoc* results show that the P2 amplitudes of the musicians were greater than those of the non-musicians for 10% ($p = 0.005$) and 25% ($p = 0.022$) frequency changes. In addition, the P2 amplitudes for the 25% frequency change were greater than those for the 10% frequency change in both the musician group ($p = 0.009$) and the non-musician group ($p = 0.029$). No group differences were found for P2 latency.

### Dipole Source Activity

The grand average N1 dipole source waveforms as a function of frequency changes for the 250 and 4,000 Hz base frequencies are shown in **Figure 5**. Using two symmetric single equivalent dipoles, the N1/P2 dipoles were fitted, and amplitudes and latencies of N1/P2 sources waveforms

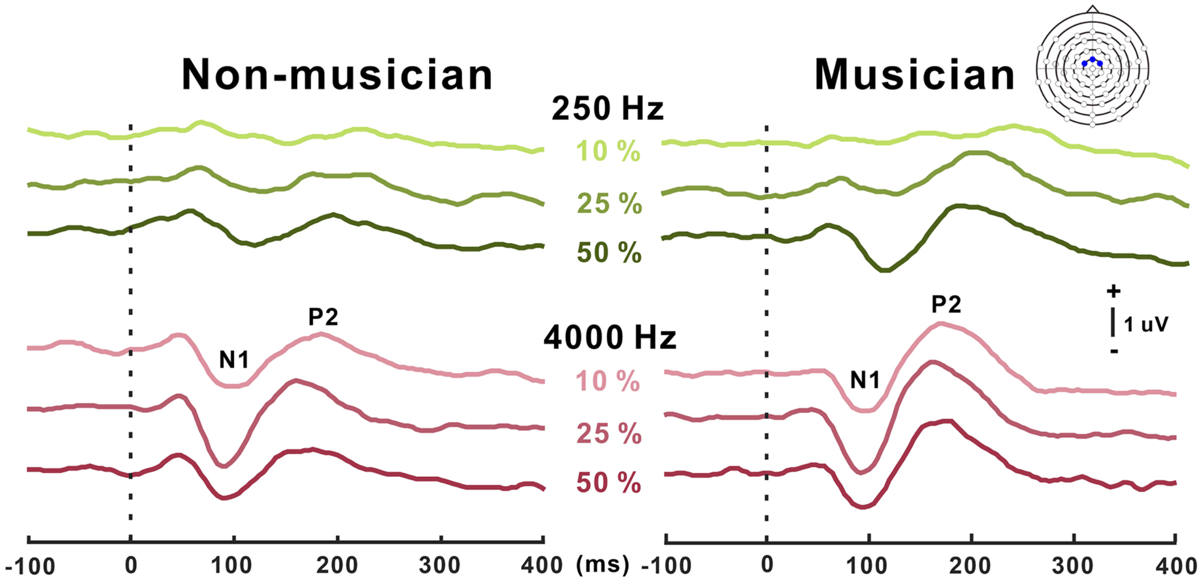



**FIGURE 3 |** Grand average N1/P2 responses to frequency changes. N1/P2 cortical potentials to frequency change stimuli in non-musicians (left) and musicians (right). Green and red color waveforms represent cortical responses for 250 and 4,000 Hz base frequencies, respectively. The amount of frequency change is indicated as a percentage. In the top right, a figure shows an equidistant cap layout indicating the frontal central electrodes (blue dots).

were averaged for each hemisphere. The overall morphology of the N1 dipole waveforms was similar to the N1 scalp-recorded waveforms in that N1 activity increased as the frequency change became greater, which was more apparent at 4,000 Hz than 250 Hz. P2 dipole source analysis showed that the musician group had greater P2 dipole activity than the non-musician group.

A significant frequency change × base frequency × hemisphere interaction was found for N1 dipole amplitude ($F_{(1,22)} = 6.45$; $p = 0.019$). The *post hoc* results show that the N1 dipole amplitudes in the right hemisphere were greater than those in the left hemisphere for 25% ($p = 0.019$) and 50% ($p = 0.031$) changes at 4,000 Hz. Similar to the dipole amplitude, a significant frequency change × base frequency × hemisphere interaction for N1 dipole latency was revealed ($F_{(1,22)} = 6.63$; $p = 0.017$). The results show that the dipole latencies in the right hemisphere were shorter than those in the left hemisphere for 4,000 Hz with the 50% frequency change ($p = 0.03$).

For P2 dipole amplitude, two interactions including frequency change × hemisphere ($F_{(1,22)} = 9.43$; $p = 0.006$) and frequency change × base frequency ($F_{(1,22)} = 114.04$; $p < 0.0001$) were found. The P2 dipole amplitudes were greater in the right hemisphere than in the left hemisphere for 25% ($p = 0.053$) and 50% change ($p = 0.019$). Also, the P2 dipole amplitudes to 4,000 Hz were greater compared to 250 Hz for 10% ($p = 0.005$), 25% ($p = 0.007$), and 50% frequency changes ($p = 0.029$). For P2 dipole latency, a significant base frequency × frequency change interaction was found ($F_{(1,22)} = 3{,}159$; $p < 0.0001$) such that the P2 latencies for a 25% frequency change were significantly shorter than those for 10% at 4,000 Hz ($p = 0.006$). The P2 dipole latencies for 250 Hz were prolonged compared to 4,000 Hz for 25% ($p = 0.002$) and 50% ($p = 0.004$) frequency changes.

The effect of musical training on hemispheric asymmetry for spectral processing was examined by comparing left- and right-hemispheric activation separately between the musician and non-musician groups. For the group comparison, we conducted a two-sample *t*-test and found significant group differences for both left ($p = 0.001$) and right hemispheres ($p = 0.013$). The results indicate that P2 dipole source activities in both hemispheres of the musicians were larger than in non-musicians.

### Relationship Between N1/P2 Cortical Response and Behavioral Performance/Duration of Musical Training

Pearson's correlation results showed that the N1 amplitudes to 4,000 Hz were positively correlated with behavioral thresholds for 4,000 Hz with +5 SNR TEN ($p = 0.0036$, corrected for multiple comparisons; **Figure 6A**). Moreover, the P2 amplitudes to 4,000 Hz base tone were associated with the duration of musical training ($p = 0.0028$, corrected for multiple comparisons; **Figure 6B**). However, none of the latency measures were correlated with behavioral performance and musical training.

## DISCUSSION

Our aim in this study was to examine the effect of musical training on behavioral frequency discrimination as well as N1/P2 cortical responses. These are elicited by tones with frequency change, and their relationships with the

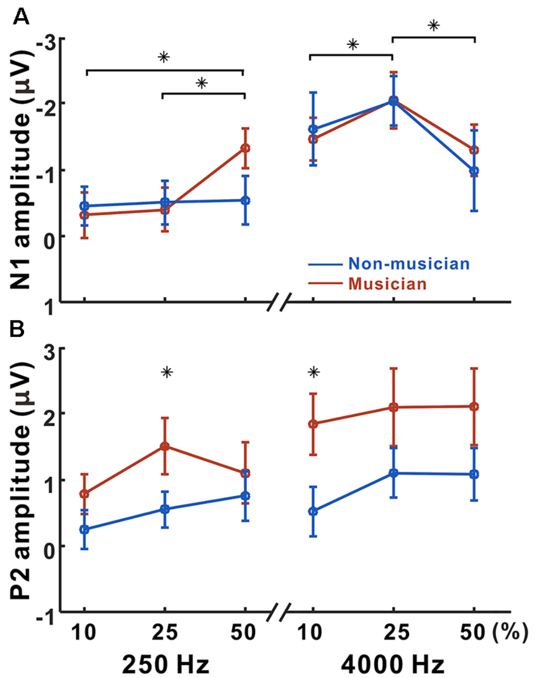



**FIGURE 4** | Mean N1 and P2 amplitudes as a function of frequency change. Mean of N1 **(A)** and P2 **(B)** amplitudes in musicians and non-musicians are shown. Note that significant differences among listening conditions are revealed for the N1 (250 Hz: 10 vs. 50, 25 vs. 50%, 4,000 Hz: 10 vs. 25, 25 vs. 50%), whereas the differences between musicians and non-musicians are found for the P2 amplitude (250 Hz: 25%, 4,000 Hz: 10%). Asterisks (*) indicate significant differences ($p < 0.05$).

threshold for frequency discrimination and duration of musical training were assessed. Our results demonstrate that P2 was increased in musicians compared to non-musicians whereas N1 revealed more stimulus-dependent characteristics in that it was modulated by frequency change. The results of the dipole source analysis show that the N1/P2 dipole activity in response to the frequency change stimuli was greater in the right hemisphere, and the P2 dipole source activities in musicians were larger than those in the non-musicians for both hemispheres. Finally, the N1 and P2 amplitudes were related to behavioral performances and the duration of musical training, respectively.

## Effect of Musical Training on Behavioral Frequency Discrimination

In the behavioral frequency discrimination test, the thresholds of the musicians were lower than those of the non-musicians in the no TEN condition. These results indicate that the musicians were able to discriminate smaller spectral differences that the non-musicians could not, especially under quiet listening conditions. Previous studies assessing pitch discrimination in quiet conditions have reported relatively consistent results that musicians outperformed non-musicians in discriminating spectral features of stimuli, thereby confirming the better pitch perception of the former (Tervaniemi et al., 2005; Micheyl et al., 2006; Liang et al., 2016). Indeed, musical training leads to an enhancement in the ability to track frequency change and detect spectral cues in sounds. On the other hand, in the noise condition, the threshold for frequency discrimination in the musicians was not different from that in the non-musicians, which is similar to recent studies reporting that any advantage incurred by musical training on sound perception is questionable in the presence of noise-masking (Ruggles et al., 2014; Boebinger et al., 2015; Madsen et al., 2019). Meanwhile, it is still controversial whether musician advantage for auditory perception in noise exists or not. Studies in which behavioral tests were conducted on musicians have shown that musical training can improve speech-in-noise perception (Parbery-Clark et al., 2009a, 2012; Yoo and Bidelman, 2019). Furthermore, those studies have provided neurological evidence of better speech-in-noise perception by musicians (Musacchia et al., 2007; Parbery-Clark et al., 2011, 2012; Zendel et al., 2015; and reviewed in Coffey et al., 2017). However, in the current study, musical expertise for noise perception was not evident. One possible reason for this is related to the test paradigm and stimulus type used to evoke a response. In a study using speech with multiple maskers varied in content and similarity to speech, improved performances by musicians in a frequency discrimination task have been revealed, although this does not carry over to speech-in-noise perception (Boebinger et al., 2015). Similarly, Micheyl et al. (2006) and Ruggles et al. (2014) reported that musicians have an advantage in pitch discrimination that is not present for perceiving masked sounds. Another explanation for no effect of music training on sound in noise processing is that the musician benefits on the noise perception can be restricted to the specific sounds which are more linguistically and cognitively demanding. Several studies have suggested that the musician's advantage in noise perception is dependent on the complexity of target sounds or tasks (Krizman et al., 2017; Yoo and Bidelman, 2019). For example, Yoo and Bidelman reported that musicians revealed improved sentences in noise perception, but the musician advantage was not applied for words in noise processing. In summary, the results of these studies suggest that the possible advantage of sound perception incurred by musical training is questionable in the presence of noise-masking, and it would be dependent on the complexity of the task (Ruggles et al., 2014; Boebinger et al., 2015; Madsen et al., 2019).

Investigating the neural overlap between pitch perception and perceiving sound in noise could uncover a mechanism to explain the perceptual advantages observed in musicians. Musical practice is a complex form of training consisting of dozens of perceptual and cognitive skills drawing on hearing, selective attention, and auditory memory. However, previous works examining the relationship between musical experience and cognitive/perceptual skills have shown that musical training is only related to specific musical features such as pitch, melody, and rhythm perception (Ruggles et al., 2014). Thus, selective listening related to the perception of masked sounds may not be a crucial aspect linked to musical training. Moreover, it has been suggested that the outcomes of musical training may not always be generalizable beyond the tasks that are closely related to musical perception

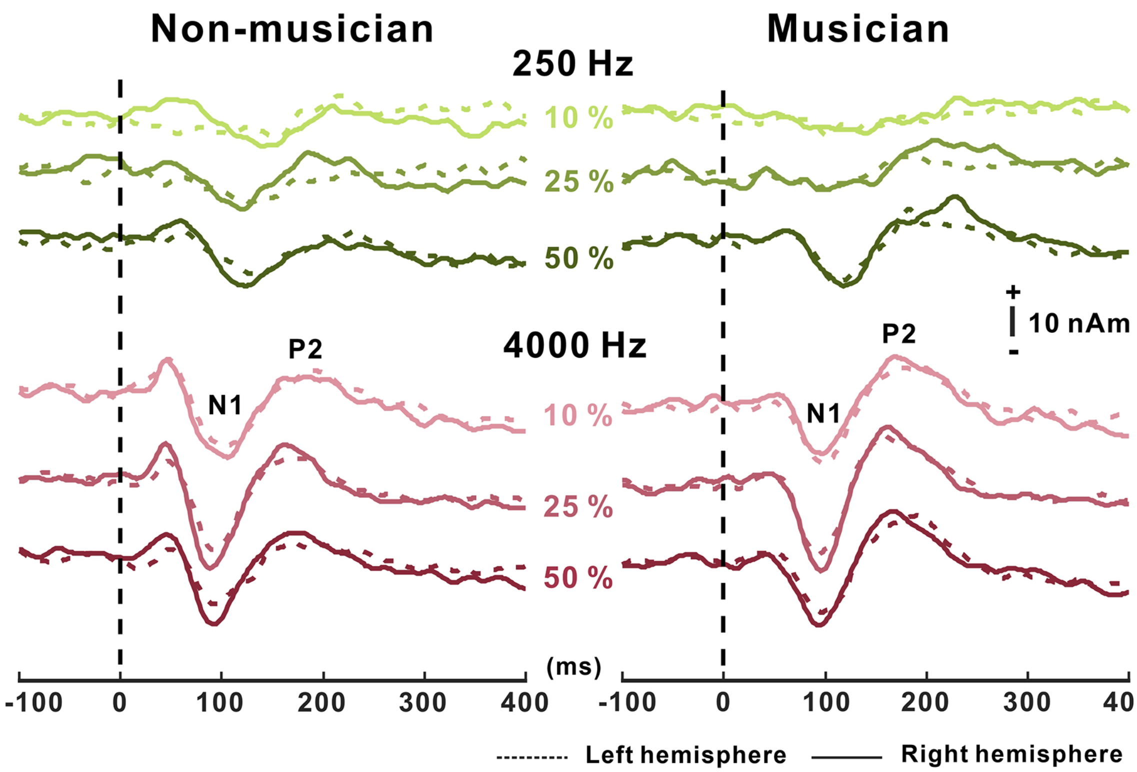



**FIGURE 5 |** Dipole source waveforms to frequency changes in non-musician and musician groups. N1 dipole source waveforms to frequency change stimuli in non-musicians (left) and musicians (right). Green and red color waveforms represent dipole activity for 250 and 4,000 Hz base frequencies, respectively. Dashed lines of waveform represent dipole activity in the left hemisphere and solid lines indicate activity in the right hemisphere.

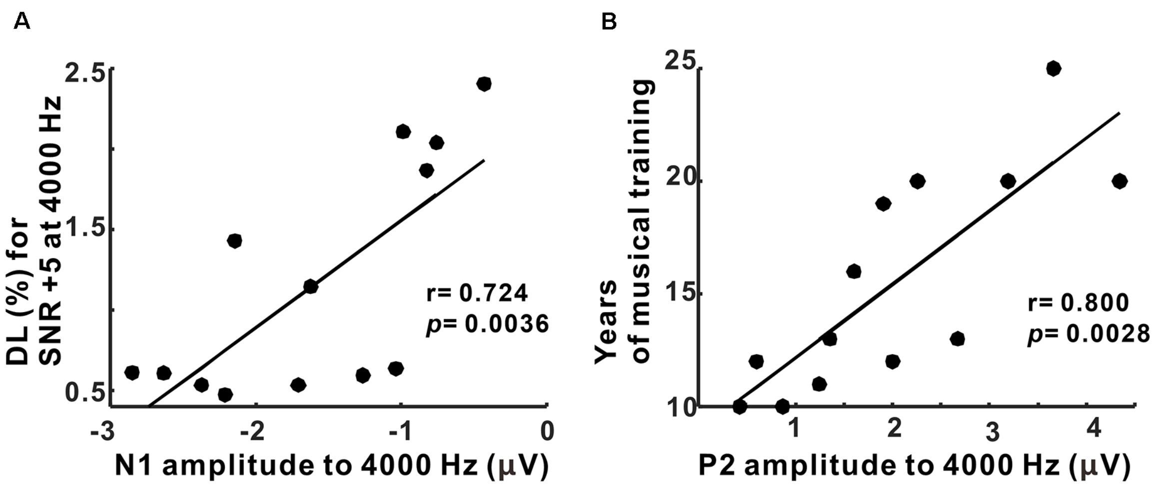



**FIGURE 6 |** Correlations between behavioral performance/duration of musical training and N1/P2 amplitudes in musicians. **(A)** N1 amplitudes to 4,000 Hz condition are significantly related to frequency discrimination thresholds for SNR +5 at 4,000 Hz. **(B)** A significant relationship between P2 amplitudes to 4,000 Hz and the duration of musical training is revealed.

(Geiser et al., 2009; Okada and Slevc, 2018). Rather, high sensitivity to sound in noise seems to be associated with daily music-related behavior, such as listening to music in everyday life (Kliuchko et al., 2015). Alternatively, the neurological basis for music perception is another possibility for the result as the cortical areas governing music and speech may not completely share their neural origins (Albouy et al., 2020; reviewed in Peretz et al., 2015).

## Effects of Musical Training on N1/P2 Cortical Potentials

### N1 Modulation as a Function of Frequency Change

We found that N1 is modulated as a function of frequency change to a greater degree than P2, whereas P2 reflects the musical training-induced enhancements in the musicians. Previously, it has been suggested that the N1/P2 responses are evoked by acoustic changes in a sound: either amplitude (Han and Dimitrijevic, 2015), intensity (Dimitrijevic et al., 2009) or frequency (Shahin et al., 2003; Dimitrijevic et al., 2008; Pratt et al., 2009). The N1 was evoked by stimuli with changes in frequency which is close to the level of behavioral thresholds in frequency discrimination. In turn, the N1 may not be elicited by the sound with frequency change which is not detected by listeners perceptually (Martin and Boothroyd, 2000; Jones and Perez, 2001). In general, the N1 amplitude is modulated by an increase in frequency that is more apparent for stimuli with frequencies higher than 1,000 Hz (Picton, 2011). Enhanced amplitude with frequency increase has been attributed to the level of neuronal activation relating to the range in basilar membrane deflection (Rose et al., 1967; Picton, 2011).

In our study, N1/P2 amplitudes were larger for the higher frequency relative to the lower one. Although these were larger for the lower frequency compared to the higher one, some studies using the mismatch negativity (MMN) paradigm have reported similar results to our findings. For example, using a frequency change as a deviant stimulus, Novitski et al. (2004) found larger MMN responses to a higher frequency than a lower one. A possible reason for the increased cortical responses to the higher frequency could be related to the frequency change experiment including change stimuli embedded in the ongoing tones. This is similar to the MMN paradigm (Lavikainen et al., 1995) in which infrequent deviant stimuli presented with repetitive standard sounds. From this point of view, we speculate that the listening condition of the frequency change activates the neuronal populations in a similar way to the MMN paradigm (see the review in Alho, 1995).

### P2 Response Reflecting Musical-Training Induced Plasticity

In contrast to N1, our results show that P2 responses to frequency changes in musicians are more robust than in non-musicians. Human electrophysiological studies have also shown enhanced P2 cortical activity in individuals with short-term auditory training (reviewed in Tremblay, 2007; Tremblay et al., 2014), language experience (Wagner et al., 2013), and short/long-term musical training (Atienza et al., 2002; Shahin et al., 2003; Tremblay, 2007; Tong et al., 2009). More recently, results showing increased P2 to trained pitch sounds during passive listening infer that training-induced cortical plasticity is related to permanent perception changes rather than the effect of selective attention (Wisniewski et al., 2020). There is growing neuroimaging evidence to support the notion that the neural representations of complex sounds such as music at the peripheral and central levels are influenced extensively by experience or training (Parbery-Clark et al., 2012; Sankaran et al., 2020). Perceiving music requires listeners to integrate sources of information, including amplitude, timbre, and pitch, each of which can provide cues for perceiving music. Neuroanatomically, the understanding of music requires systematic processing as a set of hierarchical neural representations in different areas of the brain. Previous studies have proposed that music perception requires the activation of multiple areas of the brain involved with not only sound discrimination but also cognitive/perceptual skills (Okada and Slevc, 2018). Among the long-latency responses, P2 is related to neural processes mediating cognitive/perceptual aspects of sound processing (Näätänen et al., 1993; Alain et al., 2007). Therefore, we assume that P2 represents training-induced cortical plasticity due to the characteristic of being sensitive to acoustic features importantly contributing to music perception.

In Neuromusicology, there has been a long-standing debate about whether the central auditory processing to musical features is altered by musical training (nurture) or preexisting factors (nature). Previous findings have supported the idea of experience-driven plasticity of musicians by showing that training changes the neural representation of acoustic features in individuals with extensive musical experience (Pantev et al., 2001; Shahin et al., 2008). It has been suggested that skilled musicians exhibit enhanced cortical representations of musical timbres associated with the instrument they have trained with (Pantev et al., 2001; Pantev and Herholz, 2011). Such timbre specificity constrained to the principal instrument supports the theory that changes in neural activity in musicians are mainly driven by experience (reviewed in Pantev and Herholz, 2011). Furthermore, in studies investigating the effect of short-term musical training on cortical plasticity, Tremblay et al. (2001) and Atienza et al. (2002) found that P2 responses were enhanced by short-term intensive training in non-musicians. These results indicate that auditory cortical responses can be altered regardless of the musical training duration of the non-musicians. Given that professional musicians receive much longer training than non-musicians, it can be inferred that the cortical plasticity induced by the training should be greater in musicians. Numerous studies have reported the musical training effects both the behavioral level (Shahin et al., 2003, 2008; Tervaniemi et al., 2005; Liang et al., 2016; Intartaglia et al., 2017) and the perceptual levels (Gaser and Schlaug, 2003; Bermudez et al., 2009; Hyde et al., 2009) of the auditory processing. Meanwhile, an attempt to explain the musician's advantage of innate properties has been made. These studies have supported the view that genetic factors could be involved in the etiology of musical properties including absolute pitch (Gregersen et al., 1999, 2001; Theusch et al., 2009; Theusch and Gitschier, 2011), congenital amusia (Peretz et al., 2007), and music perception (Drayna et al., 2001; Pulli et al., 2008; Ukkola et al., 2009; Ukkola-Vuoti et al., 2013; Oikkonen et al., 2015). In a study assessing non-musicians, individuals with superior musical ability showed enhanced neural encoding of speech. Moreover, they were less susceptible to noise in a similar way to what appeared in professional musicians (Mankel and Bidelman, 2018). Swaminathan and Schellenberg (2018) examined relationships among musical training and non-musical factors and musical ability to find a marker for musical competence. In this study, non-musical factors such as socioeconomic status, short-term memory, general cognitive ability, and personality were indirectly associated with the musical ability along with the musical training, suggesting that the musical competence would be established by complex interactions between nature and nurture traits. It seems difficult to make a conclusion of nature vs. nurture debate at this point. To clarify the issue, further studies are necessary to compare multiple factors relating to the musical ability in a large group of musicians.

### N1/P2 Correlation With Behavioral Performance and Musical Training Experience

In our study, N1 is correlated with the perceptual change to frequency information in musician whereas the duration of musical training is related to P2. The lack of a consistent N1 relationship with musical training may be accounted for by the notion that N1 is related to neural processing for frequency information in sound rather than a musical experience. In particular, the relationship between N1 and behavioral performance was found between frequency discrimination thresholds in noise and N1 amplitudes to frequency change (see **Figure 6**). This finding is related to the previous finding that spectral processing is associated to sound perception in noise (Fu et al., 1998; Won et al., 2007), and difficulty with sound in noise has been attributed to a reduction in the ability to distinguish acoustic signals from noise (Gaudrain et al., 2007). For P2, we found that the amplitude increased with longer duration of musical training but not with age at the onset of the training (data not shown). Relationships between musical training and P2 evoked by auditory stimuli have been reported in previous studies on adults (Atienza et al., 2002; Shahin et al., 2003; Choi et al., 2014) as well as children (Shahin et al., 2003, 2004). Moreover, the P2 amplitude elicited by musical tones is correlated with musical training (Choi et al., 2014). These results suggest that continuous musical training may help to maintain cortical synaptic plasticity regardless of when musical training started. Meanwhile, a previous study comparing behavioral thresholds for speech discrimination and objective/cognitive properties has shown that the P2 threshold is associated with cognitive factors such as non-verbal IQ but not with musical experience (Boebinger et al., 2015); the authors suggested that the musician's advantage may be accounted for by co-variation in higher-order cognitive factors with musicianship. To better understand the complex relationships among musical training and cognitive and perceptual processing, more studies are necessary to compare perceptual measures of sound processing, cortical activity, and cognitive factors interconnected through both bottom-up and top-down auditory pathways.

## Asymmetrical Hemispheric Activation to Frequency Change

We investigated whether hemispheric asymmetry in the processing of frequency change exists at the cortical level. The findings from the N1/P2 dipole source analysis showed that source activation in response to frequency changes was greater in the right hemisphere than in the left hemisphere. This result is consistent with previous reports (Shahin et al., 2003, 2007; Dimitrijevic et al., 2008; Pratt et al., 2009; Okamoto and Kakigi, 2015) showing that the processing of frequency information is lateralized to the right hemisphere. The right hemisphere dominance for the processing of frequency change seems to be based on fundamental brain mechanisms that are closely related to the functional specialization of the right hemisphere for pitch perception (Zatorre and Belin, 2001). Research on a large sample of musicians has reported that the musicians were sensitive to pitch change and their behavioral sensitivity was associated with the right-ward asymmetry for pitch processing (Schneider et al., 2005). Moreover, a lesion-related study reported abnormal pitch discrimination in patients who had undergone the removal of the right Heschl's gyrus (Johnsrude et al., 2000). Zatorre and Belin (2001) also confirmed that spectral processing recruits anterior superior temporal regions bilaterally, with greater activation in the right hemisphere (Zatorre and Belin, 2001).

By comparing the left- and right-hemispheric activities separately in musicians and non-musicians, we found that the dipole source activity in musicians evoked by frequency changes was larger than that in non-musicians in both hemispheres. Increased bilateral engagement of the hemispheres in the musician was mainly attributed to the group difference, and the effects of musical training on hemispheric reorganization were only observed for the P2 dipole. Indeed, increased bilateral hemispheric activation following long-term musical experience has previously been reported. Using near-infrared spectroscopy, Gibson et al. (2009) found greater bilateral frontal activity in musicians compared to non-musicians during a cognitively demanding task; they suggested that extensive musical experience yields the symmetrical activities in the musicians. Also, Tremblay et al. (2009) reported that short-term auditory training evoked a different pattern of hemispheric asymmetry such that the P2 dipole sources to training-specific stimuli increased in the left hemisphere. This is consistent with our results showing that musical training enhances cortical activity in the left hemisphere. Furthermore, all of the musicians except for the vocalists in our study require both hands to play their instruments. Musicians can incorporate auditory feedback to play instruments and appropriately alter their motor response in both hands in a very short period. Given that this auditory-motor interaction interplays between the left and right hemispheres, this process may strengthen the direct connections between the hemispheres (reviewed inZatorre et al., 2007).

## CONCLUSIONS

In the present study, we showed that the effect of frequency change was more apparent for N1, while P2 responses are closely related to musical training. An enhanced N1 response to frequency changes is associated with better frequency discrimination whereas P2 responses are positively related to the duration of musician training, indicating training-induced cortical plasticity. Also, musicians had more robust P2 source activation in both hemispheres, which indicates musical experience may alter the hemispheric lateralization for processing of frequency change more symmetrically. Given that enhanced P2 activity with frequency change reflects changes in the summation of postsynaptic field potentials in the auditory cortex, our findings infer that neural plasticity evoked by long-term musical training can alter the cortical representation of a change in frequency even when passively listening to sounds. In future studies, we will examine the cortical activity to frequency change with noise-masking to compare with quiet listening to define a neural overlap between pitch perception and sound in noise perception. Also, the effect of attention on spectral processing is worth investigating in that the selective attention in musicians increases the neural encoding of sound and suppresses background noise to enhance their speech-in-noise perception ability (Strait and Kraus, 2011).

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The study involving human participants were reviewed and approved by the ethics committee of the Hallym University Sacred Hospital. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

JL and J-HH contributed to the conception, design of the study, and wrote the manuscript. JL performed the experiment and statistical analysis. JL, J-HH, and H-JL contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

This project was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1I1A1A0107091411 and 2019R1A2B5B01070129), the Center for Women in Science, Engineering and Technology (WISET) Grant funded by the Ministry of Science ICT and Future Planning of Korea (MSIP) under the Program for Returners into R&D (WISET-2020-203), and by the Hallym University Research Fund (HURF-2017-79).

## ACKNOWLEDGMENTS

We would like to thank Dr. Andrew Dimitrijevic for providing MATLAB codes to generate stimuli and perform behavioral frequency discrimination test.

## REFERENCES


Alain, C., Snyder, J. S., He, Y., and Reinke, K. S. (2007). Changes in auditory cortex parallel rapid perceptual learning. *Cereb. Cortex* 17, 1074–1084. doi: 10.1093/cercor/bhl018

Albouy, P., Benjamin, L., Morillon, B., and Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science* 367, 1043–1047. doi: 10.1126/science.aaz3468

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNM) elicited by sound changes. *Ear Hear.* 16, 38–51. doi: 10.1097/00003446-199502000-00004

Atienza, M., Cantero, J. L., and Dominguez-marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learn. Mem.* 9, 138–150. doi: 10.1101/lm.46502

Bell, A. (2013). Randomized or fixed order for studies of behavioral syndromes? *Behav. Ecol.* 24, 16–20. doi: 10.1093/beheco/ars148

Bermudez, P., Lerch, J. P., Evans, A. C., and Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb. Cortex* 19, 1583–1596. doi: 10.1093/cercor/bhn196

Bidelman, G. M., Davis, M. K., and Pridgen, M. H. (2018). Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hear. Res.* 367, 149–160. doi: 10.1016/j.heares.2018.05.018

Bidelman, G. M., Weiss, M. W., Moreno, S., and Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* 40, 2662–2673. doi: 10.1111/ejn.12627

Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., and Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *J. Acoust. Soc. Am.* 137, 378–387. doi: 10.1121/1.4904537

Choi, I., Bharadwaj, H. M., Bressler, S., Loui, P., Lee, K., and Shinn-Cunningham, B. G. (2014). Automatic processing of abstract musical tonality. *Front. Hum. Neurosci.* 8:988. doi: 10.3389/fnhum.2014.00988

Coffey, E. B. J., Mogilever, N. B., and Zatorre, R. J. (2017). Speech-in-noise perception in musicians: a review. *Hear. Res.* 352, 49–69. doi: 10.1016/j.heares.2017.02.006

Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., et al. (2012). Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: behavioral, electrophysiological and psychoacoustic study. *Brain Res.* 1455, 75–89. doi: 10.1016/j.brainres.2012.03.034

Dimitrijevic, A., Lolli, B., Michalewski, H. J., Pratt, H., Zeng, F. G., and Starr, A. (2009). Intensity changes in a continuous tone: auditory cortical potentials comparison with frequency changes. *Clin. Neurophysiol.* 120, 374–383. doi: 10.1016/j.clinph.2008.11.009

Dimitrijevic, A., Michalewski, H. J., Zeng, F.-G., Pratt, H., and Starr, A. (2008). Frequency changes in a continuous tone: auditory cortical potentials. *Clin. Neurophysiol.* 119, 2111–2124. doi: 10.1016/j.clinph.2008.06.002

Dochtermann, N. A. (2010). Behavioral syndromes: Carryover effects, false discovery rates and *a priori* hypotheses. *Behav. Ecol.* 21, 437–439. doi: 10.1093/beheco/arq021

Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., and Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science* 291, 1969–1972. doi: 10.1126/science.291.5510.1969

Fu, Q.-J., Shannon, R. V., and Wang, X. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing. *J. Acoust. Soc. Am.* 104, 3586–3596. doi: 10.1121/1.423941

Gaser, C., and Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *J. Neurosci.* 23, 9240–9245. doi: 10.1523/JNEUROSCI.23-27-09240.2003

Gaudrain, E., Grimault, N., Healy, E. W., and Béra, J. C. (2007). Effect of spectral smearing on the perceptual segregation of vowel sequences. *Hear. Res.* 231, 32–41. doi: 10.1016/j.heares.2007.05.001

Geiser, E., Ziegler, E., Jancke, L., and Meyer, M. (2009). Early electrophysiological correlates of meter and rhythm processing in music perception. *Cortex* 45, 93–102. doi: 10.1016/j.cortex.2007.09.010

Gibson, C., Folley, B. S., and Park, S. (2009). Enhanced divergent thinking and creativity in musicians: a behavioral and near-infrared spectroscopy study. *Brain Cogn.* 69, 162–169. doi: 10.1016/j.bandc.2008.07.009

Gregersen, P. K., Kowalsky, E., Kohn, N., and Marvin, E. W. (1999). Absolute pitch: prevalence, ethnic variation and estimation of the genetic component. *Am. J. Hum. Genet.* 65, 911–913. doi: 10.1086/302541

Gregersen, P. K., Kowalsky, E., Kohn, N., and Marvin, E. W. (2001). Early childhood music education and predisposition to absolute pitch: teasing apart genes and environment. *Am. J. Med. Genet.* 98, 280–282. doi: 10.1002/1096-8628(20010122)98:3<280::aid-ajmg1083>3.0.co;2-6

Han, J. H., and Dimitrijevic, A. (2015). Acoustic change responses to amplitude modulation: a method to quantify cortical temporal processing and hemispheric asymmetry. *Front. Neurosci.* 9:38. doi: 10.3389/fnins.2015.00038

Han, J. H., Zhang, F., Kadis, D. S., Houston, L. M., Samy, R. N., Smith, M. L., et al. (2016). Auditory cortical activity to different voice onset times in cochlear implant users. *Clin. Neurophysiol.* 127, 1603–1617. doi: 10.1016/j.clinph.2015.10.049

Hutka, S., Bidelman, G. M., and Moreno, S. (2015). Pitch expertise is not created equal: cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music. *Neuropsychologia* 71, 52–63. doi: 10.1016/j.neuropsychologia.2015.03.019

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., et al. (2009). Musical training shapes structural brain development. *J. Neurosci.* 29, 3019–3025. doi: 10.1523/jneurosci.5118-08.2009

Intartaglia, B., White-Schwoch, T., Kraus, N., and Schön, D. (2017). Music training enhances the automatic neural processing of foreign speech sounds. *Sci. Rep.* 7:12631. doi: 10.1038/s41598-017-12575-1

Johnsrude, I. S., Penhune, V. B., and Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* 123, 155–163. doi: 10.1093/brain/123.1.155

Jones, S. J., and Perez, N. (2001). The auditory "C-process": analyzing the spectral envelope of complex sounds. *Clin. Neurophysiol.* 112, 965–975. doi: 10.1016/s1388-2457(01)00515-6

Kliuchko, M., Heinonen-Guzejev, M., Monacis, L., Gold, B. P., Heikkilä, K. V., Spinosa, V., et al. (2015). The association of noise sensitivity with music listening, training and aptitude. *Noise Health* 17, 350–357. doi: 10.4103/1463-1741.165065

Koelsch, S., Schröger, E., and Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport* 10, 1309–1313. doi: 10.1097/00001756-199904260-00029

Kokoska, S., and Zwillinger, D. (2000). *CRC Standard Probability and Statistics Tables and Formulae, Student Edition.* Boca Raton: CRC Press, 200.

Krizman, J., Bradlow, A. R., Lam, S. S. Y., and Kraus, N. (2017). How bilinguals listen in noise: linguistic and non-linguistic factors. *Biling.* 20, 834–843. doi: 10.1017/s1366728916000444

Lavikainen, J., Huotilainen, M., Ilmoniemi, R. J., Simola, J. T., and Näätänen, R. (1995). Pitch change of a continuous tone activates two distinct processes in human auditory cortex: a study with whole-head magnetometer. *Electroencephalogr. Clin. Neurophysiol.* 96, 93–96. doi: 10.1016/0013-4694(94)00283-q

Liang, C., Earl, B., Thompson, I., Whitaker, K., Cahn, S., Xiang, J., et al. (2016). Musicians are better than non-musicians in frequency change detection: behavioral and electrophysiological evidence. *Front. Neurosci.* 10:464. doi: 10.3389/fnins.2016.00464

Logue, D. M., Mishra, S., McCaffrey, D., Ball, D., and Cade, W. H. (2009). A behavioral syndrome linking courtship behavior toward males and females predicts reproductive success from a single mating in the hissing cockroach, Gromphadorhina portentosa. *Behav. Ecol.* 20, 781–788. doi: 10.1093/beheco/arp061

Madsen, S. M. K., Marschall, M., Dau, T., and Oxenham, A. J. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Sci. Rep.* 9:10404. doi: 10.1038/s41598-019-46728-1

Mankel, K., and Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proc. Natl. Acad. Sci. U S A* 115, 13129–13134. doi: 10.1073/pnas.1811793115

Marques, C., Moreno, S., Castro, S. L., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463. doi: 10.1162/jocn.2007.19.9.1453

Martin, B. A., and Boothroyd, A. (2000). Cortical, auditory, evoked potentials in response to changes of spectrum and amplitude. *J. Acoust. Soc. Am.* 107, 2155–2161. doi: 10.1121/1.428556

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hear. Res.* 219, 36–47. doi: 10.1016/j.heares.2006.05.004

Moore, B. C. J. (2008). The role of temporal fine structure processing in pitch perception, masking and speech perception for normal-hearing and hearing-impaired people. *J. Assoc. Res. Otolaryngol.* 9, 399–406. doi: 10.1007/s10162-008-0143-x

Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U S A* 104, 15894–15898. doi: 10.1073/pnas.0701498104

Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., and Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology* 30, 436–450. doi: 10.1111/j.1469-8986.1993.tb02067.x

Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C. C., Dong, Q., et al. (2018). Piano training enhances the neural processing of pitch and improves speech perception in mandarin-speaking children. *Proc. Natl. Acad. Sci. U S A* 115, E6630–E6639. doi: 10.1073/pnas.1808412115

Novitski, N., Tervaniemi, M., Huotilainen, M., and Näätänen, R. (2004). Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. *Cogn. Brain Res.* 20, 26–36. doi: 10.1016/j.cogbrainres.2003.12.011

Oikkonen, J., Huang, Y., Onkamo, P., Ukkola-Vuoti, L., Raijas, P., Karma, K., et al. (2015). A genome-wide linkage and association study of musical aptitude identifies loci containing genes related to inner ear development and neurocognitive functions. *Mol. Psychiatry* 20, 275–282. doi: 10.1038/mp.2014.8

Okada, B. M., and Slevc, L. R. (2018). Individual differences in musical training and executive functions: a latent variable approach. *Mem. Cognit.* 46, 1076–1092. doi: 10.3758/s13421-018-0822-8

Okamoto, H., and Kakigi, R. (2015). Encoding of frequency-modulation (FM) rates in human auditory cortex. *Sci. Rep.* 5:18143. doi: 10.1038/srep18143

Oxenham, A. J. (2012). Pitch perception. *J. Neurosci.* 32, 13335–13338. doi: 10.1523/JNEUROSCI.3815-12.2012

Pantev, C., and Herholz, S. C. (2011). Plasticity of the human auditory cortex related to musical training. *Neurosci. Biobehav. Rev.* 35, 2140–2154. doi: 10.1016/j.neubiorev.2011.06.010

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., and Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport* 12, 169–174. doi: 10.1097/00001756-200101220-00041

Pantev, C., Ross, B., Berg, P., Elbert, T., and Rockstroh, B. (1998). Study of the human auditory cortices using a whole-head magnetometer: left vs. right hemisphere and ipsilateral vs. contralateral stimulation. *Audiol. Neurootol.* 3, 183–190. doi: 10.1159/000013789

Parbery-Clark, A., Anderson, S., Hittner, E., and Kraus, N. (2012). Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. *Front. Aging Neurosci.* 4:30. doi: 10.3389/fnagi.2012.00030

Parbery-Clark, A., Marmel, F., Bair, J., and Kraus, N. (2011). What subcortical-cortical relationships tell us about processing speech in noise. *Eur. J. Neurosci.* 33, 549–557. doi: 10.1111/j.1460-9568.2010.07546.x

Parbery-Clark, A., Skoe, E., and Kraus, N. (2009a). Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* 29, 14100–14107. doi: 10.1523/jneurosci.3256-09.2009

Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009b). Musician enhancement for speech-in-noise. *Ear Hear.* 30, 653–661. doi: 10.1097/AUD.0b013e3181b412e9

Peretz, I., Cummings, S., and Dubé, M. P. (2007). The genetics of congenital amusia (tone deafness): a family-aggregation study. *Am. J. Hum. Genet.* 81, 582–588. doi: 10.1086/521337

Peretz, I., Vuvan, D., Lagrois, M.-É., and Armony, J. L. (2015). Neural overlap in processing music and speech. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140090. doi: 10.1098/rstb.2014.0090

Picton, T. (2011). *Human Auditory Evoked Potentials*. 1st Edn. (San Diego, CA: Plural Publishing Inc).

Pratt, H., Starr, A., Michalewski, H. J., Dimitrijevic, A., Bleich, N., and Mittelman, N. (2009). Auditory-evoked potentials to frequency increase and decrease of high- and low-frequency tones. *Clin. Neurophysiol.* 120, 360–373. doi: 10.1016/j.clinph.2008.10.158

Pulli, K., Karma, K., Norio, R., Sistonen, P., Göring, H. H. H., Järvelä, I., et al. (2008). Genome-wide linkage scan for loci of musical aptitude in finnish families: evidence for a major locus at 4q22. *J. Med. Genet.* 45, 451–456. doi: 10.1136/jmg.2007.056366

Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (1967). Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *J. Neurophysiol.* 30, 769–793. doi: 10.1152/jn.1967.30.4.769

Ruggles, D. R., Freyman, R. L., and Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PLoS One* 9:e86980. doi: 10.1371/journal.pone.0086980

Sankaran, N., Carlson, T. A., and Thompson, W. F. (2020). The rapid emergence of musical pitch structure in human cortex. *J. Neurosci.* 40, 2108–2118. doi: 10.1523/JNEUROSCI.1399-19.2020

Schellenberg, E. (2019). Music training, music aptitude and speech perception. *Proc. Natl. Acad. Sci. U S A* 116, 2783–2784. doi: 10.1073/pnas.1821109116

Schellenberg, E. G. (2015). Music training and speech perception: a gene-environment interaction. *Ann. N Y Acad. Sci.* 1337, 170–177. doi: 10.1111/nyas.12627

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* 5, 688–694. doi: 10.1038/nn871

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., et al. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nat. Neurosci.* 8, 1241–1247. doi: 10.1038/nn1530

Seppänen, M., Hämäläinen, J., Pesonen, A. K., and Tervaniemi, M. (2012). Music training enhances rapid neural plasticity of N1 and P2 source activation for unattended sounds. *Front. Hum. Neurosci.* 6:43. doi: 10.3389/fnhum.2012.00043

Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., and Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *NeuroImage* 41, 113–122. doi: 10.1016/j.neuroimage.2008.01.067

Shahin, A. J., Roberts, L. E., Pantev, C., Aziz, M., and Picton, T. W. (2007). Enhanced anterior-temporal processing for complex tones in musicians. *Clin. Neurophysiol.* 118, 209–220. doi: 10.1016/j.clinph.2006.09.019

Shahin, A., Bosnyak, D. J., Trainor, L. J., and Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J. Neurosci.* 23, 5545–5552. doi: 10.1523/jneurosci.23-13-05545.2003

Shahin, A., Roberts, L. E., and Trainor, L. J. (2004). Enhancement of auditory cortical development by musical experience in children. *Neuroreport* 15, 1917–1921. doi: 10.1097/00001756-200408260-00017

Stangor, C., and Walinga, J. (2014). Introduction to Psychology - 1st Canadian Edition (p. 709). Ch5.3 Hearing Retrieved from. Available online at https://opentextbc.ca/introductiontopsychology/.

Strait, D. L., and Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Front. Psychol.* 2:113. doi: 10.3389/fpsyg.2011.00113

Swaminathan, S., and Schellenberg, E. G. (2018). Musical competence is predicted by music training, cognitive abilities and personality. *Sci. Rep.* 8:9223. doi: 10.1038/s41598-018-27571-2

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., and Schröger, E. (2005). Pitch discrimination accuracy in musicians vs. nonmusicians: an event-related potential and behavioral study. *Exp. Brain Res.* 161, 1–10. doi: 10.1007/s00221-004-2044-5

Theusch, E., and Gitschier, J. (2011). Absolute pitch twin study and segregation analysis. *Twin Res. Hum. Genet.* 14, 173–178. doi: 10.1375/twin.14.2.173

Theusch, E., Basu, A., and Gitschier, J. (2009). Genome-wide study of families with absolute pitch reveals linkage to 8q24.21 and locus heterogeneity. *Am. J. Hum. Genet.* 85, 112–119. doi: 10.1016/j.ajhg.2009.06.010

Tong, Y., Melara, R. D., and Rao, A. (2009). P2 enhancement from auditory discrimination training is associated with improved reaction times. *Brain Res.* 1297, 80–88. doi: 10.1016/j.brainres.2009.07.089

Tremblay, K. L. (2007). Training-related changes in the brain: evidence from human auditory-evoked potentials. *Semin. Hear.* 28, 120–132. doi: 10.1055/s-2007-973438

Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K., and Collet, G. (2014). Is the auditory evoked P2 response a biomarker of learning? *Front. Syst. Neurosci.* 8:28. doi: 10.3389/fnsys.2014.00028

Tremblay, K. L., Shahin, A. J., Picton, T., and Ross, B. (2009). Auditory training alters the physiological detection of stimulus-specific cues in humans. *Clin. Neurophysiol.* 120, 128–135. doi: 10.1016/j.clinph.2008.10.005

Tremblay, K., Kraus, N., McGee, T., Ponton, C., and Otis, B. (2001). Central auditory plasticity: changes in the N1–P2 complex after speech-sound training. *Ear Hear.* 22, 79–90. doi: 10.1097/00003446-200104000-00001

Ukkola, L. T., Onkamo, P., Raijas, P., Karma, K., and Järvelä, I. (2009). Musical aptitude is associated with AVPR1A-haplotypes. *PLoS One* 4:e5534. doi: 10.1371/journal.pone.0005534

Ukkola-Vuoti, L., Kanduri, C., Oikkonen, J., Buck, G., Blancher, C., Raijas, P., et al. (2013). Genome-wide copy number variation analysis in extended families and unrelated individuals characterized for musical aptitude and creativity in music. *PLoS One* 8:e56356. doi: 10.1371/journal.pone.0056356

Vonck, B. M. D., Lammers, M. J. W., van der Waals, M., van Zanten, G. A., and Versnel, H. (2019). Cortical auditory evoked potentials in response to frequency changes with varied magnitude, rate and direction. *J. Assoc. Res. Otolaryngol.* 20, 489–498. doi: 10.1007/s10162-019-00726-2

Wagner, M., Shafer, V. L., Martin, B., and Steinschneider, M. (2013). The effect of native-language experience on the sensory-obligatory components, the P1–N1-P2 and the T-complex. *Brain Res.* 1522, 31–37. doi: 10.1016/j.brainres.2013.04.045

Wisniewski, M. G., Ball, N. J., Zakrzewski, A. C., Iyer, N., Thompson, E. R., and Spencer, N. (2020). Auditory detection learning is accompanied by plasticity in the auditory evoked potential. *Neurosci. Lett.* 721:134781. doi: 10.1016/j.neulet.2020.134781

Won, J. H., Clinard, C. G., Kwon, S., Dasika, V. K., Nie, K., Drennan, W. R., et al. (2011). Relationship between behavioral and physiological spectral-ripple discrimination. *J. Assoc. Res. Otolaryngol.* 12, 375–393. doi: 10.1007/s10162-011-0257-4

Won, J. H., Drennan, W. R., and Rubinstein, J. T. (2007). Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *J. Assoc. Res. Otolaryngol.* 8, 384–392. doi: 10.1007/s10162-007-0085-8

Yoo, J., and Bidelman, G. M. (2019). Linguistic, perceptual and cognitive factors underlying musicians' benefits in noise-degraded speech perception. *Hear. Res.* 377, 189–195. doi: 10.1016/j.heares.2019.03.021

Zatorre, R. J., and Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11, 946–953. doi: 10.1093/cercor/11.10.946

Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* 8, 547–558. doi: 10.1038/nrn2152

Zendel, B. R., Tremblay, C.-D., Belleville, S., and Peretz, I. (2015). The impact of musicianship on the cortical mechanisms related to separating speech from background noise. *J. Cogn. Neurosci.* 27, 1044–1059. doi: 10.1162/jocn_a_00758

Zuk, J., Ozernov-Palchik, O., Kim, H., Lakshminarayanan, K., Gabrieli, J. D. E., Tallal, P., et al. (2013). Enhanced syllable discrimination thresholds in musicians. *PLoS One* 8:e80546. doi: 10.1371/journal.pone.0080546

**Conflict of Interest**: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 25 August 2020
doi: 10.3389/fnhum.2020.00311

# Physiological and Behavioral Factors in Musicians' Performance Tempo

*Shannon E. Wright* and Caroline Palmer**

*Department of Psychology, McGill University, Montreal, QC, Canada*

Musicians display individual differences in their spontaneous performance rates (tempo) for simple melodies, but the factors responsible are unknown. Previous research suggests that musical tempo modulates listeners' cardiovascular activity. We report an investigation of musicians' melody performances measured over a 12-h day and subsequent changes in the musicians' physiological activity. Skilled pianists completed four testing sessions in a single day as cardiac activity was recorded during an initial 5 min of baseline rest and during performances of familiar and unfamiliar melodies. Results indicated slower tempi for familiar and unfamiliar melodies at early testing times. Performance rates at 09 h were predicted by differences in participants' alertness and musical training; these differences were not explained by sleep patterns, chronotype, or cardiac activity. Individual differences in pianists' performance tempo were consistent across testing sessions: participants with a faster tempo at 09 h maintained a faster tempo at later testing sessions. Cardiac measures at early testing times indicated increased heart rates and more predictable cardiac dynamics during music performance than baseline rest, and during performances of unfamiliar melodies than familiar melodies. These findings provide the first evidence of cardiac dynamics that are unique to music performance contexts.

**Keywords: circadian rhythms, music performance, cardiac dynamics, alertness, recurrence quantification analysis, chronotype**

OPEN ACCESS

***Edited by:***
*Marta Olivetti Belardinelli,*
*Sapienza University of Rome, Italy*

***Reviewed by:***
*Riccardo Brunetti,*
*Università Europea di Roma, Italy*
*Elizabeth Tricomi,*
*Rutgers, The State University of New Jersey, United States*

****Correspondence:***
*Shannon E. Wright*
*shannon.eilyce.wright@mail.mcgill.ca*
*Caroline Palmer*
*caroline.palmer@mcgill.ca*

***Specialty section:***
*This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience*

***Received:*** *29 April 2020*
***Accepted:*** *14 July 2020*
***Published:*** *25 August 2020*

***Citation:***
*Wright SE and Palmer C (2020) Physiological and Behavioral Factors in Musicians' Performance Tempo. Front. Hum. Neurosci. 14:311. doi: 10.3389/fnhum.2020.00311*

## INTRODUCTION

The ways in which musical behaviors interact with human cognition and action have been of great interest to psychologists. For example, models of musical rhythm perception have posited networks of electrophysiological activity, based on populations of neuronal oscillators that fire in synchrony with musical rhythms (Large et al., 2015); these proposals suggest a tight link between musical behaviors and physiological activity. Several studies have focused on the effects of music perception on physiological measures such as heart rate and heart rate variability (see Koelsch and Jäncke, 2015 for a review). Less is known about influences of music performance on physiological processes that underlie cognition and action. We report an investigation of musicians' melody performances measured over a 12-h day and subsequent changes in the performers' physiological activity.

### Circadian Effects on Cognitive and Motor Performance

Several studies have documented time-of-day (circadian) effects on motor and cognitive performance. Circadian rhythms refer to approximately 24 h biological oscillations entrained to the light-dark cycle. For example, body temperature is known to fluctuate predictably over a 24 h cycle (Czeisler and Klerman, 1999), serving as a robust marker of circadian phase, and peaks and troughs in alertness tend to follow the body temperature curve (Dijk et al., 1992). Heart rate and heart rate variability (HRV) also fluctuate predictably over a 24 h period: Heart rate tends to rise in the early morning and decrease in the evening, whereas HRV is typically highest at night and lowest during the day (Bonnemeier et al., 2003; Vandewalle et al., 2007). Edwards et al. (2007) reported that participants' improved accuracy on a simple task of flicking a counter into a target coincided with their late-afternoon peak in body temperature and alertness. A similar finding was reported by Reilly et al. (2007) for soccer-specific motor skills. Rhythmic motor tasks such as pedaling a bike (Moussay et al., 2002; Atkinson et al., 2005) and tapping a steady rhythm with one's finger (Dosseville et al., 2002) have also been shown to have time of day effects, with peak rates of movement occurring in the late afternoon. Furthermore, Dosseville et al. (2002) found that un-cued rhythmic finger tapping rates increase as heart rate increases. Overall, these studies suggest that rhythmic motor performance is influenced by time of day effects and cardiac activity, which shows circadian rhythmicity. We address whether music performance is influenced by circadian fluctuations in physiology similar to other sequential motor activities.

Motor performance is also influenced by circadian-linked individual differences in chronotype, sleep habits, and alertness (Waterhouse et al., 2007; Tamm et al., 2009; Vitale et al., 2015). Chronotype, which depends on the phase of entrainment of one's circadian rhythms to the light-dark cycle (Roenneberg et al., 2003a), refers to the timing of one's sleep and wake in a 24 h period. The commonly known phenomenon of being an "early bird" or a "night owl" (Roenneberg et al., 2003b) refers to differences in the timing of the peaks and troughs of one's circadian rhythms (Baehr et al., 2000): Early birds wake up and go to sleep earlier than night owls, and early birds are more alert in the morning and night owls more alert in the evening. Van Vugt et al. (2013) found that night owl pianists performed scales with greater temporal stability in the evening relative to the morning, while early bird pianists performed scales with more stability in the morning relative to the evening. Differences in chronotype, sleep habits, and alertness may influence performing musicians, who often work in the evening (Gjermunds et al., 2019).

## Individual Variability in Music Performance

Large individual differences in music performances of the same musical works have been documented (Palmer, 1996; Repp, 2005). One common difference across performing musicians is the tempo at which they perform a given piece. Tempo is a factor that differentiates individuals as they speak, walk, tap, and perform other rhythmic movements. These natural movement rates reflect the rate at which individuals comfortably execute a performance in the absence of external stimulus cues. Individual differences in natural movement rates have been observed not only in music performance (Loehr and Palmer, 2011; Zamm et al., 2015, 2016; Scheurich et al., 2018; Palmer et al., 2019) but also in a wide range of rhythmic movements such as walking (Hoyt and Taylor, 1981; Nessler and Gilliland, 2009), speaking (Jungers et al., 2003; Ding et al., 2017), biking (Moussay et al., 2002), and finger tapping (Fraisse, 1982; Dosseville et al., 2002). Individual differences in musicians' spontaneous rates for simple melodies tend to be consistent within individuals but differ widely across individuals (Loehr and Palmer, 2011; Zamm et al., 2015; Scheurich et al., 2018; Palmer et al., 2019). Performers tend to drift toward their spontaneous rate in solo performances when they are initially cued at different rates (Zamm et al., 2018). Moreover, these individual differences in spontaneous rates play an important role in coordinating performances with others: pianists with similar spontaneous rates showed more synchronous performance in duets than pianists with dissimilar rates, in a variety of novel musical works (Zamm et al., 2016). Mechanisms that account for individual differences in musicians' performance rates for the same musical works remain largely unknown; we test whether circadian-related variations in physiology can explain some of these individual differences.

## Cardiac Activity During Music Behaviors

Both the rhythms of cardiac activity and of musical behaviors form long time series of interrelated events; a few studies have addressed how heart rate modulations and musical tempo change together over time. For example, passive listening to music has shown decreased heart rate in response to slower-tempo music (Van Dyck et al., 2017) and increased heart rate during fast-tempo music (Gomez and Danuser, 2007). Heart rate variability during music listening changes less predictably; da Silva et al. (2014) found no difference in HRV between rest (baseline) and music listening, whereas Bretherton et al. (2019) reported that only some tempo manipulations elicited HRV changes relative to a rest condition. Fewer studies have examined changes in musicians' cardiac activity as they perform. De Manzano et al. (2010) found increased heart rate as pianists played familiar music for which they reported large amounts of "flow." Studies of performance anxiety have shown that musicians' heart rate increased when they performed in front of an audience as compared to alone (Brotons, 1994; LeBlanc et al., 1997; Vellers et al., 2015). These studies did not, however, compare resting baseline conditions to music performance. Moreover, the impact of music performance on cardiac activity may be affected by time of day, as cardiac activity shows a circadian rhythm (Bonnemeier et al., 2003; Vandewalle et al., 2007). We investigate how cardiac activity is modulated by music performance, within and across times of day.

Despite the unfolding nature of time series for both cardiac activity and musical behaviors, most studies of heart rate and musical tempo tend to rely on linear measures that fail to capture the non-linear dynamics of the cardiovascular system and of human musical behaviors. The time series formed by music performances and cardiac activity are plausibly more complex than can be captured with a single mean value for beat-to-beat intervals or a standard deviation of those intervals. Recent studies have used non-linear methods of recurrence quantification analysis (RQA) to capture aberrant cardiac activity over time in cardiovascular patient populations (for examples, see Javorka et al., 2008, 2009; Arcentales et al., 2011). Other studies of cardiac dynamics in healthy control populations during sit-to-stand transition tasks show greater cardiac predictability during the more physically demanding standing task (Schlenker et al., 2016). Konvalinka et al. (2011) used RQA techniques to measure cardiac dynamics during a 30-min firewalking ritual during which music was heard. The cardiac dynamics became more predictable (recurrent) during the ritual than during a 30-min pre-ritual baseline measure. Goshvarpour and Goshvarpour (2012) similarly found greater predictability in cardiac dynamics during meditation than during a resting baseline state. Based on these findings, we expect that the predictability of cardiac dynamics may increase during music performance, relative to rest.

The current study had three aims. First, we investigated time-of-day effects on music performance rates by measuring musicians' performances of simple melodies across a 12 h day while measuring their cardiac activity. To disentangle musical familiarity effects from time-of-day effects, performances of both familiar (previously learned) as well as unfamiliar (novel) melodies were measured. Second, we examined influences of circadian rhythms on individual differences in performance tempo. Based on previous findings, we hypothesized that performers with slower spontaneous rates may show slower heart rates and lower alertness than individuals with faster spontaneous rates (within the same time of day). Based on Van Vugt et al.'s (2013) study, early chronotypes were predicted to show less variable performance rates in the morning, whereas late chronotypes should show less variable performance in the evening, respectively.

Third, we investigated how the time series formed by music performance and the accompanying cardiac dynamics changed, by comparing cardiac activity during music performance with cardiac activity during a rest period. We predicted that linear measures of heart rate would be faster and HRV would be lower during music performance relative to rest. Non-linear measures of performers' cardiac dynamics were expected to show more predictability during music performance than during rest. We also examined whether performances of unfamiliar music generated more predictable dynamics than performances of familiar music, based on previous findings of increased cardiac patterning during more demanding tasks (Javorka et al., 2008; Konvalinka et al., 2011) and increased temporal patterning in novices' (non-musician) productions of musical rhythms than in musicians' productions (Scheurich et al., 2018).

## MATERIALS AND METHODS

### Participants

Thirty-two trained pianists with at least 6 years of private piano instruction from the Montreal community participated in the study (mean years of private instruction = 10.6; range = 6–16). Sample size was based on studies of musicians' spontaneous performance tempo that reported moderate effect sizes for comparable samples (Palmer et al., 2019, $n$ = 32 musicians; Zamm et al., 2016, $n$ = 40 musicians). Participants' mean age was 19.5 years (range = 18–27, male = 7). Twenty eight participants were right handed. Exclusion criteria included diagnosed hearing problems or sleep disorders, doing overnight shift work, habitually drinking more than three cups of coffee per day, or having taken a transcontinental flight within the 3 week period prior to participating in the study. Additionally, participants had normal hearing for the range of frequencies used in the music stimuli (<30 dB HL threshold for 125–750 Hz frequencies), as determined by audiometry screening, and had to memorize and perform short melodies without errors. Six additional participants were excluded from the study due to an inability to perform the melodies correctly from memory (3), equipment issues in collecting cardiac data (2), and having fewer than 6 years of private piano instruction (1). Participants received a small honorarium for their participation, and the study was reviewed by the Institutional Review Board of McGill University.

### Stimulus Materials and Equipment

Two musical melodies, primarily isochronous, were included in the study: Frère Jacques ("Twinkle, Twinkle," C Major) and a Canon by Thomas Tallis (D Major). The Frère Jacques theme, composed in the 18th century, was chosen for its familiarity, whereas the Tallis canon, composed in the 16th century, was chosen for its unfamiliarity. Both musical pieces contained eight measures composed in binary (4/4) meter with the majority of quarter-note beat durations. Frère Jacques contains a few eighth notes and half notes in addition. Pianists performed each melody with their right hand, and they were provided with suggested fingerings.

Participants performed melodies on a Roland RD-700 keyboard. Participants' auditory feedback from the keyboard was received directly through AKG K271 Studio headphones. Tones were sounded with a classical piano timbre, and the volume was set by participants to a comfortable listening level. MIDI keystroke information from the performances was recorded with FTAP (Finney, 2001) on a Dell T3600 PC running Linux (Fedora 16).

Cardiac activity was recorded with a Polar H10 heart rate monitor connected via Bluetooth to the application Elite HRV (Personal Pro) run on an iPad Mini. Sublingual temperature was measured with a digital oral thermometer (Personelle Digital Thermometer), following suggestions that sublingual temperature is a reasonable and pragmatic proxy to core body temperature under specific guidelines (Taylor et al., 2014). The temperature measures followed guidelines of a minimum measuring period of 5 min as well as ensuring the mouth is closed for the whole duration of the measurement (Pušnik and Miklavec, 2009; Taylor et al., 2014).

Alertness measures included the Psychomotor Vigilance Task (PVT) and a Visual Analog Scale (VAS). The PVT is a computer-based reaction time task in which participants are asked to click the mouse button as soon as a visual stimulus appears on the computer screen (Dinges and Powell, 1985). The 3-min version of the PVT was used, which has been previously validated (Basner et al., 2011), and presents visual stimuli at randomly varying interstimulus intervals ranging from 1 to 4 s. The PVT measures were collected on a Dell T5810 computer with a HyperX Pulsefire gaming mouse (1000 Hz polling rate) that recorded reaction times. The VAS task (Folstein and Luria, 1973; Monk, 1987) consisted of participants indicating their current level of alertness by making a vertical tick mark on a 10 cm line.

Participants completed a series of questionnaires about their sleep habits, including the Epworth Sleepiness Scale (ESS; Johns, 1991), the Pittsburgh Sleep Quality Index

(PSQI; Buysse et al., 1989), and a sleep diary from Carney et al. (2012). Chronotype was measured with the Munich Chronotype Questionnaire (MCTQ; Roenneberg et al., 2003b). All participants completed the Edinburgh Handedness Inventory and a musical background questionnaire. Participants also completed a short questionnaire about their activities in the hour preceding each laboratory session that might affect alertness, body temperature, or cardiac measures.

## Design

Participants came to the lab for four testing sessions (09, 13, 17, and 21 h) in a single day. The order of testing sessions remained constant across participants (each pianist's first session began at 9 h). Baseline physiological recordings and melody performance tasks were completed at each testing session by all participants, making this a within-subjects 4 (Testing Time) by 2 (Task: 5-min Rest/Music performance) repeated-measures design. The task order was always rest first, followed by music performance. Within the music performance task, the ordering of the familiar and unfamiliar melody performances was alternated between participants and testing sessions: Half of the participants performed the Familiar melody first at the 09 h testing session, and the other half began with the Unfamiliar melody. At subsequent testing sessions, participants alternated which melody they performed first. Each participant performed a total of 32 melody performance trials (4 times of day × 2 melodies × 3 trials) over the course of the experiment.

The main behavioral dependent variables from the melody performances were spontaneous production rate (SPR, mean interonset interval, IOI in ms) and variability of interonset intervals (measured by the coefficient of variation, SD/mean IOI). Primary physiological dependent variables included sublingual temperature (°C), heart rate (mean inter- heartbeat interval, RR), heart rate variability (measured by the standard deviation of normal-to-normal intervals, SDNN), alertness (PVT reaction times and VAS subjective scores), chronotype, and sleep deprivation measures computed from the sleep diary (described below).

## Procedure

Participants were first screened for eligibility via e-mail; if eligible, electronic copies of the musical notation (without melody titles) for the melodies used in the study were sent to participants, and participants were asked to memorize the melodies before their participation in the study. Participants also received a sleep diary which they were asked to complete for the week preceding the laboratory session.

Upon arrival at the lab, participants read and signed a consent form before completing an audiometry screening in which pure tones were presented over closed headphones (Maico MA40), to ensure they could hear the range of frequencies involved in the music performance task at a threshold of <30 db. Participants who passed the audiometry screening were invited to continue to a melody memorization task. First, participants were presented with a melody in notation. After practicing the first melody (Familiar or Unfamiliar) both with and without musical notation, participants were given up to three practice trials to perform the melody from memory without pitch errors. Then the participants repeated the task with the second melody. All participants performed the melodies without pitch errors in the memorization phase.

Next, participants attached the heart rate monitor around their chest. A 5-min baseline sublingual temperature and heart rate recording was taken during the Rest task while participants were seated and completing questionnaires. To ensure correct temperature readings, participants were instructed to insert the thermometer under their tongue and breathe normally through their nose; they were instructed to keep movement to a minimum and to avoid crossing their legs so as not to influence heart rate measures. During this time, participants marked their current alertness level in the VAS task. At the end of the 5-min rest period, participants removed the thermometer but kept the heart rate monitor on for the rest of the testing session.

Participants then completed the Psychomotor Vigilance Task. They were instructed that red numbers would appear on a black screen, and they were to click the mouse as soon as, but not before, they saw the red numbers appear. If participants clicked the mouse before the red numbers appeared, the letters "fs" appeared on the screen to inform the participant they had made a false start. A new trial was then begun. Each trial continued until participants clicked the mouse.

Participants then sat at the piano keyboard and were presented with the first melody in music notation. They were instructed to perform a practice trial consisting of four repetitions at a steady, comfortable rate without pauses. The experimenter removed the music notation and participants repeated a practice trial of the same length from memory. Once participants were comfortable with the task, they moved on to the experimental trials. Each experimental trial consisted of 4 repetitions of the melody performed from memory (in the absence of music notation) without pauses at a comfortable, steady rate. After completing all trials of the first melody, participants filled out a brief questionnaire about their activities prior to the testing session; then the same practice and experimental trials were repeated for the second melody. At the end of the melody performance task, participants removed the heart rate monitor and received a small honorarium. The same procedure was repeated at each testing session with the addition of a debriefing period at the end of the final session. The duration of the first testing session (which included the audiometric screening and memorization practice) was approximately 45 min; subsequent testing sessions took approximately 25 min.

## Data Analysis

Pitch errors in melody performances were identified by comparing the recorded MIDI data with the contents of the musical score, using the MIDI Matcher Toolbox in Matlab (Large, 1993). Repetitions containing a pitch error were excluded from analysis as timing errors are likely to co-occur with pitch errors (Drake and Palmer, 2000); 0.03% of all repetitions were excluded from analysis. The half-note durations in Frère Jacques were interpolated at the quarter-note level, and eighth notes that did not align with the quarter-note beat were excluded from the analyses. Interonset intervals (IOI), coinciding with quarter-note beats in both melodies, were computed. IOI's greater or less than 3 standard deviations away from the mean IOI for that trial were excluded from behavioral analyses (0.13% of all IOIs).

Each participant's Spontaneous Production Rate (SPR) was computed on the IOIs from the middle two of four melody repetitions in each trial, similar to previous studies (Zamm et al., 2016; Palmer et al., 2019), as the middle of each trial tends to show more stable tempo due to musicians' tendencies to slow down at phrase boundaries at beginnings and endings of trials (Palmer, 1989; Repp, 1990). Participants' SPR for each melody was then calculated from the mean IOI of the middle 2 repetitions of each trial and averaged across trials within melody. Similarly, the mean Coefficients of Variation (CV) were calculated from the same IOIs in the middle two repetitions and a mean CV was computed across trials.

Linear analyses of cardiac data were completed using Kubios (HRV Standard, 3.1.0). Mean RR intervals and the SDNN were computed for each 5-min baseline recording as well as during the total duration of melody performances, including practice and experimental trials, in order to have the longest consecutive measurement period possible. Recurrence quantification analysis (RQA) was also conducted on cardiac data using the CRP Toolbox 5.22 [Marwan, 2019, run with MATLAB 2018a (v9.4.0)]. RQA is a non-linear analysis technique, often used on behavioral and cardiac data (Javorka et al., 2008; Demos et al., 2011; Marwan et al., 2013), that identifies recurrent states in a dynamical system using Takens's (1981) method of higher-dimensional reconstruction (Webber and Zbilut, 2005; Nayak et al., 2018). Time-delayed copies of the cardiac signals are generated and projected into multidimensional phase space (Konvalinka et al., 2011) with the parameter tau denoting the time delay. For each resting period (baseline) and music performance, tau was chosen based on the first local minimum of the average mutual information function. Tau therefore varied across participants and within participants by testing session and task (Javorka et al., 2009), and the resulting range was 2–12. The False Nearest Neighbor (FNN) method was used to select an embedding dimension; FNN values close to zero indicate that the signal is projected into a sufficient number of dimensions (Webber and Zbilut, 2005; Nayak et al., 2018). Embedding dimensions were chosen on an individual basis and ranged from 4 to 8. A Theiler window fixed to the time delay (Javorka et al., 2009) was applied to the data, as cardiac signals tend to show high autocorrelation (Martin-Gonzalez et al., 2018). Recurrence rate, the percentage of recurrent points in the system, was fixed to 5% as per previous RQA studies of cardiac signals (Javorka et al., 2008, 2009).

Recurrence plots, 2-dimensional representations of the recurrent points in a system, were generated to visualize the cardiac dynamics. Each point in the plot represents a system state that is recurrent with a previous state (Webber and Zbilut, 2005). The time series signal is plotted against itself such that the recurrence plot is symmetric across the diagonal. Two parameters were used to quantify the observed recurrence. First, determinism (DET) measured the percentage of points in the recurrence plot forming diagonal lines (excluding the line of identity), where the minimum number of points required to be considered a line was set to 2 (Equation 1). Determinism is a measure of the predictability of a system over time (Webber and Zbilut, 1994). Second, laminarity (LAM) captures the percentage of points forming vertical (or horizontal) lines in the recurrence plot (Equation 2) and is an indicator of the extent to which a system "gets stuck" in a specific state (Nayak et al., 2018).

$$\%DET = 100 * \frac{\sum_{l=lmin}^{N} lP(l)}{\sum_{l=1}^{N} lP(l)} \quad (1)$$

$$\%LAM = 100 * \frac{\sum_{v=vmin}^{N} vP(v)}{\sum_{v=1}^{N} vP(v)} \quad (2)$$

Chronotype was determined from the MCTQ which estimates an individual's mid-sleep point based on self-reported times of sleep onset and wake for both work and free days (Roenneberg et al., 2003b). As imposed social schedules may mask an individual's natural mid-sleep point and lead to sleep debt (Wittmann et al., 2006), an adjusted value of mid-sleep on work-free (weekend) days (MSF) that accounts for possible sleep debt was used to estimate one's chronotype (Roenneberg et al., 2004). The adjusted value ($MSF_{sc}$) is derived according to the following equation:

$$MSF_{sc} = MSF - 0.5[TS_F - (5(TS_w) + 2(TS_F)/7)], \quad (3)$$

where $TS_w$ is the average total sleep duration (in minutes) on work days and $TS_F$ is the average total sleep duration (in minutes) on free days. This equation yields a time (ex. 04:00 h) corresponding to the midpoint of the individual's sleep cycle. Midpoints earlier than 05:00 h typically denote an earlier chronotype and later midpoints denote a later chronotype (Roenneberg et al., 2003b).

Alertness scores were derived for each testing session. Mean reaction times on the PVT were calculated per participant for correct response trials. A score from 1 to 10 on the VAS at each testing session per participant was analyzed, with higher scores indicating greater alertness.

## RESULTS

### Time of Day Effects in Music Performance

To test for differences in mean SPR values across the day, a two-way ANOVA on mean SPR by Time of Day (09, 13, 17, 21 h) and Melody (Familiar, Unfamiliar) was performed. This analysis indicated significant main effects of Time of Day [$F(3,93) = 17.42$, $p < 0.01$, $\eta_p^2 = 0.36$), and of Melody [$F(1,31) = 41.73$, $p < 0.01$, $\eta_p^2 = 0.57$], and no significant interaction. Shown in **Figure 1** (top), mean SPR was significantly slower at 09 h than at all other testing sessions, and was slower at 13 h than at 21 h (Tukey's $HSD = 13.72$, $p < 0.05$). SPR was faster for the Unfamiliar melody performances (mean = 362.61 ms) than the Familiar performances (mean = 397.84 ms); this finding is not surprising as the Familiar melody's rhythm contained half and quarter notes which constrained the fastest rate possible, whereas the Unfamiliar melody contained only quarter notes.

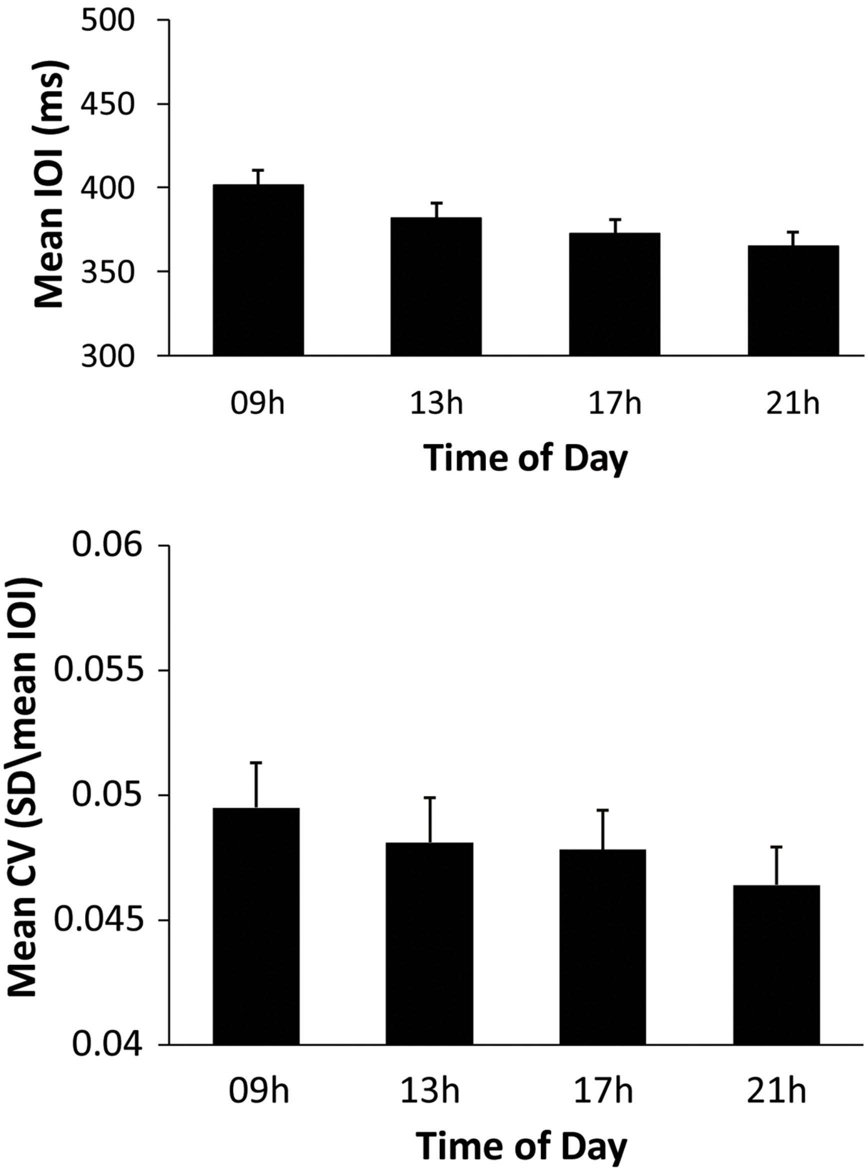



**FIGURE 1 |** Performers' mean Spontaneous production rates (ms) by Time of Day **(top)** and mean CV by Time of Day **(bottom)** for all melody performances.

To test whether the stability of music performance changed over the day, the same ANOVA was conducted on mean CV. There were significant main effects of Time of Day [$F(3,93) = 3.827$, $p = 0.012$, $\eta_p^2 = 0.11$] and of Melody [$F(1,31) = 9.200$, $p = 0.005$, $\eta_p^2 = 0.23$], and no significant interaction. **Figure 1** (bottom) shows the CV values; the CV at 09 h was significantly larger than at 17 and 21 h, and the CV at 13 h was significantly larger than at 21 h (Tukey $HSD = 0.002$, $p < 0.05$). Paralleling the findings of mean SPR becoming faster across the day, pianists became more stable in their performances across the day. The mean CV for Familiar melody performances (mean = 0.05) was greater than for Unfamiliar melody performances (mean = 0.045), consistent with the varying rhythmic structure of the Familiar melody compared with the isochronous rhythm of the Unfamiliar melody. Overall, these findings suggest a 09 h effect on SPR and CV that diminished over the day.

To examine whether performers' alertness levels varied over the testing sessions, we tested participants' reaction times on correct trials in the Psychomotor Vigilance Task in a one-way ANOVA by Time of Day (09, 13, 17, and 21 h). Mean reaction times varied significantly across the day [$F(3,93) = 3.70$, $p < 0.01$, $\eta_p^2 = 0.11$]. Mean reaction times at 09 h were significantly slower (mean = 233.57 ms) than mean reaction times at 21 h (mean = 224.50 ms) ($HSD = 7.29$, $p < 0.05$). No other time-of-day comparisons were significant. In line with the primarily late chronotype sample, these findings suggest that participants were less alert at 09 h than at 21 h. Mean subjective alertness scores (Visual Analog Scale, VAS) did not show significant effects of time of day.

The sublingual body temperatures were assessed with a one-way ANOVA by Time of Day (09, 13, 17, and 21 h). There was a significant main effect [$F(3,93) = 6.28$, $p = 0.001$, $\eta_p^2 = 0.17$]; *post hoc* analyses indicated that body temperature at 09h was significantly higher (mean = 36.73°Celsius) than body temperature at 13 and 17 h ($HSD = 0.237$, $p < 0.05$), with no other comparisons differing significantly. Consistent with previous work (Christie and McBrearty, 1979; Monk, 2005) participants' sublingual temperature decreased slightly in the middle portion of the day and rose again through the evening.

## Individual Differences in Performance Tempo

Next, we examined individual differences in spontaneous production rate (SPR). **Figure 2** shows the mean spontaneous rates of individuals' Familiar melody performances at each testing session, ordered in each graph from fastest to slowest individual at 09 h. The similarity of the faster-to-slower patterns across the four graphs suggests that the individual differences in performance tempo were consistent. To test whether the SPR values were stable across times of day, Spearman's rank order correlations were applied to test whether the ordering of individuals at the 09 h session matched the ordering at the 13, 17, and 21 h sessions. The rank-ordered SPR values held from the 09h session to each testing session [13 h $\rho = 0.87$, $p < 0.01$; 17 h $\rho = 0.85$, $p < 0.01$; 21 h $\rho = 0.82$, $p < 0.01$). **Figure 3** shows the same pattern of individuals' SPR values across testing sessions for the Unfamiliar melody performances, where each graph is again ordered by fastest to slowest individual at 09 h. Similar to the Familiar melodies, the individual differences at 09 h were significantly retained across all testing sessions (13 h $\rho = 0.84$, $p < 0.01$; 17 h $\rho = 0.88$, $p < 0.01$; 21 h $\rho = 0.81$, $p < 0.01$). These findings suggest that large individual differences in spontaneous rates existed for both familiar and unfamiliar melodies, and the individual differences were consistent across times of day.

To address whether individual differences in SPR were related to chronotype, we computed participants' chronotype from the Munich Chronotype Questionnaire (MCTQ), following Equation 3. The mean and median midsleep point on free days ($MSF_{sc}$) were 05 h14 and 05 h08, respectively (range = 03 h13 to 09 h21). Individuals with $MSF_{sc}$ values later than about 05 h are typically considered a late chronotype or night owl (Roenneberg et al., 2003b). The present sample $MSF_{sc}$ was positively skewed, with 3 of 32 participants in the <04 h range, 23 of 32 participants in the 04–05 h range, and 6 of 32 participants in the ≥06 h range. Although chronotype appears nearly normally distributed in the general population, the overrepresentation of night owls in the present sample is consistent with previous findings for this age group (Roenneberg et al., 2003b) as well as for musicians (Gjermunds et al., 2019). Due to the lack of variability

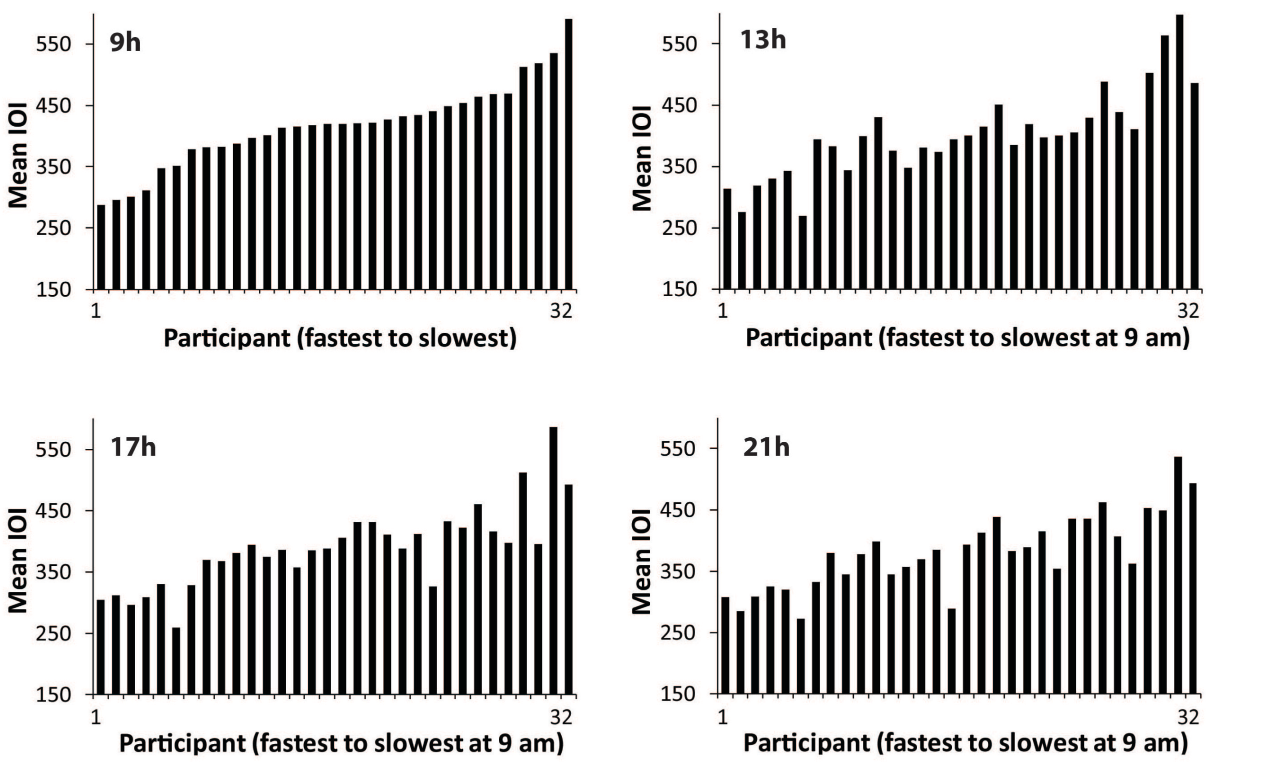



**FIGURE 2 |** Distributions of performers' mean SPR values (mean IOI in ms) for the Familiar melody, ordered at each testing session from fastest to slowest performer according to 09 h. Each bar = one performer.

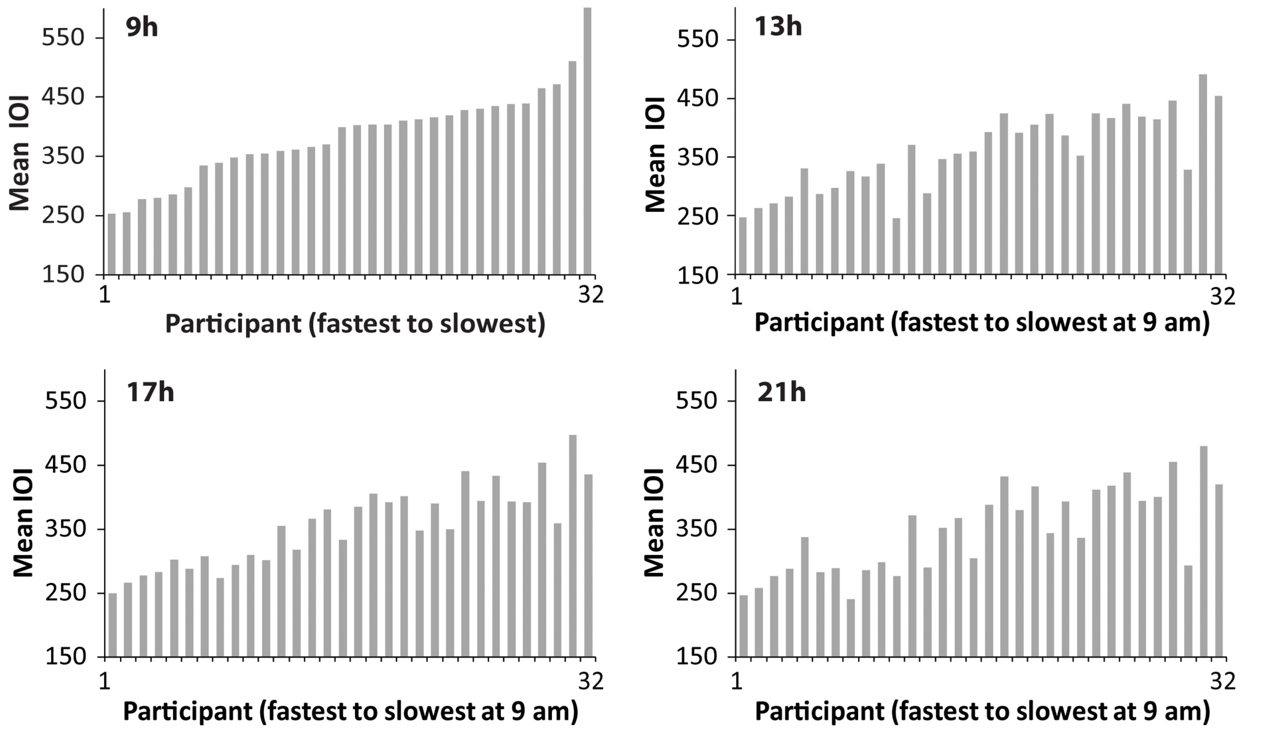



**FIGURE 3 |** Distributions of performers' mean SPR values (mean IOI in ms) for the Unfamiliar melody, ordered at each testing session from fastest to slowest performer according to 09 h. Each bar = one performer.

in chronotype and the overrepresentation of night owls, the relationship between chronotype and SPR could not be assessed; the three earliest chronotypes and the six latest chronotypes did not show SPR patterns that differed from the remaining cohort.

We next examined the individual differences in spontaneous rates (mean SPR) in terms of amount of musical training and alertness (reaction times on correct trials from the PVT) using a multiple regression model that predicted mean SPR from years of musical training and reaction time (RT). The multiple regression fits for the Familiar melody performances, predicting mean SPR from RT and Musical Training, were significant at 09 h ($R = 0.523$, $p < 0.01$) and at 13 h ($R = 0.51$, $p < 0.05$). Semi-partial correlations indicated significant contributions to the SPR of both RT (standardized coefficient = 0.35, $p = 0.035$) and Musical Training (standardized coefficient = −0.378, $p = 0.02$) at 09 h. The semi-partial correlations at 13 h indicated similar contributions of RT (standardized coefficient = 0.35, $p = 0.04$) and Musical Training (standardized coefficient = −0.3305, $p = 0.04$). At both 09 and 13 h, individuals' slower tempi were associated with longer RT values in the PVT (lower alertness) and with less musical training. The same multiple regression model did not predict individuals' SPR values at 17 h or at 21 h. The same multiple regression model fit to mean SPR values for the Unfamiliar melody performances showed similar influences of alertness (RT) but not of musical training. The multiple regression fit reached significance at 09 h ($R = 0.45$, $p = 0.04$) but not at any other testing session. The semi-partial correlations indicated significant contributions of RT at 09 h (standardized coefficient = 0.3958, $p = 0.024$). Consistent with performances of the Familiar piece, participants with lower alertness scores (higher RT values) performed the Unfamiliar melody at a slower tempo at the first session of the day.

There was no significant relationship between acute sleep deprivation (average duration of sleep in 1 week – duration of single night sleep preceding laboratory session) and individual SPR values at any testing session, for Familiar or Unfamiliar melody performances, suggesting that individual differences in SPRs were not accounted for by differences in acute sleep deprivation.

## Cardiac Dynamics During Music Performance

Linear cardiac measures (RR interval and SDNN) were examined to identify whether cardiac activity varied across the day and across music and rest. A two-way within-subjects ANOVA on mean RR interval by Time of Day (09, 13, 17, and 21 h) and Task (Baseline rest, Music Performance) showed a significant main effect of Task [$F(1,31) = 13.51$, $p = 0.001$, $\eta_p^2 = 0.30$), and no main effect of Time of Day or interactions. RR interval was shorter during music performance (mean = 712.57 ms) than during baseline rest (mean = 734.12 ms), indicating that pianists' heart rate increased from baseline to music performance. To examine the two melodies performed at each testing session, a follow-up two-way ANOVA on mean RR interval by Time of Day (09, 13, 17, 21 h) and Melody (Familiar, Unfamiliar) was performed. There was a significant main effect of Melody [$F(1,31) = 6.27$, $p = 0.02$, $\eta_p^2 = 0.17$] and a significant Time of Day × Melody interaction [$F(3,93) = 3.20$, $p = 0.03$, $\eta_p^2 = 0.09$]. As seen in **Figure 4**, participants' RR intervals were shorter during Unfamiliar melody performances than during Familiar melody performances at 09, 13, and 17 h, but not at 21 h (*HSD* = 4.36, $p < 0.05$). Participants' heart rate increased during the Unfamiliar melody performance relative to the Familiar melody performance earlier in the day but not later in the evening. Similar analyses on mean SDNN values showed no significant effects of time of day or type of melody, and no interaction.

Non-linear RQA measures of cardiac activity evaluated the predictability of performers' heart rate measures (R–R intervals, in ms). A two-way ANOVA on mean determinism (%DET, measuring predictability) by Time of Day (09, 13, 17, and 21 h) and Task (Baseline rest, Music Performance) showed no main effect of Time of Day, a significant main effect of Task [$F(1,31) = 4.15$, $p = 0.05$, $\eta_p^2 = 0.12$] and a significant Time of Day × Task interaction [$F(3,93) = 6.48$, $p < 0.001$, $\eta_p^2 = 0.17$]. There was greater determinism (predictability) during music performance (mean %DET = 45.699) than during baseline rest (mean %DET = 42.959). **Figure 5** (top) demonstrates that the cardiac activity showed significantly greater determinism during music performance at 09 and 13 h (*HSD* = 0.053, $p < 0.05$) but not at 17 and 21 h. Recurrence plots for a single participant at 09 h in **Figure 6** demonstrate the greater amount of determinism or predictability during music performance than during baseline rest.

A follow-up two-way ANOVA on mean %DET by Time of Day (09, 13, 17, and 21 h) and Melody (Familiar, Unfamiliar) showed a significant main effect of Melody [$F(1,31) = 6.348$, $p = 0.017$, $\eta_p^2 = 0.17$], and no main effects or interactions with Time of Day. Specifically, %DET values were larger during the Unfamiliar melody performances (mean = 46.99) than the Familiar melody performances (mean = 44.41). **Figure 7** shows a pair of recurrence plots illustrating this difference for a single subject, where a greater proportion of recurrent points form diagonal lines in the plot on the right (Unfamiliar performance). Thus, greater determinism (predictability) in cardiac activity was seen during music performances compared to rest, and for Unfamiliar melody performances compared to Familiar melody performances.

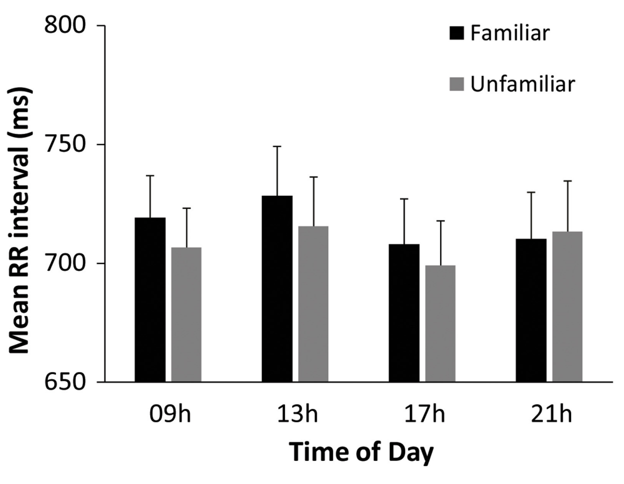



**FIGURE 4 |** Performer's mean heartbeat interval (RR interval, in ms) by Time of Day and Melody.

The same analyses were performed to identify whether laminarity in the cardiac system (how much the system got stuck in a recurrent state) changed with Time of day and Task (Baseline rest and Music Performance). The mean laminarity (LAM) values indicated a significant main effect of Task [$F(1,31) = 5.415$, $p = 0.027$, $\eta_p^2 = 0.15$], no main effect of Time of Day, and a significant Time of Day × Task interaction [$F(3,93) = 3.678$, $p = 0.015$, $\eta_p^2 = 0.11$]. Recurrence plots for a single participant in **Figure 8** show that a greater proportion of points form vertical/horizontal lines during the melody performances (mean %LAM = 53.74) than during baseline rest (mean %LAM = 50.25). *Post hoc* comparisons of the interaction showed that mean laminarity values were significantly greater during music performance than baseline only at 09 h (*HSD* = 0.055, $p < 0.05$), also shown in **Figure 5** (bottom). A follow-up ANOVA on mean LAM value by Time of Day and Melody (Familiar, Unfamiliar) showed a significant main effect of Time of Day [$F(3,93) = 4.107$, $p = 0.009$, $\eta_p^2 = 0.17$] and no effects or interactions with Melody. Overall, there was greater laminarity and determinism (predictability) in cardiac rhythms during music performance than during baseline rest; that difference was larger at earlier testing sessions. In addition, there was greater determinism during Unfamiliar melody performances than during Familiar melody performances, controlling for time of day.

## DISCUSSION

This study examined time-of-day effects on musicians' performance tempo for simple melodies, and whether circadian effects on physiology could account for individual differences in performance tempo. Trained pianists' performance rates for familiar and unfamiliar melodies were recorded at four testing sessions in a single day (09, 13, 17, and 21 h) while cardiac activity was recorded. Resting measures of performers' cardiac activity, alertness, and body temperature were recorded at each testing session. Additionally, this study utilized a non-linear analysis technique (RQA) to investigate cardiac dynamics during music performance both within and across times of day.

Overall, musicians' spontaneous performance rates were slower and more variable at 09 h and became slightly faster and less variable at later testing sessions. The largest difference in SPR and variability of performances was between 09 and 21 h, similar to previous findings on spontaneous motor rates of tapping (Dosseville et al., 2002) and cycling (Moussay et al., 2002), which have shown slowest rates in the morning and fastest rates in the evening. These results suggest that melody performances increased in tempo and in temporal regularity from the morning to the evening, a finding that is somewhat consistent with a sample of largely night-owl chronotypes (Van Vugt et al., 2013). Participants completed all testing sessions in the same order

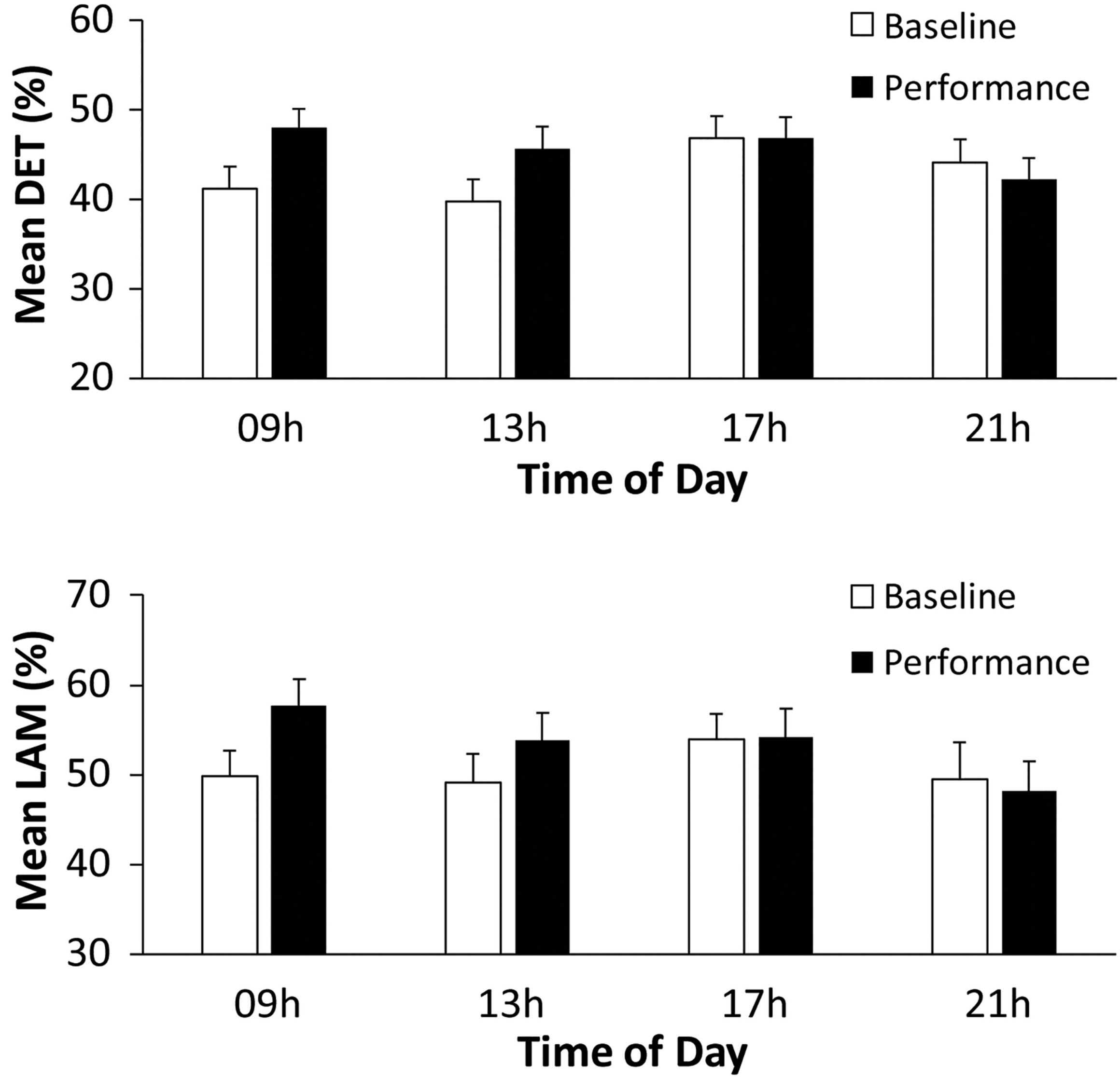



**FIGURE 5 |** Performers' mean% Determinism **(top)** and mean% Laminarity **(bottom)**, by Time of Day and Task (Baseline rest/Music performance).

in this study (to control for sleep differences between testing sessions); therefore, it is possible that some changes in melody performance rate and temporal variability were attributable to practice effects over the session trials. In the context of motor sequencing, performing repeated trials of specific finger sequences in a blocked (rather than randomized) fashion typically results in faster learning rates (Fogel et al., 2017; Caramiaux et al., 2018). The observed changes in participants' melody performances across times of day were similar for unfamiliar and familiar melodies, which is consistent with practice effects over trials (as opposed to familiarity with the musical melodies).

Musicians showed large individual differences in spontaneous performance rates (SPR), replicating previous studies on natural movement rates in music performance (Zamm et al., 2015; Palmer et al., 2019) and tapping tasks (Scheurich et al., 2018). Importantly, the individual differences in pianists' performance tempo were consistent across the day for both familiar and unfamiliar melodies: Pianists who performed quickly in the morning also performed quickly in the evening, and the same was true for pianists with slower rates. These findings are consistent with dynamical systems theory predictions that an individual's natural movement rate, a property of a periodic oscillatory system (Kelso, 1997), may serve as an attractor state at which movement efficiency is maximized (Hoyt and Taylor, 1981; Zamm et al., 2018). Indeed, neuromuscular fatigue has been shown to be minimized at cyclists' spontaneous (uncued) pedaling rates for a given load resistance (Takaishi et al., 1996; Moussay et al., 2002), and reduced kinetic energy expenditure in pianists' finger movements is associated with increased temporal accuracy of performance (Goebl and Palmer, 2014). Our finding of consistency across testing sessions in individuals' performance tempo suggests that one's spontaneous production rate may be

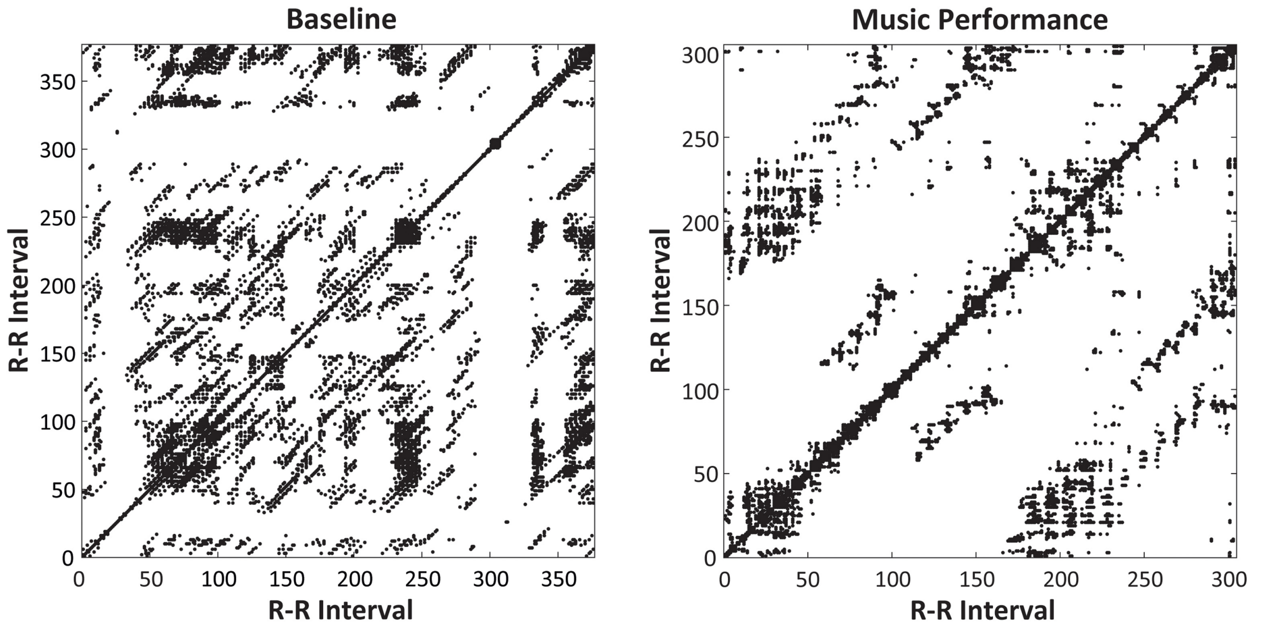



**FIGURE 6 |** Recurrence plots for a single participant's heartbeat series (RR intervals, in ms, plotted on *x*- and *y*-axes) at 09 h. Plots include 5-mine Baseline rest period **(left)** and Unfamiliar melody performance **(right)**. The melody performance showed greater determinism (%DET = 58.88) than did the Baseline Rest (%DET = 39.62).

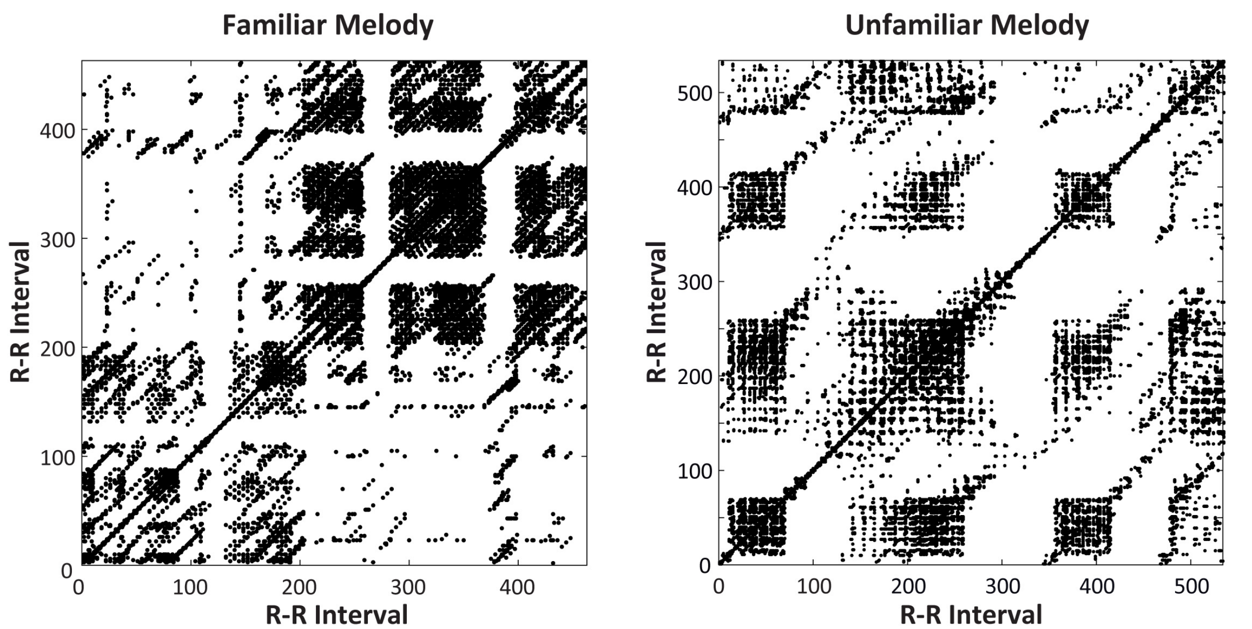



**FIGURE 7 |** Recurrence plots for a single participant's heartbeat series (RR intervals, in ms, plotted on *x*- and *y*-axes) at 13 h. Plots include Familiar melody performance **(left)** and Unfamiliar melody performance **(right)**. The Unfamiliar melody performance showed greater Determinism (=53.31%) than did the Familiar melody performance (= 44.04%).

an energy-efficient state for melody performance that transcends time of day effects or familiarity with the melody.

Alertness measures also showed time of day effects and explained some of the individual variability in performance rates; participants who performed melodies at a slower rate at 09 and 13 h had slower reaction times on the PVT task at these times. Lower alertness in the morning is not surprising for the later chronotype sample of musicians tested here (Roenneberg et al., 2003a). At early testing times, musicians' spontaneous performance rates were influenced by

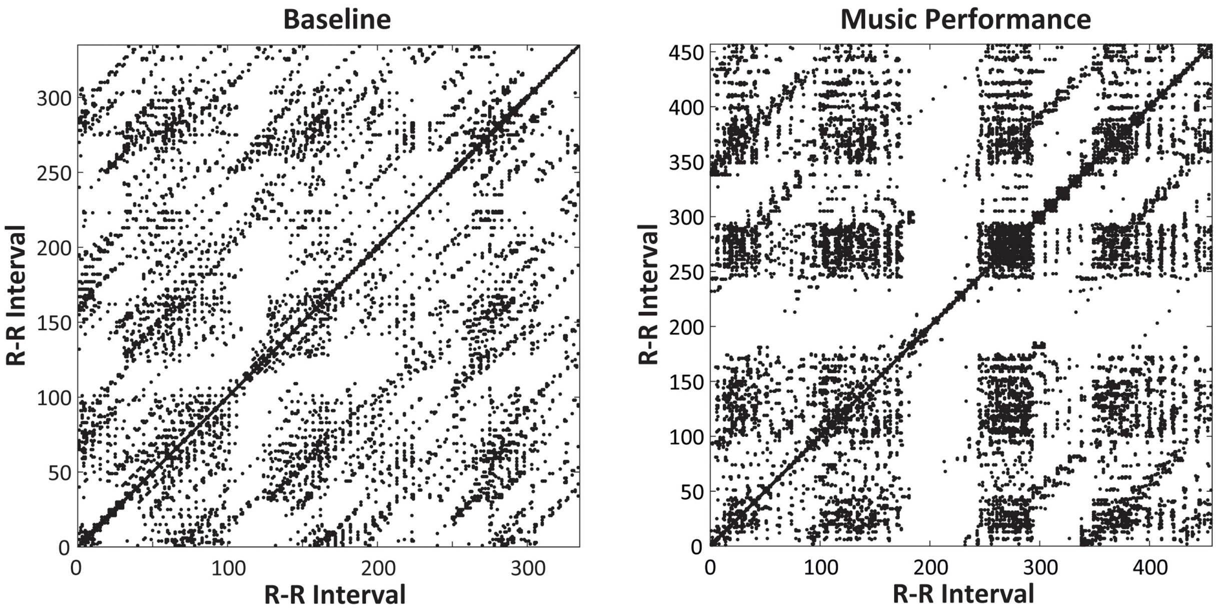



**FIGURE 8 |** Recurrence plots for a single participant's heartbeat series (RR intervals, in ms, plotted on *x*- and *y*-axes) at 09 h. Plots include Baseline rest **(left)** and Unfamiliar melody performance **(right)**. The Melody performance showed greater Laminarity (=51.6%) than did the Baseline Rest (= 27.0%).

both physiological (alertness) and behavioral (musical training) variables. Participants with faster reaction times in the PVT task and more years of formal piano training tended to show faster performance rates at 09 and 13 h. Interestingly, neither physiological nor behavioral variables predicted performance rates later in the day. Alertness and musical training may have greater effects on melody performance when musicians are less comfortable with a musical task (for example at the first 09 h testing session), an interpretation consistent with the general increased temporal stability reported for musicians with increased training (Scheurich et al., 2018). This hypothesis could be addressed by randomizing participants' first testing session to begin at different times of day in future studies.

Finally, the complexity of musicians' cardiac activity was compared between 5-min rest periods and music performances, as well as between performances of familiar and unfamiliar melodies. Both linear and non-linear measures of heart rate (R–R intervals) indicated significant differences from rest to music performance, with faster and more patterned (deterministic) heart rates during music performance than during rest, across times of day. The largest differences between music performance and rest were seen at 09 h and at 13 h. In addition, heart rates were faster during performances of unfamiliar melodies than familiar melodies, and laminarity (recurring patterns) of cardiac activity was greater for unfamiliar melodies than for familiar melodies. Increased predictability of cardiac signals has been observed during increases in task difficulty for both physical (Javorka et al., 2009; Konvalinka et al., 2011; Schlenker et al., 2016) as well as cognitive behaviors (Goshvarpour and Goshvarpour, 2012). Overall, the differences in cardiac dynamics between rest and music performance, and between performance of familiar and unfamiliar melodies, suggest that increased predictability and stability of cardiac signals may be a physiological marker of increased behavioral difficulty.

The current findings were limited by the simple musical materials used, and the chronotype sample of musicians obtained. Two simple melodies were included to reduce the memorization demands on participants; those melodies had simple but not identical rhythmic structures. Future research may examine the roles of musical performance styles and rhythmic complexity in performance rates and cardiac rhythms. Furthermore, the chronotype of the obtained musician sample was biased toward night owls, in line with previous research (Gjermunds et al., 2019). It is possible that decreases in SPR and increases in performance stability over the day were specific to the night owl chronotype, as late chronotypes perform better on strength tasks (Tamm et al., 2009) and music performance tasks (Van Vugt et al., 2013) in the evening relative to the morning. Future research may extend these findings to a more diverse sampling of chronotypes.

In sum, pianists' rates of melody performances increased and variability decreased across the 12 h day, similar to circadian influences on other motor skills. Time of day may be an important relationship for musicians to consider; there may be ideal times of day to practice or perform. Individual differences in performance rates early in the day were predicted by both alertness and musical training. In addition, large individual differences in the musicians' performance rates remained consistent across the 12-h time period. Finally, pianists' cardiac dynamics became more predictable and recurred more during music performance than during a baseline rest interval, as well as during performances of an unfamiliar melody than a familiar melody. To our knowledge, these findings provide the first evidence that performing music affects non-linearities of cardiac dynamics in specific and replicable ways within individuals. Overall, these discoveries of performers' cardiac dynamics suggest possible applications to music therapy; the time of day at which music is performed, as well as the familiarity of the music, may influence music's ability to modulate physiological systems.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, upon reasonable request.

## ETHICS STATEMENT

The studies involving human participants were reviewed by the McGill University Research Ethics Board. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

SW and CP designed the experiments, wrote and edited the manuscript. SW conducted the experiments and analyzed the data. All the authors contributed to the article and approved the submitted version.

## FUNDING

This research was funded in part by an NSERC-CREATE scholarship to SW, and by NSERC Grant 298173 and a Canada Research Chair to CP.

## ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of Sebastian Andric, Jocelyne Chan, Amy Friedman, Nour Saba, Jessica Slater, and Frances Spidle.

## REFERENCES


Arcentales, A., Giraldo, B. F., Caminal, P., Benito, S., and Voss, A. (2011). Recurrence quantification analysis of heart rate variability and respiratory flow series in patients on weaning trials. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2011, 2724–2727.

Atkinson, G., Todd, C., Reilly, T., and Waterhouse, J. (2005). Diurnal variation in cycling performance: influence of warm-up. *J. Sports Sci.* 23, 321–329. doi: 10.1080/02640410410001729919

Baehr, E. K., Revelle, W., and Eastman, C. I. (2000). Individual differences in the phase and amplitude of the human circadian temperature rhythm: with an emphasis on morningness–eveningness. *J. Sleep Res.* 9, 117–127. doi: 10.1046/j.1365-2869.2000.00196.x

Basner, M., Mollicone, D., and Dinges, D. F. (2011). Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronautica* 69, 949–959. doi: 10.1016/j.actaastro.2011.07.015

Bonnemeier, H., Wiegand, U. K., Brandes, A., Kluge, N., Katus, H. A., Richardt, G., et al. (2003). Circadian profile of cardiac autonomic nervous modulation in healthy subjects: differing effects of aging and gender on heart rate variability. *J. Cardiovasc. Electrophysiol.* 14, 791–799. doi: 10.1046/j.1540-8167.2003.03078.x

Bretherton, B., Deuchars, J., and Windsor, W. L. (2019). Effects of controlled tempo manipulations on cardiovascular function. *Music Sci.* 2, 1–14.

Brotons, M. (1994). Effects of performing conditions on music performance anxiety and performance quality. *J. Music Ther.* 31, 63–81. doi: 10.1093/jmt/31.1.63

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., and Kupfer, D. J. (1989). The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res.* 28, 193–213. doi: 10.1016/0165-1781(89)90047-4

Caramiaux, B., Bevilacqua, F., Wanderley, M. M., and Palmer, C. (2018). Dissociable effects of practice variability on learning motor and timing skills. *PLoS One* 13:e0193580. doi: 10.1371/journal.pone.0193580

Carney, C. E., Buysse, D. J., Ancoli-Israel, A., Edinger, J. D., Krystal, A. D., Lichstein, K. L., et al. (2012). The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep* 35, 287–302. doi: 10.5665/sleep.1642

Christie, M. J., and McBrearty, E. M. T. (1979). Psychophysiological investigations of post lunch state in male and female subjects. *Ergonomics* 22, 307–323. doi: 10.1080/00140137908924615

Czeisler, C. A., and Klerman, E. B. (1999). Circadian and sleep-dependent regulation of hormone release in humans. *Recent Prog. Hormone Res.* 54, 97–130.

da Silva, S. A. F., Guida, H. L., Dos Santos Antonio, A. M., de Abreu, L. C., Monteiro, C. B., and Ferreira, C. (2014). Acute auditory stimulation with different styles of music influences cardiac autonomic regulation in men. *Int. Cardiovasc. Res. J.* 8, 105–110.

De Manzano, Ö, Theorell, T., Harmat, L., and Ullén, F. (2010). The psychophysiology of flow during piano playing. *Emotion* 10, 301–311. doi: 10.1037/a0018432

Demos, A., Frank, T., and Chaffin, R. (2011). "Understanding movement during performance: a recurrence quantization approach," in *International Symposium on Performance Sciences*, eds A. Williamon, D. Edwards, and L. Bartel (Utrecht: AEC), 259–264.

Dijk, D. J., Duffy, J. F., and Czeisler, C. A. (1992). Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance. *J. Sleep Res.* 1, 112–117. doi: 10.1111/j.1365-2869.1992.tb00021.x

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., and Poeppel, D. (2017). Temporal modulations in speech and music. *Neurosci. Biobehav. Rev.* 81, 181–187. doi: 10.1016/j.neubiorev.2017.02.011

Dinges, D. F., and Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behav. Res. Methods* 17, 652–655. doi: 10.3758/bf03200977

Dosseville, F., Moussay, S., Larue, J., Gauthier, A., and Davenne, D. (2002). Physical exercise and time of day: influences on spontaneous motor tempo. *Percept. Mot. Skills* 95, 965–972. doi: 10.1177/003151250209500301

Drake, C., and Palmer, C. (2000). Skill acquisition in music performance: relations between planning and temporal control. *Cognition* 10, 1–32. doi: 10.1016/s0010-0277(99)00061-x

Edwards, B., Waterhouse, J., and Reilly, T. (2007). The effects of circadian rhythmicity and time-awake on a simple motor task. *Chronobiol. Int.* 24, 1109–1124. doi: 10.1080/07420520701795316

Finney, S. A. (2001). FTAP: a Linux-based program for tapping and music experiments. *Behav. Res. Methods Instr. Comput.* 33, 65–72. doi: 10.3758/bf03195348

Fogel, S., Vien, C., Karni, A., Benali, H., Carrier, J., and Doyon, J. (2017). Sleep spindles: a physiological marker of age-related changes in gray matter in brain regions supporting motor skill memory consolidation. *Neurobiol. Aging* 49, 154–164. doi: 10.1016/j.neurobiolaging.2016.10.009

Folstein, M. L., and Luria, R. (1973). Reliability, validity, and clinical application of the visual analogue mood scale. *Psychol. Med.* 3, 479–486. doi: 10.1017/s0033291700054283

Fraisse, P. (1982). "Rhythm and tempo," in *The Psychology of Music*, ed. D. Deutsch (San Diego: Academic Press), 149–1880.

Gjermunds, N., Brechan, I., Johnsen, S. ÅK., and Watten, R. G. (2019). Musicians: larks, owls or hummingbirds? *J. Circ. Rhythms* 17, 1–4.

Goebl, W., and Palmer, C. (2014). Temporal control and hand movement efficiency in skilled music performance. *PLoS One* 8:e50901. doi: 10.1371/journal.pone.0050901

Gomez, P., and Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion* 7, 377–387. doi: 10.1037/1528-3542.7.2.377

Goshvarpour, A., and Goshvarpour, A. (2012). Recurrence plots of heart rate signals during meditation. *Int. J. Image Graph. Signal Process.* 4, 44–50. doi: 10.5815/ijigsp.2012.02.07

Hoyt, D. F., and Taylor, C. R. (1981). Gait and the energetics of locomotion in horses. *Nature* 292, 239–240. doi: 10.1038/292239a0

Javorka, M., Trunkvalterova, Z., Tonhajzerova, I., Lazarova, Z., Javorkova, J., and Javorka, K. (2008). Recurrences in heart rate dynamics are changed in patients with diabetes mellitus. *Clin. Physiol. Funct. Imaging* 28, 326–331. doi: 10.1111/j.1475-097x.2008.00813.x

Javorka, M., Turianikova, Z., Tonhajzerova, I., Javorka, M., and Baumert, M. (2009). The effect of orthostasis on recurrence quantification analysis of heart rate and blood pressure dynamics. *Institute Phys. Eng. Med.* 30, 29–41. doi: 10.1088/0967-3334/30/1/003

Johns, M. W. (1991). A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 14, 540–545. doi: 10.1093/sleep/14.6.540

Jungers, M. K., Palmer, C., and Speer, S. (2003). Time after time: the coordinating influence of tempo in music and speech. *Cogn. Process.* 1, 21–35.

Kelso, J. A. S. (1997). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press.

Koelsch, S., and Jäncke, L. (2015). Music and the heart. *Eur. Heart J.* 36, 3043–3048.

Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjødt, U., Jegindø, E., Wallot, S., et al. (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *PNAS* 108, 8514–8519. doi: 10.1073/pnas.1016955108

Large, E. W. (1993). Dynamic programming for the analysis of serial behaviors. *Behav. Res. Methods Instr. Comput.* 25, 238–241. doi: 10.3758/bf03204504

Large, E. W., Herrera, M. J., and Velasco, J. A. (2015). Neural networks for beat perception in musical rhythm. *Front. Syst. Neurosci.* 9:159. doi: 10.3389/fnsys.2015.00159

LeBlanc, A., Jin, Y. C., Obert, M., and Siivola, C. (1997). Effect of audience on music performance anxiety. *J. Res. Music Educ.* 34, 480–496. doi: 10.2307/3345541

Loehr, J. D., and Palmer, C. (2011). Temporal coordination between performing musicians. *Q. J. Exp. Psychol.* 64, 2153–2167. doi: 10.1080/17470218.2011.603427

Martin-Gonzalez, S., Navarro-Mesa, J. L., Julia-Serda, G., Ramirez-Avila, G. M., and Ravelo-Garcia, A. G. (2018). Improving the understanding of sleep apnea characterization using Recurrence Quantification Analysis by defining overall acceptable values for the dimensionality of the system, the delay, and the distance threshold. *PLoS One* 13:e0194462. doi: 10.1371/journal.pone.0194462

Marwan, N. (2019). *Cross recurrence plot toolbox for Matlab version 7 release 32.4.*

Marwan, N., Zou, Y., Wessel, N., Riedl, M., and Kurths, J. (2013). Estimating coupling directions in the cardiorespiratory system using recurrence properties. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 371:20110624. doi: 10.1098/rsta.2011.0624

Monk, T. H. (1987). Subjective ratings of sleepiness- the underlying circadian mechanisms. *Sleep* 10, 343–353. doi: 10.1093/sleep/10.4.343

Monk, T. H. (2005). The post-lunch dip in performance. *Clin. Sports Med.* 24, 15–23.

Moussay, S., Dosseville, F., Gauthier, A., Larue, J., Sesboüe, B., and Davenne, D. (2002). Circadian rhythms during cycling exercise and finger-tapping task. *Chronobiol. Int.* 19, 1137–1149. doi: 10.1081/cbi-120015966

Nayak, S. K., Bit, A., Dey, A., Mohapatra, B., and Pal, K. (2018). A review on the nonlinear dynamical system analysis of electrocardiogram signal. *J. Healthc. Eng.* 2018, 1–19. doi: 10.1155/2018/6920420

Nessler, J. A., and Gilliland, S. J. (2009). Interpersonal synchronization during side by side treadmill walking is influenced by leg length differential and altered sensory feedback. *Hum. Mov. Sci.* 28, 772–785. doi: 10.1016/j.humov.2009.04.007

Palmer, C. (1989). Mapping musical thought to music performance. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 331–346. doi: 10.1037/0096-1523.15.2.331

Palmer, C (1996). Music performance. *Ann. Rev. Psychol.* 48, 115–138. doi: 10.1146/annurev.psych.48.1.115

Palmer, C., Spidle, F., Koopmans, E., and Schubert, P. (2019). Ears, heads and eyes: when singers synchronize. *Q. J. Exp. Psychol.* 72, 2272–2287. doi: 10.1177/1747021819833968

Pušnik, I., and Miklavec, A. (2009). Dilemmas in measurement of human body temperature. *Instrument. Sci. Technol.* 37, 516–530. doi: 10.1080/10739140903149061

Reilly, T., Atkinson, G., Edwards, B., Waterhouse, J., Farrelly, K., and Fairhurst, E. (2007). Diurnal variation in temperature, mental and physical performance, and tasks specifically related to football (soccer). *Chronobiol. Int.* 24, 507–519. doi: 10.1080/07420520701420709

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen pianists. *J. Acoust. Soc. Am.* 88, 622–641. doi: 10.1121/1.399766

Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* 12, 969–992. doi: 10.3758/BF03206433

Roenneberg, T., Daan, S., and Merrow, M. (2003a). The art of entrainment. *J. Biol. Rhythms* 18, 183–194. doi: 10.1177/0748730403018003001

Roenneberg, T., Kuehnle, T., Pramstaller, P. P., Ricken, J., Havel, M., Guth, A., et al. (2004). A marker for the end of adolescence. *Curr. Biol.* 14:R1038.

Roenneberg, T., Wirz-Justice, A., and Merrow, M. (2003b). Life between clocks: daily temporal patterns of human chronotypes. *J. Biol. Rhythms* 18, 80–90. doi: 10.1177/0748730402239679

Scheurich, R., Zamm, A., and Palmer, C. (2018). Tapping into rate flexibility: musical training facilitates synchronization around spontaneous production rates. *Front. Psychol.* 9:458. doi: 10.3389/fpsyg.2018.00458

Schlenker, J., Socha, V., Riedlbauchová, L., Nedelka, T., Schlenker, A., Potocková, V., et al. (2016). Recurrence plot of heart rate variability signal in patients with vasovagal syncopes. *Biomed. Signal Process. Control* 25, 1–11. doi: 10.1016/j.bspc.2015.10.007

Takaishi, T., Yasuda, Y., Ono, T., and Moritani, T. (1996). Optimal pedaling rate estimated from neuromuscular fatigue for cyclists. *Med. Sci. Sports Exerc.* 28, 1492–1497. doi: 10.1097/00005768-199612000-00008

Takens, F. (1981). "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, Vol. 898, eds D. Rand, and L. Young (Berlin: Springer), 366–381. doi: 10.1007/bfb0091924

Tamm, A. S., Lagerquist, O., Ley, A. L., and Collins, D. F. (2009). Chronotype influences diurnal variations in the excitability of the human motor cortex and the ability to generate torque during a maximum voluntary contraction. *J. Biol. Rhythms* 24, 211–224. doi: 10.1177/0748730409334135

Taylor, N. A., Tipton, M. J., and Kenny, G. P. (2014). Considerations for the measurement of core, skin and mean body temperatures. *J. Therm. Biol.* 46, 72–101. doi: 10.1016/j.jtherbio.2014.10.006

Van Dyck, E., Six, J., Soyer, E., Denys, M., Bardijn, I., and Leman, M. (2017). Adopting a music-to-heart rate alignment strategy to measure the impact of music and its tempo on human heart rate. *Music. Sci.* 21, 390–404. doi: 10.1177/1029864917700706

Van Vugt, F. T., Treutler, K., Altenmüller, E., and Jabusch, H.-C. (2013). The influence of chronotype on making music: circadian fluctuations in pianists' fine motor skills. *Front. Hum. Neurosci.* 7:347. doi: 10.3389/fnhum.2013.00347

Vandewalle, G., Middleton, B., Rajaratnam, S. M., Stone, B. M., Thorleifsdottir, B., Arendt, J., et al. (2007). Robust circadian rhythm in heart rate and its variability: influence of exogenous melatonin and photoperiod. *J. Sleep Res.* 16, 148–155. doi: 10.1111/j.1365-2869.2007.00581.x

Vellers, H. L., Irwin, C., and Lightfoot, J. T. (2015). Heart rate response of professional musicians when playing music. *Med. Probl. Perform. Art.* 30, 100–105. doi: 10.21091/mppa.2015.2017

Vitale, J., Caliguri, G., and Weydahl, A. (2015). Influence of chronotype on responses to a standardized, self-paced walking task in the morning vs afternoon: a pilot study. *Percept. Mot. Skills* 116, 1020–1028.

Waterhouse, J., Atkinson, G., Edwards, B., and Reilly, T. (2007). The role of a short post-lunch nap in improving cognitive, motor, and sprint performance in participants with partial sleep deprivation. *J. Sports Sci.* 25, 1557–1566. doi: 10.1080/02640410701244983

Webber, C. L., and Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* 76, 965–973. doi: 10.1152/jappl.1994.76.2.965

Webber, C. L. J., and Zbilut, J. P. (2005). "Recurrence quantification analysis of nonlinear dynamical systems," in *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, eds M. Riley, and G. van Order (Washington, DC: National Science Foundation), 26–94.

Wittmann, M., Dinich, J., Merrow, M., and Roenneberg, T. (2006). Social jetlag: misalignment of biological, and social time. *Chronobiol. Int.* 23, 497–509. doi: 10.1080/07420520500545979

Zamm, A., Pfordresher, P. Q., and Palmer, C. (2015). Temporal coordination in joint music performance: effects of endogenous rhythms and auditory feedback. *Exp. Brain Res.* 233, 607–615. doi: 10.1007/s00221-014-4140-5

Zamm, A., Wang, Y., and Palmer, C. (2018). Musicians' natural frequencies of performance display optimal temporal stability. *J. Biol. Rhythms* 33, 432–440. doi: 10.1177/0748730418783651

Zamm, A., Wellman, C., and Palmer, C. (2016). Endogenous rhythms influence interpersonal synchrony. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 611–616. doi: 10.1037/xhp0000201

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

REVIEW
published: 26 August 2020
doi: 10.3389/fnhum.2020.00350

# Links Between the Neurobiology of Oxytocin and Human Musicality

*Alan R. Harvey**

*School of Human Sciences, The University of Western Australia, Perron Institute for Neurological and Translational Science, Perth, WA, Australia*

OPEN ACCESS

**Edited by:**
*Franco Delogu,*
*Lawrence Technological University,*
*United States*

**Reviewed by:**
*Kai Lutz,*
*Center for Neurology and Rehabilitation (CERENEO), Switzerland*
*Madhavi Rangaswamy,*
*Christ University, India*

***Correspondence:**
*Alan R. Harvey*
*alan.harvey@uwa.edu.au*

**Specialty section:**
*This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience*

**Received:** *01 May 2020*
**Accepted:** *04 August 2020*
**Published:** *26 August 2020*

**Citation:**
*Harvey AR (2020) Links Between the Neurobiology of Oxytocin and Human Musicality. Front. Hum. Neurosci. 14:350. doi: 10.3389/fnhum.2020.00350*

The human species possesses two complementary, yet distinct, universal communication systems—language and music. Functional imaging studies have revealed that some core elements of these two systems are processed in closely related brain regions, but there are also clear differences in brain circuitry that likely underlie differences in functionality. Music affects many aspects of human behavior, especially in encouraging prosocial interactions and promoting trust and cooperation within groups of culturally compatible but not necessarily genetically related individuals. Music, presumably *via* its impact on the limbic system, is also rewarding and motivating, and music can facilitate aspects of learning and memory. In this review these special characteristics of music are considered in light of recent research on the neuroscience of the peptide oxytocin, a hormone that has both peripheral and central actions, that plays a role in many complex human behaviors, and whose expression has recently been reported to be affected by music-related activities. I will first briefly discuss what is currently known about the peptide's physiological actions on neurons and its interactions with other neuromodulator systems, then summarize recent advances in our knowledge of the distribution of oxytocin and its receptor (OXTR) in the human brain. Next, the complex links between oxytocin and various social behaviors in humans are considered. First, how endogenous oxytocin levels relate to individual personality traits, and then how exogenous, intranasal application of oxytocin affects behaviors such as trust, empathy, reciprocity, group conformity, anxiety, and overall social decision making under different environmental conditions. It is argued that many of these characteristics of oxytocin biology closely mirror the diverse effects that music has on human cognition and emotion, providing a link to the important role music has played throughout human evolutionary history and helping to explain why music remains a special prosocial human asset. Finally, it is suggested that there is a potential synergy in combining oxytocin- and music-based strategies to improve general health and aid in the treatment of various neurological dysfunctions.

**Keywords: oxytocin, music, dance, reward, empathy, trust, therapy**

## INTRODUCTION

The human species has evolved two universal systems of inter-personal communication, language, and music. These communication streams possess some common elements, for example, a requirement for processing certain aspects of pitch, rhythm, and syntax; however, there are also well-established differences in neural circuitry that are linked to differences in functionality. The possible evolutionary origin of musical behaviors in our species has been discussed elsewhere (e.g., Brown, 2000; Mithen, 2005; Fitch, 2006; Patel, 2008; Morley, 2013; Richter and Ostovar, 2016; Harvey, 2017) and is not considered in detail here. Language plays an essential role in cognition; it is the primary means by which modern humans communicate thoughts and ideas, it facilitates the sharing of learned information and knowledge within *and between* generations, it permits intuitive reasoning, foresight, and planning, and it likely co-evolved with our capacity to imagine times and places not personally experienced in our lifetime (Harvey, 2017). Language and the emergence and continued development of human culture seem to be closely intertwined, but then why do we also communicate and enjoy music and its partner dance? Why does music continue as a human universal and what is its significance to the species?

Music affects many aspects of human behavior, behaviors that may have had (and still have) adaptive benefits that presumably contribute to the ongoing existence of musicality in humans (e.g., Cross, 2009; Harvey, 2018). These benefits, which are by no means mutually exclusive, are thought to include the attraction and selection of mates, the facilitation of attachment between caregivers and preverbal infants, aiding the development of perceptual, cognitive and motor skills, and encouraging trust, social bonding, and mutual cooperation. In a group context music-related activities, including dance (Laland et al., 2016; Richter and Ostovar, 2016), encourage the formation of bigger social networks, help to define cultural identity, and may represent a "safe haven" in which individuals can interact and share experiences without revealing their innermost thoughts and fears. Evidence supporting the important role that music plays in promoting the development and maintenance of cooperative, prosocial behaviors comes from an increasing number of studies in children and in adults (Freeman, 2000; Kirschner and Tomasello, 2009; Tarr et al., 2014; Pearce et al., 2015; Schellenberg et al., 2015). Music, *via* its impact on various regions within the limbic system, is also rewarding, motivating, and facilitates aspects of learning and memory (Zatorre and Salimpoor, 2013; Koelsch, 2018). Lastly, and no less important, it is increasingly appreciated that musical activities are useful therapeutic tools, aiding in the treatment of some developmental disorders (Quintin, 2019), and capable of ameliorating behavioral and psychological symptoms in several neurodegenerative conditions (e.g., Abraha et al., 2017; Zhang et al., 2017; Särkämö and Sihvonen, 2018; Groussard et al., 2019; Pereira et al., 2019).

In this review article, these special characteristics of music are considered in light of recent research on the neurobiology of the peptide oxytocin. Oxytocin is a hormone, synthesized in the hypothalamus that has both peripheral and central actions. Peripherally, oxytocin has important roles before and after childbirth, acting on the uterus during labor and stimulating lactation. Centrally, oxytocinergic systems are thought to influence many complex human social behaviors including, for example, pair bonding, attachment and social memory, emotional empathy, trust and generosity, and suppression of anxiety. The first part of the review focuses on what is currently known about the physiological actions of oxytocin on cells in the mammalian central nervous system (CNS) and the peptide's interactions with other neuromodulator systems including the closely related pituitary hormone arginine vasopressin (AVP), the stress-related hormone cortisol, and neurotransmitters such as dopamine and serotonin. The second section summarizes recent advances in our knowledge of the distribution of the peptide and its receptor in the human brain, the relationship between endogenous oxytocin levels and complex behavioral traits typical of *Homo sapiens*, and then reviews the diverse effects of intranasal oxytocin administration on human behavior. The final section discusses links between music-related activities and oxytocin expression, documenting the similarities between the generally prosocial behaviors engendered by oxytocin and the many positive effects that music has on human cognition, memory, and mental health. Oxytocin and music can also have beneficial effects on cardiovascular and immune systems, and it is argued that a better understanding of the multiple actions of the oxytocinergic system may lead to its synergistic use with music in a range of therapeutic applications in psychology and neurology.

## THE NEUROSCIENCE OF OXYTOCIN—ANIMAL STUDIES

Oxytocin is a nine amino-acid peptide that is enzymatically derived from a larger peptide precursor made from the oxytocin gene. This peptide, or closely related versions of it, is involved in reproductive functions across almost all vertebrate species (Carter, 2014; Ebitz and Platt, 2014; Grinevich et al., 2016; Feldman, 2017; Jurek and Neumann, 2018) and its peripheral and central actions have been the subject of increasing interest in recent years (Jurek and Neumann, 2018)—as of 1st April 2020 there were more than 27,000 articles, including 3,700 reviews, listed on the NIH PubMed search engine.

In the mammalian brain, oxytocin is synthesized predominantly by magnocellular neurons in the supraoptic (SON) and paraventricular (PVN) nuclei of the hypothalamus, by some parvocellular neurons in PVN, and in accessory magnocellular nuclei of the hypothalamus. There is also some expression in peripheral tissues such as the gonads, kidney, and pancreas although the oxytocin generated there is unlikely to enter the CNS (Jirikowski, 2019). Many oxytocin-expressing neurons likely bifurcate, with a "traditional" endocrine-related projection to the posterior pituitary for systemic release into the bloodstream and a second central branch projecting to about 50 brain regions, including the sensory and prefrontal cortex, nucleus accumbens in the ventral striatum, amygdala, hippocampus, hypothalamus and ventral tegmentum (Grinevich and Stoop, 2018). In the circulation, oxytocin has a half-life of only a few minutes before it is metabolized in the liver and kidneys. This is an important point that will be returned to later in this review when discussing how best to measure and interpret peripheral oxytocin levels in humans.

In animal models, a variety of methods have been used to investigate the molecular and cellular properties of oxytocin and its receptor (OXTR), and to better understand the nature of the relationship between the physiology and pharmacology of oxytocin signaling and overt behavior (Jurek and Neumann, 2018; Mitre et al., 2018; Cilz et al., 2019; Neumann and Landgraf, 2019; Tan et al., 2019; Raam, 2020). These methods include neuroanatomical pathway tracing, immunohistochemistry, electron microscopy. receptor autoradiography, *in situ* hybridization, electrophysiology, microdialysis, and functional magnetic resonance imaging (fMRI). Experimental interventions have also been used to perturb the oxytocinergic system such as intracerebral infusion of the peptide, the use of receptor agonists or antagonists, antisense methods, optogenetic and chemogenetic stimulation of oxytocinergic neurons, conditional deletion of OXTR, and the use of genetically engineered reporter mice.

## The Oxytocin Receptor

Oxytocin binds with high affinity to its specific receptor OXTR and can initiate an array of intracellular signaling cascades and transcriptional events (Chatterjee et al., 2016; Busnelli and Chini, 2017; Jurek and Neumann, 2018). There is also significant crosstalk with structurally related AVP receptors, most particularly the AVP1a receptor (AVPR1a; Bakos et al., 2018; Grinevich and Stoop, 2018; Song and Albers, 2018). In turn, AVP can also bind to OXTR, however, the specificity of action largely remains, probably due to differences in the distribution of oxytocin vs. AVP containing axons (Grinevich and Stoop, 2018; Rogers et al., 2018; Song and Albers, 2018; Pekarek et al., 2020). Nonetheless, there are potential sites of interaction in some brain regions (Smith et al., 2019), and there is evidence of functionally relevant spillover of oxytocin into the extracellular space beyond traditional synaptic sites (Busnelli and Chini, 2017; Chini et al., 2017; Song and Albers, 2018). OXTRs are widely distributed and found in many neuronal types, expressed on cell bodies, dendrites, and axon terminals, and the receptor is also expressed by astrocytes (Wang et al., 2017; Bakos et al., 2018; Young and Song, 2020). In different species, the receptor seems to be specifically enriched in those sensory/perceptual systems that are most relevant to conspecific maternal as well as more general socially interactive behaviors (Grinevich and Stoop, 2018; Pekarek et al., 2020).

## The Physiology of Oxytocin

In the CNS, oxytocin can affect various ion channels, increase intracellular calcium ion concentrations, alter membrane excitability and enhance long-term potentiation (LTP) in neurons (Tomizawa et al., 2003; Lee et al., 2015; Lin and Hsu, 2018; Tirko et al., 2018). Oxytocin signaling also increases the expression of neurotrophic factors such as brain-derived neurotrophic factor (BDNF; Bakos et al., 2018; Zhang et al., 2020), of relevance to later discussion focussed on oxytocin, social learning/memory, and hippocampal function. The peptide can also act presynaptically to affect neurotransmitter secretion (Dölen et al., 2013; Bakos et al., 2018). Overall, from a physiological perspective, oxytocin influences cell viability, synaptic and structural plasticity in neurons (Bakos et al., 2018; Jurek and Neumann, 2018; Pekarek et al., 2020), and modulates the balance of excitatory and inhibitory activity in regions such as the cerebral cortex and hippocampus (e.g., Mitre et al., 2016; Grinevich and Stoop, 2018; Lin and Hsu, 2018; Lopatina et al., 2018; Tirko et al., 2018; Cilz et al., 2019; Maniezzi et al., 2019; Tan et al., 2019), amygdala (Crane et al., 2020), and nucleus accumbens (Moaddab et al., 2015; Cox et al., 2017). Rodents lacking oxytocin or OXTR display impaired sociability and social memory (Ferguson et al., 2000). and conditional deletion of OXTR in the hippocampus negatively affects LTP and impairs long-term social recognition memory (Lin et al., 2018).

Likely increasing its diversity of action, oxytocin also interacts with several other receptors and neuromodulatory systems. For example, the peptide: (i) potentiates excitatory dopamine-mediated synaptic transmission (Li et al., 2020); (ii) interacts with a class of serotonin receptor (Chruścicka et al., 2019) and affects serotonin release (Yoshida et al., 2009); (iii) modulates signaling mediated by opioid receptors (dal Monte et al., 2017; Meguro et al., 2018; Salighedar et al., 2019); and (iv) activates TRPV2 channels (Van den Burg et al., 2015). There are also dynamic interactions with steroids (Jirikowski et al., 2018) and oxytocin levels are negatively correlated with cortisol, significantly modifying responses to stress (Lee et al., 2015; Schladt et al., 2017; Latt et al., 2018; Masis-Calvo et al., 2018; Neumann and Landgraf, 2019).

The foregoing section has, of necessity, over-simplified the physiological effects of oxytocin on neural tissue in animals, and more in-depth reviews are available (e.g., Bakos et al., 2018; Jurek and Neumann, 2018; Mitre et al., 2018; Neumann and Landgraf, 2019). However, some discussion of animal-based research is warranted because, given that the peptide is highly conserved in evolution it is likely that similar wide-ranging molecular and cellular mechanisms are operative in the human brain (see also Grinevich and Neumann, 2020). From a behavioral perspective, animal studies reveal that oxytocin has an important role in pair-bonding and maternal attachment, in moderating affiliative behaviors and conspecific social recognition, and in modulating the formation and maintenance of episodic memories, whether they be positive or negative. The next section will show that oxytocin has generally similar effects on human social behavior, but these effects would seem to be more subtle and complex in cognitively advanced members of *Homo sapiens*, extending to personality traits, emotional empathy, trust, altruism, reciprocity, group conformity, social decision making and so on.

# OXYTOCIN IN HUMANS

## Oxytocinergic Networks

In humans, immunoreactive oxytocinergic fibers are sparse but present in all cortical layers of the orbitofrontal cortex and anterior cingulate (Rogers et al., 2018). The fibers were found to have large varicosities usually associated with *en passant* boutons—likely sites of oxytocin release into the surrounding neuropil (Busnelli and Chini, 2017; Chini et al., 2017; Song and Albers, 2018). Fibers immunoreactive for AVP were also seen in these cortical regions and in the insular and olfactory cortices. Using antibodies to the receptor, OXTR was first identified in parts of the amygdala, anterior cingulate cortex, hypothalamus, and preoptic area, olfactory nucleus, and some brainstem nuclei (Boccia et al., 2013). A more recent extensive survey of the oxytocin system analyzed the distribution of the gene encoding OXTR as well as the gene encoding the oxytocin prepropeptide and the gene encoding CD38, a transmembrane protein needed for oxytocin secretion (Quintana et al., 2019). OXTR gene expression was widespread throughout the brain, significantly higher in olfactory bulbs, but also higher in the caudate, putamen, pallidum, and hypothalamus; levels were also greater than average in the hippocampus, parahippocampal region, amygdala, parts of the temporal lobe and anterior cingulate cortex. Expression of the gene was "reproducible, regardless of individual differences, such as ethnicity and sex" (Quintana et al., 2019).

The pattern of expression was essentially similar for the CD38 gene, with significantly increased expression in caudate, putamen, pallidum, thalamus, and anterior cingulum. For both genes, expression was significantly lower in the cerebellum. Interestingly, there was co-expression with several genes involved in dopaminergic and muscarinic cholinergic signaling, suggesting potential pathway interactions perhaps similar to those suggested for the opioids (dal Monte et al., 2017). Co-expression with genes involved in the regulation of metabolism and appetite was also seen. According to Quintana et al. (2019), "the oxytocin pathway gene maps correspond with the processing of anticipatory, appetitive, and aversive cognitive states." Interaction with dopaminergic and cholinergic systems is likely to add to the broad impact of oxytocin on social behaviors, motivation, reward, desire, anxiety, and the processing of emotions.

## Receptor Polymorphisms and Behavior

In children, adolescents, and adults, genetic variants of the OXTR gene are linked to an individual's response to stress (Rodrigues et al., 2009) and altered prosocial/affiliative behaviors and empathy. The need for pleasant social company is increased after a stressful event, a need that varies depending on which alleles of OXTR are present (Sicorello et al., 2020). Anatomically, there are subtle changes in structure and inter-connectivity of hypothalamus and parts of the limbic system, and mutations have been implicated in a range of highly maladaptive, sometimes psychopathic traits (e.g., Israel et al., 2008; Tost et al., 2010; Dadds et al., 2014; Aspé-Sánchez et al., 2016; Feldman et al., 2016; Gedeon et al., 2019; Poore and Waldman, 2020). A recent neuroimaging study examining the effect of OXTR alleles on resting-state networks reported that receptor genotype affected connectivity between the right hippocampus, medial prefrontal cortex, dorsal anterior cingulate cortex, amygdala, basal ganglia and thalamus (Luo et al., 2020). The functional impact that alleles of OXTR have on social behavior is however complex and not always consistent across studies, and is affected by factors such as gender, age, upbringing, and culture (Tost et al., 2010; Feldman et al., 2016; Fujiwara et al., 2019; Plasencia et al., 2019; Poore and Waldman, 2020). Environmental epigenetic influences on the OXTR function that influence social interactions must also be considered (Chen et al., 2020), and as described earlier there may be differential interactions with other neuromodulatory systems such as AVP, the opioids, steroids, and various catecholamines.

## Measurement of Endogenous Oxytocin

As yet it has not proved possible to measure oxytocin levels in the living human brain, thus endogenous oxytocin measurements are obtained from either plasma, saliva, or urine. Interpretation of these peripheral measures of oxytocin is however difficult for several reasons (Ebstein et al., 2012; Leng and Ludwig, 2016; Mitre et al., 2016; Valstad et al., 2017; Jurek and Neumann, 2018). First, peripheral oxytocin levels are related to the release of the peptide from the posterior pituitary and do not necessarily reflect levels of the peptide within specific regions of the brain that contain neurons expressing OXTR. Second, even when undertaking peripheral measurements, compared to saliva there is, in animals at least, a more consistent relationship between blood plasma levels of oxytocin and levels of the peptide found in cerebrospinal fluid (Valstad et al., 2017). Third, as pointed out by Jurek and Neumann (2018): "basal plasma or brain oxytocin levels might strongly depend on individual events occurring within the last hour(s) before sampling (e.g., fear of hospital or laboratory, prior eating, rushing to the laboratory, or sex) or on the time of the day." And all is compounded by the fact that circulating levels of oxytocin are normally low, even exogenous peptide is rapidly eliminated 1–2 h after intranasal delivery, and to measure native (unbound) oxytocin levels requires sophisticated techniques for specificity and accuracy of analysis (Franke et al., 2019, 2020). Nonetheless, and given these caveats, some intriguing and important observations have come from peripheral endogenous oxytocin measurements in humans.

## Nurturing and Bonding

Plasma and/or salivary oxytocin levels rise postpartum when mothers interact and bond with their infants (Matthiesen et al., 2001; Feldman et al., 2007; Feldman, 2012; Gordon et al., 2010). These interactive behaviors include gaze, facial expression, vocalizing using the preverbal maternal-infant communication known as "motherese," affectionate touch, and so on (Kerr et al., 2019). The increase in serum oxytocin when mothers interact and bond with their own smiling/happy infants is higher in mothers rated as having "secure" attachment to their offspring (Strathearn et al., 2009) and with a sensitive temperament (Strathearn et al., 2012). In these mothers, fMRI revealed greater activity in the hypothalamus/pituitary region and in reward centers in the ventral striatum. The rise in oxytocin is, at least in part, related to the amount of maternal gaze directed towards the child (Kim et al., 2014). Intranasal oxytocin also increases a father's neural response to images of their young children, with increased activity in caudate, anterior cingulate, and visual cortex (Li et al., 2017a).

Lullabies are a universal way of soothing infants (Mehr et al., 2018), and it is thus of interest that vocalization by mothers increases levels of salivary oxytocin (and reduces cortisol) in their children, although admittedly these were older girls aged between 7 and 12 years old (Seltzer et al., 2010). In comparing maternal vs. paternal changes in endogenous oxytocin during early parent-infant bonding, both mothers and fathers showed increases but there were dimorphic differences that depended on the type of interaction (Gordon et al., 2010). The development of affiliative behaviors between caregiver and infant, linked especially to oxytocin, leads to plasticity and adaptations in both the parent and infant (Feldman, 2015). Remarkably, the basal level of oxytocin measured in the saliva, and certain polymorphisms in the OXTR gene, are transgenerationally associated with the type of parental care that is given, influencing affiliative and social behaviors across as many as three generations within a family (Fujiwara et al., 2019).

### Oxytocin in Adolescents and Adults

Endogenous plasma oxytocin concentrations vary with age and there are differences between males and females, young women having the highest and old men the lowest levels (Plasencia et al., 2019). Experimentally, levels of salivary oxytocin were higher when female subjects were confronted with a novel situation, associated with reduced stress and greater trust, compared with a later familiarization session (Tops et al., 2013). Higher levels of plasma oxytocin in women but not men were linked to indicators of relationship stress and attachment anxiety (Taylor et al., 2010; Weisman et al., 2013; Moons et al., 2014). In young males, lower urinary oxytocin (but not AVP) levels were linked to lower measures of empathy and trust, presumably associated with a greater propensity for aggressive behavior (Malik et al., 2012; Weisman et al., 2013; de Jong and Neumann, 2018; Berends et al., 2019), and in males social cognitive ability was correlated with plasma oxytocin concentrations (Deuse et al., 2019; Strauss et al., 2019).

### Effects of Exogenous Administration of Oxytocin

Exogenous delivery of oxytocin affects neural processing and has consistently been reported to influence a wide range of interactive human behaviors. These behaviors have been described in different ways and using different terminologies. They include pair bonding, attachment, and social learning/memory, social salience and emotional empathy, recognition and interpretation of emotions, behavioral synchrony, familiarization and within group co-operation, altruism, generosity and trust, reward sensitivity, calmness and reduction of stress, and amelioration of anxiety (anxiolytic effects; e.g., Kosfeld et al., 2005; Baumgartner et al., 2008; Ditzen et al., 2009; Strathearn et al., 2009; Hurlemann et al., 2010; De Dreu, 2012; Fischer-Shofty et al., 2012; Tops et al., 2013; Bethlehem et al., 2014; Preckel et al., 2014; Shamay-Tsoory and Abu-Akel, 2016; Feldman, 2017; Fineberg and Ross, 2017; Leppanen et al., 2017; Wang et al., 2017; Ellenbogen, 2018; Geng et al., 2018; Jurek and Neumann, 2018; Rilling et al., 2018; Alos-Ferrer and Farolfi, 2019; Liu et al., 2019; Tillman et al., 2019; Sicorello et al., 2020; Wu et al., 2020). Analyses of the impact of oxytocin on neural activity imaged in the brains of healthy subjects generally reflect this, with altered activity in interconnected structures associated with valence, salience, trust, prosocial behavior and mentalizing, including the amygdala, insula, nucleus accumbens, lateral septum, anterior cingulate, hippocampus, caudate, tempero-parietal cortex, dorsomedial and dorsolateral prefrontal cortex (e.g., Kirsch et al., 2005; Rilling and Sanfey, 2011; Bethlehem et al., 2013; Eckstein et al., 2017; Wang et al., 2017; Rilling et al., 2018; Kumar et al., 2020; Wu et al., 2020).

The great majority of studies emphasize positive, prosocial behavioral outcomes after exogenous oxytocin administration; however, it is important to emphasize that not all studies describe these effects (Keech et al., 2018; Tabak et al., 2019; Erdozain and Peñagarikano, 2020) and some antisocial outcomes have been reported, including increased competitive and aggressive tendencies, particularly in males (Fischer-Shofty et al., 2012; Alcorn et al., 2015; Ne'eman et al., 2016; de Jong and Neumann, 2018; Gedeon et al., 2019). Others have also reported differential effects of exogenous oxytocin on women compared to men (e.g., Rilling et al., 2012, 2018; Preckel et al., 2014; Feng et al., 2015; Chen et al., 2016; Bredewold and Veenema, 2018; Bartz et al., 2019; Xu et al., 2020).

Overall, variation in the effects of administered oxytocin is linked to: (i) gender differences; (ii) whether individuals possess intrinsic pro- or antisocial personality traits—sometimes related to polymorphisms in, or the methylation state of, OXTR; (iii) early environmental experience; and/or (iv) the social cues and psychological context when testing is undertaken (e.g., Guastella and MacLeod, 2012; Evans et al., 2014; Nishina et al., 2015; Chen et al., 2016; Feldman et al., 2016; Lambert et al., 2017; Aydogan et al., 2018; de Jong and Neumann, 2018; Wagner and Echterhoff, 2018; Fragkaki and Cima, 2019; Gedeon et al., 2019; Liu et al., 2019; Sicorello et al., 2020). As mentioned earlier, oxytocin and AVP can activate each other's receptors, although the differential distribution of fibers and receptors may limit crosstalk (Rogers et al., 2018; Song and Albers, 2018). There may be little interaction under normal conditions of endogenous release, but perhaps there is more crosstalk after intranasal application of higher concentrations of oxytocin which may contribute to some of the complexity of behavioral outcomes.

## THE LINKS BETWEEN OXYTOCIN AND MUSIC

Oxytocin is an ancient peptide, in mammals universally involved in reproductive biology, modulating social learning and affiliative behaviors, as well as modifying responses to adverse conditions (Ebitz and Platt, 2014; Feldman et al., 2016; de Jong and Neumann, 2018). In *Homo sapiens*, it is conceivable that the unique prosocial, harmonizing activities of music and dance incorporated, perhaps even required, elements of this pre-existing oxytocinergic network. Music encourages affiliative interactions in infancy and adulthood, aids in the development of perceptual, cognitive, and motor skills, promotes trust and reduces a sense of social vulnerability, is rewarding and motivating, and has a beneficial effect on aspects of learning and memory. Music and its evolutionary partner dance (Richter and Ostovar, 2016) also promote synchrony and social interaction, contribute to cultural identity, and encourage the formation of cooperative networks. Based on the experimental work described above, it should be apparent that many of these musical influences on human behavior are also characteristic of many of the psychological and sociological effects of oxytocin. These associations become even clearer when comparing the neural networks that are: (i) activated when listening to music perceived as being rewarding and pleasurable with; (ii) regions that process behaviors that involve social cooperation, empathy and altruism; and (iii) the distribution of oxytocinergic fibers and OXTR in the human brain (**Figure 1**).

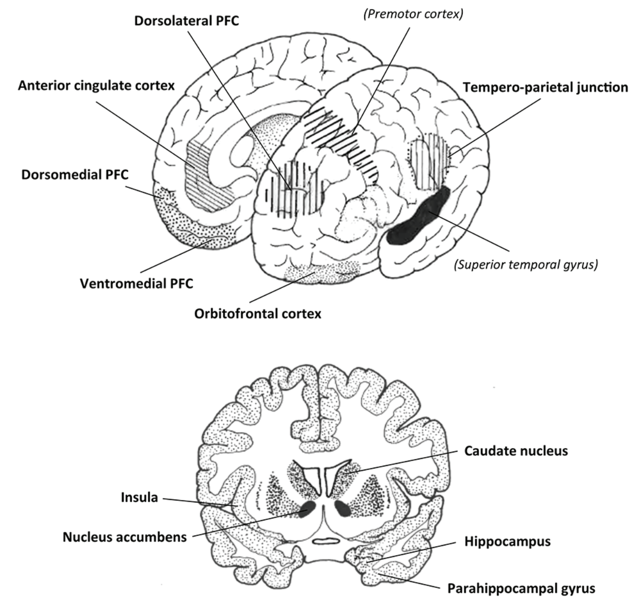



**FIGURE 1** | Two schematic views of the brain showing many of the regions in common that are reported to be involved in: (i) processing subjective/arousing/rewarding/emotional aspects of human musicality; (ii) performing or responding to a range of interactive social tasks; and (iii) oxytocinergic processing. Two other potentially relevant regions are shown in italics. Amygdala not shown. PFC, the prefrontal cortex. Both diagrams modified from Harvey (2017).

## Music Networks

As described in detail elsewhere (Harvey, 2017), and acknowledging that overlap in activity maps does not, *a priori*, mean that the same circuits are involved (Peretz et al., 2015), some of the basic elements of language and music such as pitch and rhythm share similar neural substrates, but clearer differences become apparent when more extended processing networks are considered. In right-handers at least, there is a left hemisphere bias for language and speech while the right hemisphere is biased more for music, and separate processing areas specific for these two communication streams have now been identified within secondary auditory regions in the superior temporal gyrus (Angulo-Perkins et al., 2014; Norman-Haignere et al., 2015). Most relevant to the present discussion are the limbic pathways and multiple cortical regions known to be activated by the overall subjective experience and emotional impact of music (Koelsch, 2018).

The limbic system, which includes the hippocampus, parahippocampal gyrus, amygdala, and cingulate cortex, is involved in several functions including learning, memory, motivation and emotional responsiveness. Music can induce activity in all these regions, while music that is perceived as arousing and is appreciated also drives dopaminergic activity in nucleus accumbens in the ventral striatum, an anticipatory and reward center (Blood and Zatorre, 2001; Menon and Levitin, 2005; Boso et al., 2006; Salimpoor et al., 2011; Zatorre and Salimpoor, 2013; Mueller et al., 2015; Ferreri et al., 2019; Gold et al., 2019; Shany et al., 2019). Music that evokes strong emotional valence is associated with altered activity not only in the superior temporal gyrus but also in the caudate nucleus, insula, thalamus, cingulate cortex, orbitofrontal, dorsomedial, dorsolateral and ventromedial prefrontal cortex, inferior frontal cortex and supplementary motor area (e.g., Blood and Zatorre, 2001; Koelsch et al., 2006; Mitterschiffthaler et al., 2007; Chapin et al., 2010; Brattico et al., 2011; Pereira et al., 2011; Khant et al., 2012; Altenmüller et al., 2014; Koelsch, 2018; Särkämö and Sihvonen, 2018; Sachs et al., 2019). Of course, music involves more than just listening, and imaging of people—alone or with others—creating and improvising jazz, rap or rock music has also revealed increased neural activity in the medial frontal lobe and altered, usually decreased, activity in the dorsolateral prefrontal cortex when compared to the same subjects playing or singing "formulaic sequences" (Limb and Braun, 2008; Liu et al., 2012; Donnay et al., 2014; Tachibana et al., 2019).

## Behavioral Networks

Numerous studies have used fMRI to image healthy subjects performing and responding to interactive social tasks that may involve altruism, empathy, trust and cooperation, norm-abiding behavior, mentalizing, and/or an appreciation of nuanced social context (**Figure 1**). These studies generally find increased functional activity in the nucleus accumbens, amygdala, parahippocampal gyrus, caudate nucleus, insula, anterior cingulate cortex, superior temporal cortex, tempero-parietal junction, and several regions in the prefrontal cortex (medial orbitofrontal, medial prefrontal, ventromedial, dorsolateral, dorsomedial; e.g., O'Doherty, 2004; Völlm et al., 2006; Rilling et al., 2008; Cooper et al., 2010; Rilling and Sanfey, 2011; Rushworth et al., 2011; Carter et al., 2012; Korn et al., 2012; Fukuda et al., 2019).

## Oxytocinergic Systems

Oxytocin fibers have been identified in the orbitofrontal cortex and anterior cingulate cortex (Rogers et al., 2018), and there are high levels of the receptor in the olfactory bulb, amygdala, hippocampus, parahippocampal gyrus, regions in the temporal lobe, anterior cingulate cortex, hypothalamus and preoptic area, and some brainstem nuclei (Boccia et al., 2013; Quintana et al., 2019). Intranasal oxytocin administration alters neural activity in structures such as the amygdala, insula, nucleus accumbens, anterior cingulate cortex, hippocampus, caudate, tempero-parietal cortex, dorsomedial and dorsolateral prefrontal cortex (e.g., Kirsch et al., 2005; Lischke et al., 2012; Bethlehem et al., 2013; Eckstein et al., 2017; Wang et al., 2017; Rilling et al., 2018; Kumar et al., 2020; Wu et al., 2020). Finally, altered OXTR genotypes have been found to correlate with altered local network metrics and functional connectivity between the hippocampus, medial prefrontal cortex, dorsal anterior cingulate cortex, amygdala, basal ganglia and thalamus (Luo et al., 2020).

Clearly then, musicality, cooperative prosocial interactions, and the oxytocinergic system are linked to neural activity in several common regions and interconnected networks in the brain of modern humans; the most consistently involved components being the hippocampus, parahippocampal gyrus, amygdala and anterior cingulate cortex, the caudate nucleus and nucleus accumbens, insula, superior temporal gyrus and orbitofrontal, ventromedial, dorsomedial and dorsolateral prefrontal cortex (**Figure 1**). Numerous examples of the close interrelationship between oxytocin and prosocial human behaviors have been presented, promoting group empathy and the participation in collective decision making, all involving a shift from personal concerns to more communal interests, including a willingness to learn from others (Zak and Berroza, 2013; Shalvi and De Dreu, 2014; De Dreu and Kret, 2016; De Wilde et al., 2017; Ten Velden et al., 2017; Schiller et al., 2020; Xu et al., 2020). But to what extent does oxytocin provide a nexus between these behaviors and music? What impact does music have on peripheral oxytocin release and OXTR expressing networks in the human brain, and can exogenous oxytocin administration synergistically affect performance and responsiveness to music?

## Experimental Studies on Music and Oxytocin

Only a few studies have directly examined the impact of music on oxytocin expression, in solo or ensemble settings. As described earlier, comforting maternal vocalizations—which can have music-like properties—by themselves have been shown to increase oxytocin levels and reduce cortisol in young daughters (Seltzer et al., 2010). It was recently reported that salivary oxytocin levels are reduced in maltreated children (Suzuki et al., 2020) and it is therefore of interest that, following a program of group drumming sessions for "emotionally disturbed" children, salivary oxytocin concentrations were increased in both boys and girls, significantly so when comparing practice and free play sessions performed by boys aged 8–12 years (Yuhi et al., 2017). In adults, salivary oxytocin levels were also found to be raised after a singing lesson, amateur singers, in particular, expressing a heightened sense of well-being (Grape et al., 2003), and raised levels were also reported after choral singing (Kreutz, 2014). In one sensory study, it was found that listening to slow relaxing music was associated with raised salivary oxytocin levels and lower heart rate, whereas fast music had little impact on oxytocin but reduced cortisol levels and increased arousal (Ooishi et al., 2017). The effect of relaxing music on moderating salivary cortisol levels after the stress has also been noted (Khalfa et al., 2003).

The nature of the musical activity is important because an increase in plasma oxytocin levels in members of a vocal jazz group was only recorded when singers were improvising together (Keeler et al., 2015), likely due to altered activity in the prefrontal cortex and enhanced affiliative and prosocial interactions (Limb and Braun, 2008; Liu et al., 2012; Donnay et al., 2014; Tachibana et al., 2019). Schladt et al. (2017) reported that salivary oxytocin levels slightly increased when subjects were solo singing but were decreased when singing in a choir. In that same study, cortisol levels were reduced in both situations, but choral participants described greater feelings of happiness and reduced worry. Of course, performing music can be stressful, perhaps especially in a solo compared to an ensemble/choral situation. Indeed, the intranasal application of oxytocin has recently been shown to increase positive interactions between performers and reduce performance anxiety (Sabino et al., 2020). The reported variability in measured levels of oxytocin and markers of stress such as cortisol reflects the complex, and highly interactive, sexually dimorphic systems that are involved (Brown et al., 2016). The other issue that should be borne in mind, discussed in detail earlier, is that measurement of salivary or urinary oxytocin levels do not necessarily reflect the concentration of the peptide in OXTR expressing regions in the brain (Leng and Ludwig, 2016; Jurek and Neumann, 2018), although there is a closer relationship between plasma oxytocin levels and those in cerebrospinal fluid (Valstad et al., 2017).

Overall, whilst there is a clear trend for increased endogenous oxytocin and reduced cortisol in subjects involved in musical activities, more controlled trials are needed in this area because communal music experiences are prime examples of human social engagement. From a physiological and psychosocial perspective, group music-making such as choral singing increases connectedness, heightens empathy, reduces depression and improves mood, is arousing and stimulates cognition, and has systemic health benefits including improved immune competency, reduced cytokine and inflammatory markers, lowered blood pressure and reduced cortisol and ACTH levels (Kuhn, 2002; Khalfa et al., 2003; Kreutz et al., 2004; Dunbar et al., 2012; Fancourt et al., 2014; Keeler et al., 2015; Pearce et al., 2015; Stewart and Lonsdale, 2016; Johnson et al., 2017; Ooishi et al., 2017; Finn and Fancourt, 2018; Kang et al., 2018; Moss et al., 2018; Perkins et al., 2018; Walker et al., 2019). The impact of exogenous oxytocin is relevant here because of the positive effect that it has on individual stress levels and the promotion of group empathy, reciprocal trust and collective social decision making, all involving a shift from personal to group agency (Zak and Berroza, 2013; Chen et al., 2016; De Dreu and Kret, 2016; De Wilde et al., 2017; Ten Velden et al., 2017; Sicorello et al., 2020; Xu et al., 2020). Music and the community associated with it may be especially important to individuals who are lonely and/or who have lower emotional empathy and exhibit fewer prosocial traits (e.g., Berends et al., 2019; Fragkaki and Cima, 2019; Liu et al., 2019; Johnson et al., 2020; Schiller et al., 2020).

## Links to Other Neuromodulatory Systems

Rhythm in music induces bodily movement and akin to music, dance is a universal human behavior (Levitin et al., 2018). From an evolutionary perspective, it has been argued that dance advantages humans "by contributing to sexual reproduction signaling, cooperation, social bonding, infant care, violence avoidance as well as embodied individual and social communication and memorization" (Richter and Ostovar, 2016). To my knowledge, there have not, to date, been any substantive reports on how oxytocin levels are affected by solo or group dance. Yet the impact of dance, especially in a coordinated group context, on increased empathy (Gujing et al., 2019) social bonding (Tarr et al., 2015, 2016), cognitive performance, general fitness and well-being (Kattenstroth et al., 2010, 2013; Zilidou et al., 2018; Douka et al., 2019) is clear. Choral singing and dance have both been reported to increase pain threshold, viewed as a surrogate for levels of circulating β-endorphin (Dunbar et al., 2012; Tarr et al., 2015; Weinstein et al., 2016). This peptide binds to μ-opioid receptors and plays a role in social networking and maintaining social bonds (Pearce et al., 2017). These observations are important, but it should be noted that many other factors can influence the perception and processing of pain (Millan, 2002), including oxytocin (Gamal-Eltrabily et al., 2020; Hilfiger et al., 2020; Schneider et al., 2020), which as described earlier positively modulates signaling mediated by opioid receptors (dal Monte et al., 2017; Meguro et al., 2018; Salighedar et al., 2019). Furthermore, depending on age and health status, the perception of pain does not necessarily reflect circulating β-endorphin levels (Bruehl et al., 2017; Ahn et al., 2019).

AVP is also thought to influence human behavior in many ways (Neumann and Landgraf, 2012; Benarroch, 2013) although significant effects are not always evident (Tabak et al., 2019). AVP and oxytocin may interact with each other to influence prosocial vs. antisocial behaviors, trust vs. aggression, fear and so on (Huber et al., 2005; Veenema and Neumann, 2018; Ebstein et al., 2012; Rilling et al., 2012; Jurek and Neumann, 2018; Song and Albers, 2018; Berends et al., 2019), at least some of which are sexually dimorphic (Rilling et al., 2014; Feng et al., 2015; Bredewold and Veenema, 2018). Comparison of endogenous AVP and oxytocin levels in plasma from young and old men and women revealed a negative correlation between all groups, higher AVP levels associated with greater "attachment anxiety" (Plasencia et al., 2019) and pair-bond distress in men (Taylor et al., 2010). Polymorphisms in the AVPR1a receptor have been linked to variability in aggression, response to stress, trust, and altruistic behaviors (Israel et al., 2008; Moons et al., 2014; Aspé-Sánchez et al., 2016; Nishina et al., 2019). Whilst there is as yet no evidence that endogenous levels of circulating AVP are altered by musical activity, AVPR1a receptor polymorphisms have been linked to musical aptitude (Pulli et al., 2008; Ukkola et al., 2009; Liu et al., 2016; Mariath et al., 2017), memory (Granot et al., 2007, 2013) and appreciation (Ukkola-Vuoti et al., 2011), as well as music and dance creativity (Bachner-Melman et al., 2005; Israel et al., 2008; Oikkonen et al., 2016). On the other hand, receptor polymorphisms were not more common in choral singers compared with people designated as non-musicians (Morley et al., 2012).

## Anxiety, Extinction, and PTSD

Participation in music is rewarding; it encourages prosocial interactions, facilitates social cognition, and promotes cooperation within groups of culturally compatible but not necessarily genetically related individuals. Ensemble music-making, and communal choral and dance activities, involve synchronized and coordinated activity with the special attribute of allowing individuals to be subsumed within a greater, living whole. Perhaps most importantly, the generally ambiguous, non-propositional nature of music provides a safe, usually risk-free space where individual thoughts and emotions, personal autobiographical memories and ambitions, can exist in a cooperative and interactive social context. Participation in musical activities can help individuals who lack self-confidence, who lack trust and may feel socially excluded, reduces fear and a sense of vulnerability, and can diminish potential conflict: "Music allows participants to explore the prospective consequences of their actions and attitudes toward others within a temporal framework that promotes the alignment of participants' sense of goals" (Cross, 2009).

This putative "safe haven" aspect of human musicality is similar to some of the behavioral effects elicited by oxytocin and further supports the proposed close links between music and oxytocinergic systems. Although not evident in all trials (Donadon et al., 2018), many studies have reported that exogenous delivery of oxytocin has anxiolytic and calming effects on human behavior (Neumann and Slattery, 2016; Wang et al., 2017; Lancaster et al., 2018; Yoon and Kim, 2020), enhancing the detection of threat (Lischke et al., 2012; Bredewold and Veenema, 2018) and facilitating the extinction of fearful or distressing memories (Kirsch et al., 2005; Hu et al., 2019; Koch et al., 2019; Triana-Del Río et al., 2019). Indeed, endogenous oxytocin levels are reduced in individuals with emotional trauma and in sufferers of posttraumatic stress disorder (PTSD; e.g., Frijling et al., 2015) and the administration of oxytocin may prove to be a useful therapeutic strategy (Giovanna et al., 2020). The acquisition and processing of autobiographical experiences, including fear and extinction, involves the hippocampus and ventromedial prefrontal cortex (Bonnici and Maguire, 2018; Dunsmoor et al., 2019) as well as interactions with the amygdala (Dunsmoor et al., 2019; Hasan et al., 2019). Activity in all these regions is associated with aspects of both musical and oxytocinergic processing. Concerning extinction, reducing the emotional impact of remembering fearful and threatening events involves substitution with novel, less impactful memories during the retrieval and reconsolidation process, a process facilitated by oxytocin (Hu et al., 2019; Triana-Del Río et al., 2019) and one that may also be aided by participation in the safe, neutral and motivating mental space evoked by communal music-related activities. In this context, music therapy has been suggested as a possible treatment for PTSD (Beck et al., 2018), and its use in association with oxytocin administration may prove even more beneficial.

Of the many endogenous opioids, β-endorphin—which may be raised by social music-making—has been implicated in resilience, stress, and PTSD (Bali et al., 2015) as has the neuropeptide nociceptin (Tollefson et al., 2017; Narendran et al., 2019). Nociceptin receptor polymorphisms have been linked to the severity of PTSD, but unlike β-endorphin, no relationship to music has been examined. Finally. there is also evidence of an important role for dopamine in the pathophysiology of PTSD (Lee et al., 2017; Torrisi et al., 2019) with potential interaction with oxytocinergic systems (Zhang et al., 2019), further strengthening the suggestion about the potential usefulness of music therapy given the known impact that music has on dopaminergic motivation and reward systems in the human brain (Chanda and Levitin, 2013; Zatorre and Salimpoor, 2013; Ferreri et al., 2019).

## Learning, Social Memory, and Hippocampal Plasticity

In children, some degree of music training has a significant impact on brain structure and plasticity as well as having a positive influence on social, empathic, cognitive and academic development (e.g., Schlaug et al., 2009; Kirschner and Tomasello, 2009; Schellenberg et al., 2015; Habibi et al., 2018; Sachs et al., 2018; de Manzano and Ullén, 2018; Guhn et al., 2020). Learning to play an instrument requires the recruitment of many sensorimotor systems and circuits, and many studies have reported that music training has beneficial effects on various executive functions and some types of memory, benefits that are maintained throughout a person's lifetime and may be protective against cognitive decline (Talamini et al., 2017; Mansens et al., 2018). Once again there are several intriguing and potentially important links between music training, music-related activities, and the neuroscience of oxytocin, in this case, the links that are relevant to memory and aging, with dance and exercise adding an additional dimension to the discussion. Oxytocin's effects on social recognition, learning, and memory are associated with activity in the hippocampus, amygdala, nucleus accumbens, and prefrontal cortex (e.g., Ferguson et al., 2002; Hurlemann et al., 2010; Mitre et al., 2016; Grinevich and Stoop, 2018; Jurek and Neumann, 2018; Lin and Hsu, 2018; Lin et al., 2018; Lopatina et al., 2018; Cilz et al., 2019; Tan et al., 2019; Raam, 2020; Xu et al., 2020). In the following discussion, the focus is primarily on the hippocampus, given its role in consolidating, integrating and retrieving personal autobiographical memories (Bonnici and Maguire, 2018; Sheldon et al., 2019). There is a huge literature on hippocampal connectivity and plasticity related to these dynamic and transformational processes—the emphasis here will be limited to several aspects of social learning and memory perhaps most relevant to a review of music and oxytocin.

Music activates diverse regions and circuits within the CNS including, depending on context and emotional valence, essentially the same limbic structures that are responsive to oxytocin (Boso et al., 2006; Koelsch, 2014, 2018). Music training and practice improves memory (Talamini et al., 2017; Mansens et al., 2018) and affects the architecture and organization of both gray and white matter in the brain (de Manzano and Ullén, 2018). Of particular relevance here is the positive effect that music training has on gray matter volume and plasticity in the hippocampus (Herdener et al., 2010), and whether this may be in some way related to increased endogenous oxytocin and reduced cortisol levels in individuals involved in musical activities—by what mechanisms could music, memory and oxytocin be linked? Acting through its receptor, oxytocin can act both pre-and postsynaptically to enhance LTP, alter the balance of excitatory and inhibitory activity, and modulate synaptic plasticity (Tomizawa et al., 2003; Lee et al., 2015; Bakos et al., 2018; Lin and Hsu, 2018; Tirko et al., 2018), These are all critical elements during the process of learning, socialization and memory consolidation (Ferguson et al., 2000; Lin et al., 2018), and intranasal application of the peptide at low doses is known to enhance social memory in human subjects (Jurek and Neumann, 2018).

## Neurogenesis

The hippocampal dentate gyrus appears to be one of the few sites in the adult mammalian CNS where new neurons are born throughout life (neurogenesis). This ongoing process is thought to be important in learning and in facilitating the addition of new memories onto similar previous experiences and knowledge, minimizing overlap in the resultant patterns of activity so that particular events can be discriminated from each other (e.g., Conçalves et al., 2016; França et al., 2017; Alam et al., 2018; Toda and Gage, 2018; Licht et al., 2020). In animals, experimental disruption of neurogenesis impairs social memory and coping with stress (Clelland et al., 2009; Garrett et al., 2015; Alam et al., 2018). Social interactions enhance new neuronal birth (Hsiao et al., 2014) whereas social isolation and stress-related changes that include increased cortisol levels lead to a reduction in neurogenesis (McEwen, 1999; Cinini et al., 2014; Opendak et al., 2016; Snyder and Drew, 2020), negatively affecting cognition, learning, and memory (Ouanes and Popp, 2019).

Oxytocin protects the hippocampus from stress-related effects including the negative impact of corticosterone treatment and directly induces neurogenesis in the adult rodent dentate gyrus (Lee et al., 2015; Sánchez-Vidaña et al., 2016; Lin et al., 2017; Lin and Hsu, 2018). The peptide also promotes the differentiation and dendritic maturation of these new neurons with associated effects on social behavior (Sánchez-Vidaña et al., 2016). This influence of oxytocin on hippocampal neurogenesis and social learning is indirectly enhanced by the peptide's actions in increasing BDNF expression (Dayi et al., 2015; Havranek et al., 2015; Zhang et al., 2020). This neurotrophin plays a key role in hippocampal plasticity and neurogenesis (Miranda et al., 2019). Its expression in the hippocampus is negatively affected by stress (Bennett and Lagopoulos, 2014; Dayi et al., 2015) but is significantly increased by physical exercise (Ding et al., 2011). In animals, increased BDNF levels are correlated with increased neurogenesis, the greater the amount of exercise the greater the proliferation of new neurons (reviewed in Liu and Nusslock, 2018). Indeed, it was recently shown that intense physical activity releases breakdown products from the muscle that act on promoters to increase BDNF gene expression and protein (Sleiman et al., 2016; Stephan and Sleiman, 2019).

There remains some controversy as to whether new neurons are born and survive within the adult human dentate gyrus (Sorrells et al., 2018; Duque and Spector, 2019); however, the weight of evidence and opinion is that neurogenesis and neuronal turnover does occur (Spalding et al., 2013; Boldrini et al., 2018; Kempermann et al., 2018; Kuhn et al., 2018; Cope and Gould, 2019; Horgusluoglu-Moloch et al., 2019; Lima and Gomes-Leal, 2019; Petrik and Encinas, 2019; Tobin et al., 2019; Lucassen et al., 2020), although estimates of the number of neurons born each day vary, and numbers may decline with age and disease (Moreno-Jiménez et al., 2019; Snyder, 2019). It is however clear that exercise and cardiovascular fitness are correlated with increased hippocampal volume and improved cognitive function (Erickson et al., 2011; Stillman et al., 2016). Furthermore, while it is not yet known if such changes are associated with enhanced neurogenesis, hippocampal size in humans is correlated with plasma BDNF levels (Erickson et al., 2011).

It is well established that neural activity in the hippocampus and other parts of the limbic system is altered by listening to music. Given this, and what is known about the effects of music on hormones such as oxytocin and cortisol, it will be of interest to determine if participation in musical activities influences human hippocampal neurogenesis (Fukui and Toyoshima, 2008), and how this might relate to the known beneficial effects of music on memory and cognition. Such activities should include movement and dance which are entrained within the diverse neural networks responsive to music (e.g., Brown et al., 2006; Phillips-Silver and Trainor, 2007; Nozaradan et al., 2011). Dance not only increases cooperation and group synchrony (Reddish et al., 2013; Karpati et al., 2016; Chauvigné et al., 2019) but improves fitness in the elderly (Douka et al., 2019). Measurement of oxytocin, cortisol, and BDNF in dancers seems warranted, and it may well be that, in addition to potential oxytocin-mediated effects, exercise and cardiovascular fitness associated with dancing are capable of adding an important extra dimension to the social, physical and mental health benefits of music appreciation and music-related activities, perhaps especially in the elderly.

### Systemic Effects—Further Links Between Oxytocin and Musicality

In addition to its physiological effects on CNS function, oxytocin has been reported to have even broader health benefits. The peptide decreases the progression of atherosclerosis and protects against cardiovascular disease (Reiss et al., 2019; Wang et al., 2019; Buemann and Uvnäs-Moberg, 2020), in association with social engagement (Ulmer-Yaniv et al., 2016; Walker et al., 2019) it has beneficial effects on the immune system, and it lowers cytokine levels and inhibits inflammation (Li et al., 2017b; Reiss et al., 2019). Oxytocin has also been reported to regulate appetite and food intake (Lawson et al., 2019; Onaka and Takayanagi, 2019; Quintana et al., 2019). Associated with these multiple beneficial effects, the peptide has been found to lower blood pressure and assist in maintaining glucose homeostasis, potentially useful as a therapeutic tool in the treatment of type 2 diabetes and obesity (Reiss et al., 2019). Again, many of these systemic oxytocinergic effects overlap those that can be elicited by listening to and/or performing music, including a reduction in blood pressure, the modification of immune responses and inflammatory markers, reduction of anxiety and stress and a moderating effect on blood glucose levels (Koelsch and Jäncke, 2015; Finn and Fancourt, 2018; Kang et al., 2018). The potentially additional benefits of music-related exercises such as dance have already been alluded to in the preceding paragraphs.

## CONCLUSION AND THERAPEUTIC IMPLICATIONS

Given what is increasingly becoming known about the neurological and systemic effects of oxytocin, it is important to analyze further how music and dance influence this peptide and its downstream pathways, and the extent to which social musical activities drive some of the interactions between oxytocin and other neuromodulatory systems such as dopamine, BDNF, and the various endogenous opioids. From an evolutionary perspective, it may clarify the extent to which evolving musical capabilities in modern humans took advantage of the ancient oxytocinergic network to facilitate prosocial interactions, promote trust and reciprocal affiliative behaviors, and help reduce levels of anxiety and individual insecurity throughout life. It will also contribute to a better understanding of the mechanisms that underlie the mental and general health benefits of music, its remarkable emotional and mnemonic power, it's capacity to alter brain architecture, and its ability to revitalize episodic memories (Zhang et al., 2017; Särkämö and Sihvonen, 2018), especially vulnerable in early stages of Alzheimer's disease (Groussard et al., 2019; Slattery et al., 2019).

From a therapeutic perspective, dancing is already a useful tool in the treatment of Parkinson's disease (Pereira et al., 2019); perhaps for some dementia sufferers, in addition to singing, dancing to favorite tunes may be even more beneficial when promoting social interactions with partners and unlocking autobiographical memories. Similarly, the use of intranasal oxytocin as a therapeutic tool in conditions such as PTSD shows promise (Giovanna et al., 2020) and may be even more effective when used with other treatments including combination with appropriate prosocial music-related activities. As another example, the combined use of music and the anti-nociceptive properties of oxytocin may enhance the therapeutic efficacy of strategies aimed at reducing the perception of chronic (Garza-Villarreal et al., 2017; Hilfiger et al., 2020; Schneider et al., 2020) or peri-operative pain (Nilsson et al., 2005, 2009; Nilsson, 2009; Van der Heijden et al., 2015). Last but not least, the use of oxytocin in the treatment of autism spectrum disorders (Yamasue and Domes, 2018) and other psychiatric conditions (Peled-Avron et al., 2020) may benefit from synergistic application with appropriate music-related therapeutic strategies (Quintin, 2019).

## AUTHOR'S NOTE

During revision of this manuscript, a review was published that emphasized the importance of studying the biology of oxytocin systems in animals to better understand the translational potential of this peptide in psychiatry and mental health (Grinevich and Neumann, 2020). The reader is encouraged to access this review to complement the more focussed emphasis on human musicality described herein.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

## REFERENCES


Abraha, I., Rimland, J. M., Trotta, F. M., Dell'Aquila, G., Cruz-Jentoft, A., Petrovic, M., et al. (2017). Systematic review of systematic reviews of non-pharmacological interventions to treat behavioural disturbances in older patients with dementia. The SENATOR-OnTop series. *BMJ Open* 7:e012759. doi: 10.1136/bmjopen-2016-012759

Ahn, H., La, J. H., Chung, J. M., Miao, H., Zhong, C., Kim, M., et al. (2019). The relationship between β-endorphin and experimental pain sensitivity in older adults with knee osteoarthritis. *Biol. Res. Nurs.* 21, 400–406. doi: 10.1177/1099800419853633

Alam, M. J., Kitamura, T., Saitoh, Y., Ohkawa, N., Kondo, T., and Inokuchi, K. (2018). Adult neurogenesis conserves hippocampal memory capacity. *J. Neurosci.* 38, 6854–6863. doi: 10.1523/jneurosci.2976-17.2018

Alcorn, J. L. III., Green, C. E., Schmitz, J., and Lane, S. D. (2015). Effects of oxytocin on aggressive responding in healthy adult men. *Behav. Pharmacol.* 8, 798–804. doi: 10.1097/fbp.0000000000000173

Alos-Ferrer, C., and Farolfi, F. (2019). Trust games and beyond. *Front. Neurosci.* 13:887. doi: 10.3389/fnins.2019.00887

Altenmüller, E., Siggel, S., Mohammadi, B., Samii, A., and Münte, T. F. (2014). Play it again, Sam: brain correlates of emotional music recognition. *Front. Psychol.* 5:114. doi: 10.3389/fpsyg.2014.00114

Angulo-Perkins, A., Aubé, W., Peretz, I., Barios, F. A., Armony, J. L., and Concha, L. (2014). Music listening engages specific cortical regions within the temporal lobes: differences between musicians and non-musicians. *Cortex* 59, 126–137. doi: 10.1016/j.cortex.2014.07.013

Aspé-Sánchez, M., Moreno, M., Rivera, M. I., Rossi, A., and Ewer, J. (2016). Oxytocin and vasopressin receptor gene polymorphisms: role in social and psychiatric traits. *Front. Neurosci.* 9:510. doi: 10.3389/fnins.2015.00510

Aydogan, G., Jobst, A., Loy, F., Dehning, S., Zill, P., Müller, N., et al. (2018). The effect of oxytocin on group formation and strategic thinking in men. *Horm. Behav.* 100, 100–106. doi: 10.1016/j.yhbeh.2018.02.003

Bachner-Melman, R., Dina, C., Zohar, A. H., Constantini, N., Lerer, E., Hoch, S., et al. (2005). AVPR1a and SLC6A4 gene polymorphisms are associated with creative dance performance. *PLoS Genet.* 1:e42. doi: 10.1371/journal.pgen.0010042

Bakos, J., Srancikova, A., Havranek, T., and Bacova, Z. (2018). Molecular mechanisms of oxytocin signaling at the synaptic connection. *Neural Plast.* 2018:4864107. doi: 10.1155/2018/4864107

Bali, A., Randhawa, P. K., and Jaggi, A. S. (2015). Stress and opioids: role of opioids in modulating stress-related behavior and effect of stress on morphine conditioned place preference. *Neurosci. Biobehav. Rev.* 51, 138–150. doi: 10.1016/j.neubiorev.2014.12.018

Bartz, J. A., Nitschke, J. P., Krol, S. A., and Tellier, P. P. (2019). Oxytocin selectively improves empathic accuracy: a replication in men and novel insights in women. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 4, 1042–1048. doi: 10.1016/j.bpsc.2019.01.014

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and adaptation in humans. *Neuron* 58, 639–650. doi: 10.1016/j.neuron.2008.04.009

Beck, B. D., Lund, S. T., Søgaard, U., Simonsen, E., Tellier, T. C., Cordtz, T. O., et al. (2018). Music therapy versus treatment as usual for refugees diagnosed with posttraumatic stress disorder (PTSD): study protocol for a randomized controlled trial. *Trials* 19:301. doi: 10.1186/s13063-018-2662-z

Benarroch, E. E. (2013). Oxytocin and vasopressin: social neuropeptides with complex modulatory functions. *Neurology* 80, 1521–1528. doi: 10.1212/wnl.0b013e31828cfb15

Bennett, M. R., and Lagopoulos, J. (2014). Stress and trauma: BDNF control of dendritic-spine formation and regression. *Prog. Neurobiol.* 112, 80–99. doi: 10.1016/j.pneurobio.2013.10.005

Berends, Y. R., Tulen, J. H. M., Wierdsma, A. I., van Pelt, J., Kushner, S. A., and van Marle, H. J. C. (2019). Oxytocin, vasopressin and trust: associations with aggressive behavior in healthy young males. *Physiol. Behav.* 204, 180–185. doi: 10.1016/j.physbeh.2019.02.027

Bethlehem, R. A., Baron-Cohen, S., Van Honk, J., Auyeung, B., and Bos, P. A. (2014). The oxytocin paradox. *Front. Behav. Neurosci.* 8:48. doi: 10.3389/fnbeh.2014.00048

Bethlehem, R. A., van Honk, J., Auyeung, B., and Baron-Cohen, S. (2013). Oxytocin, brain physiology and functional connectivity: a review of intranasal oxytocin fMRI studies. *Psychoneuroendocrinology* 38, 962–974. doi: 10.1016/j.psyneuen.2012.10.011

Blood, A. J., and Zatorre, R. L. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. U S A* 98, 11818–11823. doi: 10.1073/pnas.191355898

Boccia, M. L., Petrusz, P., Suzuki, K., Marson, L., and Pederson, C. A. (2013). Immunohistochemical localization of oxytocin receptors in human brain. *Neuroscience* 253, 155–164. doi: 10.1016/j.neuroscience.2013.08.048

Boldrini, M., Fulmore, C. A., Tartt, A. N., Simeon, L. R., Pavlova, I., Poposka, V., et al. (2018). Human hippocampal neurogenesis persists throughout aging. *Cell Stem Cell* 22, 589.e5–599.e5. doi: 10.1016/j.stem.2018.03.015

Bonnici, H. M., and Maguire, E. A. (2018). Two years later - revisiting autobiographical memory representations in vmPFC and hippocampus. *Neuropsychologia* 110, 159–169. doi: 10.1016/j.neuropsychologia.2017.05.014

Boso, M., Politi, P., Barale, F., and Enzo, E. (2006). Neurophysiology and neurobiology of the musical experience. *Funct. Neurol.* 21, 187–191.

Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., et al. (2011). A functional MRI study of happy and sad emotions in music with and without lyrics. *Front. Psychol.* 2:308. doi: 10.3389/fpsyg.2011.00308

Bredewold, R., and Veenema, R. H. (2018). Sex differences in the regulation of social and anxiety-related behaviors: insights from vasopressin and oxytocin brain systems. *Curr. Opin. Neurobiol.* 49, 132–140. doi: 10.1016/j.conb.2018.02.011

Brown, S. (2000). "The 'musilanguage' model of human evolution," in *Origins of Music*, eds N. Wallin, B. Merker and S. Brown (Cambridge, MA: MIT Press), 271–300.

Brown, C. A., Cardoso, C., and Ellenbogen, M. A. (2016). A meta-analytic review of the correlation between peripheral oxytocin and cortisol concentrations. *Front. Neuroendocrinol.* 43, 19–27. doi: 10.1016/j.yfrne.2016.11.001

Brown, S., Martinez, M. J., and Parsons, L. M. (2006). The neural basis of dance. *Cereb. Cortex* 16, 1157–1167. doi: 10.1093/cercor/bhj057

Bruehl, S., Burns, J. W., Gupta, R., Buvanendran, A., Chont, M., Orlowska, D., et al. (2017). Do resting plasma β-endorphin levels predict responses to opioid analgesics? *Clin. J. Pain.* 33, 12–20. doi: 10.1097/AJP.0000000000000389

Buemann, B., and Uvnäs-Moberg, K. (2020). Oxytocin may have a therapeutical potential against cardiovascular disease. Possible pharmaceutical and behavioral approaches. *Med. Hypoth.* 138:109597. doi: 10.1016/j.mehy.2020.109597

Busnelli, M., and Chini, B. (2017). "Molecular basis of oxytocin receptor signalling in the brain: what we know and what we need to know," in *Behavioral Pharmacology of Neuropeptides: Oxytocin. Current Topics in Behavioral Neurosciences.* (Vol. 35), eds R. Hurlemann and V. Grinevich (Cham: Springer). doi: 10.1007/7854_2017_6

Carter, C. S. (2014). Oxytocin pathways and the evolution of human behavior. *Annu. Rev. Psychol.* 65, 17–39. doi: 10.1146/annurev-psych-010213-115110

Carter, R. M., Bowling, D. L., Reeck, C., and Huettel, S. A. (2012). A distinct role of the temporal-parietal junction in predicting socially guided behaviour. *Science* 337, 109–111. doi: 10.1126/science.1219681

Chanda, M. L., and Levitin, D. J. (2013). The neurochemistry of music. *Trends Cogn. Sci.* 17, 179–193. doi: 10.1016/j.tics.2013.02.007

Chapin, H., Jantzen, K., Kelso, J. A., Steinberg, F., and Large, E. (2010). Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLoS One* 5:e13812. doi: 10.1371/journal.pone.0013812

Chatterjee, O., Patil, K., Sahu, A., Gopalakrishnan, L., Mol, P., Advani, J., et al. (2016). An overview of the oxytocin-oxytocin receptor signaling network. *J. Cell Commun. Signal.* 10, 355–360. doi: 10.1007/s12079-016-0353-7

Chauvigné, L. A. S., Walton, A., Richardson, M. J., and Brown, S. (2019). Multi-person and multisensory synchronization during group dancing. *Hum. Mov. Sci.* 63, 199–208. doi: 10.1016/j.humov.2018.12.005

Chen, X., Hackett, P. D., DeMarco, A. C., Feng, C., Stair, S., Haroon, E., et al. (2016). Effects of oxytocin and vasopressin on the neural response to unreciprocated cooperation within brain regions involved in stress and anxiety in men and women. *Brain Imaging Behav.* 10, 581–593. doi: 10.1007/s11682-015-9411-7

Chen, X., Nishitani, S., Haroon, E., Smith, A. K., and Rilling, J. K. (2020). OXTR methylation modulates exogenous oxytocin effects on human brain activity during social interaction. *Genes Brain Behav.* 19:e12555. doi: 10.1111/gbb.12555

Chini, B., Verhage, M., and Grinevich, V. (2017). The action radius of oxytocin release in the mammalian CNS: from single vesicles to behavior. *Trends Pharmacol. Sci.* 38, 982–991. doi: 10.1016/j.tips.2017.08.005

Chruścicka, B., Wallace Fitzsimons, S. E., Borroto-Escuela, D. O., Druelle, C., Stamou, P., Nally, K., et al. (2019). Attenuation of oxytocin and serotonin 2A receptor signaling through novel heteroreceptor formation. *ACS Chem. Neurosci.* 10, 3225–3240. doi: 10.1021/acschemneuro.8b00665

Cilz, N. I., Cymerblit-Sabba, A., and Young, W. S. (2019). Oxytocin and vasopressin in the rodent hippocampus. *Genes Brain Behav.* 18:e12535. doi: 10.1111/gbb.12535

Cinini, S. M., Barnabe, G. F., Galváo-Coelho, N., de Medeiros, M. A., Perez-Mendes, P., Sousa, M. B. C., et al. (2014). Social isolation disrupts hippocampal neurogenesis in young non-human primates. *Front. Neurosci.* 8:45. doi: 10.3389/fnins.2014.00045

Clelland, C. D., Choi, M., Romberg, C., Clemenson, G. D. Jr., Fragniere, A., Tyers, P., et al. (2009). A functional role for adult hippocampal neurogenesis in spatial pattern separation. *Science* 325, 210–213. doi: 10.1126/science.1173215

Conçalves, J. T., Schafer, S. T., and Gage, F. H. (2016). Adult neurogenesis in the hippocampus: from stem cells to behavior. *Cell* 167, 897–914. doi: 10.1016/j.cell.2016.10.021

Cooper, J. C., Kreps, T. A., Wiebe, T., Pirkl, T., and Knutson, B. (2010). When giving is good: ventromedial prefrontal cortex activation for others' intentions. *Neuron* 67, 511–521. doi: 10.1016/j.neuron.2010.06.030

Cope, E. C., and Gould, E. (2019). Adult neurogenesis, glia and the extracellular matrix. *Cell Stem Cell* 24, 690–705. doi: 10.1016/j.stem.2019.03.023

Cox, B. M., Bentzley, B. S., Regen-Tuero, H., See, R. E., Reichel, C. M., and Aston-Jones, G. (2017). Oxytocin acts in nucleus accumbens to attenuate methamphetamine seeking and demand. *Biol. Psych.* 81, 949–958. doi: 10.1016/j.biopsych.2016.11.011

Crane, J. W., Holmes, N. M., Fam, J., Westbrook, R. F., and Delaney, A. J. (2020). Oxytocin increases inhibitory synaptic transmission and blocks the development of long-term potentiation in the lateral amygdala. *J. Neurophysiol.* 123, 587–599. doi: 10.1152/jn.00571.2019

Cross, I. (2009). The evolutionary nature of musical meaning. *Music Sci.* 13, 179–200. doi: 10.1177/1029864909013002091

Dadds, M. R., Moul, C., Cauchi, A., Dobson-Stone, C., Hawes, D. J., Brennan, J., et al. (2014). Polymorphisms in the oxytocin receptor gene are associated with the development of psychopathy. *Dev. Psychopathol.* 26, 21–31. doi: 10.1017/s0954579413000485

dal Monte, O., Piva, M., Anderson, K. M., Tringides, M., Holmes, A. J., and Chang, S. W. C. (2017). Oxytocin under opioid antagonism leads to supralinear enhancement of social attention. *Proc. Natl. Acad. Sci. U S A* 114, 5247–5252. doi: 10.1073/pnas.1702725114

Dayi, A., Cetin, F., Sisman, A. R., Aksu, I., Tas, A., Gönenc, S., et al. (2015). The effects of oxytocin on cognitive defect caused by chronic restraint stress applied to adolescent rats and on hippocampal VEGF and BDNF levels. *Med. Sci. Monit.* 21, 69–75. doi: 10.12659/msm.893159

De Dreu, C. K. (2012). Oxytocin modulates cooperation within and competition between groups: an integrative review and research agenda. *Horm. Behav.* 61, 419–428. doi: 10.1016/j.yhbeh.2011.12.009

De Dreu, C. K., and Kret, M. E. (2016). Oxytocin conditions intergroup relations through upregulated in-group empathy, cooperation, conformity, and defense. *Biol. Psychiatry* 79, 165–173. doi: 10.1016/j.biopsych.2015.03.020

de Jong, T. R., and Neumann, I. D. (2018). Oxytocin and aggression. *Curr. Top. Behav. Neurosci.* 35, 175–192. doi: 10.1007/7854_2017_13

de Manzano, Ö., and Ullén, F. (2018). Same genes, different brains: neuroanatomical differences between monozygotic twins discordant for musical training. *Cereb. Cortex* 28, 387–394. doi: 10.1093/cercor/bhx299

Deuse, L., Wudarczyk, O., Rademacher, L., Kaleta, P., Karges, W., Kacheva, S., et al. (2019). Peripheral oxytocin predicts higher-level social cognition in men regardless of empathy quotient. *Pharmacopsychiatry* 52, 148–154. doi: 10.1055/a-0590-4850

De Wilde, T. R., Ten Velden, F. S., and De Dreu, C. K. (2017). The neuropeptide oxytocin enhances information sharing and group decision making quality. *Sci. Rep.* 7:40622. doi: 10.1038/srep40622

Ding, Q., Ying, Z., and Gómez-Pinilla, F. (2011). Exercise influences hippocampal plasticity by modulating brain-derived neurotrophic factor processing. *Neuroscience* 192, 773–780. doi: 10.1016/j.neuroscience.2011.06.032

Ditzen, B., Schaer, M., Gabriel, B., Bodenmann, G., Ehlert, U., and Heinrichs, M. (2009). Intranasal oxytocin increases positive communication and reduces cortisol levels during couple conflict. *Biol. Psychiatry* 65, 728–731. doi: 10.1016/j.biopsych.2008.10.011

Dölen, G., Darvishzadeh, A., Huang, K. W., and Malenka, R. C. (2013). Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. *Nature* 510, 179–184. doi: 10.1038/nature12518

Donadon, M. F., Martin-Santos, R., and Osório, F. L. (2018). The associations between oxytocin and trauma in humans: a systematic review. *Front. Pharmacol.* 9:154. doi: 10.3389/fphar.2018.00154

Donnay, G. F., Rankin, S. K., Lopez-Gonzalez, M., Jiradejvong, P., and Limb, C. J. (2014). Neural substrates of interactive musical improvisation: an fMRI study of 'trading fours' in jazz. *PLoS One* 9:e88665. doi: 10.1371/journal.pone.0088665

Douka, S., Zilidou, V. I., Lilou, O., and Manou, V. (2019). Traditional dance improves the physical fitness and well-being of the elderly. *Front. Aging Neurosci.* 11:75. doi: 10.3389/fnagi.2019.00075

Dunbar, R. I. M., Kaskatis, K., MacDonald, I., and Barra, V. (2012). Performance of music elevates pain threshold and positive affect: implications for the evolutionary function of music. *Evol. Psychol.* 10, 688–702. doi: 10.1177/147470491201000403

Dunsmoor, J. E., Kroes, M. C. W., Li, J., Daw, N. D., Simpson, H. B., and Phelps, E. A. (2019). Role of human ventromedial prefrontal cortex in learning and recall of enhanced extinction. *J. Neurosci.* 39, 3264–3276. doi: 10.1523/JNEUROSCI.2713-18.2019

Duque, A., and Spector, R. (2019). A balanced evaluation of the evidence for adult neurogenesis in humans: implication for neuropsychiatric disorders. *Brain Struct. Funct.* 224, 2281–2295. doi: 10.1007/s00429-019-01917-6

Ebitz, R. B., and Platt, M. L. (2014). An evolutionary perspective on the behavioral consequences of exogenous oxytocin application. *Front. Behav. Neurosci.* 7:225. doi: 10.3389/fnbeh.2013.00225

Ebstein, R. P., Knafo, A., Mankuta, D., Chew, S. H., and Lai, P. S. (2012). The contributions of oxytocin and vasopressin pathway genes to human behavior. *Horm. Behav.* 61, 359–379. doi: 10.1016/j.yhbeh.2011.12.014

Eckstein, M., Markett, S., Kendrick, K. M., Ditzen, B., Liu, F., Hurlemann, R., et al. (2017). Oxytocin differentially alters resting state functional connectivity between amygdala subregions and emotional control networks: inverse correlation with depressive traits. *NeuroImage* 149, 458–467. doi: 10.1016/j.neuroimage.2017.01.078

Ellenbogen, M. A. (2018). Oxytocin and facial emotion recognition. *Curr. Top. Behav. Neurosci.* 35, 349–374. doi: 10.1007/7854_2017_20

Erdozain, A. M., and Peñagarikano, O. (2020). Oxytocin as a treatment for social cognition, not there yet. *Front. Psychiatry* 10:930. doi: 10.3389/fpsyt.2019.00930

Erickson, K. I., Miller, D. L., and Roecklein, K. A. (2011). The aging hippocampus: interactions between exercise, depression and BDNF. *Neuroscientist* 18, 82–97. doi: 10.1177/1073858410397054

Evans, S. L., Dal Monte, O., Noble, P., and Averbeck, B. B. (2014). Intranasal oxytocin effects on social cognition: a critique. *Brain Res.* 1580, 69–77. doi: 10.1016/j.brainres.2013.11.008

Fancourt, D., Ockelford, A., and Belai, A. (2014). The psychoneuroimmunological effects of music: a systematic review and a new model. *Brain Behav. Immun.* 36, 15–26. doi: 10.1016/j.bbi.2013.10.014

Feldman, R. (2012). Oxytocin and social affiliation in humans. *Horm. Behav.* 61, 380–391. doi: 10.1016/j.yhbeh.2012.01.008

Feldman, R. (2015). The adaptive human parental brain: implications for children's social development. *Trends Neurosci.* 38, 387–399. doi: 10.1016/j.tins.2015.04.004

Feldman, R. (2017). The neurobiology of human attachments. *Trends Cogn. Sci.* 21, 80–99. doi: 10.1016/j.tics.2016.11.007

Feldman, R., Monakhov, M., Pratt, M., and Epstein, R. P. (2016). Oxytocin pathway genes: evolutionary ancient system impacting on human affiliation, sociality and psychopathology. *Biol. Psychiatry* 79, 174–184. doi: 10.1016/j.biopsych.2015.08.008

Feldman, R., Weller, A., Zagoory-Sharon, O., and Levine, A. (2007). Evidence for a neuroendocrinological foundation of human affiliation: plasma oxytocin levels across pregnancy and the postpartum period predict mother-infant bonding. *Psychol. Sci.* 18, 965–970. doi: 10.1111/j.1467-9280.2007.02010.x

Feng, C., Hackett, P. D., DeMarco, A. C., Chen, X., Stair, S., Haroon, E., et al. (2015). Oxytocin and vasopressin effects on the neural response to social cooperation are modulated by sex in humans. *Brain Imaging Behav.* 9, 754–764. doi: 10.1007/s11682-014-9333-9

Ferguson, J. N., Young, L. J., Hearn, E. F., Matzuk, M. M., Insel, T. R., and Winslow, J. T. (2000). Social amnesia in mice lacking the oxytocin gene. *Nat. Genet.* 25, 284–288. doi: 10.1038/77040

Ferguson, J. N., Young, L. J., and Insel, T. R. (2002). The neuroendocrine basis of social recognition. *Front. Neuroendocrinol.* 23, 200–224. doi: 10.1006/frne.2002.0229

Ferreri, L., Mas-Herrero, E., Zatorre, R. J., Ripollés, P., Gomez-Andres, A., Alicart, H., et al. (2019). Dopamine modulates the reward experiences elicited by music. *Proc. Natl. Acad. Sci. U S A* 116, 3793–3798. doi: 10.1073/pnas.1811878116

Fineberg, S. K., and Ross, D. A. (2017). Oxytocin and the social brain. *Biol. Psychiatry* 81, e19–e21. doi: 10.1016/j.biopsych.2016.11.004

Finn, S., and Fancourt, D. (2018). The biological impact of listening to music in clinical and nonclinical settings: a systematic review. *Prog. Brain Res.* 237, 173–200. doi: 10.1016/bs.pbr.2018.03.007

Fischer-Shofty, M., Levkovitz, Y., and Shamay-Tsoory, S. G. (2012). Oxytocin facilitates accurate perception of competition in men and kinship in women. *Soc. Cogn. Affect. Neurosci.* 8, 313–317. doi: 10.1093/scan/nsr100

Fitch, T. W. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215. doi: 10.1016/j.cognition.2005.11.009

Fragkaki, I., and Cima, M. (2019). The effect of oxytocin administration on empathy and emotion recognition in residential youth: a randomized, within-subjects trial. *Horm. Behav.* 114:104561. doi: 10.1016/j.yhbeh.2019.104561

França, T. F. A., Bitencourt, A. M., Maximilla, N. R., Barros, D. M., and Monserrat, J. M. (2017). Hippocampal neurogenesis and pattern separation: a meta-analysis of behavioral data. *Hippocampus* 27, 937–950. doi: 10.1002/hipo.22746

Franke, A. A., Li, X., Dabalos, C., and Lai, J. F. (2020). Improved oxytocin analysis from human serum and urine by orbitrap ESI-LC-HRAM-MS. *Drug Test Anal.* 12, 846–852. doi: 10.1002/dta.2783

Franke, A. A., Li, X., Menden, A., Lee, M. R., and Lai, J. F. (2019). Oxytocin analysis from human serum, urine and saliva by orbitrap liquid chromatography-mass spectrometry. *Drug Test Anal.* 11, 119–128. doi: 10.1002/dta.2475

Freeman, W. (2000). "A neurological role of music in social bonding," in *The Origins of Music* eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 411–424.

Frijling, J. L., van Zuiden, M., Nawijn, L., Koch, S. B., Neumann, I. D., Veltman, D. J., et al. (2015). Salivary oxytocin and vasopressin levels in police officers with and without post-traumatic stress disorder. *J. Neuroendocrinol.* 27, 743–751. doi: 10.1111/jne.12300

Fujiwara, T., Weisman, O., Ochi, M., Shirai, K., Matsumoto, K., Noguchi, E., et al. (2019). Genetic and peripheral markers of the oxytocin system and parental care jointly support the cross-generational transmission of bonding across three generations. *Psychoneuroendocrinology* 102, 172–181. doi: 10.1016/j.psyneuen.2018.12.004

Fukuda, H., Ma, N., Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., et al. (2019). Computing social value conversion in the human brain. *J. Neurosci.* 39, 5153–5172. doi: 10.1523/JNEUROSCI.3117-18.2019

Fukui, H., and Toyoshima, K. (2008). Music facilitate the neurogenesis, regeneration and repair of neurons. *Med. Hypotheses* 71, 765–769. doi: 10.1016/j.mehy.2008.06.019

Gamal-Eltrabily, M., de los Monteros-Zúñiga, E. A., Manzano-García, A., Martínez-Lorenzana, G., Condés-Lara, M., and González-Hernández, A. (2020). The rostral agranular insular cortex, a new site of oxytocin to induce antinociception. *J. Neurosci.* 40, 5669–5680. doi: 10.1523/jneurosci.0962-20.2020

Garrett, L., Zhang, J., Zimprich, A., Niedermeier, K. M., Fuchs, H., Gailus-Durner, V., et al. (2015). Conditional reduction of adult born doublecortin-positive neurons reversibly impairs selective behaviors. *Front. Behav. Neurosci.* 9:302. doi: 10.3389/fnbeh.2015.00302

Garza-Villarreal, E. A., Pando, P., Vuust, P., and Parsons, C. (2017). Music-induced analgesia in chronic pain conditions: a systematic review and meta-analysis. *Pain Physician* 20, 597–610.

Gedeon, T., Parry, J., and Völlm, B. (2019). The role of oxytocin in antisocial personality disorders: a systematic review of the literature. *Front. Psychiatry* 10:76. doi: 10.3389/fpsyt.2019.00076

Geng, Y., Zhao, W., Zhou, F., Ma, X., Yao, S., Hurlemann, R., et al. (2018). Oxytocin enhancement of emotional empathy: generalization across cultures and effects on amygdala activity. *Front. Neurosci.* 12:512. doi: 10.3389/fnins.2018.00512

Giovanna, G., Damiani, S., Fusar-Poli, L., Rocchetti, M., Brondino, N., de Cagna, F., et al. (2020). Intranasal oxytocin as a potential therapeutic strategy in post-traumatic stress disorder: a systematic review. *Psychoneuroendocrinology* 115:104605. doi: 10.1016/j.psyneuen.2020.104605

Gold, B. P., Mas-Herrero, E., Zeighami, Y., Benovoy, M., Dagher, A., and Zatorre, R. J. (2019). Musical reward prediction errors engage the nucleus accumbens and motivate learning. *Proc. Natl. Acad. Sci. U S A* 116, 3310–3315. doi: 10.1073/pnas.1809855116

Gordon, I., Zagoory-Sharon, O., Leckman, J. F., and Feldman, R. (2010). Oxytocin and the development of parenting in humans. *Biol. Psych.* 68, 377–382. doi: 10.1016/j.biopsych.2010.02.005

Granot, R., Frankel, Y., Gritsenko, V., Lerer, E., Gritsenko, I., Bachner-Melman, R., et al. (2007). Provisional evidence that the arginine vasopressin 1a receptor gene is associated with musical memory. *Evol. Hum. Behav.* 28, 313–318. doi: 10.1016/j.evolhumbehav.2007.05.003

Granot, R. Y., Uzefovsky, F., Bogopolsky, H., and Ebstein, R. P. (2013). Effects of arginine vasopressin on musical working memory. *Front. Psychol.* 4:712. doi: 10.3389/fpsyg.2013.00712

Grape, C., Sandgren, M., Hansson, L. O., Ericson, M., and Theorell, T. (2003). Does singing promote well-being? An empirical study of professional and amateur singers during a singing lesson. *Integr. Phys. Behav. Sci.* 38, 65–74. doi: 10.1007/bf02734261

Grinevich, V., Knobloch-Bollmann, H. S., Eliava, M., Busnelli, M., and Chini, B. (2016). Assembling the puzzle: pathways of oxytocin signaling in the brain. *Biol. Psychiatry* 79, 155–164. doi: 10.1016/j.biopsych.2015.04.013

Grinevich, V., and Neumann, I. D. (2020). Brain oxytocin: how puzzle stones from animal studies translate into psychiatry. *Mol. Psychiatry* doi: 10.1038/s41380-020-0802-9 [Epub ahead of print].

Grinevich, V., and Stoop, R. (2018). Interplay between oxytocin and sensory systems in the orchestration of socio-emotional behaviors. *Neuron* 99, 887–904. doi: 10.1016/j.neuron.2018.07.016

Groussard, M., Chan, T. G., Coppalle, R., and Patel, H. (2019). Preservation of musical memory throughout the progression of Alzheimer's Disease? Toward a reconciliation of theoretical, clinical and neuroimaging evidence. *Alzheimers Dis.* 68, 857–883. doi: 10.3233/jad-180474

Guastella, A. J., and MacLeod, C. (2012). A critical review of the influence of oxytocin nasal spray on social cognition in humans: evidence and future directions. *Horm. Behav.* 61, 410–418. doi: 10.1016/j.yhbeh.2012.01.002

Guhn, M., Emerson, S. D., and Gouzouasis, P. (2020). A population-level analysis of associations between school music participation and academic achievement. *J. Educat. Psychol.* 112, 308–328. doi: 10.1037/edu0000376

Gujing, L., Hui, H., Xin, L., Lirong, Z., Yutong, Y., Guofeng, Y., et al. (2019). Increased insular connectivity and enhanced empathic ability associated with dance/music training. *Neural Plast.* 2019:9693109. doi: 10.1155/2019/9693109

Habibi, A., Damasio, A., Ilari, B., Elliott Sachs, M., and Damasio, H. (2018). Music training and child development: a review of recent findings from a longitudinal study. *Ann. N Y Acad. Sci.* doi: 10.1111/nyas.13606 [Epub ahead of print].

Harvey, A. R. (2017). *Music, Evolution and the Harmony of Souls.* Oxford: Oxford University Press.

Harvey, A. R. (2018). Music and the meeting of human minds. *Front. Psychol.* 9:762. doi: 10.3389/fpsyg.2018.00762

Hasan, M. T., Althammer, F., Silva da Gouveia, M., Goyon, S., Eliava, M., Lefevre, A., et al. (2019). A fear memory engram and its plasticity in the hypothalamic oxytocin system. *Neuron* 103, 133.e8–146.e8. doi: 10.1016/j.neuron.2019.04.029

Havranek, T., Zatkova, M., Lestanova, Z., Bacova, Z., Mravec, B., Hodosy, J., et al. (2015). Intracerebroventricular oxytocin administration in rats enhances object recognition and increases expression of neurotrophins, microtubule-associated protein 2 and synapsin I. *J. Neurosci. Res.* 93, 893–901. doi: 10.1002/jnr.23559

Herdener, M., Esposito, F., di Salle, F., Boller, C., Hilti, C. C., Habermeyer, B., et al. (2010). Musical training induces functional plasticity in human hippocampus. *J. Neurosci.* 30, 1377–1384. doi: 10.1523/jneurosci.4513-09.2010

Hilfiger, L., Zhao, Q., Kerspern, D., Inquimbert, P., Andry, V., Goumon, Y., et al. (2020). A nonpeptide oxytocin receptor agonist for a durable relief of inflammatory pain. *Sci. Rep.* 10:3017. doi: 10.1038/s41598-020-59929-w

Horgusluoglu-Moloch, E., Risacher, S. L., Crane, P. K., Hibar, D., Thompson, P. M., Saykin, A. J., et al. (2019). Genome-wide association analysis of hippocampal volume identifies enrichment of neurogenesis-related pathways. *Sci. Rep.* 9:14498. doi: 10.1038/s41598-019-50507-3

Hsiao, Y.-H., Hung, H.-C., Chen, S.-H., and Gean, P.-W. (2014). Social interaction rescues memory deficit in an animal model of Alzheimer's Disease by increasing BDNF-dependent hippocampal neurogenesis. *J. Neurosci.* 34, 16207–16219. doi: 10.1523/JNEUROSCI.0747-14.2014

Hu, J., Wang, Z., Feng, X., Long, C., and Schiller, D. (2019). Post-retrieval oxytocin facilitates next day extinction of threat memory in humans. *Psychopharmacology* 236, 293–301. doi: 10.1007/s00213-018-5074-6

Huber, D., Veinante, P., and Stoop, R. (2005). Vasopressin and oxytocin excite distinct neuronal populations in the central amygdala. *Science* 308, 245–248. doi: 10.1126/science.1105636

Hurlemann, R., Patin, A., Onur, O. A., Cohen, M. X., Baumgartner, T., Metzler, S., et al. (2010). Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. *J. Neurosci.* 30, 4999–5007. doi: 10.1523/JNEUROSCI.5538-09.2010

Israel, S., Lerer, E., Shalev, I., Uzefovsky, F., Reibold, M., Bachner-Melman, R., et al. (2008). Molecular genetic studies of the arginine vasopressin 1a receptor (AVPR1a) and the oxytocin receptor (OXTR) in human behaviour: from autism to altruism with some notes in between. *Prog. Brain Res.* 170, 435–449. doi: 10.1016/S0079-6123(08)00434-2

Jirikowski, G. F. (2019). Diversity of central oxytocinergic projections. *Cell Tissue Res.* 375, 41–48. doi: 10.1007/s00441-018-2960-5

Jirikowski, G. F., Ochs, S. D., and Caldwell, J. D. (2018). Oxytocin and steroid actions. *Curr. Top. Behav. Neurosci.* 35, 77–95. doi: 10.1007/7854_2017_9

Johnson, J. K., Louhivuori, J., and Silijander, E. (2017). Comparison of well-being of older adult choir singers and the general population in Finland: a case-control study. *Music Sci.* 21, 178–194. doi: 10.1177/1029864916644486

Johnson, J. K., Stewart, A. L., Acree, M., Nápoles, A. M., Flatt, J. D., Max, W. B., et al. (2020). A community choir intervention to promote well-being among diverse older adults: results from the community of voices trial. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 75, 549–559. doi: 10.1093/geronb/gby132

Jurek, B., and Neumann, I. D. (2018). The oxytocin receptor: from intracellular signaling to behavior. *Physiol. Rev.* 98, 1805–1908. doi: 10.1152/physrev.00031.2017

Kang, J., Scholp, A., and Jiang, J. J. (2018). A review of the physiological effects and mechanisms of singing. *J. Voice* 32, 390–395. doi: 10.1016/j.jvoice.2017.07.008

Karpati, F. J., Giacosa, C., Foster, N. E., Penhune, V. B., and Hyde, K. L. (2016). Sensorimotor integration is enhanced in dancers and musicians. *Exp. Brain Res.* 234, 893–903. doi: 10.1007/s00221-015-4524-1

Kattenstroth, J. C., Kalisch, T., Holt, S., Tegenthoff, M., and Dinse, H. R. (2013). Six months of dance intervention enhances postural, sensorimotor and cognitive performance in elderly without affecting cardio-respiratory functions. *Front. Aging Neurosci.* 5:5. doi: 10.3389/fnagi.2013.00005

Kattenstroth, J. C., Kolankowska, I., Kalisch, T., and Dinse, H. R. (2010). Superior sensory, motor and cognitive performance in elderly individuals with multi-year dancing activities. *Front. Aging Neurosci.* 2:31. doi: 10.3389/fnagi.2010.00031

Keech, B., Crowe, S., and Hocking, D. R. (2018). Intranasal oxytocin, social cognition and neurodevelopmental disorders: a meta-analysis. *Psychoneuroendocrinology* 87, 9–19. doi: 10.1016/j.psyneuen.2017.09.022

Keeler, J. R., Roth, E. A., Neuser, B. L., Spitsbergen, J. M., Waters, D. J., and Vianney, J. M. (2015). The neurochemistry and social flow of singing: bonding and oxytocin. *Front. Hum. Neurosci.* 9:518. doi: 10.3389/fnhum.2015.00518

Kempermann, G., Gage, F. H., Aigner, L., Song, H., Curtis, M. A., Thuret, S., et al. (2018). Human adult neurogenesis: evidence and remaining questions. *Cell Stem Cell* 23, 25–30. doi: 10.1016/j.stem.2018.04.004

Kerr, F., Wiechula, R., Feo, R., Schultz, T., and Kitson, A. (2019). Neurophysiology of human touch and eye gaze in therapeutic relationships and healing: a scoping review. *JBI Database System Rev. Implement. Rep.* 17, 209–247. doi: 10.11124/JBISRIR-2017-003549

Khalfa, S., Bella, S. D., Roy, M., Peretz, I., and Lupien, S. J. (2003). Effects of relaxing music on salivary cortisol level after psychological stress. *Ann. N Y Acad. Sci.* 999, 374–376. doi: 10.1196/annals.1284.045

Khant, T., Chang, L. J., Park, S. Q., Heinzle, J., and Haynes, J.-D. (2012). Connectivity-based parcellation of the human orbitofrontal cortex. *J. Neurosci.* 32, 6240–6250. doi: 10.1523/JNEUROSCI.0257-12.2012

Kim, S., Fonagy, P., Koos, O., Dorsett, K., and Strathearn, L. (2014). Maternal oxytocin response predicts mother-to-infant gaze. *Brain Res.* 1580, 133–142. doi: 10.1016/j.brainres.2013.10.050

Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., et al. (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. *J. Neurosci.* 25, 11489–11493. doi: 10.1523/JNEUROSCI.3984-05.2005

Kirschner, S., and Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child. Psychol.* 102, 299–314. doi: 10.1016/j.jecp.2008.07.005

Koch, S. B. J., van Zuiden, M., Nawijn, L., Frijling, J. L., Veltman, D. J., and Olff, M. (2019). Effects of intranasal oxytocin on distraction as emotion regulation strategy in patients with post-traumatic stress disorder. *Eur. Neuropsychopharmacol.* 29, 266–277. doi: 10.1016/j.euroneuro.2018.12.002

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666

Koelsch, S. (2018). Investigating the neural encoding of emotion with music. *Neuron* 98, 1075–1079. doi: 10.1016/j.neuron.2018.04.029

Koelsch, S., Fritz, T., v Cramon, D. Y., Müller, K., and Friederici, A. D. (2006). Investigating emotion with music: an fMRI study. *Hum. Brain Mapp.* 27, 239–250. doi: 10.1002/hbm.20180

Koelsch, S., and Jäncke, L. (2015). Music and the heart. *Eur. Heart J.* 36, 3043–3049. doi: 10.1093/eurheartj/ehv430

Korn, C. W., Prehn, K., Park, S. Q., Walter, H., and Heekeren, H. R. (2012). Positively based processing of self-relevant social feedback. *J. Neurosci.* 32, 16832–16844. doi: 10.1523/JNEUROSCI.3016-12.2012

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., and Fehr, E. (2005). Oxytocin increases trust in humans. *Nature* 435, 673–676. doi: 10.1038/nature03701

Kreutz, G. (2014). Singing and social bonding. *Music Med.* 6, 51–60.

Kreutz, G., Bongard, S., Rohrmann, S., Hodapp, V., and Grebe, D. (2004). Effects of choir singing or listening on secretory immunoglobulin A, cortisol, and emotional state. *J. Behav. Med.* 27, 623–635. doi: 10.1007/s10865-004-0006-9

Kuhn, D. (2002). The effects of active and passive participation in musical activity on the immune system as measured by salivary immunoglobulin A (SIgA). *J. Music Ther.* 39, 30–39. doi: 10.1093/jmt/39.1.30

Kuhn, H. G., Toda, T., and Gage, F. H. (2018). Adult hippocampal neurogenesis: a coming of age story. *J. Neurosci.* 38, 10401–10410. doi: 10.1523/JNEUROSCI.2144-18.2018

Kumar, J., Iwabuchi, S. J., Völlm, B. A., and Palaniyappan, L. (2020). Oxytocin modulates the effective connectivity between the precuneus and the dorsolateral prefrontal cortex. *Eur. Arch. Psychiatry Clin. Neurosci.* 270, 567–576. doi: 10.1007/s00406-019-00989-z

Laland, K., Wilkins, C., and Clayton, N. (2016). The evolution of dance. *Curr. Biol.* 26, R5–R9. doi: 10.1016/j.cub.2015.11.031

Lambert, B., Declerck, C. H., Boone, C., and Parizel, P. M. (2017). A functional MRI study on how oxytocin affects decision making in social dilemmas: cooperate as long as it pays off, aggress only when you think you can win. *Horm. Behav.* 94, 145–152. doi: 10.1016/j.yhbeh.2017.06.011

Lancaster, K., Goldbeck, L., Pournajafi-Nazarloo, H., Connelly, J. J., Carter, C. S., and Morris, J. P. (2018). The role of endogenous oxytocin in anxiolysis: structural and functional correlates. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 3, 618–625. doi: 10.1016/j.bpsc.2017.10.003

Latt, H. M., Matsushita, H., Morino, M., Koga, Y., Michiue, H., Nishiki, T., et al. (2018). Oxytocin inhibits corticosterone-induced apoptosis in primary hippocampal neurons. *Neuroscience* 379, 383–389. doi: 10.1016/j.neuroscience.2018.03.025

Lawson, E. A., Olszewski, P. K., Weller, A., and Blevins, J. E. (2019). The role of oxytocin in regulation of appetitive behaviour, body weight and glucose homeostasis. *J. Neuroendocrinol.* 28:e12805. doi: 10.1111/jne.12805

Lee, J. H., Lee, S., and Kim, J. H. (2017). Amygdala circuits for fear memory: a key role for dopamine regulation. *Neuroscientist* 23, 542–553. doi: 10.1177/1073858416679936

Lee, S. Y., Park, S. H., Chung, C., Kim, J. J., Choi, S. Y., and Han, J. S. (2015). Oxytocin protects hippocampal memory and plasticity from uncontrollable stress. *Sci. Rep.* 5:18540. doi: 10.1038/srep18540

Leng, G., and Ludwig, M. (2016). Intranasal oxytocin: myths and delusions. *Biol. Psychiatry* 79, 243–250. doi: 10.1016/j.biopsych.2015.05.003

Leppanen, J., Ng, K. W., Tchanturia, K., and Treasure, J. (2017). Meta-analysis of the effects of intranasal oxytocin on interpretation and expression of emotions. *Neurosci. Biobehav. Rev.* 78, 125–144. doi: 10.1016/j.neubiorev.2017.04.010

Levitin, D. J., Grahn, J. A., and London, J. (2018). The psychology of music: rhythm and movement. *Annu. Rev. Psychol.* 69, 51–75. doi: 10.1146/annurev-psych-122216-011740

Li, T., Chen, X., Mascaro, J., Haroon, E., and Rilling, J. K. (2017a). Intranasal oxytocin, but not vasopressin, augments neural responses to toddlers in human fathers. *Horm. Behav.* 93, 193–202. doi: 10.1016/j.yhbeh.2017.01.006

Li, T., Wang, P., Wang, S. C., and Wang, Y. F. (2017b). Approaches mediating oxytocin regulation of the immune system. *Front. Immunol.* 7:693. doi: 10.3389/fimmu.2016.00693

Li, Q., Zhang, B., Cao, H., Liu, W., Guo, F., Shen, F., et al. (2020). Oxytocin exerts antidepressant-like effect by potentiating dopaminergic synaptic transmission in the mPFC. *Neuropharmacology* 162:107836. doi: 10.1016/j.neuropharm.2019.107836

Licht, T., Kreisel, T., Biala, Y., Mohan, S., Yaari, Y., Anisimov, A., et al. (2020). Age-dependent remarkable regenerative potential of the dentate gyrus provided by intrinsic stem cells. *J. Neurosci.* 40, 974–995. doi: 10.1523/jneurosci.1010-19.2019

Lima, S. M. A., and Gomes-Leal, W. (2019). Neurogenesis in the hippocampus of adult humans: controversy "fixed" at last. *Neural Regen. Res.* 14, 1917–1918. doi: 10.4103/1673-5374.259616

Limb, C. J., and Braun, A. R. (2008). Neural substrates of spontaneous musical performance: an FMRI study of jazz improvisation. *PLoS One* 3:e1679. doi: 10.1371/journal.pone.0001679

Lin, Y. T., and Hsu, K. S. (2018). Oxytocin receptor signaling in the hippocampus: role in regulating neuronal excitability, network oscillatory activity, synaptic plasticity and social memory. *Prog. Neurobiol.* 171, 1–14. doi: 10.1016/j.pneurobio.2018.10.003

Lin, Y. T., Chen, C. C., Huang, C. C., Nishimori, K., and Hsu, K. S. (2017). Oxytocin stimulates hippocampal neurogenesis *via* oxytocin receptor expressed in CA3 pyramidal neurons. *Nat. Commun.* 8:537. doi: 10.1038/s41467-017-00675-5

Lin, Y.-T., Hsieh, T.-Y., Tsai, T.-C., Chen, C.-C., Huang, C.-C., and Hsu, K.-S. (2018). Conditional deletion of hippocampal CA2/CA3a oxytocin receptors impairs the persistence of long-term social recognition memory in mice. *J. Neurosci.* 38, 1218–1231. doi: 10.1523/jneurosci.1896-17.2017

Lischke, A., Gamer, M., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., et al. (2012). Oxytocin increases amygdala reactivity to threatening scenes in females. *Psychoneuroendocrinology* 37, 1431–1438. doi: 10.1016/j.psyneuen.2012.01.011

Liu, S., Chow, H. M., Xu, Y., Erkkinen, M. G., Swett, K. E., Eagle, M. W., et al. (2012). Neural correlates of lyrical improvisation: an FMRI study of freestyle rap. *Sci. Rep.* 2:834. doi: 10.1038/srep00834

Liu, X., Kanduri, C., Oikkonen, J., Karma, K., Raijas, P., Ukkola-Vuoti, L., et al. (2016). Detecting signatures of positive selection associated with musical aptitude in the human genome. *Sci. Rep.* 6:21198. doi: 10.1038/srep21198

Liu, Y., Li, S., Lin, W., Li, W., Yan, X., Wang, X., et al. (2019). Oxytocin modulates social value representations in the amygdala. *Nat. Neurosci.* 22, 633–641. doi: 10.1038/s41593-019-0351-1

Liu, P. Z., and Nusslock, R. (2018). Exercise-mediated neurogenesis in the hippocampus *via* BDNF. *Front. Neurosci.* 12:52. doi: 10.3389/fnins.2018.00052

Lopatina, O. L., Komleva, Y. K., Gorina, Y. V., Olovyannikova, R. Y., Trufanova, L. V., Hashimoto, T., et al. (2018). Oxytocin and excitation/inhibition balance in social recognition. *Neuropeptides* 72, 1–11. doi: 10.1016/j.npep.2018.09.003

Lucassen, P. J., Fitzsimonsm, C. P., Salta, E., and Maletic-Savatic, M. (2020). Adult neurogenesis, human after all (again): classic, optimized and future approaches. *Behav. Brain Res.* 381:112458. doi: 10.1016/j.bbr.2019.112458

Luo, S., Zhu, Y., Fan, L., Gao, D., and Han, S. (2020). Resting-state brain network properties mediate the association between the oxytocin receptor gene and interdependence. *Soc. Neurosci.* 15, 296–310. doi: 10.1080/17470919.2020.1714718

Malik, A. I., Zai, C. C., Abu, Z., Nowrouzi, B., and Beitchman, J. H. (2012). The role of oxytocin and oxytocin receptor gene variants in childhood-onset aggression. *Genes Brain Behav.* 11, 545–551. doi: 10.1111/j.1601-183x.2012.00776.x

Maniezzi, C., Talpo, F., Spaiardi, P., Toselli, M., and Biella, G. (2019). Oxytocin increases phasic and tonic GABAergic transmission in CA1 region of mouse hippocampus. *Front. Cell. Neurosci.* 13:178. doi: 10.3389/fncel.2019.00178

Mansens, D., Deeg, D. J. H., and Comijs, H. C. (2018). The association between singing and/or playing a musical instrument and cognitive functions in older adults. *Aging Ment. Health* 22, 964–971. doi: 10.1080/13607863.2017.1328481

Mariath, L. M., Silva, A. M. D., Kowalski, T. W., Gattino, G. S., Araujo, G. A., Figueiredo, F. G., et al. (2017). Music genetics research: association with musicality of a polymorphism in the AVPRA gene. *Genet. Mol. Biol.* 40, 421–429. doi: 10.1590/1678-4685-gmb-2016-0021

Masis-Calvo, M., Schmidtner, A. K., de Moura Oliveira, V. E., Grossmann, C. P., de Jong, T. R., and Neumann, I. D. (2018). Animal models of social stress: the dark side of social interactions. *Stress* 21, 417–432. doi: 10.1080/10253890.2018.1462327

Matthiesen, A. S., Ransjo-Arvidson, A. B., Nissen, E., and Uvnas- Moberg, K. (2001). Postpartum maternal oxytocin release by newborns: effects of infant hand massage and sucking. *Birth* 28, 13–19. doi: 10.1046/j.1523-536x.2001.00013.x

McEwen, B. S. (1999). Stress and hippocampal plasticity. *Annu. Rev. Neurosci.* 12, 105–122.

Meguro, Y., Miyano, K., Hirayama, S., Yoshida, Y., Ishibashi, N., Ogino, T., et al. (2018). Neuropeptide oxytocin enhances μ opioid receptor signaling as a positive allosteric modulator. *J. Pharmacol. Sci.* 137, 67–75. doi: 10.1016/j.jphs.2018.04.002

Mehr, S. A., Singh, M., York, H., Glowacki, L., and Krasnow, M. M. (2018). Form and function in human song. *Curr. Biol.* 28, 356.e5–368.e5. doi: 10.1016/j.cub.2017.12.042

Menon, V., and Levitin, D. J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *NeuroImage* 28, 175–184. doi: 10.1016/j.neuroimage.2005.05.053

Millan, M. J. (2002). Descending control of pain. *Prog. Neurobiol.* 66, 355–474. doi: 10.1016/s0301-0082(02)00009-6

Miranda, M., Morici, J. F., Zanoni, M. B., and Bekinschtein, P. (2019). Brain-derived neurotrophic factor: a key molecule for memory in the healthy and the pathological brain. *Front. Cell. Neurosci.* 13:363. doi: 10.3389/fncel.2019.00363

Mithen, S. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body.* London: Orion.

Mitre, M., Marlin, B. J., Schiavo, J. K., Morina, E., Norden, S. E., Hackett, T. A., et al. (2016). A distributed network for social cognition enriched for oxytocin receptors. *J. Neurosci.* 36, 2517–2535. doi: 10.1523/jneurosci.2409-15.2016

Mitre, M., Minder, J., Morina, E. X., Chao, M. V., and Froemke, R. C. (2018). Oxytocin modulation of neural circuits. *Curr. Top. Behav. Neurosci.* 35, 31–53. doi: 10.1007/7854_2017_7

Mitterschiffthaler, M. T., Fu, C. H. Y., Dalton, J. A., Andrew, C. M., and Williams, S. C. R. (2007). A functional MRI study of happy and sad affective states induced by classical music. *Hum. Brain Mapp.* 28, 1150–1162. doi: 10.1002/hbm.20337

Moaddab, M., Hyland, B. I., and Brown, C. H. (2015). Oxytocin excites nucleus accumbens shell neurons *in vivo*. *Mol. Cell. Neurosci.* 68, 323–330. doi: 10.1016/j.mcn.2015.08.013

Moons, W. G., Way, B. M., and Taylor, S. E. (2014). Oxytocin and vasopressin receptor polymorphisms interact with circulating neuropeptides to predict human emotional reactions to stress. *Emotion* 14, 562–572. doi: 10.1037/a0035503

Moreno-Jiménez, E. P., Flor-García, M., Terreros-Roncal, J., Rábano, A., Cafini, F., Pallas-Bazarra, N., et al. (2019). Adult hippocampal neurogenesis is abundant in neurologically healthy subjects and drops sharply in patients with Alzheimer's disease. *Nat. Med.* 25, 554–560. doi: 10.1038/s41591-019-0375-9

Morley, I. (2013). *The Prehistory of Music: Human Evolution, Archaeology and the Origins of Musicality.* Oxford: Oxford University Press.

Morley, A. P., Narayanan, M., Mines, R., Molokhia, A., Baxter, S., Craig, G., et al. (2012). AVPR1A and SLC6A4 polymorphisms in choral singers and non-musicians: a gene association study. *PLoS One* 7:e31763. doi: 10.1371/journal.pone.0031763

Moss, H., Lynch, J., and O'Donoghue, J. (2018). Exploring the perceived health benefits of singing in a choir: an international cross-sectional mixed-methods study. *Perspect. Public Health* 138, 160–168. doi: 10.1177/1757913917739652

Mueller, K., Fritz, T., Mildner, T., Richter, M., Schulze, K., Lepsien, J., et al. (2015). Investigating the dynamics of the brain response to music: a central role of the ventral striatum/nucleus accumbens. *NeuroImage* 116, 68–79. doi: 10.1016/j.neuroimage.2015.05.006

Narendran, R., Tollefson, S., Fasenmyer, K., Paris, J., Himes, M. L., Lopresti, B., et al. (2019). Decreased nociceptin receptors are related to resilience and recovery in college women who have experienced sexual violence: therapeutic implications for posttraumatic stress disorder. *Biol. Psych.* 85, 1056–1064. doi: 10.1016/j.biopsych.2019.02.017

Ne'eman, R., Perach-Barzilay, N., Fischer-Shofty, M., Atias, A., and Shamay-Tsoory, S. G. (2016). Intranasal administration of oxytocin increases human aggressive behaviour. *Horm. Behav.* 80, 125–131. doi: 10.1016/j.yhbeh.2016.01.015

Neumann, I. D., and Landgraf, R. (2012). Balance of brain oxytocin and vasopressin: implications for anxiety, depression, and social behaviors. *Trends Neurosci.* 35, 649–659. doi: 10.1016/j.tins.2012.08.004

Neumann, I. D., and Landgraf, R. (2019). Tracking oxytocin functions in the rodent brain during the last 30 years: from push-pull perfusion to chemogenetic silencing. *J. Neuroendocrinol.* 31:e12695. doi: 10.1111/jne.12695

Neumann, I. D., and Slattery, D. A. (2016). Oxytocin in general anxiety and social fear: a translational approach. *Biol. Psych.* 79, 213–221. doi: 10.1016/j.biopsych.2015.06.004

Nilsson, U. (2009). Soothing music can increase oxytocin levels during bed rest after open-heart surgery: a randomised control trial. *J. Clin. Nurs.* 18, 2153–2161. doi: 10.1111/j.1365-2702.2008.02718.x

Nilsson, S., Kokinsky, E., Nilsson, U., Sidenvall, B., and Enskär, K. (2009). School-aged children's experiences of postoperative music medicine on pain, distress and anxiety. *Paediatr. Anaesth.* 19, 1184–1190. doi: 10.1111/j.1460-9592.2009.03180.x

Nishina, K., Takagishi, H., Inoue-Murayama, M., Takahashi, H., and Yamagishi, T. (2015). Polymorphism of the oxytocin receptor gene modulates behavioral and attitudinal trust among men but not women. *PLoS One* 10:e0137089. doi: 10.1371/journal.pone.0137089

Nishina, K., Takagishi, H., Takahashi, H., Sakagami, M., and Inoue-Murayama, M. (2019). Association of polymorphism of arginine-vasopressin receptor 1A (*AVPR1a*) gene with trust and reciprocity. *Front. Hum. Neurosci.* 13:230. doi: 10.3389/fnhum.2019.00230

Nilsson, U., Unosson, M., and Rawal, N. (2005). Stress reduction and analgesia in patients exposed to calming music postoperatively: a randomized controlled trial. *Eur. J. Anaesthesiol.* 22, 96–102. doi: 10.1017/s0265021505000189

Norman-Haignere, S., Kanwisher, N. G., and McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88, 1281–1296. doi: 10.1016/j.neuron.2015.11.035

Nozaradan, S., Peretz, I., Missal, M., and Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* 31, 10234–10240. doi: 10.1523/jneurosci.0411-11.2011

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.* 14, 769–776. doi: 10.1016/j.conb.2004.10.016

Oikkonen, J., Kuusi, T., Peltonen, P., Raijas, P., Ukkola-Vuoti, L., Karma, K., et al. (2016). Creative activities in music—a genome-wide linkage analysis. *PLoS One* 11:e0148679. doi: 10.1371/journal.pone.0148679

Onaka, T., and Takayanagi, Y. (2019). Role of oxytocin in the control of stress and food intake. *J. Neuroendocrinol.* 31:e12700. doi: 10.1111/jne.12700

Ooishi, Y., Mukai, H., Watanabe, K., Kawato, S., and Kashino, M. (2017). Increase in salivary oxytocin and decrease in salivary cortisol after listening to relaxing slow-tempo and exciting fast-tempo music. *PLoS One* 12:e0189075. doi: 10.1371/journal.pone.0189075

Opendak, M., Offit, L., Monari, P., Schoenfeld, T. J., Sonti, A. N., Cameron, H. A., et al. (2016). Lasting adaptations in social behavior produced by social disruption and inhibition of adult neurogenesis. *J. Neurosci.* 36, 7027–7038. doi: 10.1523/jneurosci.4435-15.2016

Ouanes, S., and Popp, J. (2019). High cortisol and the risk of dementia and Alzheimer's disease: a review of the literature. *Front. Aging Neurosci.* 11:43. doi: 10.3389/fnagi.2019.00043

Patel, A. D. (2008). *Music, Language and the Brain.* New York, NY: Oxford University Press.

Pearce, E., Launay, J., and Dunbar, R. I. M. (2015). The ice-breaker effect: singing mediates fast social bonding. *R. Soc. Open Sci.* 2:150221. doi: 10.1098/rsos.150221

Pearce, E., Wlodarski, R., Machin, A., and Dunbar, R. I. M. (2017). Variation in the β-endorphin, oxytocin and dopamine receptor genes is associated with different dimensions of human sociality. *Proc. Natl. Acad. Sci. U S A* 114, 5300–5305. doi: 10.1073/pnas.1700712114

Pekarek, B. T., Hunt, P. J., and Arenkiel, B. R. (2020). Oxytocin and sensory network plasticity. *Front. Neurosci.* 14:30. doi: 10.3389/fnins.2020.00030

Peled-Avron, L., Abu-Akel, A., and Shamay-Tsoory, S. (2020). Exogenous effects of oxytocin in five psychiatric disorders: a systematic review, meta-analyses and a personalized approach through the lens of the social salience hypothesis. *Neurosci. Biobehav. Rev.* 114, 70–95. doi: 10.1016/j.neubiorev.2020.04.023

Pereira, A. P. S., Marinho, V., Gupta, D., Magalhães, F., Ayres, C., and Teixeira, S. (2019). Music therapy and dance as gait rehabilitation in patients with Parkinson disease: a review of evidence. *J. Geriatr. Psychiatry Neurol.* 32, 49–56. doi: 10.1177/0891988718819858

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., and Brattico, E. (2011). Music and emotions in the brain: familiarity matters. *PLoS One* 6:e27241. doi: 10.1371/journal.pone.0027241

Peretz, I., Vuvan, D., Lagrois, M-E., and Armony, J. L. (2015). Neural overlap in processing music and speech. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140090. doi: 10.1098/rstb.2014.0090

Perkins, R., Yorke, S., and Fancourt, D. (2018). How group singing facilitates recovery from the symptoms of postnatal depression: a comparative qualitative study. *BMC Psychol.* 6:41. doi: 10.1186/s40359-018-0253-0

Petrik, D., and Encinas, J. M. (2019). Perspective: of mice and men - how widespread is adult neurogenesis? *Front. Neurosci.* 13:923. doi: 10.3389/fnins.2019.00923

Phillips-Silver, J., and Trainor, L. J. (2007). Hearing what the body feels: auditory encoding of rhythmic movement. *Cognition* 105, 533–546. doi: 10.1016/j.cognition.2006.11.006

Plasencia, G., Luedicke, J. M., Nazarloo, H. P., Carter, C. S., and Ebner, N. C. (2019). Plasma oxytocin and vasopressin levels in young and older men and women: functional relationships with attachment and cognition. *Psychoneuroendocrinology* 110:104419. doi: 10.1016/j.psyneuen.2019.104419

Poore, H. E., and Waldman, I. D. (2020). The association of oxytocin receptor gene (OXTR) polymorphisms with antisocial behavior: a meta-analysis. *Behav. Genet.* 50:174. doi: 10.1007/s10519-020-09998-4

Preckel, K., Scheele, D., Kendrick, K. M., Maier, W., and Hurlemann, R. (2014). Oxytocin facilitates social approach behaviour in women. *Front. Behav. Neurosci.* 8:191. doi: 10.3389/fnbeh.2014.00191

Pulli, K., Karma, K., Norio, R., Sistonen, P., Göring, H. H., and Järvelä, I. (2008). Genome-wide linkage scan for loci of musical aptitude in Finnish families: evidence for a major locus at 4q22. *J. Med. Genet.* 45, 451–456. doi: 10.1136/jmg.2007.056366

Quintana, D. S., Rokicki, J., van der Meer, D., Alnæs, D., Kaufmann, T., Córdova-Palomera, A., et al. (2019). Oxytocin pathway gene networks in the human brain. *Nat. Commun.* 10:668. doi: 10.1038/s41467-019-08503-8

Quintin, E. M. (2019). Music-evoked reward and emotion: relative strengths and response to intervention of people with ASD. *Front. Neural Circuits* 13:49. doi: 10.3389/fncir.2019.00049

Raam, T. (2020). Oxytocin-sensitive neurons in prefrontal cortex gate social recognition memory. *J. Neurosci.* 40, 1194–1196. doi: 10.1523/jneurosci.1348-19.2019

Reddish, P., Fischer, R., and Bulbalia, L. (2013). Let's dance together: synchrony, shared intentionality and cooperation. *PLoS One* 8:e71182. doi: 10.1371/journal.pone.0071182

Reiss, A. B., Glass, D. S., Lam, E., Glass, A. D., De Leon, J., and Kasselman, L. J. (2019). Oxytocin: potential to mitigate cardiovascular risk. *Peptides* 117:170089. doi: 10.1016/j.peptides.2019.05.001

Richter, J., and Ostovar, R. (2016). "It don't mean a thing if it ain't got that swing"-an alternative concept for understanding the evolution of dance and music in human beings. *Front. Hum. Neurosci.* 10:485. doi: 10.3389/fnhum.2016.00485

Rilling, J. K., Chen, X., Chen, X., and Haroon, E. (2018). Intranasal oxytocin modulates neural functional connectivity during human social interaction. *Am. J. Primatol.* 80:e22740. doi: 10.1002/ajp.22740

Rilling, J. K., DeMarco, A. C., Hackett, P. D., Chen, X., Gautam, P., Stair, S., et al. (2014). Sex differences in the neural and behavioral response to intranasal oxytocin and vasopressin during human social interaction. *Psychoneuroendocrinology* 39, 237–248. doi: 10.1016/j.psyneuen.2013.09.022

Rilling, J. K., DeMarco, A. C., Hackett, P. D., Thompson, R., Ditzen, B., Patel, R., et al. (2012). Effects of intranasal oxytocin and vasopressin on cooperative behavior and associated brain activity in men. *Psychoneuroendocrinology* 37, 447–461. doi: 10.1016/j.psyneuen.2011.07.013

Rilling, J. K., King-Cass, B., and Sanfey, A. G. (2008). The neurobiology of social decision-making. *Curr. Opin. Neurobiol.* 18, 159–165. doi: 10.1016/j.conb.2008.06.003

Rilling, J. K., and Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annu. Rev. Psychol.* 62, 23–48. doi: 10.1146/annurev.psych.121208.131647

Rodrigues, S. M., Saslow, L. R., Garcia, N., John, O. P., and Keltner, D. (2009). Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans. *Proc. Natl. Acad. Sci. U S A* 106, 21437–21441. doi: 10.1073/pnas.0909579106

Rogers, C. N., Ross, A. P., Sahu, S. P., Siegel, E. R., Dooyema, J. M., Cree, M. A., et al. (2018). Oxytocin- and arginine vasopressin-containing fibers in the cortex of humans, chimpanzees and rhesus macaques. *Am. J. Primatol.* 80:e22875. doi: 10.1002/ajp.22875

Rushworth, M. F. S., Noonan, M. P., Boorman, E. D., Walton, M. E., and Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron* 70, 1054–1069. doi: 10.1016/j.neuron.2011.05.014

Sabino, A. D. V., Chagas, M. H. N., and Osório, F. L. (2020). Acute effects of oxytocin in music performance anxiety: a crossover, randomized, placebo-controlled trial. *Psychopharmacology* 237, 1757–1767. doi: 10.1007/s00213-020-05493-0

Sachs, M. E., Habibi, A., and Damasio, A. (2018). Reflections on music, affect and sociality. *Prog. Brain Res.* 237, 153–172. doi: 10.1016/bs.pbr.2018.03.009

Sachs, M. E., Habibi, A., Damasio, A., and Kaplan, J. T. (2019). Dynamic intersubject neural synchronization reflects affective responses to sad music. *NeuroImage* 218:116512. doi: 10.1016/j.neuroimage.2019.116512

Salighedar, R., Erfanparast, A., Tamaddonfard, E., and Soltanalinejad, F. (2019). Medial prefrontal cortex oxytocin-opioid receptors interaction in spatial memory processing in rats. *Physiol. Behav.* 209:112599. doi: 10.1016/j.physbeh.2019.112599

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion in music. *Nat. Neurosci.* 14, 257–264. doi: 10.1038/nn.2726

Sánchez-Vidaña, D. I., Chan, N. M., Chan, A. H., Hui, K. K., Lee, S., Chan, H. Y., et al. (2016). Repeated treatment with oxytocin promotes hippocampal cell proliferation, dendritic maturation and affects socio-emotional behavior. *Neuroscience* 333, 65–77. doi: 10.1016/j.neuroscience.2016.07.005

Särkämö, T., and Sihvonen, A. J. (2018). Golden oldies and silver brains: deficits, preservation, learning and rehabilitation effects of music in ageing-related neurological disorders. *Cortex* 109, 104–123. doi: 10.1016/j.cortex.2018.08.034

Schellenberg, E. G., Corrigall, K. A., Dys, S. P., and Malti, T. (2015). Group music training and children's prosocial skills. *PLoS One* 10:e0141449. doi: 10.1371/journal.pone.0141449

Schiller, B., Domes, G., and Heinrichs, M. (2020). Oxytocin changes behavior and spatio-temporal brain dynamics underlying inter-group conflict in humans. *Eur. Neuropsychopharmacol.* 31, 119–130. doi: 10.1016/j.euroneuro.2019.12.109

Schladt, T. M., Nordmann, G. C., Emilius, R., Kudielka, B. M., de Jong, T. R., and Neumann, I. D. (2017). Choir versus solo singing: effects on mood and salivary oxytocin and cortisol concentrations. *Front. Hum. Neurosci.* 11:430. doi: 10.3389/fnhum.2017.00430

Schlaug, G., Forgeard, M., Zhu, L., Norton, A., Norton, A., and Winner, E. (2009). Training-induced neuroplasticity in young children. The neurosciences and music III - disorders and plasticity. *Ann. N Y Acad. Sci.* 1169, 205–208. doi: 10.1111/j.1749-6632.2009.04842.x

Schneider, I., Schmitgen, M. M., Boll, S., Roth, C., Nees, F., Usai, K., et al. (2020). Oxytocin modulates intrinsic neural activity in patients with chronic low back pain. *Eur. J. Pain* 24, 945–955. doi: 10.1002/ejp.1543

Seltzer, L. J., Ziegler, T. E., and Pollak, S. D. (2010). Social vocalizations can release oxytocin in humans. *Proc. Biol. Sci.* 277, 2661–2666. doi: 10.1098/rspb.2010.0567

Shalvi, S., and De Dreu, C. K. (2014). Oxytocin promotes group-serving dishonesty. *Proc. Natl. Acad. Sci. U S A* 111, 5503–5507. doi: 10.1073/pnas.1400724111

Shamay-Tsoory, S. G., and Abu-Akel, A. (2016). The social salience hypothesis of oxytocin. *Biol. Psych.* 79, 194–202. doi: 10.1016/j.biopsych.2015.07.020

Shany, O., Singer, N., Gold, B. P., Jacoby, N., Tarrasch, R., Hendler, T., et al. (2019). Surprise-related activation in the nucleus accumbens interacts with music-induced pleasantness. *Soc. Cogn. Affect. Neurosci.* 14, 459–470. doi: 10.1093/scan/nsz019

Sheldon, S., Fenerci, C., and Gurguryan, L. (2019). A neurocognitive perspective on the forms and functions of autobiographical memory retrieval. *Front. Syst. Neurosci.* 13:4. doi: 10.3389/fnsys.2019.00004

Sicorello, M., Dieckmann, L., Moser, D., Lux, V., Luhmann, M., Schlotz, W., et al. (2020). Oxytocin and the stress buffering effect of social company: a genetic study in daily life. *Soc. Cogn. Affect. Neurosci.* 15, 293–301. doi: 10.1093/scan/nsaa034

Slattery, C. F., Agustus, J. L., Paterson, R. W., McCallion, O., Foulkes, A. J. M., Macpherson, K., et al. (2019). The functional neuroanatomy of musical memory in Alzheimer's disease. *Cortex* 115, 357–370. doi: 10.1016/j.cortex.2019.02.003

Sleiman, S. F., Henry, J., Al-Haddad, R., El Hayek, L., Abou Haidar, E., Stringer, T., et al. (2016). Exercise promotes the expression of brain derived neurotrophic factor (BDNF) through the action of the ketone body β-hydroxybutyrate. *Elife* 5:e15092. doi: 10.7554/elife.15092

Smith, C. J. W., Di Benedictis, B. T., and Veenema, A. H. (2019). Comparing vasopressin and oxytocin fiber and receptor density patterns in the social behavior neural network: implications for cross-system signaling. *Front. Neuroendocrinol.* 53:100737. doi: 10.1016/j.yfrne.2019.02.001

Snyder, J. S. (2019). Recalibrating the relevance of adult neurogenesis. *Trends Neurosci.* 42, 164–178. doi: 10.1016/j.tins.2018.12.001

Snyder, J. S., and Drew, M. R. (2020). Functional neurogenesis over the years. *Behav. Brain Res.* 382:112470. doi: 10.1016/j.bbr.2020.112470

Song, Z., and Albers, H. E. (2018). Cross-talk among oxytocin and arginine-vasopressin receptors: relevance for basic and clinical studies of the brain and periphery. *Front. Neuroendocrinol.* 51, 14–24. doi: 10.1016/j.yfrne.2017.10.004

Sorrells, S. F., Paredes, M. F., Cebrian-Silla, A., Sandoval, K., Qi, D., Kelley, K. W., et al. (2018). Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults. *Nature* 555, 377–381. doi: 10.1038/nature25975

Spalding, K. L., Bergmann, O., Alkass, K., Bernard, S., Salehpour, M., Huttner, H. B., et al. (2013). Dynamics of hippocampal neurogenesis in adult humans. *Cell* 153, 1219–1227. doi: 10.1016/j.cell.2013.05.002

Stephan, J. S., and Sleiman, S. F. (2019). Exercise factors as potential mediators of cognitive rehabilitation following traumatic brain injury. *Curr. Opin. Neurol.* 32, 808–814. doi: 10.1097/wco.0000000000000754

Stewart, N. A. J., and Lonsdale, A. J. (2016). It's better together: the psychological benefits of singing in a choir. *Psychol. Music* 44, 1240–1254. doi: 10.1177/0305735615624976

Stillman, C. M., Cohen, J., Lehman, M. E., and Erickson, K. I. (2016). Mediators of physical activity on neurocognitive function: a review at multiple levels of analysis. *Front. Hum. Neurosci.* 10:626. doi: 10.3389/fnhum.2016.00626

Strathearn, L., Fonagy, P., Amico, J., and Montague, P. R. (2009). Adult attachment predicts maternal brain and oxytocin response to infant cues. *Neuropsychopharmacol.* 34, 2655–2666. doi: 10.1038/npp.2009.103

Strathearn, L., Iyengar, U., Fonagy, P., and Kim, S. (2012). Maternal oxytocin response during mother-infant interaction: associations with adult temperament. *Horm. Behav.* 61, 429–435. doi: 10.1016/j.yhbeh.2012.01.014

Strauss, G. P., Chapman, H. C., Keller, W. R., Koenig, J. I., Gold, J. M., Carpenter, W. T., et al. (2019). Endogenous oxytocin levels are associated with impaired social cognition and neurocognition in schizophrenia. *J. Psychiatr. Res.* 112, 38–43. doi: 10.1016/j.jpsychires.2019.02.017

Suzuki, S., Fujisawa, T. X., Sakakibara, N., Fujioka, T., Takiguchi, S., and Tomoda, A. (2020). Development of social attention and oxytocin levels in maltreated children. *Sci. Rep.* 10:7407. doi: 10.1038/s41598-020-64297-6

Tabak, B. A., Teed, A. R., Castle, E., Dutcher, J. M., Meyer, M. L., Bryan, R., et al. (2019). Null results of oxytocin and vasopressin administration across a range of social cognitive and behavioral paradigms: evidence from a randomized controlled trial. *Psychoneuroendocrinology* 107, 124–132. doi: 10.1016/j.psyneuen.2019.04.019

Tachibana, A., Noah, J. A., Ono, Y., Taguchi, D., and Ueda, S. (2019). Prefrontal activation related to spontaneous creativity with rock music improvisation: a functional near-infrared spectroscopy study. *Sci. Rep.* 9:16044. doi: 10.1038/s41598-019-52348-6

Talamini, F., Altoè, G., Carretti, B., and Grassi, M. (2017). Musicians have better memory than nonmusicians: a meta-analysis. *PLoS One* 2:e0186773. doi: 10.1371/journal.pone.0186773

Tan, Y., Singhal, S. M., Harden, S. W., Cahill, K. M., Nguyen, D. M., Colon-Perez, L. M., et al. (2019). Oxytocin receptors are expressed by glutamatergic prefrontal cortical neurons that selectively modulate social recognition. *J. Neurosci.* 39, 3249–3263. doi: 10.1523/jneurosci.2944-18.2019

Tarr, B., Launay, J., Cohen, E., and Dunbar, R. (2015). Synchrony and exertion during dance independently raise pain threshold and encourage social bonding. *Biol. Lett.* 11:20150767. doi: 10.1098/rsbl.2015.0767

Tarr, B., Launay, J., and Dunbar, R. I. M. (2014). Music and social bonding: "self-other" merging and neurohormonal mechanisms. *Front. Psychol.* 5:1096. doi: 10.3389/fpsyg.2014.01096

Tarr, B., Launay, J., and Dunbar, R. I. (2016). Silent disco: dancing in synchrony leads to elevated pain thresholds and social closeness. *Evol. Hum. Behav.* 37, 343–349. doi: 10.1016/j.evolhumbehav.2016.02.004

Taylor, S. E., Saphire-Bernstein, S., and Seeman, T. E. (2010). Are plasma oxytocin in women and plasma vasopressin in men biomarkers of distressed pair-bond relationships? *Psychol. Sci.* 21, 3–7. doi: 10.1177/0956797609356507

Ten Velden, F. S., Daughters, K., and De Dreu, C. K. W. (2017). Oxytocin promotes intuitive rather than deliberated cooperation with the in-group. *Horm. Behav.* 92, 164–171. doi: 10.1016/j.yhbeh.2016.06.005

Tillman, R., Gordon, I., Naples, A., Rolison, M., Leckman, J. F., Feldman, R., et al. (2019). Oxytocin enhances the neural efficiency of social perception. *Front. Hum. Neurosci.* 13:71. doi: 10.3389/fnhum.2019.00071

Tirko, N. N., Eyring, K. W., Carcea, I., Mitre, M., Chao, M. V., Froemke, R. C., et al. (2018). Oxytocin transforms firing mode of CA2 hippocampal neurons. *Neuron* 100, 593.e3–608.e3. doi: 10.1016/j.neuron.2018.09.008

Tobin, M. K., Musaraca, K., Disouky, A., Shetti, A., Bheri, A., Honer, W. G., et al. (2019). Human hippocampal neurogenesis persists in aged adults and Alzheimer's disease patients. *Cell Stem Cell* 24, 974.e3–982.e3. doi: 10.1016/j.stem.2019.05.003

Toda, T., and Gage, F. H. (2018). Review: adult neurogenesis contributes to hippocampal plasticity. *Cell Tissue Res.* 373, 693–709. doi: 10.1007/s00441-017-2735-4

Tollefson, S., Himes, M., and Narendran, R. (2017). Imaging corticotropin-releasing-factor and nociceptin in addiction and PTSD models. *Int. Rev. Psych.* 29, 567–579. doi: 10.1080/09540261.2017.1404445

Tomizawa, K., Iga, N., Lu, Y.-F., Moriwaki, A., Matsushita, M., Li, S.-T., et al. (2003). Oxytocin improves long-lasting spatial memory during motherhood through MAP kinase cascade. *Nat. Neurosci.* 6, 384–390. doi: 10.1038/nn1023

Tops, M., Huffmeijer, R., Linting, M., Grewen, K. M., Light, K. C., Koole, S. L., et al. (2013). The role of oxytocin in familiarization-habituation responses to social novelty. *Front. Psychol.* 4:761. doi: 10.3389/fpsyg.2013.00761

Torrisi, S. A., Leggio, G. M., Drago, F., and Salamone, S. (2019). Therapeutic challenges of post-traumatic stress disorder: focus on the dopaminergic system. *Front. Pharmacol.* 10:404. doi: 10.3389/fphar.2019.00404

Tost, H., Kolachana, B., Hakimi, S., Lemaitre, H., Verchinski, B. A., Mattay, V. S., et al. (2010). A common allele in the oxytocin receptor gene (OXTR) impacts prosocial temperament and human hypothalamic-limbic structure and function. *Proc. Natl. Acad. Sci. U S A* 107, 13936–13941. doi: 10.1073/pnas.1003296107

Triana-Del Río, R., van den Burg, E., Stoop, R., and Hegoburu, C. (2019). Acute and long-lasting effects of oxytocin in cortico-limbic circuits: consequences for fear recall and extinction. *Psychopharmacology* 236, 339–354. doi: 10.1007/s00213-018-5030-5

Ukkola, L. T., Onkamo, P., Raijas, P., Karma, K., and Järvelä, I. (2009). Musical aptitude is associated with AVPR1A-haplotypes. *PLoS One* 4:e5534. doi: 10.1371/journal.pone.0005534

Ukkola-Vuoti, L., Oikkonen, J., Onkamo, P., Karma, K., Raijas, P., and Järvelä, I. (2011). Association of the arginine vasopressin receptor 1A (AVPR1A) haplotypes with listening to music. *J. Hum. Genet.* 56, 324–329. doi: 10.1038/jhg.2011.13

Ulmer-Yaniv, A., Avitsur, R., Kanat-Maymon, Y., Schneiderman, I., Zagoory-Sharon, O., and Feldman, R. (2016). Affiliation, reward and immune biomarkers coalesce to support social synchrony during periods of bond formation in humans. *Brain Behav. Immun.* 56, 130–139. doi: 10.1016/j.bbi.2016.02.017

Valstad, M., Alvares, G. A., Egknud, M., Matziorinis, A. M., Andreassen, O. A., Westlye, L. T., et al. (2017). The correlation between central and peripheral oxytocin concentrations: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 78, 117–124. doi: 10.1016/j.neubiorev.2017.04.017

Van den Burg, E. H., Stindl, J., Grund, T., Neumann, I. D., and Strauss, O. (2015). Oxytocin stimulates extracellular Ca2+ influx through TRPV2 channels in hypothalamic neurons to exert its anxiolytic effects. *Neuropsychopharmacol.* 40, 2938–2947. doi: 10.1038/npp.2015.147

Van der Heijden, M. J. E., Araghi, S., van Dijk, M., Jeekel, J., and Hunink, M. G. M. (2015). The effects of perioperative music interventions in pediatric surgery: a systematic review and meta-analysis of randomized controlled trials. *PLoS One* 10:e0133608. doi: 10.1371/journal.pone.0133608

Veenema, Y. H., and Neumann, I. D. (2018). Central vasopressin and oxytocin release: regulation of complex social behaviours. *Prog. Brain Res.* 170, 261–276. doi: 10.1016/S0079-6123(08)00422-6

Völlm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., et al. (2006). Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *NeuroImage* 29, 90–98. doi: 10.1016/j.neuroimage.2005.07.022

Wagner, U., and Echterhoff, G. (2018). When does oxytocin affect human memory encoding? The role of social context and individual attachment style. *Front. Hum. Neurosci.* 12:349. doi: 10.3389/fnhum.2018.00349

Walker, E., Ploubidis, G., and Fancourt, D. (2019). Social engagement and loneliness are differentially associated with neuro-immune markers in older age: time-varying associations from the English Longitudinal Study of Ageing. *Brain Behav. Immun.* 82, 224–229. doi: 10.1016/j.bbi.2019.08.189

Wang, P., Wang, S. C., Yang, H., Lv, C., Jia, S., Liu, X., et al. (2019). Therapeutic ootential of oxytocin in atherosclerotic cardiovascular disease: mechanisms and signaling pathways. *Front. Neurosci.* 13:454. doi: 10.3389/fnins.2019.00454

Wang, D., Yan, X., Li, M., and Ma, Y. (2017). Neural substrates underlying the effects of oxytocin: a quantitative meta-analysis of pharmaco-imaging studies. *Soc. Cogn. Affect. Neurosci.* 12, 1565–1573. doi: 10.1093/scan/nsx085

Weinstein, D., Launay, J., Pearce, E., Dunbar, R. I., and Stewart, L. (2016). Group music performance causes elevated pain thresholds and social bonding in small and large groups of singers. *Evol. Hum. Behav.* 37, 152–158. doi: 10.1016/j.evolhumbehav.2015.10.002

Weisman, O., Zagoory-Sharon, O., Schneiderman, I., Gordon, I., and Feldman, R. (2013). Plasma oxytocin distributions in a large cohort of women and men and their gender specific associations with anxiety. *Psychoneuroendocrinology* 38, 694–701. doi: 10.1016/j.psyneuen.2012.08.011

Wu, H., Feng, C., Lu, X., Liu, X., and Liu, Q. (2020). Oxytocin effects on the resting-state mentalizing brain network. *Brain Imaging Behav.* doi: 10.1007/s11682-019-00205-5 [Epub ahead of print].

Xu, L., Becker, B., Luo, R., Zheng, X., Zhao, W., Zhang, Q., et al. (2020). Oxytocin amplifies sex differences in human mate choice. *Psychoneuroendocrinology* 112:104483. doi: 10.1016/j.psyneuen.2019.104483

Yamasue, H., and Domes, G. (2018). Oxytocin and autism spectrum disorders. *Curr. Top. Behav. Neurosci.* 35, 449–465. doi: 10.1007/7854_2017_24

Yoon, S., and Kim, Y. K. (2020). The role of the oxytocin system in anxiety disorders. *Adv. Exp. Med. Biol.* 1191, 103–120. doi: 10.1007/978-981-32-9705-0_7

Yoshida, M., Takayanagi, Y., Inoue, K., Kimura, T., Young, L. J., Onaka, T., et al. (2009). Evidence that oxytocin exerts anxiolytic effects *via* oxytocin receptor expressed in serotonergic neurons in mice. *J. Neurosci.* 29, 2259–2271. doi: 10.1523/jneurosci.5593-08.2009

Young, W. S., and Song, J. (2020). Characterization of oxytocin receptor expression within various neuronal populations of the mouse dorsal hippocampus. *Front. Mol. Neurosci.* 13:40. doi: 10.3389/fnmol.2020.00040

Yuhi, T., Kyuta, H., Mori, H. A., Murakami, C., Furuhara, K., Okuno, M., et al. (2017). Salivary oxytocin concentration changes during a group drumming intervention for maltreated school children. *Brain Sci.* 7:E152. doi: 10.3390/brainsci7110152

Zak, P. J., and Berroza, J. A. (2013). The neurobiology of collective action. *Front. Neurosci.* 7:211. doi: 10.3389/fnins.2013.00211

Zatorre, R. J., and Salimpoor, V. N. (2013). From perception to pleasure: music and its neural substrates. *Proc. Natl. Acad. Sci. U S A* 110, 10430–10437. doi: 10.1073/pnas.1301228110

Zhang, K., Li, G., Wang, L., Cao, C., Fang, R., Luo, S., et al. (2019). An epistasis between dopaminergic and oxytocinergic systems confers risk of post-traumatic stress disorder in a traumatized Chinese cohort. *Sci. Rep.* 9:19252. doi: 10.1038/s41598-019-55936-8

Zhang, S., Liu, D., Ye, D., Li, H., and Chen, F. (2017). Can music-based movement therapy improve motor dysfunction in patients with Parkinson's disease? Systematic review and meta-analysis. *Neurol. Sci.* 38, 1629–1636. doi: 10.1007/s10072-017-3020-8

Zhang, T. Y., Shahrokh, D., Hellstrom, I. C., Wen, X., Diorio, J., Breuillaud, L., et al. (2020). Brain-derived neurotrophic factor in the nucleus accumbens mediates individual differences in behavioral responses to a natural, social reward. *Mol. Neurobiol.* 57, 290–301. doi: 10.1007/s12035-019-01699-2

Zilidou, V. I., Frantzidis, C. A., Romanopoulou, E. D., Paraskevopoulos, E., Douka, S., and Bamidis, P. D. (2018). Functional re-organization of cortical networks of senior citizens after a 24-week traditional dance program. *Front. Aging Neurosci.* 10:422. doi: 10.3389/fnagi.2018.00422

**Conflict of Interest**: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 31 August 2020
doi: 10.3389/fpsyg.2020.02110

# The Impact of Emotion on Musical Long-Term Memory

*Clémence Nineuil[1], Delphine Dellacherie[1,2] and Séverine Samson[1,3]**

[1]Univ. Lille, ULR 4072 - PSITEC - Psychologie: Interactions, Temps, Emotions, Cognition, Lille, France,
[2]Department of Pediatric Neurology, University Hospital of Lille, Lille, France, [3]AP-HP, GH Pitié-Salpêtrière-Charles Foix, Unité d'Epileptologie, Paris, France

The influence of emotional dimensions such as arousal and valence on memory has been a topic of particularly intense inquiry. As stimuli go, music is capable of provoking strong emotional responses from listeners, which can in turn influence memory. However, few studies have examined the effect of musical emotions on memory, and even fewer the effect of valence and arousal. In order to shed light on the ways in which emotional dimensions affect musical memory as study-test delay intervals increase, we tested recognition after a short delay and after a long delay. In line with the literature, we hypothesized an emotional enhancement of music memory induced by post-encoding processes leading to better recognition of musical excerpts in delayed condition, as compared to the immediate condition. The effects of arousal and valence were expected to become exaggerated after a long delay. We also predicted that the two emotional dimensions would be differently affected by the study-test intervals. Our results showed that the emotional enhancement of memory depends upon the valence, with remembering of positive and negative stimuli being differently affected by the duration of the study-test delay interval. Furthermore, our data demonstrated that musical excerpts were better recognized after a long delay than after a short delay, illustrating that memory consolidation for musical information is taking place during the long study-test interval. Moreover, musical memory consolidation is strongly related to the characteristics of the positive valence, which have been discussed in relation to its pleasantness. This original finding provides new insights into the modulatory effects of emotional valence on memory consolidation and could offer promising therapeutic possibilities for the rehabilitation of memory disorders.

**Keywords: emotion, musical memory, consolidation, valence, arousal**

OPEN ACCESS

***Edited by:***
*Franco Delogu,*
*Lawrence Technological University, United States*

***Reviewed by:***
*Frank A. Russo,*
*Ryerson University, Canada*
*José Lino Oliveira Bueno,*
*University of São Paulo, Brazil*

****Correspondence:***
*Séverine Samson*
*severine.samson@univ-lille.fr*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology*

***Received:*** *30 April 2020*
***Accepted:*** *29 July 2020*
***Published:*** *31 August 2020*

***Citation:***
*Nineuil C, Dellacherie D and Samson S (2020) The Impact of Emotion on Musical Long-Term Memory. Front. Psychol. 11:2110. doi: 10.3389/fpsyg.2020.02110*

## INTRODUCTION

Emotional events from our life are more likely to be later recollected than similar, non-emotional events. This emotional enhancement of memory has been extensively studied using words and pictures (for review, see Yonelinas and Ritchey, 2015). Fewer studies have however examined the influence of emotion on musical memory. And yet, the emotional power of music is well-established (Koelsch, 2014), and musical memory may in fact benefit from such emotional enhancements, explaining why certain pieces of music frequently become unforgettable. Is this related to the particularly efficient post-encoding processes involved in memorizing musical information? In order to investigate this question, we examined musical memory abilities in adult listeners by manipulating the emotional characteristics of musical excerpts and the lengths of time separating the study phase and a recognition test.

Emotion theorists often take the position that affective experience can be described according to two orthogonal dimensions, namely arousal and valence (Russell, 1980). Arousal refers to a continuum ranging from calm to excitement, whereas valence is measured along a continuum ranging from positive to negative. The impact of these emotional dimensions on different forms of memory, including declarative (explicit) memory, has been explored at great length. In particular, evidence has shown that high-arousal information is better remembered than low-arousal information (for review, see Hamann, 2001) in immediate (Bradley et al., 1992) or delayed memory tests administered after 1 day (Hu et al., 2006), a year (Bradley et al., 1992) or several years (Wagner et al., 2006). By increasing attention and elaboration at time of encoding, high-arousal stimuli are thought to be more deeply processed than low-arousal stimuli. Their memory trace would be subsequently enhanced by post-encoding processes, including stress hormone release, leading to better consolidation over a long period of time (Hamann, 2001; Sharot and Phelps, 2004; Payne et al., 2008; Sharot and Yonelinas, 2008). Emotional valence has also been reported to influence memory, with stimuli with a negative or positive valence being better memorized than neutral stimuli (Ely et al., 1999; Cabeza and Dolcos, 2002; Mickley and Kensinger, 2008). Furthermore, remembering negative information yielded even higher memory performance than remembering positive information, whether using pictures (Comblain et al., 2004; Waring and Kensinger, 2009) or words (Inaba et al., 2005; for review, see Kensinger, 2007). This effect would persist over days (Denburg et al., 2003; Kensinger, 2007; Waring and Kensinger, 2009) or weeks (Ely et al., 1999; Ochsner, 2000; Pierce and Kensinger, 2011), suggesting a negativity bias in emotional memory, at least in young adults. Few studies have conjointly manipulated arousal and valence within a single experiment. In some cases, valence and arousal had their own respective effects on memory. Arousing pictures were better remembered than non-arousing ones and negative pictures were better memorized than positive ones (Comblain et al., 2004). In other investigations, the effect of arousal on memory performance was found to be modulated by valence (Kensinger, 2008; Li et al., 2018). However, although studies show interaction effects, the nature of the interaction differs. This interaction and its impact on memory performance are still debated in the behavioral literature (Ochsner, 2000; Dolcos et al., 2004). However, the joint effect of valence and arousal on recognition has been demonstrated in numerous neuroimaging studies, suggesting that these two dimensions influence memory through distinct neural mechanisms (Dolcos et al., 2004; Kensinger and Corkin, 2004). These results therefore justify the importance of considering these two dimensions in studies exploring the impact of emotions on memory.

Interestingly, prior investigations suggest that the emotional characteristics of stimuli may shape the retention of information differently over long delays, compared with short delays. In some studies, emotional enhancement was observed after a long delay, but not a short delay (Yonelinas and Ritchey, 2015). In other cases, the effects of arousal on memory were enhanced after a long delay, relative to a shorter one (Sharot and Phelps, 2004; Payne et al., 2008; Sharot and Yonelinas, 2008; Kensinger, 2009). Finally, Waring and Kensinger (2009) reported a complex interaction between all these factors in a visual scene recognition task. After a short delay, they showed that the level of arousal modulated the effect of valence on memory. Thus, an emotional enhancement of memory was observed in all conditions, except for positive low-arousal scenes, at least in young adults. After a long delay, the pattern of results was different. The enhancement of memory was greater for negative than for positive scenes, and greater for arousing than for non-arousing scenes underlying once more the negativity bias and the effect of arousal in emotional memory. Furthermore, the emotional enhancement effect was greater after a long delay than a short delay, emphasizing the role of post-encoding processes in memory consolidation of emotional information. Taken as a whole, extant literature underlines the complex interplay between emotional features of stimuli and study-test intervals, and how memory performance can be shaped by post-encoding actions.

An absence of agreement persists among musical domain scholars about how these factors contribute to the emotional memory for music. Indeed, as an enjoyable human activity that is present in all cultures, capable of generating strong and varied emotions, it seems to be a privileged medium for studying this effect. One series of studies used the exact same prototypal clips intended to express happy, fearful, peaceful, or sad emotions, emotional characteristics which had been validated previously (Gosselin et al., 2005; Vieillard et al., 2008). Musical recognition of these four emotional categories of computer-generated MIDI musical excerpts was generally assessed after a short delay (Aubé et al., 2013; Vieillard and Gilet, 2013; Narme et al., 2016), though sometimes after a longer delay of 24 h (Samson et al., 2009). Except for one study that did not report any difference in explicit recognition, at least in young adults (Narme et al., 2016), the other investigations showed better recognition of musical clips expressing fear and, to some extent, happiness (Samson et al., 2009; Aubé et al., 2013; Vieillard and Gilet, 2013), suggesting a benefit for high-arousal stimuli. However, this finding should be interpreted cautiously, as it could be explained by different levels of difficulty between recognition of high and low-arousing musical excerpts, or of each prototypical emotional category. As demonstrated by a control study (Samson et al., 2009), the perceptual distinctiveness (or dissimilarity) between the scary music was larger than the distinctiveness between the three other categories, explaining, at least in part, the superior recognition scores obtained with scary musical excerpts. However, the high recognition score obtained with happy musical excerpts, in particular after a 24-h delay, might nonetheless reflect the effect of arousal and/or valence on consolidation of musical memory, though it is too early to draw any firm conclusions.

Other studies reported in the literature have used instrumental classical music. Eschrich et al. (2005) examined the relationship between emotion and music recognition.

In this study, piano pieces by Bach were rated in terms of arousal (from very pacifying to very arousing) and valence (from negative to positive valence) during encoding. Recognition tests conducted 2 weeks later revealed that well-recognized pieces were associated with higher arousal ratings and received a higher positive valence rating. In a subsequent study, symphonic film music eliciting positive feelings (i.e., little positive to very positive) were repeated twice over different days (Eschrich et al., 2008). The authors confirmed an effect of valence on recognition with better memory for very positive music excerpts than for less positive ones, although no effect of arousal on recognition was obtained in this case. These findings suggest that emotional valence, as rated by the participants, appears to influence long-term memory for music, underlying once more the impact of positive valence music on memory consolidation. However, the lack of negative valence music in this study limits the interpretation of the results. By addressing again this question in a neuroimaging study, the authors failed to replicate their previous finding (Altenmüller et al., 2014). Participants' single exposure to the musical pieces during encoding may have been insufficient to induce a memory enhancement, thereby explaining the lack of behavioral results. Based on all these data, it remains unclear whether only positive valence or both positive and negative valence provides a memory advantage.

Prior investigations into emotional memory in music have assessed memory at only one delay. To our knowledge, in the music domain, only one study has examined the effects of delay interval upon emotional memory (Alonso et al., 2015). Making use of the parsimonious model of emotion, which defines emotional spectrum into two dimensions, its authors manipulated valence and arousal levels in symphonic musical excerpts. Participants were requested to rate these two emotional feelings induced by listening to music before taking two recognition memory tests administered immediately after the study and 24 h after the study session. The results indicated that arousal and valence interacted differently with memory performance at each study-test delay. In immediate recognition, the effect of valence varied as a function of arousal. Whereas valence did not interfere with the remembering of high-arousal excerpts, it did modulate recognition of low-arousal stimuli such that positive excerpts were better recognized than negative excerpts. In contrast, in the delayed condition, the results revealed no interaction between the two emotional dimensions. Only independent effects of arousal and valence were reported, such that high-arousal excerpts and negative excerpts were better memorized than low-arousal and positive ones, respectively, confirming the memory advantage for high arousing and negative stimuli already reported in non-musical domains (Ochsner, 2000; Waring and Kensinger, 2009). Unlike findings obtained with words and pictures, there was no loss of memory in delayed, as compared to immediate recognition, indicating no deleterious impact of delay interval in music. Yet this recognition test, frequently used in psychology, presents a methodological bias. By presenting the target stimuli once again in the delayed recognition test, this condition benefits from an additional exposure compared to the immediate condition. It is therefore difficult to disentangle the effect of delay from the effect had by number of presentations in post-encoding processes.

To overcome this methodological limit and to clarify the ways emotional dimensions (i.e., valence and arousal) affect musical memory as study-test delay intervals increase, we designed a new study using the symphonic musical excerpts selected by Alonso et al. (2015). However, we manipulated the study-test interval while keeping the number of exposures constant. For this purpose, the encoding phase was distributed over two distinct sessions, one session on day 1 allowing the first half of the target stimuli to be encoded and the other session on day 2 allowing the other half to be encoded. Immediately after this second session on day 2, a recognition test including all target stimuli mixed with foils was presented. The recognition of the target stimuli presented just before (on day 2) or after 24 h (on day 1) provided respective memory performances after a short delay and a long-delay retention without changing the number of exposures to the targets. In line with the literature, we hypothesized an emotional enhancement of music memory induced by post-encoding processes leading to better recognition of musical excerpts in delayed condition as compared to immediate one. The effects of arousal and valence should become exaggerated after a long delay. Finally, we also predicted that the two emotional dimensions would be differently affected by the study-test intervals.

## MATERIALS AND METHODS

### Participants

Eighty native French speakers took part in this study (mean age = 34.09 ± 7.20; 37 females and 43 men). In accordance with the Music Expertise Questionnaire (Ehrlé, 1998), 43 participants had no musical expertise and 37 were musicians. The quality and quantity of sleep was also assessed twice using the St. Mary's Hospital Sleep Questionnaire (Ellis et al., 1981; mean duration of sleep, night 1 = 7.48 ± 1.61, night 2 = 7.57 ± 1.41). No participants reported a history of psychiatric or neurological disorders, alcoholism, or present treatment with centrally acting medications. Participants were screened for the presence of mood disorders using the Profile of Mood State (POMS; Shacham, 1983; mean score for anxiety = 4.42 ± 3.57, depression = 2.65 ± 2.55, confusion = 2.00 ± 3.79, anger = 4.03 ± 3.15, tiredness = 6.15 ± 3.70, and vigor = 11.60 ± 3.55) and for the presence of attention deficit using standardized psychometric tests: the Paced Auditory Serial Addition Test (PASAT; Mazza and Naegele, 2004; mean score = 49.66 ± 6.15) and the Wechsler adult intelligence scale–Fourth Edition (WAIS-IV) coding subtest (Wechsler, 2014; mean scaled score = 11.25 ± 12.67). All participants have signed informed consent and the study was carried out following the Declaration of Helsinki principles.

### Stimuli

The musical material consisted of 32 symphonic excerpts with a duration of 5 s (±1-s fade and in fade out) already used in a previous study (Alonso et al., 2015). These excerpts were taken from different symphonies written by composers between 1830 and 1954 and were normalized to a maximal amplitude of 1.2 dB. To control the effect of familiarity, the most famous composers and symphonies of this period were excluded. Musical stimuli had been previously rated for stimulus valence (positive vs. negative) and arousal levels (high vs. low) to create four different emotional combinations in a two by two design: high-arousal and positive (A+; V+); high-arousal and negative (A+; V−); low-arousal and positive (A−; V+); and low-arousal and negative (A−; V+). We selected 16 target stimuli (four per emotional combination) and 16 distractors for the recognition test. Thus, to each target was matched to a distractor from the same emotional combination and melodic style, composed by the same composer or even excerpts from the same symphony as the target (the list of the symphonies is presented in Alonso et al., 2015).

## Procedure

Participants sat in front of a laptop in a quiet room and listened to the musical excerpts using stereophonic headphones. PsychoPy software was used to run the experiment and to record recognition ratings. All participants were tested individually over two different sessions, with the encoding phase being split over two consecutive days. On day 1, they listened to eight different musical excerpts (two stimuli per emotional combination). Participants were instructed to listen carefully to the musical excepts because their memory would be tested later (intentional encoding). To reinforce the encoding process, the studied items were presented three times in random order. Two questionnaires were administered: the St. Mary's Hospital Sleep Questionnaire (Ellis et al., 1981) and the POMS (Shacham, 1983).

On day 2, approximately 24 h later, each participant listened to eight additional musical excerpts following the same procedure used on day 1. As with the first encoding phase, the items were presented three times. On both days, participants completed questionnaires either before or after the encoding task in a randomized manner, on day 2 the St. Mary's Hospital Sleep Questionnaire (Ellis et al., 1981) and the Music Expertise Questionnaire (Ehrlé, 1998) were administered. Then, 15 min after the second encoding phase, the recognition phase began. Sixteen targets, including eight musical excerpts encoded on day 1 and eight musical excerpts encoded on day 2, were randomly intermixed with 16 distractors. Participants were told that items from the same-day study session and the study session a day earlier were intermixed with novel items. This recognition phase assessed memory recognition after short (15-min) and long (24-h) delays within a single test. After each presentation, participants had to decide whether they had already heard a given excerpt (yes/no) and to assign their responses confidence ratings (sure/unsure). Participants were not given a time limit to respond. During the 15-min delay, the participants performed standardized cognitive tests to assess auditory attention and working memory using the PASAT (Mazza and Naegele, 2004) and the WAIS–IV coding subtest (Wechsler, 2014).

## Data Analysis

Based on the accuracy of the responses and the confidence ratings, receiver operating characteristic (ROC; Yonelinas, 1994) curves have been calculated for each arousal and valence combination. Preliminary analyses were carried out to test the effects of age, sex, musical expertise and duration of sleep on the global recognition score. Then, a three-way repeated-measure analysis of variance (ANOVA) with arousal (high/low), valence (positive/negative), study-to-test delay period (short/long) as well as subsequent *post hoc* analysis was carried out.

# RESULTS

## Preliminary Analysis

These analyses showed there was no correlation between the recognition of the musical material and participants' age ($r = -0.201$; $p = 0.073$) or duration of sleep ($r = -0.075$; $p = 0.510$). Furthermore, we found sex and musical expertise not to have an effect on the recognition test (all $ps > 0.05$). Finally, the completion of the questionnaires before or after the encoding tasks had no effect on the musical recognition ($U = 744$; $df = 78$; $p = 0.593$).

## Recognition Task

The ANOVA with three within-subjects factors (i.e., valence, arousal, and delay) was carried out on the areas under the ROC curves. The analysis showed a valence by delay interaction ($F_{(1,79)} = 23.74$; $p < 0.001$; $\eta_p^2 = 0.231$), as depicted in **Figure 1**. *Post hoc* analyses showed that after a short delay, the mean area under the curve was higher for the negatively valenced (mean ± SD = 0.80 ± 0.24) than for positively valenced excerpts (mean ± SD = 0.74 ± 0.24; $p < 0.05$). In contrast, after a long delay, the mean area under the curve was higher for positively valenced (mean ± SD = 0.89 ± 0.18) than for negatively valenced excerpts (mean ± SD = 0.82 ± 0.21; $p < 0.002$).

This ANOVA also revealed a main effect of delay ($F_{(1,79)} = 43.63$; $p < 0.001$; $\eta_p^2 = 0.356$), with the mean area under the curve being significantly higher after long delay (mean ± SD = 0.85 ± 0.11) than after short delay (mean ± SD = 0.76 ± 0.13). No main other effect was had by arousal ($F_{(1,79)} = 0.08$; $p > 0.05$), valence ($F_{(1,79)} = 0.39$; $p > 0.05$), or interaction.

# DISCUSSION

The objective of this study was to investigate the impact of emotional dimensions on musical memory by comparing the effects of valence (i.e., positive and negative) and arousal (i.e., low and high) on the recognition of musical excerpts. To improve our understanding about how emotional dimensions (i.e., valence and arousal) affect musical memory as study-test delay intervals increase, we compared memory performance after a short (15 min) and a long (24 h) retention delay. We showed that the emotional enhancement of memory

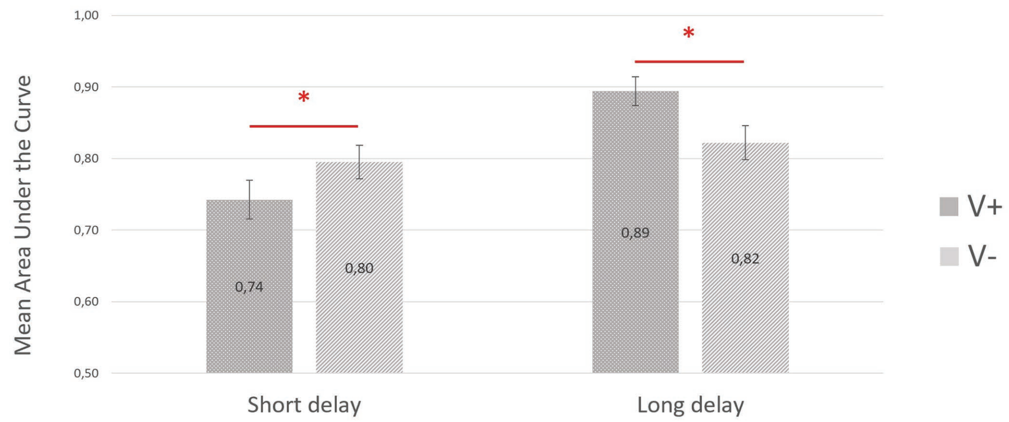



**FIGURE 1 |** Mean areas under the receiver operating characteristic (ROC) curve for the recognition of positive and negative musical excerpts as a function of delay (short vs. long). Origin is set at chance level (0.50). The error bars correspond to the standard error of the mean.

depends upon valence, with musical memory of positive and negative stimuli being differently affected by the duration of the study-test delay interval. Furthermore, our data demonstrated that musical excerpts were better recognized after a long delay than after a short delay, illustrating that memory consolidation for musical information is taking place during the long study-test interval.

The main finding borne out by our study is that emotion strengthens musical memory. More specifically, we showed that emotional valence differently affected the recognition of musical excerpts after a short- and a long-delay interval. Whereas remembering of musical excerpts is higher for negative than for positive stimuli after a short delay, the reverse was observed after a long delay. Thus, the memory performance for stimuli associated with positive valence was not only higher than for negative ones, but also improved after a 24-h delay, which was not the case for the negative stimuli. The fact that emotional enhancement on musical memory varies as a function of study-test delay intervals confirmed our predictions and is compatible with previously reported results (Alonso et al., 2015). It suggests that post-encoding processes, including consolidation that transforms newly formed memories from a fragile state to a more permanent state, modulated the effects of emotion upon musical memory. However, the discrepancies between the pattern of results obtained in the current study and that of our previous study with Alonso et al. (2015) raised several questions worth discussing.

By examining the impact of aging on emotional musical memory, Alonso et al. (2015) found that valence interacted with arousal when recognition was assessed after a short delay. Whereas valence had no impact on the recognition of high-arousal stimuli, it did affect low-arousal stimuli, with positive valence music being better remembered than negative valence music. After a long delay, only the effects of valence and arousal were demonstrated, with negative valence stimuli being better recognized than positive ones, and high-arousal stimuli better recognized than low-arousal stimuli. The different result profiles obtained after short and long-delay intervals in Alonso et al.'s study and in the present study are surprising, since both used the exact same set of stimuli. The apparent discrepancies between the studies might nonetheless be due to methodological differences. In Alonso et al. (2015), ratings of emotional dimensions were requested during the first exposure to the musical excerpts, whereas in the present study, no specific instructions were given at the time of encoding. By asking participants to rate valence and arousal, attention devoted to the emotional meaning of the stimuli might have induced deeper encoding processes than in the present study. As previously discussed in the literature (Everaert et al., 2011; Greenberg et al., 2012), when participants are particularly attentive to emotion, the effects of arousal on memory are more likely to appear. Such a memory enhancement for arousing stimuli can be even greater after long study-test delay periods than after short ones (Sharot and Phelps, 2004; Sharot and Yonelinas, 2008), at least when emotional stimuli (i.e., words or pictures) were compared with neutral stimuli. Taken as a whole, these results underline the importance of the depth of encoding in emotional memory. Unlike verbal or visual information, the concept of neutral emotion does not really exist in music, which is intrinsically emotional. Thus, music is defined as the language of emotion and is rarely perceived as non-emotional. Rather than "neutral" music, Dellacherie et al. (2008) have proposed "ambiguous" music, that is, when the emotional cues are insufficient to evoke a specific emotion. Since the concept of neutral emotion in music is questionable, we deliberately used only emotional musical excerpts in the current study as in Alonso et al. (2015). However, a fruitful avenue for future research may be to modify the encoding instructions in order to improve use of controlled cognitive mechanisms and provide an opportunity for sustained attention effects to influence memory. It is possible that under these conditions, an effect from arousal on musical memory would be noted, as in Alonso et al. (2015) and Eschrich et al. (2005), who both asked participants to judge arousal and valence at encoding during an intentional memory task.

Another important difference between the present investigation and Alonso et al. (2015) concerns procedure. In Alonso et al. (2015), all target stimuli were presented within a single session on day 1, followed by two recognition tests, one proposed after a short delay and the other after a long delay. In this case, as in many classical recognition paradigms, delayed recognition performance benefited not only from a longer consolidation time but also from an additional exposure to the target stimuli, relative to immediate recognition. To overcome this methodological bias in the present study, we split the presentation of the target stimuli into two study phase sessions, with half of them being presented on day 1 and the other half on day 2. The memory test presented on day 2 after the second study phase allowed us to test recognition after a short and a long delay while keeping the number of exposures constant. Instead of making participants learn the whole series of 16 target stimuli and apply them in a single block presentation, we distributed the 16 targets across two sessions, each with only eight targets. Finally, Alonso et al. (2015) investigated a group of young adults (mean age = 22 years) with a group of elderly people (mean age = 75 years). Since they found no effect of age on performance, they mixed all the 30 participants. As a result, their participants displayed a larger age range and were globally older than participants from the present study. Despite these methodological differences in terms of the number of items to be learned, the number of exposures, the different instructions and subjects' characteristics, it appears that negative valence stimuli were similarly recognized after long delay in both studies (ROC = 0.82 in the present study; ROC = 0.80 in Alonso et al., 2015). However, positive valence stimuli were better recognized in the present study (ROC = 0.89) than in Alonso et al. (2015) (ROC = 0.74). This finding firstly indicates that the consolidation process improves musical memory when the number of stimuli to be learned is limited. Secondly, the decrease of performance in recognizing positive valence stimuli in Alonso et al. (2015) can also be explained by the age of the participants. However, this hypothesis seems rather improbable since the age does not seem to affect the recognition of negative valence stimuli. Moreover, it would contradict the positivity bias found in the aging process, which is described in the literature (for review, see Reed et al., 2014). Finally, the advantage of positive valence over negative valence on recognition performance after a long-delay interval in the present study underlines the importance of emotional valence in memory consolidation of music. This final result seems to be in agreement with Eschrich et al. (2008) who showed a better recognition, after 48 h, of very positive music compared to less positive music. Thus, our study, by responding to the limitations of Alonso et al. (i.e., the difference in number of presentations, the heterogeneous population, and the small sample size) brings new evidence on the impact of the emotional dimensions and in particular the positive valence on the consolidation of musical memory.

The better consolidation of positive vs. negative stimuli reported here suggests that emotion does not have a uniform effect on memory. Moreover, it is noteworthy that previously reported studies also demonstrated a positive valence advantage on musical memory after long time periods (after a 24-h delay for Samson et al., 2009 and after several days or weeks for Eschrich et al., 2005, 2008). In these cases as well, memory was assessed at only one delay interval. Furthermore, this result confirms that memory enhancement is particularly pronounced over time (Yonelinas and Ritchey, 2015). While the memory traces of neutral information are gradually affected by a phenomenon of forgetfulness (Ebbinghaus, 1885), emotional stimuli are more resistant and even better recognized after a long period of time thanks to the consolidation process. Emotions generated by the musical material make musical memory more resistant to forgetting, a finding consistent with the literature on the emotional enhancement effect of memory (Jäncke, 2008). The novel aspect of this current study, however, is the finding that musical memory consolidation is strongly related to the characteristics of the positive valence, which is not the case in studies in the non-musical field. Using pictures and words, several studies demonstrated a negative valence effect on memory (for review, see Kensinger, 2007). Indeed, according to evolutionary theories, the main function of emotion would be to guide our behavior: fear, for example, allows us to avoid danger and prepares us to act (Darwin, 1872). It is therefore logical that our attention should focus more on the threatening elements, thus allowing us to better memorize negative information (Kensinger, 2007). Given that music is a predominantly hedonic activity, this negativity bias would not be found for musical stimuli.

Another particularity of the current study is to have used exclusively non-familiar musical excerpts to avoid any confounding factors related to familiarity. Indeed, when musical excerpts are known, such as the piano pieces by Bach used in the study by Eschrich et al. (2005), the emotion associated with the excerpts would be more often linked to an autobiographical event (Janata et al., 2007). As a result, these excerpts could lead participants to remember memories and convey more intense emotions than when faced with unfamiliar music. This could also explain the effect of the arousal found in the study by Eschrich et al. (2005) that we were not able to demonstrate in the current study. Although we were not in control of the encoding strategies used by the listeners, we can assume that performing an elaborative rehearsal with non-familiar music is more complex, and probably impossible. Unlike pictures or words, which are generally associated with preexisting knowledge, new pieces of music are difficult to relate to already existing information. In this case, difficulties in thinking about the meaning of an item to be remembered, or making connections between that item and prior knowledge, may have limited the use of elaborative rehearsal known to improve memory abilities. Therefore, when exposed to new music, listeners can only rely on their own emotional feeling, on the attractiveness or the pleasantness of the musical pieces or on their familiarity to the musical style rather than to a specific event. One recent study demonstrated musical memory to be strongly related to pleasantness (Ferreri and Rodriguez-Fornells, 2017). The authors showed that musical excerpts rated as very pleasant were better recognized than musical excerpts rated as less pleasant. Listeners may indeed experience greater pleasure from listening to positive valence music than negative valence music.

Based on all previously reported findings, we therefore propose that the time-dependent effect of emotion attributed to a process of emotional consolidation is mainly predicated on the positive or pleasurable experience of music. It remains to be clarified whether dispensing instructions to rate pleasantness will further improve memory consolidation for music. Moreover, a relationship between arousal and pleasure has already been discussed by Berlyne's hedonic model (Berlyne, 1971). According to this model, very low or high arousal values would lead to low pleasure whereas medium arousal values would lead to high pleasure (Berlyne, 1971). This suggests the existence of a complex relationship between arousal and pleasure that needs to be further investigated in the musical domain. Another point to be clarified in the future would be the impact of musical characteristics that might have contributed to emotion and to musical memory. For instance, it has been reported that variations in intrinsic musical characteristics such as timbre or tempo modulate musical memory (Halpern and Müllensiefen, 2008). Even if the musical excerpts used in our study were very homogeneous in terms of musical characteristics, such as rhythm, mode, and tempo since targets and distractors were selected from small set symphonies written by composers between 1830 and 1954, we cannot exclude the impact of intrinsic musical features in recognition.

To conclude, the present study succeeded in demonstrating the impact of emotion on musical long-term memory, thus improving our understanding of the specific role played by emotional valence and arousal on memory consolidation. The positivity effect observed in young adults after a long study-test interval may be related to the effect of pleasantness in remembering musical excerpts recently reported in the literature (Ferreri and Rodriguez-Fornells, 2017), lending credence to arguments in favor of considering memory differently according to valence (i.e., positive vs. negative). To our knowledge, our study is the first to demonstrate a clear improvement in remembering positive musical excerpts resulting from a consolidation process. Further study will be necessary to generalize this emotional enhancement of musical memory to other musical styles and different populations, from children to the elderly. If positive music has such a strong influence on our cognitive system and in particular on our memory abilities, it raises questions as to whether the memory-enhancing effect of emotional music could be used to enhance cognitive performance in general and clinical settings.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

CN, DD, and SS contributed to the design and implementation of the research, to the analysis of the results, and to the writing of the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

Univ. Lille, ULR 4072 - PSITEC - Psychologie: Interactions Temps, Émotions, Cognition, F-59000 Lille, France supports all publication fees (2950 USD). This work was supported by the Conseil Régional des Hauts-de-France to CN, the University of Lille to CN, and the Institut Universitaire de France to SS.

## ACKNOWLEDGMENTS

The authors are grateful to Margaux Boutin, Laura Creuza, Gretel Dutriaux and Laurie Piffeteau for their help in data collection.

## REFERENCES

Alonso, I., Dellacherie, D., and Samson, S. (2015). Emotional memory for musical excerpts in young and older adults. *Front. Aging Neurosci.* 7:23. doi: 10.3389/fnagi.2015.00023

Altenmüller, E. O., Siggel, S., Mohammadi, B., Samii, A., and Münte, T. F. (2014). Play it again sam: brain correlates of emotional music recognition. *Front. Psychol.* 5:114. doi: 10.3389/fpsyg.2014.00114

Aubé, W., Peretz, I., and Armony, J. L. (2013). The effects of emotion on memory for music and vocalisations. *Memory* 21, 981–990. doi: 10.1080/09658211.2013.770871

Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts.

Bradley, M. M., Greenwald, M. K., Petry, M. C., and Lang, P. J. (1992). Remembering pictures: pleasure and arousal in memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 379–390. doi: 10.1037//0278-7393.18.2.379

Cabeza, R., and Dolcos, F. (2002). Event-related potentials of emotional memory: encoding pleasant, unpleasant, and neutral pictures. *Cogn. Affect. Behav. Neurosci.* 2, 252–263. doi: 10.3758/CABN.2.3.252

Comblain, C., D'Argembeau, A., Van der Linden, M., and Aldenhoff, L. (2004). The effect of ageing on the recollection of emotional and neutral pictures. *Memory* 12, 673–684. doi: 10.1080/09658210344000477

Darwin, C. (1872). *The expression of the emotions in man and animals. 3rd Edn.* New York: Oxford University Press.

Dellacherie, D., Ehrlé, N., and Samson, S. (2008). Is the neutral condition relevant to study musical emotion in patients? *Music Percept.* 25, 285–294. doi: 10.1525/mp.2008.25.4.285

Denburg, N. L., Buchanan, T. W., Tranel, D., and Adolphs, R. (2003). Evidence for preserved emotional memory in normal older persons. *Emotion* 3, 239–253. doi: 10.1037/1528-3542.3.3.239

Dolcos, F., LaBar, K. S., and Cabeza, R. (2004). Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fMRI study. *NeuroImage* 23, 64–74. doi: 10.1016/j.neuroimage.2004.05.015

Ebbinghaus, H. (1885). *Über das gedächtnis: Untersuchungen zur experimentellen psychologie*. Leipzig: Duncker.

Ehrlé, N. (1998). Traitement temporel de l'information auditive et lobe temporal (Temporal processing of auditory information in mediotemporal lobe). dissertation/Ph.D. thesis. Reims, France: Université de Reims Champagne-Ardenne.

Ellis, B. W., Johns, M. W., Lancaster, R., Raptopoulos, P., Angelopoulos, N., and Priest, R. G. (1981). The St. Mary's Hospital Sleep Questionnaire: a study of reliability. *Sleep* 4, 93–97. doi: 10.1093/sleep/4.1.93

Ely, T. D., Grafton, S. T., Kilts, C. D., and Hamann, S. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nat. Neurosci.* 2, 289–293. doi: 10.1038/6404

Eschrich, S., Münte, T. F., and Altenmüller, E. O. (2005). Remember bach: an investigation in episodic memory for music. *Ann. N. Y. Acad. Sci.* 1060, 438–442. doi: 10.1196/annals.1360.045

Eschrich, S., Münte, T. F., and Altenmüller, E. O. (2008). Unforgettable film music: the role of emotion in episodic long-term memory for music. *BMC Neurosci.* 9:48. doi: 10.1186/1471-2202-9-48

Everaert, T., Spruyt, A., and De Houwer, J. (2011). On the (un) conditionality of automatic attitude activation: the valence proportion effect. *Can. J. Exp. Psychol.* 65, 125–132. doi: 10.1037/a0022316

Ferreri, L., and Rodriguez-Fornells, A. (2017). Music-related reward responses predict episodic memory performance. *Exp. Brain Res.* 235, 3721–3731. doi: 10.1007/s00221-017-5095-0

Gosselin, N., Peretz, I., Noulhiane, M., Hasboun, D., Beckett, D., Baulac, M., et al. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain* 128, 628–640. doi: 10.1093/brain/awh420

Greenberg, S. N., Tokarev, J., and Estes, Z. (2012). Affective orientation influences memory for emotional and neutral words. *Am. J. Psychol.* 125, 71–80. doi: 10.5406/amerjpsyc.125.1.0071

Halpern, A. R., and Müllensiefen, D. (2008). Effects of timbre and tempo change on memory for music. *Q. J. Exp. Psychol.* 61, 1371–1384. doi: 10.1080/17470210701508038

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends Cogn. Sci.* 5, 394–400. doi: 10.1016/S1364-6613(00)01707-1

Hu, P., Stylos-Allan, M., and Walker, M. P. (2006). Sleep facilitates consolidation of emotional declarative memory. *Psychol. Sci.* 17, 891–898. doi: 10.1111/j.1467-9280.2006.01799.x

Inaba, M., Nomura, M., and Ohira, H. (2005). Neural evidence of effects of emotional valence on word recognition. *Int. J. Psychophysiol.* 57, 165–173. doi: 10.1016/j.ijpsycho.2005.01.002

Janata, P., Tomic, S. T., and Rakowski, S. K. (2007). Characterisation of music-evoked autobiographical memories. *Memory* 15, 845–860. doi: 10.1080/09658210701734593

Jäncke, L. (2008). Music, memory and emotion. *J. Biol.* 7:21. doi: 10.1186/jbiol82

Kensinger, E. A. (2007). Negative emotion enhances memory accuracy: behavioral and neuroimaging evidence. *Curr. Dir. Psychol. Sci.* 16, 213–218. doi: 10.1111/j.1467-8721.2007.00506.x

Kensinger, E. A. (2008). Age differences in memory for arousing and nonarousing emotional words. *J. Gerontol. Ser. B* 63, P13–P18. doi: 10.1093/geronb/63.1.P13

Kensinger, E. A. (2009). Remembering the details: effects of emotion. *Emot. Rev.* 1, 99–113. doi: 10.1177/1754073908100432

Kensinger, E. A., and Corkin, S. (2004). Two routes to emotional memory: distinct neural processes for valence and arousal. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3310–3315. doi: 10.1073/pnas.0306408101

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666

Li, X., Li, X., Chen, S., Zhu, J., Wang, H., Tian, Y., et al. (2018). Effect of emotional enhancement of memory on recollection process in young adults: the influence factors and neural mechanisms. *Brain Imaging Behav.* 14, 119–129. doi: 10.1007/s11682-018-9975-0

Mazza, S., and Naegle, B. (2004). *Test d'attention soutenue : PASAT modifié*. Marseille: Solal.

Mickley, K. R., and Kensinger, E. A. (2008). Emotional valence influences the neural correlates associated with remembering and knowing. *Cogn. Affect. Behav. Neurosci.* 8, 143–152. doi: 10.3758/CABN.8.2.143

Narme, P., Peretz, I., Strub, M. -L., and Ergis, A. -M. (2016). Emotion effects on implicit and explicit musical memory in normal aging. *Psychol. Aging* 31, 902–913. doi: 10.1037/pag0000116

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *J. Exp. Psychol. Gen.* 129, 242–261. doi: 10.1037/0096-3445.129.2.242

Payne, J. D., Stickgold, R., Swanberg, K., and Kensinger, E. A. (2008). Sleep preferentially enhances memory for emotional components of scenes. *Psychol. Sci.* 19, 781–788. doi: 10.1111/j.1467-9280.2008.02157.x

Pierce, B. H., and Kensinger, E. A. (2011). Effects of emotion on associative recognition: valence and retention interval matter. *Emotion* 11, 139–144. doi: 10.1037/a0021287

Reed, A. E., Chan, L., and Mikels, J. A. (2014). Meta-analysis of the age-related positivity effect: age differences in preferences for positive over negative information. *Psychol. Aging* 29, 1–15. doi: 10.1037/a0035194

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714

Samson, S., Dellacherie, D., and Platel, H. (2009). Emotional power of music in patients with memory disorders: clinical implications of cognitive neuroscience. *Ann. N. Y. Acad. Sci.* 1169, 245–255. doi: 10.1111/j.1749-6632.2009.04555.x

Shacham, S. (1983). A shortened version of the Profile of Mood States. *J. Pers. Assess.* 47, 305–306. doi: 10.1207/s15327752jpa4703_14

Sharot, T., and Phelps, E. A. (2004). How arousal modulates memory: disentangling the effects of attention and retention. *Cogn. Affect. Behav. Neurosci.* 4, 294–306. doi: 10.3758/CABN.4.3.294

Sharot, T., and Yonelinas, A. P. (2008). Differential time-dependent effects of emotion on recollective experience and memory for contextual information. *Cognition* 106, 538–547. doi: 10.1016/j.cognition.2007.03.002

Vieillard, S., and Gilet, A. -L. (2013). Age-related differences in affective responses to and memory for emotions conveyed by music: a cross-sectional study. *Front. Psychol.* 4:711. doi: 10.3389/fpsyg.2013.00711

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., and Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognit. Emot.* 22, 720–752. doi: 10.1080/02699930701503567

Wagner, U., Hallschmid, M., Rasch, B., and Born, J. (2006). Brief sleep after learning keeps emotional memories alive for years. *Biol. Psychiatry* 60, 788–790. doi: 10.1016/j.biopsych.2006.03.061

Waring, J. D., and Kensinger, E. A. (2009). Effects of emotional valence and arousal upon memory trade-offs with aging. *Psychol. Aging* 24, 412–422. doi: 10.1037/a0015526

Wechsler, D. (2014). *Wechsler adult intelligence scale–Fourth Edition (WAIS–IV)*. San Antonio, Texas: Psychological Corporation.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *J. Exp. Psychol.* 20, 1341–1354. doi: 10.1037//0278-7393.20.6.1341

Yonelinas, A. P., and Ritchey, M. (2015). The slow forgetting of emotional episodic memories: an emotional binding account. *Trends Cogn. Sci.* 19, 259–267. doi: 10.1016/j.tics.2015.02.009

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

MINI REVIEW
published: 11 September 2020
doi: 10.3389/fnhum.2020.578546

# Motor and Predictive Processes in Auditory Beat and Rhythm Perception

*Shannon Proksch, Daniel C. Comstock, Butovens Médé, Alexandria Pabst and Ramesh Balasubramaniam**

*Sensorimotor Neuroscience Laboratory, Cognitive & Information Sciences, University of California, Merced, Merced, CA, United States*

In this article, we review recent advances in research on rhythm and musical beat perception, focusing on the role of predictive processes in auditory motor interactions. We suggest that experimental evidence of the motor system's role in beat perception, including in passive listening, may be explained by the generation and maintenance of internal predictive models, concordant with the Active Inference framework of sensory processing. We highlight two complementary hypotheses for the neural underpinnings of rhythm perception: The Action Simulation for Auditory Prediction hypothesis (Patel and Iversen, 2014) and the Gradual Audiomotor Evolution hypothesis (Merchant and Honing, 2014) and review recent experimental progress supporting each of these hypotheses. While initial formulations of ASAP and GAE explain different aspects of beat-based timing–the involvement of motor structures in the absence of movement, and physical entrainment to an auditory beat respectively–we suggest that work under both hypotheses provide converging evidence toward understanding the predictive role of the motor system in the perception of rhythm, and the specific neural mechanisms involved. We discuss future experimental work necessary to further evaluate the causal neural mechanisms underlying beat and rhythm perception.

**Keywords: beat perception, motor system, motor planning, sensorimotor system, rhythm, timing**

OPEN ACCESS

***Edited by:***
*Marta Olivetti Belardinelli, Sapienza University of Rome, Italy*

***Reviewed by:***
*Bruno Laeng, University of Oslo, Norway*
*Riccardo Brunetti, European University of Rome, Italy*

****Correspondence:***
*Ramesh Balasubramaniam*
*ramesh@ucmerced.edu*

***Specialty section:***
*This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience*

***Received:*** *30 June 2020*
***Accepted:*** *18 August 2020*
***Published:*** *11 September 2020*

***Citation:***
*Proksch S, Comstock DC, Médé B, Pabst A and Balasubramaniam R (2020) Motor and Predictive Processes in Auditory Beat and Rhythm Perception. Front. Hum. Neurosci. 14:578546. doi: 10.3389/fnhum.2020.578546*

## INTRODUCTION

The coupling of action and prediction in perception has been characterized by predictive models of perception (Rao and Ballard, 1999) including classical Predictive Coding (Friston, 2002, 2005) (PC), and the more recent Active Inference Framework (Active Inference – corollary to the Free Energy Principle, Friston et al., 2009; Friston, 2010; Parr and Friston, 2019). Under classical PC, the brain is thought to utilize an internal generative model and a process of probabilistic model updating to predict the causes of its sensory input. Each level of the neural hierarchy predicts the activity at the level below, with higher levels of the hierarchy providing empirical priors, or hypotheses that constrain the generation of new priors at the level below. At each level, the top-down predictive signal is compared to the bottom-up inputs from the lower level. When there is a mismatch between incoming, bottom-up sensory information and top-down predictions, a prediction error is propagated back to the level above where it is used to revise and improve the initial hypothesis. If the prediction error cannot be minimized at the level at which it is being processed, it is relayed up to the next level above. The higher in the hierarchy the prediction error is being relayed, the more substantial the revision in the hypothesis. Perceptual experience arises as prediction error is minimized and a 'winning' hypothesis is selected. Thus, the general idea of PC is perceptual inference.

However, this classical PC/Bayesian account of perception characterizes the brain as a passive, Helmholtzian, stimulus-response machine, responsive only to the generation of prediction errors between its top-down sensory predictions and the actual sensory input from the world (Friston and Stephan, 2007; Clark, 2013). Our brains are more aptly described as embodied and enactive, enabling us to move and interact with our environment to bring about the minimization of prediction errors through our own action (Thompson, 2007; Gallagher et al., 2013; Bruineberg et al., 2018). This is the premise of Active Inference. As in PC, the brain uses an internal generative model to predict incoming sensory data. However, rather than relying on the passive accumulation of bottom-up sensory prediction errors that are minimized to create the content of perception, Active Inference formulations incorporate active engagement with the world to make the sensory inputs more predictable. Thus, in Active Inference, the prediction error minimization process which gives rise to perceptual experience is achieved through actions which conform sensory inputs to the brain's predictions (Friston et al., 2009; Hohwy, 2013; Parr and Friston, 2019).

Music perception and production are exemplar cognitive and behavioral phenomena to study these predictive processes and to evaluate the role of motor processing in sensory perception. Koelsch et al. (2019) expanded on the specific properties of music which make it an ideal paradigm for investigating predictive processes in the brain. Music, in any culture, is based on the generation of regularities, from the temporal regularities of rhythm to the predictable patterns and combinations of musical pitches. These regularities, or expectancies, generated by music have even been proposed as the properties which underlie emotional experience in music (Meyer, 1956; Huron, 2008; Juslin and Västfjäll, 2008). Cross-cultural perceptual priors may exist for some aspects of rhythm perception and production (Jacoby and McDermott, 2017), while other aspects are shaped by encluturation within a certain musical niche (Cameron et al., 2015; van der Weij et al., 2017; Polak et al., 2018). In particular, the experience of musical groove, that property of 'wanting to move' to the music, is proposed to be related to the balance between prediction and prediction errors generated by rhythmic properties of the music (Janata et al., 2012; Matthews et al., 2019, 2020). Active Inference formulations account for not only predictions related to expected stimulus input, but also predictions related to the expected accuracy–the *precision,* or uncertainty–of the original sensory prediction, in addition to counterfactual predictions related to how these prediction errors and their precision would change in response to active motor engagement with the sensory stimulus. Expected precision is modulated by sensory context and active engagement with the sensory signal. The generation of internal, predictive sensorimotor timing signals aligned to the musical beat may enhance the prediction and precision of temporal expectancies when perceiving syncopated musical rhythms, such as in musical groove (Koelsch et al., 2019). Whether or not we actually move our bodies to a musical rhythm, interactions between sensory and motor systems in our brain have been theorized to generate predictive timing signals that help us process musical rhythm (Merchant and Honing, 2014; Patel and Iversen, 2014; Vuust and Witek, 2014). These predictive timing signals are what allow for *beat induction*, or the active detection of the pulse in rhythmic time-varying stimuli such as music (Honing et al., 2014).

While predictive theories or perception are not new [indeed, they precede the age of Helmholtz, dating as far back as the 11th century works of Arab scholar al-Haytham et al. (ca. 1030; 1989)], the purpose of this review is to contextualize recent advances in the role of the motor system in rhythm and musical beat perception under more recent advances within the Active Inference framework. We then directly compare two hypotheses for the neural underpinnings of rhythm perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis (Patel and Iversen, 2014) and the Gradual Audiomotor Evolution (GAE) hypothesis (Merchant and Honing, 2014). We suggest that the both hypotheses–taken together under the umbrella of Active Inference–provide converging evidence toward understanding the predictive role of the motor system within a distributed sensorimotor network underlying the perception of rhythm.

## ACTION AND PREDICTION IN RHYTHM PERCEPTION

The role of the motor system in rhythm perception is most obviously recognized by examining how it is we engage our body with music. In addition to beat induction in passive music listening, humans – and a limited group of birds and mammals (Kotz et al., 2018; Ravignani et al., 2019) – can move in time to a musical beat. This process of rhythmic entrainment is defined as the ability to flexibly perceive and synchronize to the beat of music or other complex auditory rhythms. It is argued that rhythmic entrainment abilities are determined by the ability to perceive a beat, the underlying pulse, within rhythmic stimuli. Beat perception in humans is inherently predictive, constructive, hierarchical, and modality biased. In addition, beat perception engages the motor system, even when no movement is present (Grahn and Brett, 2007; Chen et al., 2008a,b; Gordon et al., 2018).

In humans, behavioral evidence for prediction in beat perception comes from tapping experiments that reveal negative mean asynchronies, which are not observed in other primates. Asynchronies are observed when humans tap slightly earlier or later than the beat in a rhythmic stimulus, and negative mean asynchronies are a behavioral indicator that humans actively anticipate upcoming stimuli. Mean tapping asynchronies throughout a rhythmic stimulus are usually negative in the auditory domain, but much more variable in the visual domain (Pabst and Balasubramaniam, 2018). Humans also adjust future tapping response based on temporal mismatch between their movement and the current beat (Balasubramaniam et al., 2004), and overtly tapping along to the beat aids in forming temporal predictions when compared to passively tracking a beat (Morillon and Baillet, 2017). In addition, when visual stimuli are presented in a way that indicates movement over time, e.g., apparent hand motion (Hove and Keller, 2010) or a bouncing ball (Iversen et al., 2015), predictive entrainment as demonstrated by negative mean asynchrony becomes much more successful.

According to Active Inference, the brain minimizes prediction error either by updating predictions or by taking action in the world to bring actual proprioceptive input in line with top-down predictions regarding driving sensory stimuli. In musical beat perception, this means that we either take action and *move* to the beat, or we update our predictions by suppressing actual movement and instead establishing an internal model of the beat which corresponds to the proprioceptive input we would have received had we actually been moving to the beat. The ability to flexibly adapt motor behavior in response to a mismatch between a rhythmic auditory stimulus and current motor movement (Balasubramaniam et al., 2004) can be construed as one example of this more general active inference process. Enhanced rhythmic entrainment abilities for the visual domain when visual stimuli implies movement (Hove and Keller, 2010; Iversen et al., 2015), and the improvement of temporal predictions in conjunction with overt rhythmic movement (Morillon and Baillet, 2017) can also be explained by the increase of sensory information available in order to update and modulate descending predictions about the temporal regularities of the stimulus which guide motor movements.

But this Active Inference gloss on beat perception is – by itself – vague. Plausible neural architectures have been proposed to support the classical (Helmholtzian) PC/Bayesian processing of music in general (Friston and Friston, 2013). However, an empirically detailed account of the specific neural underpinnings of *embodied Active Inference* in human musical beat perception is necessary. The motor system has been proposed to play a key role in prediction and perception of sensory information (Schubotz, 2007), and is functionally organized to enable the driving (ascending) and modulatory (descending) message passing hypothesized within the Active Inference Framework (Adams et al., 2013). This differs slightly from traditional theories of motor control, where driving signals arise from descending, top-down motor commands. Under Active Inference, top-down predictive signals from the motor system serve to modulate proprioceptive predictions regarding driving, feed-forward sensory signals (Adams et al., 2013).

Concordantly, the motor system has been found to be consistently active when listening to music, even in the absence of specific motor movement. A recent meta-analysis of fMRI studies found clusters of activations in key regions of the motor system in passive music listening, including bilateral premotor cortex and right primary motor cortex (Gordon et al., 2018). Metrical musical stimuli have also elicited activation in the basal ganglia, supplementary motor area, and cerebellum (Grahn and Rowe, 2009). Indeed, the modality bias for human beat perception and rhythmic entrainment for auditory stimuli (Pabst and Balasubramaniam, 2018), and improvements of auditory beat processing when making overt action (Morillon and Baillet, 2017) can be explained by tight connections between auditory and motor regions of the brain. But the activation of motor structures of the brain, even in the absence of overt movement, indicates that the motor system plays a more fundamental role in the formation of abstract predictive models which support sensory perception (Schubotz, 2007; Adams et al., 2013; Patel and Iversen, 2014).

Strong explanations of rhythm perception must account not only for prediction in action, but also for the role of the motor activity observed in passive music listening. Below, we provide an overview on the motor system's role in rhythm perception, and review two complementary hypotheses which highlight the causal role of the motor system in beat-based timing perception.

## MOTOR SYSTEM IN RHYTHM PERCEPTION: VIEWS FROM THE ACTION SIMULATION FOR AUDITORY PREDICTION AND THE GRADUAL AUDIOMOTOR EVOLUTION HYPOTHESES

Rhythm perception involves two types of timing perception, interval-based (absolute) timing and beat-based (relative) timing (Grube et al., 2010; Ross et al., 2016a; Iversen and Balasubramaniam, 2016). Interval-based timing refers to the ability to discriminate absolute differences in interval duration, whereas beat-based timing refers to the ability to measure the duration of time intervals relative to underlying temporal regularities such as beats (Teki et al., 2011). Beat-based timing perception is thought to be uniquely human (Merchant and Honing, 2014), and is believed to rely on the formation and maintenance of internal predictive models. According to the ASAP hypothesis (Patel and Iversen, 2014), these internal predictive models consist of periodic motor planning activity communicated via the dorsal auditory stream which allow for auditory prediction in beat-based musical timing perception. ASAP highlights the dorsal auditory stream due to its structural and functional relationship between auditory and motor planning regions, facilitating temporally-precise two-way signaling between these regions. This neural pathway involved in spatial processing of sounds (Rauschecker and Tian, 2000; Patel and Iversen, 2014) is more developed in humans than non-human primates, which is consistent with differences in beat-based timing behavioral ability (Honing, 2012; Patel and Iversen, 2014). In addition, Rauschecker (2018) postulates that the dorsal auditory stream may also be forming an "internal model of the outside world. . .[which] conver[ts] sensorimotor sequences into a unified experience"(p264–5). In the case of musical beat-based timing perception, the dorsal stream should form an internal model of the periodic musical beat.

Complementary to the ASAP hypothesis, the GAE hypothesis has been proposed to account for differences in beat-based temporal processing between primates and humans (see **Figure 1** for an overview comparison of ASAP and GAE). The GAE hypothesis (Merchant and Honing, 2014) also posits the dorsal auditory stream as a potential substrate for rhythm entrainment and perception. However, GAE claims that the evolution of rhythmic entrainment results more specifically from

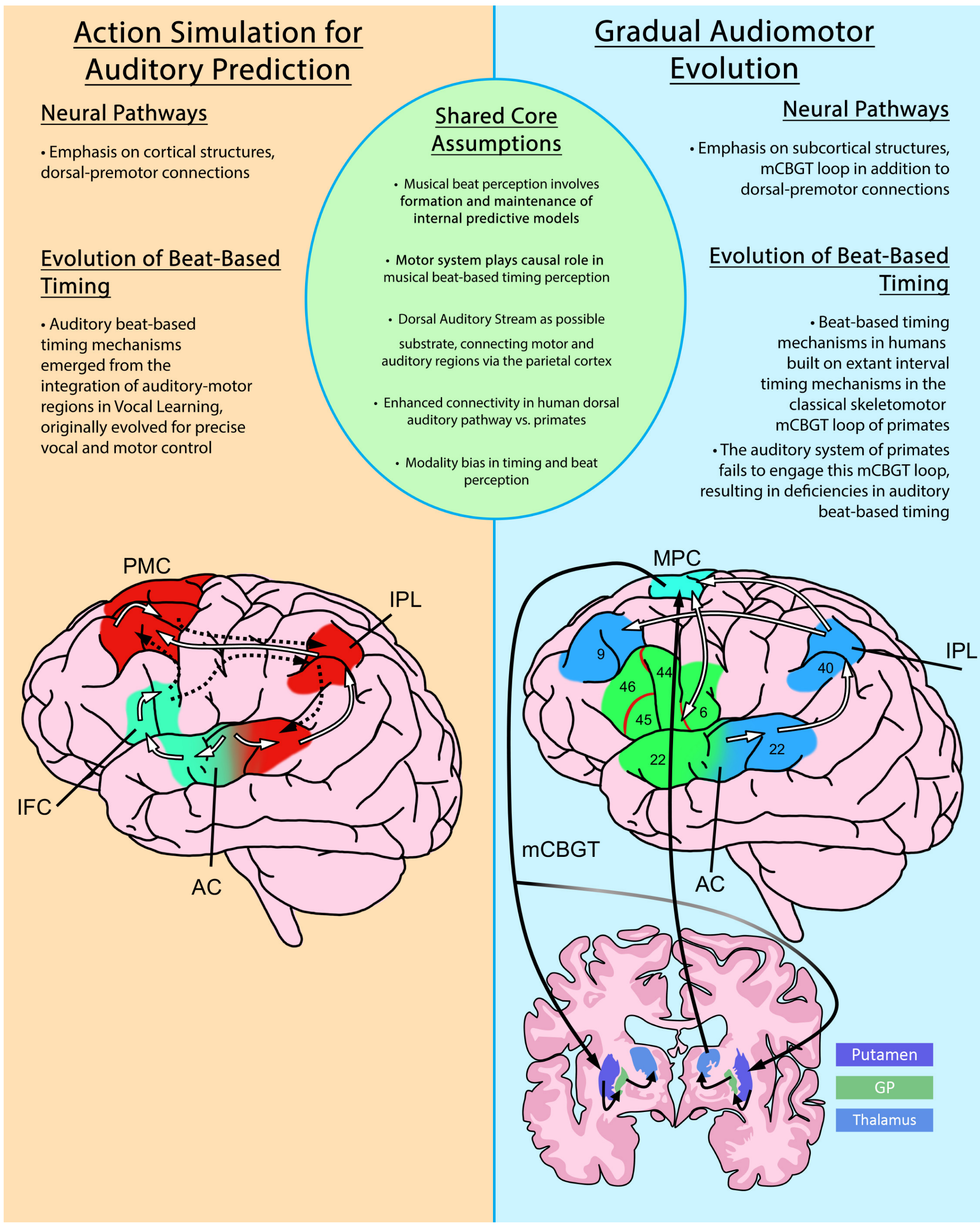



**FIGURE 1 |** An overview comparison of the Action Simulation for Auditory Prediction Hypothesis (ASAP) and the Gradual Audiomotor Evolution Hypothesis (GAE). Shared core assumptions of both hypotheses are listed at center. Brief differing emphases on neural pathways and evolutionary commitments are listed in each panel. Diagrams depict the neural pathways proposed under each hypothesis. The ASAP diagram (left), shows ascending pathways from the auditory cortex (white lines) and descending pathways back to the auditory cortex (dashed lines) in the dorsal (red) and ventral (green) streams. The GAE diagram (right) shows the dorsal auditory pathway (white lines) and dorsal (blue) and ventral (green) streams, and the motor cortico-basal ganglia-thalamo-cortical (mCBGT) circuit (black lines). PMC, primary motor cortex; IPL, inferior parietal lobule; AC, auditory cortex; IFC, inferior frontal cortex; MPC, medial premotor cortex; GP, globus pallidus. Figures adapted from Merchant and Honing (2014) and Patel and Iversen (2014).

adaptations to the motor cortico-basal ganglia thalamo-cortical circuit (mCBGT). This specification arises from observations that the mCBGT is found to be active in sequential and temporal processing and movement in Macaques (Tanji, 2001; Merchant et al., 2013; Perez et al., 2013) and humans (Grafton et al., 1995; Harrington et al., 2010), including, for humans, the processing of musical rhythms (Grahn and Brett, 2007). Explicitly including the mCBGT loop in the evolution of rhythmic entrainment accounts for the fact that interval-timing ability appears preserved in macaques (Merchant et al., 2013) and is shared among primates, including humans. This indicates a shared neural circuitry for single interval-based timing, upon which GAE hypothesizes human beat-based timing mechanisms would have evolved to enable beat-based rhythmic entrainment. It is gradual changes to this foundational neural pathway, in addition to strengthening connections to auditory cortices via the dorsal auditory pathway, that have enabled the human mCBGT to develop beat-based timing mechanisms that can process the hierarchical properties of beat-based, rhythmic stimuli, such as music. Although focusing on slightly different neural pathways, both ASAP and GAE highlight the predictive role of the *motor system* in the perception of rhythm, and support growing consensus on the role of motor pathways in the formation of internal predictive models in perception more generally.

One important difference between the ASAP and GAE hypotheses is that ASAP purports to explain the presence of motor activity in beat perception even in the absence of overt movement, while GAE explains how evolution within motor pathways enables physical entrainment–synchronized movement–to a rhythmic stimulus. ASAP claims that beat-perception in humans arose with the emergence of *vocal learning* abilities, which strengthened tight audio-motor connections in the dorsal auditory stream underlying rhythmic entrainment along the primate lineage.

In contrast, GAE favors a gradual strengthening of these connections over evolutionary time, building on specific interval-timing mechanisms already extant in the mCGBT circuit of the primate brain. The result being the formation of an additional beat-based mechanism with enhanced connection of the mCGBT to the auditory cortex via that same dorsal auditory stream in the human brain (Merchant and Honing, 2014). Recent neurophysiological evidence highlights the interconnectedness of interval and beat-based timing mechanisms proposed by GAE, indicating that even in passive listening, monkeys are able to detect isochrony in rhythm, due in part to extant interval-based timing mechanisms of the monkey motor system, but that monkeys cannot detect the underlying beat in a rhythmic stimulus, which requires auditory-motor beat-based timing mechanisms present in humans (Honing et al., 2018).

## Evidence for Prediction and Motor Activity in GAE and ASAP

In addition to fMRI observation of motor activation in music listening and rhythm processing, the predictive and causal roles of specific motor structures highlighted by the ASAP and GAE hypotheses have been experimentally tested via electroencephalography (EEG) and transcranial magnetic stimulation (TMS). Specific Event Related Potentials (ERPs) relating to prediction errors evoked by rhythmic deviations in musical stimuli include the mismatch negativity (MMN) and P3a (Honing et al., 2018; Koelsch et al., 2019). These auditory event related components indicate violation of temporal expectations in oddball paradigms, with early responses related to bottom-up sensory processing and later responses reflecting top-down cortical processes (Garrido et al., 2007) and (perhaps conscious) attention to deviant stimuli (Sussman et al., 2003). EEG studies provide insight into the neural mechanisms of beat-perception while removing the limitations of behavioral response (Honing, 2012). The MMN and P3a components have been observed in response to rhythmic violations in adult humans, as well as infants and monkeys (Ladinig et al., 2009; Winkler et al., 2009; Honing et al., 2018). However more recent research in monkeys comparing ERPs in passive listening to jittered and isochronous stimuli with occasional deviants have demonstrated that monkeys might be able to detect isochrony in rhythm – which could rely on extant interval timing mechanisms in the primate brain; but not the beat – which relies on more evolved beat-based timing mechanisms, while humans are able to detect both isochrony and the beat (Bouwer et al., 2016; Honing et al., 2018). This collection of experiments supports the gradual evolution of beat-based timing mechanisms hypothesized by GAE.

Action Simulation for Auditory Prediction has been further supported by TMS research, demonstrating causal links between specific types of beat processing and regions of the dorsal auditory stream. A set of TMS experiments evaluated the role of the posterior parietal cortex (PPC), which is thought to serve as an interface for bidirectional communication between auditory and motor regions of the brain, and the dorsal pre-motor cortex (dPMC), which is also part of the dorsal auditory stream and is associated with movement planning and synchronization to auditory stimuli (Chen et al., 2009; Giovannelli et al., 2014). By down-regulating neural activity in left PPC according to the Huang et al. (2005) protocol, Ross et al. (2018b) showed that left PPC may be involved in one aspect of beat-based timing–phase shift detection–but not tempo detection or discrete interval discrimination. Ross et al. (2018a) down-regulated activity in left dPMC, showing that left dPMC may be involved in tempo detection, but not phase shift detection or discrete interval discrimination. Additionally, measures of Motor Evoked Potentials (MEPs) in single pulse TMS over the motor cortex have indicated that musical groove modulates cortical excitability in the motor cortex. High levels of musical groove are characterized by syncopated rhythms, enhanced energy in the bass line, and the phenomenological property of 'wanting to move' with the music (Janata et al., 2012; Stupacher et al., 2013; Ross et al., 2016b). High-groove music has been shown to more strongly activate the motor system (resulting in higher MEPs) when compared with low-groove music (Stupacher et al., 2013). These results indicate the bidirectionality of auditory-motor interactions, as causally down-regulating activity in the motor cortex can impair auditory perception of aspects of musical rhythm, and varying degrees of rhythmic information in auditory stimuli (i.e., syncopation and bass frequencies in musical groove) can change aspects of motor cortical function.

## Mechanisms for Timing and Rhythm Prediction

While there is growing consensus that the motor system is causally involved in timing and rhythm perception, and that the neural substrate includes cortical structures of the dorsal auditory stream and subcortical structures within the motor-cortico basal ganglia thalamo-cortical loop, the specific neural mechanisms which enable timing and rhythm perception within these substrates remains an open question. For some years, cognitive scientists have been looking for how internal timing can be instantiated by patterns of temporal stimuli via, e.g., clock-based or oscillatory mechanisms (Povel and Essens, 1985; Large and Jones, 1999). Given the amount of neuroscientific evidence pointing to a distributed timing network in the brain (Buonomano, 2014), mechanisms of entrainment to patterns of temporal stimuli have received significant attention. The striatal beat frequency model was suggested to support a clock-based mechanism based on banks of oscillators (Matell and Meck, 2000, 2004). In contrast, Large et al. (2015) describe an oscillatory model of pulse perception called Neural Resonance Theory (NRT), which provides a plausible mechanism of adaptive entrainment and beat-based timing without requiring an internal clock mechanism. According to NRT, rhythmic stimuli are encoded in sensory networks which interact with motor networks thus entraining them to the pulse frequency. Neural entrainment is induced to the pulse, even when the rhythmic stimulus itself lacks physical information at the location of the pulse–such as silences found 'on the beat' within syncopated rhythms–demonstrating the influence of top-down effects on pulse perception (Large et al., 2015; Tal et al., 2017). The cerebellum has also been shown to play a prominent role in absolute timing (Nozaradan et al., 2017)–but not beat-based timing–with proposed mechanisms including an oscillatory pacemaker based on regular oscillations found within the inferior olive (Ashe and Bushara, 2014), and a state-spaced based mechanism, in which the timing of a stimulus can be inferred from the state of a relevant cortical network over time (Buonomano, 2014). In various cortical areas, ramping activity of neural firing rates has been proposed as a mechanism for interval-based timing–where interval duration is encoded in the modulation of neural spiking thresholds or by varying the slope of ramping activity preceding threshold (Durstewitz, 2003). However, in the Macaque brain, ramping activity has also been implicated for relative timing in coordination with multidimensional state space models as part of a multilayer timing system involving two neural populations (Merchant et al., 2014). These two neural populations are differentially associated with absolute and relative timing, and are observed in the medial motor cortex, consistent with the proposed role of the motor system under the GAE hypothesis (Crowe et al., 2014; Merchant et al., 2014).

Continuous state-space models have also been proposed in Active Inference accounts for the generation of predictive models in action and sensory processing more generally, neurally mediated by the balance of pre- and post-synaptic activity (Friston et al., 2017) and neuronal firing rates in, e.g., medial or lateral intraparietal areas (de Lafuente et al., 2015). Striatal dopamine in particular has been proposed to code for both prediction error and certainty in response to sensory stimuli (Sarno et al., 2017) across a variety of timescales (Schultz, 2007). Dopaminergic activity also plays a role in rhythmic motor control (Koshimori and Thaut, 2018) and is responsive to rhythmic auditory stimulation (Koshimori et al., 2019), positioning dopamine as a crucial facilitator of the motor system's role in auditory-motor interactions underlying beat-based timing perception. The motor system's predictive role in music and rhythm perception is only one component of larger networks of sensorimotor processing, namely the dorsal auditory pathway and the mCBGT. Further experimental and computational work is necessary to determine whether and how the specific neural mechanisms of the human motor cortex processes timing information within the cortical and subcortical networks proposed by ASAP and GAE. To facilitate the generation of experimental and computational hypotheses, we have compiled an overview of recent experimental and theoretical research on the motor and distributed brain areas and mechanisms within the dorsal auditory pathway and the mCBGT–including the dopaminergic system–which are involved in the predictive processing of auditory-motor beat and rhythm perception in **Table 1**.

## CONCLUSION AND FUTURE DIRECTIONS

In this paper, we reviewed recent advances in research on rhythm perception, focusing on the role of predictive processes in auditory motor interactions in beat-processing. We highlighted two complementary hypotheses for the neural underpinnings of rhythm perception: The ASAP hypothesis (Patel and Iversen, 2014) and the GAE hypothesis (Merchant and Honing, 2014) and reviewed recent experimental progress supporting each of these hypotheses. While initial formulations of ASAP and GAE explain different aspects of beat-based timing–the involvement of motor structures in the absence of movement, and physical entrainment to an auditory beat respectively–both theories have moved us closer to understanding the predictive role of the motor system in the perception of rhythm and the specific neural mechanisms involved. In fact, recent computational formulations of ASAP have further incorporated the subcortical structures proposed to be involved in the evolution of beat-based timing perception by GAE. Cannon and Patel (2019, preprint), have proposed the CBGT loop as responsible for the resetting of relative timing mechanisms via a hyper direct pathway from the SMA. In addition, they hypothesize a role for striatal dopamine in the maintenance of internal rhythmic timing models by tracking confidence (uncertainty) in the beat, consistent with Predictive Coding and Active Inference accounts of rhythm perception and perception more generally.

Future work in understanding the neural, cognitive, and behavioral dynamics of musical beat perception in humans should investigate not only the sensorimotor processes responsible for the perception of rhythm, but also the specific neural mechanisms by which top-down predictions serve to modulate driving proprioceptive sensations arising from concrete actions of the body or abstract activity of the motor systems. While EEG experiments (e.g., Ladinig et al., 2009; Winkler et al., 2009; Honing et al., 2018) point to the neural mechanisms of internal predictive models in beat-based timing perception, EEG alone cannot provide causal evidence for

**TABLE 1 |** The motor system's predictive role in music and rhythm perception is only one component of larger networks of sensorimotor processing, namely the dorsal auditory pathway and the motor cortico-basal ganglia-thalamo-cortical circuit.

| ***Brain area*** | **Authors** | **Proposed role of each brain area** | **Experimental task and stimulus type** | **Type of data** |
|---|---|---|---|---|
| *Cerebellum* | Ivry and Schlerf, 2008 | Dedicated timing mechanism; coordination of movement, internal timing mechanisms involved with sub-second timing | Theoretical Paper/Review | |
| | Bastian, 2006 | Predictive models of movement | Theoretical Paper/Review | |
| | Nozaradan et al., 2017 | Tracking beats in rhythms with fast tempos; more prominent role in absolute timing vs. relative timing | **Passive Listening.** Auditory rhythms designed to induce a beat – syncopated and unsyncopated. | EEG |
| | Gordon et al., 2018 | Meta-analysis of fMRI studies of recruitment of motor system during music listening | **Meta-analysis.** Various listening tasks – auditory rhythms or music. | fMRI |
| *Basal Ganglia* | Merchant et al., 2013 | Interacts with the cortico-thalamic-striatal circuit in a context dependent manner | Theoretical Paper/Review | |
| | Coull and Nobre, 2008 | Perceptual temporal expectation; explicit timing | Theoretical Paper/Review | |
| | Nozaradan et al., 2017 | Tracking beats in complex rhythm sequences | **Passive Listening** Auditory rhythms designed to induce a beat – syncopated and unsyncopated. | EEG |
| | Grahn, 2009 | Internal beat generation; more prominent role in relative vs. absolute timing | **Discrimination task, same or different judgment of two auditory stimuli.** Auditory rhythms – beat-based structure and non-beat-based structure; Accents- duration or volume accented (externally generated) or unaccented (internally generated) beats. | fMRI/Behavioral |
| | Grahn et al., 2011 | Internal representation of auditory rhythms that support cross-modal interactions in beat perception and generation | **Discrimination task, rhythmic tempo change.** Auditory tone metronome and visual flashing metronome. Two groups: one with auditory first visual second, and the other vice versa. | fMRI/Behavioral |
| | Grahn and Rowe, 2009 | Internal beat generation: part of cortico-subcortical network involved in beat perception and generation | **Indicate the strength of the perceived beat.** Auditory rhythms of varying complexity and some with volume accents. | fMRI/Behavioral |
| | Grahn and Rowe, 2013 | Putamen activity in beat prediction, but not beat finding | **Attentive listening; occasionally indicate level of feeling of the beat.** Auditory rhythms of varying intervals and rates, beat and non-beat (jittered) rhythms. | fMRI/Behavioral |
| | Grahn and Brett, 2007 | Higher activity for rhythms with integer ratio relationships between intervals and with regular perceptual accents | *1st experiment (behavioral)* **reproduce auditory rhythms.** *2nd experiment (fMRI)* **indicate if the rhythm played matched previous rhythms**. Metered auditory rhythms of varying integer intervals and complexity. | fMRI/Behavioral |
| | Teki et al., 2011 | Striato-thalamo-cortical network involved in beat-based timing, while an olivocerebellar network involved in duration-based timing | **Judge duration matches in a set of tones.** Auditory tones, either isochronous or jittered, arranged in either rhythm-based or absolute duration-based sets. | fMRI/Behavioral |

*(Continued)*

**TABLE 1** | Continued

| *Brain area* | Authors | Proposed role of each brain area | Experimental task and stimulus type | Type of data |
|---|---|---|---|---|
| | Araneda et al., 2017 | Hearing, feeling or seeing a beat recruits a supramodal network in the auditory dorsal stream | **Discrimination task, between beat and non-beat rhythms**. Auditory, visual, and vibrotactile rhythms. | fMRI/Behavioral |
| *Primary and premotor cortices* | Kilavik et al., 2014 | Movement preparation, cue anticipation | Theoretical Paper/Review | |
| | Schubotz, 2007 | Predictive processing of external events, even in the absence of proprioceptive or interoceptive information | Theoretical Paper/Review | |
| | Morillon and Baillet, 2017 | Beta and delta oscillations directed to auditory cortex encode temporal predictions | **Passive listening (listen condition); active tapping with the beat (tracking condition).** Auditory melody – different tones either on beat, anti-phase, or quasi-phase with the beat. | MEG/Behavioral |
| | Gordon et al., 2018 | Meta-analysis of fMRI studies of recruitment of motor system during music listening | **Meta-analysis.** Various listening tasks – Auditory rhythms or music. | fMRI |
| *Premotor cortex* | Grahn and Rowe, 2009 | Cortico-cortical coupling with SMA and auditory cortex in duration beat perception; part of cortico-subcortical network involved in beat perception and generation | **Indicate the strength of the perceived beat.** Auditory rhythms of varying complexity and some with volume accents. | fMRI/Behavioral |
| | Teki et al., 2011 | Striato-thalamo-cortical network involved in beat-based timing, while an olivocerebellar network involved in duration based timing | **Judge duration matches in a set of tones.** Auditory tones, either isochronous or jittered, arranged in either rhythm-based or absolute duration-based sets. | fMRI/Behavioral |
| | Chen et al., 2008a | Motor regions recruited while listening to music rhythms | *Experiment 1:* **Listen to rhythm passively then tap along with rhythm.** *Experiment 2:* **Listen to rhythm passively then tap along to rhythm without foreknowledge of being asked to tap with the rhythm** auditory tones in simple, complex, or ambiguous rhythms. | fMRI/Behavioral |
| *Supplementary motor area* | Coull et al., 2016b | Perceptual and motor timing; Comparing the duration of perceptual events, error monitoring | Theoretical Paper/Review | |
| | Ross et al., 2018b | Not causally implicated in perceptual auditory interval timing | **Discrimination task – same/different judgment of auditory intervals; detection task – identification of tempo or phase shifted metronome click.** Auditory intervals of pairs of tones; metronome click track over musical stimuli. | Behavioral (pre/post TMS down-regulatory stimulation) |
| | Grahn and Brett, 2007 | Higher activity for rhythms with integer ratio relationships between intervals and with regular perceptual accents; in musicians: higher activity for all rhythms when compared to rest | *Experiment 1 (behavioral):* **reproduce auditory rhythms.** *Experiment 2 (fMRI):* **indicate if the rhythm played matched previous rhythms.** Metered auditory rhythms of varying integer intervals and complexity. | fMRI/Behavioral |
| | Grahn and McAuley, 2009 | Stronger activity in strong beat-perceivers vs. weak beat-perceivers, no correlation with musicianship | **Discrimination task, rhythmic tempo change.** Auditory isochronous rhythms. | fMRI/Behavioral |

*(Continued)*

**TABLE 1 |** Continued

| ***Brain area*** | **Authors** | **Proposed role of each brain area** | **Experimental task and stimulus type** | **Type of data** |
|---|---|---|---|---|
| | Grahn and Rowe, 2009 | Coupling with STG in beat perception for musicians; part of cortico-subcortical network involved in beat perception and generation | **Indicate the strength of the perceived beat.** Auditory rhythms of varying complexity and some with volume accents. | fMRI/Behavioral |
| | Teki et al., 2011 | Striato-thalamo-cortical network involved in beat-based time, while an olivocerebellar network involved in duration-based timing | **Judge duration matches in a set of tones.** Auditory tones, either isochronous or jittered, arranged in either rhythm-based or absolute duration-based sets. | fMRI/Behavioral |
| | Chen et al., 2008a | Motor regions recruited while listening to music rhythms | *Experiment 1:* **Listen to rhythm passively then tap along with rhythm.** *Experiment 2:* **Listen to rhythm passively then tap along to rhythm without foreknowledge of being asked to tap with the rhythm** auditory tones in simple, complex, or ambiguous rhythms. | fMRI/Behavioral |
| | Araneda et al., 2017 | Hearing, feeling or seeing a beat recruits a supramodal network in the auditory dorsal stream | **Discrimination task, between beat and non-beat rhythms.** Auditory, visual, and vibrotactile rhythms. | fMRI/Behavioral |
| *Medial premotor cortex* | Merchant et al., 2014 | Absolute and relative timing mechanisms within two separate neural populations | Theoretical Paper/Review | |
| | Crowe et al., 2014 | Absolute and relative timing mechanisms within two separate neural populations | **Synchronization Continuation Task.** Isochronous visual stimuli or auditory tones. | Behavioral; Extracellular activity of single neurons (in Macaca mulatta) |
| | Grahn and McAuley, 2009 | Stronger activity in strong beat-perceivers vs. weak beat-perceivers, no correlation with musicianship | **Discrimination task, rhythmic tempo change.** Auditory isochronous rhythms. | fMRI/Behavioral |
| *Parietal Cortex* | Rauschecker, 2011; Merchant and Honing, 2014; Patel and Iversen, 2014 | Interface between motor and auditory cortices, sensorimotor integration | Theoretical Papers/Reviews | |
| | Coull and Nobre, 2008 | Perceptual temporal expectation; implicit timing | Theoretical Paper/Review | |
| | Coull et al., 2016a | Temporal predictability via fixed or dynamic predictions | **Cued reaction time task.** Visual cue that predicted target presentation time (temporal condition), or provided no information for target presentation (neutral condition) with variable intervals between cue and target. | fMRI/Behavioral |
| | Ross et al., 2018b | Causally implicated in perceptual beat-based timing | **Discrimination task – same/different judgment of auditory intervals; detection task – identification of tempo or phase shifted metronome click.** Auditory intervals of pairs of tones; metronome click track over musical stimuli. | Behavioral (pre/post TMS down-regulatory stimulation) |

*(Continued)*

**TABLE 1** | Continued

| *Brain area* | Authors | Proposed role of each brain area | Experimental task and stimulus type | Type of data |
|---|---|---|---|---|
| *Auditory Cortex* | Koelsch et al., 2019 | Event related potentials associated with predictive processes in music | Theoretical Paper/Review | |
| | Fujioka et al., 2012 | Beta-band activity predicts onset of beats in music | **Passive listening, while watching silent videos**. Auditory isochronous rhythms of several tempos and one irregular rhythm. | MEG |
| | Fujioka et al., 2015 | Beta-band activity represents timing information being translated for auditory-motor coordination | **Passive listening** to metered rhythms, followed by **attentive listening** to un-metered rhythms that the participants were asked to imagine as metered. March and Waltz metered rhythms | MEG |
| | Auksztulewicz et al., 2019 | Temporal prediction of rhythm and beats | **Identify target chords.** Auditory rhythmic or jittered sequences of distractor chords preceding target chords. | MEG/EEG/Behavioral |
| | Honing et al., 2018 | Event related potentials to perceptual deviants in rhythmic stimuli | **Passive listening.** Auditory oddball paradigm with isochronous or jittered rhythms. | EEG (of Macaca mulatta) |
| | Bouwer et al., 2016 | Event related potentials to perceptual deviants in rhythmic stimuli; ERPs modulated by attention in musicians | **Passive or attentive listening.** Auditory oddball paradigm with isochronous or jittered rhythms. | EEG |
| *Dopaminergic System/Striatal Dopamine* | Schultz, 2007 | Multiple time courses of dopamine changes mediate multiple time courses of behavioral processes | Theoretical Paper/Review | |
| | Friston et al., 2009 | Reward learning, encoding of precision | Theoretical Paper/Review | |
| | Friston et al., 2012; FitzGerald et al., 2015 | Reward learning, encoding of precision | Theoretical Papers/Computational Models | Simulated dopaminergic responses |
| | Sarno et al., 2017 | Temporal expectation of perceptual cues; reward prediction error and (un)certainty | **Detect weak vibrotactile stimuli.** Variable interval durations between tactile start cue and vibrotactile stimuli. | Intracellular recording, monkey brain |
| | Koshimori et al., 2019 | Rhythmic auditory stimulation (RAS) attenuates dopaminergic response | **Synchronization task,** RAS and no-RAS conditions; various auditory rhythms, single auditory beats or metronome clicks over instrumental music. | Behavioral/MRI/PET |
| | Brodal et al., 2017 | Rhythmic music reduces connectivity between basal ganglia and reward system | **Passive listening.** Electronic dance music in a continuous-stimulation design. | fMRI |

*This table provides an overview of the brain areas and mechanisms which make up these networks and are involved in the predictive processing of auditory beat and rhythm perception. Each brain area is introduced with one or more Theoretical or Review Papers contextualizing that brain area's proposed role, followed by a non-exhaustive list of supporting experimental work. This table is intended to serve as a tool for new or continuing researchers engaging in work on rhythm and musical beat perception.*

the role of specific brain structures. Similarly, while TMS experiments (e.g., Stupacher et al., 2013; Ross et al., 2018a,b) have lended causal evidence for the role of specific structures in beat-based timing perception, the mentioned experiments do not provide direct evidence for the presence of *internal predictive models* of beat-based timing. If motor activity is causally involved in the formation of auditory predictions, then causal TMS manipulation to down-regulate activity in, e.g., parietal cortex or dPMC should result in the reduction of MMN and P3a event related responses to perceptual deviants in rhythmic stimuli, and this response might differ based on whether the stimuli contains timing deviants related to tempo or phase. Future research should include stimuli designed to elicit specific prediction errors with perceptual deviants, such as in oddball paradigms, while measuring event-related potentials associated with predictive processes in combined EEG and causal TMS experiments. Results from these experiments could extend and strengthen already emerging support for GAE and ASAP, as well as further contextualize the role of Active Inference in music and beat-based timing perception.

## AUTHOR CONTRIBUTIONS

SP and RB conceptualized the manuscript. BM, DC, and AP contributed to the writing and the exhaustive analysis of the literature. All authors contributed to the article and approved the submitted version.

## REFERENCES


Adams, R. A., Shipp, S., and Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Struct. Funct.* 218, 611–643. doi: 10.1007/s00429-012-0475-5

al-Haytham, I. A. (ca. 1030; 1989). *The Optics of Ibn al-Haytham. Translated by A. I. Sabra*. London: Warburg Institute.

Araneda, R., Renier, L., Ebner-Karestinos, D., Dricot, L., and De Volder, A. G. (2017). Hearing, feeling or seeing a beat recruits a supramodal network in the auditory dorsal stream. *Eur. J. Neurosci.* 45, 1439–1450. doi: 10.1111/ejn.13349

Ashe, J., and Bushara, K. (2014). The olivo-cerebellar system as a neural clock. *Adv. Exp. Med. Biol.* 829, 155–165. doi: 10.1007/978-1-4939-1782-2_9

Auksztulewicz, R., Myers, N. E., Schnupp, J. W., and Nobre, A. C. (2019). Rhythmic temporal expectation boosts neural activity by increasing neural gain. *J. Neurosci.* 39, 9806–9817. doi: 10.1523/jneurosci.0925-19.2019

Balasubramaniam, R., Wing, A., and Daffertshofer, A. (2004). Keeping with the beat: movement trajectories contribute to movement timing. *Exp. Brain Res.* 159, 129–134. doi: 10.1007/s00221-004-2066-z

Bastian, A. J. (2006). Learning to predict the future: the cerebellum adapts feedforward movement control. *Curr. Opin. Neurobiol.* 16, 645–649. doi: 10.1016/j.conb.2006.08.016

Bouwer, F. L., Werner, C. M., Knetemann, M., and Honing, H. (2016). Disentangling beat perception from sequential learning and examining the influence of attention and musical abilities on ERP responses to rhythm. *Neuropsychologia* 85, 80–90. doi: 10.1016/j.neuropsychologia.2016.02.018

Brodal, H. P., Osnes, B., and Specht, K. (2017). Listening to rhythmic music reduces connectivity within the basal ganglia and the reward system. *Front. Neurosci.* 11:153. doi: 10.3389/fnins.2017.00153

Bruineberg, J., Kiverstein, J., and Rietveld, E. (2018). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese* 195, 2417–2444. doi: 10.1007/s11229-016-1239-1

Buonomano, D. V. (2014). Neural dynamics based timing in the subsecond to seconds range. *Adv. Exp. Med. Biol.* 829, 101–117. doi: 10.1007/978-1-4939-1782-2_6

Cameron, D. J., Bentley, J., and Grahn, J. A. (2015). Cross-cultural influences on rhythm processing: reproduction, discrimination, and beat tapping. *Front. Psychol.* 6:366. doi: 10.3389/fpsyg.2015.00366

Cannon, J. J., and Patel, A. D. (2019). A neurocomputational model of beat-based temporal processing. *bioRxiv [Preprint]* doi: 10.1101/805838

Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2008a). Listening to musical rhythms recruits motor regions of the brain. *Cereb. Cortex* 18, 2844–2854. doi: 10.1093/cercor/bhn042

Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2008b). Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J. Cogn. Neurosci.* 20, 226–239. doi: 10.1162/jocn.2008.20018

Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2009). The role of auditory and premotor cortex in sensorimotor transformations. *Ann. N. Y. Acad. Sci.* 1169, 15–34. doi: 10.1111/j.1749-6632.2009.04556.x

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Coull, J. T., Cotti, J., and Vidal, F. (2016a). Differential roles for parietal and frontal cortices in fixed versus evolving temporal expectations: Dissociating prior from posterior temporal probabilities with fMRI. *NeuroImage* 141, 40–51. doi: 10.1016/j.neuroimage.2016.07.036

Coull, J. T., Vidal, F., and Burle, B. (2016b). When to act, or not to act: that's the SMAs question. *Curr. Opin. Behav. Sci.* 8, 14–21. doi: 10.1016/j.cobeha.2016.01.003

Coull, J. T., and Nobre, A. C. (2008). Dissociating explicit timing from temporal expectation with fMRI. *Curr. Opin. Neurobiol.* 18, 137–144. doi: 10.1016/j.conb.2008.07.011

Crowe, D. A., Zarco, W., Bartolo, R., and Merchant, H. (2014). Dynamic representation of the temporal and sequential structure of rhythmic movements in the primate medial premotor cortex. *J. Neurosci.* 34, 11972–11983. doi: 10.1523/jneurosci.2177-14.2014

de Lafuente, V., Jazayeri, M., and Shadlen, M. N. (2015). Representation of accumulating evidence for a decision in two parietal areas. *J. Neurosci.* 35, 4306–4318. doi: 10.1523/jneurosci.2451-14.2015

Durstewitz, D. (2003). Self-organizing neural integrator predicts interval times through climbing activity. *J. Neurosci.* 23, 5342–5353. doi: 10.1523/jneurosci.23-12-05342.2003

FitzGerald, T. H. B., Dolan, R. J., and Friston, K. (2015). Dopamine, reward learning, and active inference. *Front. Comput. Neurosci.* 9:136. doi: 10.3389/fncom.2015.00136

Friston, K. (2002). Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annu. Rev. Neurosci.* 25, 221–250. doi: 10.1146/annurev.neuro.25.112701.142846

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K., and Stephan, K. E. (2007). Free-energy and the brain. *Synthese* 159, 417–458. doi: 10.1007/s11229-007-9237-y

Friston, K. J., Daunizeau, J., and Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLoS One* 4:7. doi: 10.1371/journal.pone.0006421

Friston, K. J., and Friston, D. A. (2013). "A free energy formulation of music generation and perception: Helmholtz revisited," in *Sound - Perception - Performance. Current Research in Systematic Musicology*, Vol. 1, ed. R. Bader (Heidelberg: Springer), 43–69. doi: 10.1007/978-3-319-00107-4_2

Friston, K. J., Parr, T., and de Vries, B. (2017). The graphical brain: Belief propagation and active inference. *Net. Neurosci.* 1, 381–414. doi: 10.1162/netn_a_00018

Friston, K. J., Shiner, T., Fitzgerald, T., Galea, J. M., Adams, R., Brown, H., et al. (2012). Dopamine, affordance and active inference. *PLoS Comput. Biol.* 8:e1002327. doi: 10.1371/journal.pcbi.1002327

Fujioka, T., Ross, B., and Trainor, L. J. (2015). Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *J. Neurosci.* 35, 15187–15198. doi: 10.1523/JNEUROSCI.2397-15.2015

Fujioka, T., Trainor, L. J., Large, E. W., and Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *J. Neurosci.* 32, 1791–1802. doi: 10.1523/jneurosci.4107-11.2012

Gallagher, S., Hutto, D. D., Slaby, J., and Cole, J. (2013). The brain as part of an enactive system. *Behav. Brain Sci.* 36, 421–422. doi: 10.1017/S0140525X12002105

Garrido, M. I., Kilner, J. M., Kiebel, S. J., and Friston, K. J. (2007). Evoked brain responses are generated by feedback loops. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20961–20966. doi: 10.1073/pnas.0706274105

Giovannelli, F., Innocenti, I., Rossi, S., Borgheresi, A., Ragazzoni, A., Zaccara, G., et al. (2014). Role of the dorsal premotor cortex in rhythmic auditory-motor entrainment: a perturbational approach by rTMS. *Cereb. Cortex* 24, 1009–1016. doi: 10.1093/cercor/bhs386

Gordon, C. L., Cobb, P. R., and Balasubramaniam, R. (2018). Recruitment of the motor system during music listening: an ALE meta-analysis of fMRI data. *PLos One* 13:e0207213. doi: 10.1371/journal.pone.0207213

Grafton, S. T., Hazeltine, E., and Ivry, R. (1995). Functional mapping of sequence learning in normal humans. *J. Cogn. Neurosci.* 7, 497–510. doi: 10.1162/jocn.1995.7.4.497

Grahn, J. A. (2009). The role of the basal ganglia in beat perception. *Ann. N. Y. Acad. Sci.* 1169, 35–45. doi: 10.1111/j.1749-6632.2009.04553.x

Grahn, J. A., and Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* 19, 893–906. doi: 10.1162/jocn.2007.19.5.893

Grahn, J. A., Henry, M. J., and McAuley, J. D. (2011). FMRI investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa. *NeuroImage* 54, 1231–1243. doi: 10.1016/j.neuroimage.2010.09.033

Grahn, J. A., and McAuley, J. D. (2009). Neural bases of individual differences in beat perception. *NeuroImage* 47, 1894–1903. doi: 10.1016/j.neuroimage.2009.04.039

Grahn, J. A., and Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J. Neurosci.* 29, 7540–7548. doi: 10.1523/jneurosci.2018-08.2009

Grahn, J. A., and Rowe, J. B. (2013). Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cereb. Cortex* 23, 913–921. doi: 10.1093/cercor/bhs083

Grube, M., Lee, K.-H., Griffiths, T. D., Barker, A. T., and Woodruff, P. W. (2010). Transcranial magnetic theta-burst stimulation of the human cerebellum distinguishes absolute, duration-based from relative, beat-based perception of subsecond time intervals. *Front. Psychol.* 1:171. doi: 10.3389/fpsyg.2010.00171

Harrington, D. L., Zimbelman, J. L., Hinton, S. C., and Rao, S. M. (2010). Neural modulation of temporal encoding, maintenance, and decision processes. *Cereb. Cortex* 20, 1274–1285. doi: 10.1093/cercor/bhp194

Hohwy, J. (2013). *The Predictive Mind*. New York, NY: Oxford University Press.

Honing, H. (2012). Without it no music: beat induction as a fundamental musical trait. *Ann. N. Y. Acad. Sci.* 1252, 85–91. doi: 10.1111/j.1749-6632.2011.06402.x

Honing, H., Bouwer, F. L., and Háden, G. P. (2014). "Perceiving temporal regularity in music: the role of auditory event-related potentials (ERPs) in probing beat perception," in *Neurobiology of Interval Timing. Advances in Experimental Medicine and Biology*, eds H. Merchant and V. de Lafuente (New York, NY: Springer), 305–323. doi: 10.1007/978-1-4939-1782-2_16

Honing, H., Bouwer, F. L., Prado, L., and Merchant, H. (2018). Rhesus monkeys (*Macaca mulatta*) sense isochrony in rhythm, but not the beat: additional support for the gradual audiomotor evolution hypothesis. *Front. Neurosci.* 12:475. doi: 10.3389/fnins.2018.00475

Hove, M. J., and Keller, P. E. (2010). Spatiotemporal relations and movement trajectories in visuomotor synchronization. *Music Percept.* 28, 15–26. doi: 10.1525/mp.2010.28.1.15

Huang, Y. Z., Edwards, M. J., Rounis, E., and Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron* 45, 201–206. doi: 10.1016/j.neuron.2004.12.033

Huron, D. (2008). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

Iversen, J. R., and Balasubramaniam, R. (2016). Synchronization and temporal processing. *Curr. Opin. Behav. Sci.* 8, 175–180. doi: 10.1016/j.cobeha.2016.02.027

Iversen, J. R., Patel, A. D., Nicodemus, B., and Emmorey, K. (2015). Synchronization to auditory and visual rhythms in hearing and deaf individuals. *Cognition* 134, 232–244. doi: 10.1016/j.cognition.2014.10.018

Ivry, R. B., and Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends Cogn. Sci.* 12, 273–280. doi: 10.1016/j.tics.2008.04.002

Jacoby, N., and McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Curr. Biol.* 27, 359–370. doi: 10.1016/j.cub.2016.12.031

Janata, P., Tomic, S. T., and Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *J. Exp. Psychol.* 141, 54–75. doi: 10.1037/a0024208

Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 751–751. doi: 10.1017/s0140525x08006079

Kilavik, B. E., Confais, J., and Riehle, A. (2014). Signs of timing in motor cortex during movement preparation and cue anticipation. *Adv. Exp. Med. Biol.* 829, 121–142. doi: 10.1007/978-1-4939-1782-2_7

Koelsch, S., Vuust, P., and Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends Cogn. Sci.* 23, 63–77. doi: 10.1016/j.tics.2018.10.006

Koshimori, Y., Strafella, A. P., Valli, M., Sharma, V., Cho, S.-S., Houle, S., et al. (2019). Motor synchronization to rhythmic auditory stimulation (RAS) attenuates dopaminergic responses in ventral striatum in young healthy adults: [11C]-()-PHNO PET study. *Front. Neurosci.* 13:106. doi: 10.3389/fnins.2019.00106

Koshimori, Y., and Thaut, M. H. (2018). Future perspectives on neural mechanisms underlying rhythm and music based neurorehabilitation in Parkinson's disease. *Aging Res. Rev.* 47, 133–129. doi: 10.1016/j.arr.2018.07.001

Kotz, S. A., Ravignani, A., and Fitch, T. C. (2018). The evolution of rhythm processing. *Trends Cogn. Sci.* 10, 896–910. doi: 10.1016/j.tics.2018.08.002

Ladinig, O., Honing, H., Háden, G., and Winkler, I. (2009). Probing attentive and pre-attentive emergent meter in adult listeners without extensive music training. *Music Percept.* 28, 377–386. doi: 10.1525/mp.2009.26.4.377

Large, E. W., Herrera, J. A., and Velasco, M. J. (2015). Neural networks for beat perception in musical rhythm. *Front. Syst. Neurosci.* 9:159. doi: 10.3389/fnsys.2015.00159

Large, E. W., and Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychol. Rev.* 106, 119–159. doi: 10.1037/0033-295X.106.1.119

Matell, M. S., and Meck, W. H. (2000). Neuropsychological mechanisms of interval timing behavior. *Bioessays* 22, 94–103. doi: 10.1002/(SICI)1521-1878(200001)22:1<94::AID-BIES14>3.0.CO;2-E

Matell, M. S., and Meck, W. H. (2004). Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Cogn. Brain Res.* 21, 139–170. doi: 10.1016/j.cogbrainres.2004.06.012

Matthews, T. E., Witek, M. A. G., Heggli, O. A., Penhune, V. B., and Vuust, P. (2019). The sensation of groove is affected by the interaction of rhythmic and harmonic complexity. *PLoS One* 14:1. doi: 10.1101/415372

Matthews, T. E., Witek, M. A. G., Lund, T., Vuust, P., and Penhune, V. B. (2020). The sensation of groove engages motor and reward networks. *NeuroImage* 214:116768. doi: 10.1016/j.neuroimage.2020.116768

Merchant, H., Bartolo, R., Pérez, O., Méndez, J. C., Mendoza, G., Gámez, J., et al. (2014). "Neurophysiology of timing in the hundreds of milliseconds: multiple layers of neuronal clocks in the medial premotor areas," in *Neurobiology of Interval Timing. Advances in Experimental Medicine and Biology*, eds H. Merchant and V. de Lafuente (New York, NY: Springer), 143–154. doi: 10.1007/978-1-4939-1782-2_8

Merchant, H., Harrington, D. L., and Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annu. Rev. Neurosci.* 36, 313–336. doi: 10.1146/annurev-neuro-062012-170349

Merchant, H., and Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Front. Neurosci.* 7:274. doi: 10.3389/fnins.2013.00274

Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago, IL: The University of Chicago Press.

Morillon, B., and Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. *Proc. Natl. Acad. Sci. U.S.A.* 114, E8913–E8921. doi: 10.1073/pnas.1705373114

Nozaradan, S., Schwartze, M., Obermeier, C., and Kotz, S. A. (2017). Specific contributions of basal ganglia and cerebellum to the neural tracking of rhythm. *Cortex* 95, 156–168. doi: 10.1016/j.cortex.2017.08.015

Pabst, A., and Balasubramaniam, R. (2018). Trajectory formation during sensorimotor synchronization and syncopation to auditory and visual metronomes. *Exp. Brain Res.* 236, 2847–2856. doi: 10.1007/s00221-018-5343-y

Parr, T., and Friston, K. J. (2019). Generalised free energy and active inference. *Biol. Cybernet.* 113, 495–513. doi: 10.1007/s00422-019-00805-w

Patel, A. D., and Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Front. Syst. Neurosci.* 8:57. doi: 10.3389/fnsys.2014.00057

Perez, O., Kass, R., and Merchant, H. (2013). Trial time warping to discriminate stimulus-related from movement-related neural activity. *J. Neurosci. Methods* 212, 203–210. doi: 10.1016/j.jneumeth.2012.10.019

Polak, R., Jacoby, N., Fischinger, T., Goldberg, D., Holzapfel, A., and London, J. (2018). Rhythmic prototypes across cultures: a comparative study of tapping synchronization. *Music Percept.* 36, 1–23. doi: 10.1525/mp.2018.36.1.1

Povel, D. J., and Essens, P. (1985). Perception of temporal patterns. *Music Percept.* 2, 411–440. doi: 10.2307/40285311

Rao, R. P. N., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear. Res.* 271, 16–25. doi: 10.1016/j.heares.2010.09.001

Rauschecker, J. P. (2018). Where, when, and how: are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex* 98, 262–268. doi: 10.1016/j.cortex.2017.10.020

Rauschecker, J. P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11800–11806. doi: 10.1073/pnas.97.22.11800

Ravignani, A., Dalla-Bella, S., Falk, S., and Kello, C. (2019). Rhythm in speech and animal vocalizations: a cross-species perspective: speech rhythm across species. *Ann. N. Y. Acad. Sci.* 1452, 79–98. doi: 10.1111/nyas.14166

Ross, J. M., Iversen, J., and Balasubramaniam, R. (2018a). Dorsal premotor contributions to auditory rhythm perception: causal transcranial magnetic stimulation studies of interval, tempo, and phase. *bioRxiv [Preprint]*. doi: 10.1101/368597

Ross, J. M., Iversen, J. R., and Balasubramaniam, R. (2018b). The role of posterior parietal cortex in beat-based timing perception: a continuous theta burst stimulation study. *J. Cogn. Neurosci.* 30, 634–643. doi: 10.1162/jocn_a_01237

Ross, J. M., Iversen, J. R., and Balasubramaniam, R. (2016a). Motor simulation theories of musical beat perception. *Neurocase* 22, 558–565. doi: 10.1080/13554794.2016.1242756

Ross, J. M., Warlaumont, A. S., Abney, D. H., Rigoli, L. M., and Balasubramaniam, R. (2016b). Influence of musical groove on postural sway. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 308–319. doi: 10.1037/xhp0000198

Sarno, S., de Lafuente, V., Romo, R., and Parga, N. (2017). Dopamine reward prediction error signal codes the temporal evaluation of a perceptual decision report. *Proc. Natl. Acad. Sci. U.S.A.* 114, E10494–E10503. doi: 10.1073/pnas.1712479114

Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends Cogn. Sci.* 11, 211–218. doi: 10.1016/j.tics.2007.02.006

Schultz, W. (2007). Multiple dopamine functions at different time courses. *Ann. Rev. Neurosci.* 30, 259–288. doi: 10.1146/annurev.neuro.28.061604.135722

Stupacher, J., Hove, M. J., Novembre, G., Schütz-Bosbach, S., and Keller, P. E. (2013). Musical groove modulates motor cortex excitability: a TMS investigation. *Brain Cogn.* 82, 127–136. doi: 10.1016/j.bandc.2013.03.003

Sussman, E., Winkler, I., and Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychon. Bull. Rev.* 1, 630–637. doi: 10.3758/BF03196525

Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., et al. (2017). Neural entrainment to the beat: the "missing-pulse" phenomenon. *J. Neurosci.* 37, 6331–6341. doi: 10.1523/JNEUROSCI.2500-16.2017

Tanji, J. (2001). Sequential organization of multiple movements: involvement of cortical motor areas. *Annu. Rev. Neurosci.* 24, 631–651. doi: 10.1146/annurev.neuro.24.1.631

Teki, S., Grube, M., Kumar, S., and Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *J. Neurosci.* 10, 3805–3812. doi: 10.1523/JNEUROSCI.5561-10.2011

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.

van der Weij, B., Pearce, M. T., and Honing, H. (2017). A probabilistic model of meter perception: simulation enculturation. *Front. Psychol.* 8:824. doi: 10.3389/fpsyg.2017.00824

Vuust, P., and Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Front. Psychol.* 5:1111. doi: 10.3389/fpsyg.2014.01111

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., and Honing, H. (2009). Newborn infants detect the beat in music. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2468–2471. doi: 10.1073/pnas.0809035106

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 15 September 2020
doi: 10.3389/fnhum.2020.557642

# Do Musicians Have Better Mnemonic and Executive Performance Than Actors? Influence of Regular Musical or Theater Practice in Adults and in the Elderly

**Mathilde Groussard*, Renaud Coppalle, Thomas Hinault and Hervé Platel**

*UNICAEN, PSL Research University, EPHE, INSERM, U1077, CHU de Caen, Cyceron, Neuropsychologie et Imagerie de la Mémoire Humaine, Normandie Université, Caen, France*

The effects of musical practice on cognition are well established yet rarely compared with other kinds of artistic training or expertise. This study aims to compare the possible effect of musical and theater regular practice on cognition across the lifespan. Both of these artistic activities require many hours of individual or collective training in order to reach an advanced level. This process requires the interaction between higher-order cognitive functions and several sensory modalities (auditory, verbal, visual and motor), as well as regular learning of new pieces. This study included participants with musical or theater practice, and healthy controls matched for age (18–84 years old) and education. The objective was to determine whether specific practice in these activities had an effect on cognition across the lifespan, and a protective influence against undesirable cognitive outcomes associated with aging. All participants underwent a battery of cognitive tasks that evaluated processing speed, executive function, fluency, working memory, verbal and visual long-term memories, and non-verbal reasoning abilities. Results showed that music and theater artistic practices were strongly associated with cognitive enhancements. Participants with musical practice were better in executive functioning, working memory and non-verbal reasoning, whereas participants with regular acting practice had better long-term verbal memory and fluency performance. Thus, taken together, results suggest a differential effect of these artistic practices on cognition across the lifespan. Advanced age did not seem to reduce the benefit, so future studies should focus on the hypothetical protective effects of artistic practice against cognitive decline.

**Keywords: music, theater, practice, cognition, aging, lifespan**

OPEN ACCESS

***Edited by:***
*Marta Olivetti Belardinelli, Sapienza University of Rome, Italy*

***Reviewed by:***
*Natasha Sigala, Brighton and Sussex Medical School, United Kingdom*
*Alexandra Parbery-Clark, Swedish Medical Center, United States*

****Correspondence:***
*Mathilde Groussard*
*mathilde.groussard@unicaen.fr; groussard@cyceron.fr*

***Specialty section:***
*This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience*

***Received:*** *30 April 2020*
***Accepted:*** *18 August 2020*
***Published:*** *15 September 2020*

***Citation:***
*Groussard M, Coppalle R, Hinault T and Platel H (2020) Do Musicians Have Better Mnemonic and Executive Performance Than Actors? Influence of Regular Musical or Theater Practice in Adults and in the Elderly. Front. Hum. Neurosci. 14:557642. doi: 10.3389/fnhum.2020.557642*

## INTRODUCTION

Without a doubt, musical practice has become a model for the study of neuroplasticity in cognitive neuroscience over the past 20 years (Altenmüller, 2008; Schlaug, 2015). It is now accepted that musical expertise leads to cerebral reorganizations resulting in changes in the brain anatomy of regions engaged during formal music learning, such as motor (Wan and Schlaug, 2010), auditory perception (Parbery-Clark et al., 2013; Bidelman and Alain, 2015; Zendel et al., 2019) and memory areas (Groussard et al., 2010, 2014; Fauvel et al., 2013). Musical practice also influences cognitive functioning, involving better performance on tasks that directly call upon skills explicitly learned during formal music learning (*near transfers*) but also with an effect on general cognitive functions (*far transfers)* in musicians (Fauvel et al., 2013; Schellenberg and Weiss, 2013; Schlaug, 2015). Studies reported better performance for musicians compared to non-musicians mainly in executive functioning, notably working memory, flexibility and verbal fluency (Degé et al., 2011; Criscuolo et al., 2019).

Some authors suggested that beyond musical practice, an active, socially engaged, mentally and physically stimulating lifestyle can also have a positive effect on cognitive functioning (Jung et al., 2017). Brain activity and structure are shaped by experience throughout the lifespan, even at an old age. This plasticity has often been demonstrated after long and intensive trainings, where performance in trained activity improves after practice and leads to the building of a cognitive reserve that could explain the interindividual variability regarding aging (Stern, 2009; Chan et al., 2018). This suggests that higher cognitive reserve is associated with compensatory adjustment and could slow down age-related cognitive decline (Kalpouzos et al., 2008; Hinault and Lemaire, 2020). Different factors influence the variability of this reserve among subjects, including levels of education and general lifestyle (diet and physical fitness), but also the quality of social interaction and hobbies (Scarmeas and Stern, 2003).

Like music, theatrical practice is an artistic activity that requires many hours of individual or collective training in order to reach an advanced level. This process requires the interaction between higher-order cognitive functions and several sensory-cognitive modalities (auditory, verbal, visual and motor), as well as the regular learning of new pieces (for musicians see Brown et al., 2015). However, only few studies have investigated the positive effect of artistic activities other than music, such as theater, on cognition in adulthood. While theatrical practice also seems to have potential effect on overall well-being and cognition, its effect on cognitive functions is still poorly understood. To our knowledge, only Noice's team has conducted a series of studies to specify the effect of theatrical practice on cognitive processes. These studies investigated cognitive changes following short-term theatrical interventions in older adults (Noice and Noice, 1999; Noice et al., 2004, 2014; Banducci et al., 2017 for review). They compared older participants who received theater arts training ($n$ = 44), visual arts training ($n$ = 44) or no-treatment (controls, $n$ = 36) during nine 90-min sessions over a month. The pretest and posttest comparison suggested that the performance of the theater arts intervention group was better than no-intervention group on word recall, listening span and problem-solving tasks. Compared to the visual arts group, the theater arts group performed better on problem solving only (Noice et al., 2004). Recent work from this team further expands these results, as Banducci et al. (2017) compared the cognitive benefit of an active acting program including 86 healthy aging versus 93 participants constituting the control group (history of art classroom) for 4 weeks. A cognitive battery was administered before and after intervention, and again in a 4-month follow-up. The participants of the active acting program benefited most relative to the control group in episodic recall only, with gains still evident up to 4 months after intervention. Both groups were similar in the magnitude of gains in working memory, executive function and processing speed. Due to the scarcity of work on theater practice compared to music training in the literature, it seems necessary to specify the benefits of theatrical regular practice on cognition and to better understand its effect throughout life. Many factors appear to influence an individual's aging trajectory (Raz and Kennedy, 2009; Hinault and Lemaire, 2020), suggesting that interventions could possibly slow down cognitive decline and promote healthy aging. To this end, various behavioral interventions have been proposed, such as physical activity and cognitive training (Colcombe and Kramer, 2003; Jaeggi et al., 2008; Karbach and Kray, 2009; Fauvel et al., 2013; Sprague et al., 2019 for reviews), and the benefits of arts practices for promoting health have received growing interest. Importantly, while prior studies have undoubtedly shown the association between arts engagement and well-being (Mella et al., 2017; Fancourt and Steptoe, 2019), objective measurement of the specific cognitive benefit associated with repeated and regular art practices like acting or music across the lifespan, have not been carried out to the best of our knowledge.

The literature is consistent with results obtained by the first study performed on the link between the practice of a musical instrument and cognitive functions in elderly subjects (Hanna-Pladdy and MacKay, 2011). These authors observed that elderly musicians outperformed elderly non-musicians on non-verbal working memory, naming and executive function tests. Moreover, their study suggested a correlation between musicians considered to have a high level of expertise (i.e., having at least more than 10 years of practice) and the preservation of cognitive functioning while aging. Contrarily, Mansens et al. (2018) showed that the time spent making music was not the most important criterion with respect to cognitive function compared with other practice characteristics such as current amount of time making music or age of onset of musical practice.

To our knowledge, no study has been carried out among people engaged in theater practice in order to evaluate their cognitive abilities and possible reserve effects. However, it seems logical to think that many years of theater practice could influence cognition and especially memory, as actors have to memorize new texts regularly. Similarly, it seems surprising that actors have never constituted a reference group or a comparison group with musicians. This is probably due to the fact that it is more difficult to define equivalent criteria for the level of expertise for actors whose training and practices are more heterogeneous than for musicians educated in music conservatories.

The main objective of this study was to compare the positive influence of musical and theatrical current and regular practice on cognition. Our goal was first to determine, throughout the lifespan, whether people with current musical and theatrical practice could show cognitive differences, and if the number of years of practice influence these modifications. Second, we aimed to study the effect of musical and theatrical practice on cognition in older adults, in order to assess the specific differences in cognition between actors and musicians while aging (e.g., executives and episodic memory processes).

Participants underwent a battery of cognitive tasks evaluating processing speed, executive functioning, fluency, working memory, verbal and visual long-term memories, and non-verbal reasoning abilities. Considering the literature, the main expectation was that both groups (musicians and actors) would perform higher than control subjects without any intensive and regular leisure activity on tasks evaluating both executive functions and memory. We also expected a specific effect of the type of practice on cognition, with better performance for verbal memory in actors and executive functioning and reasoning in musicians. Moreover, we expected that processes involved in reading scores in musicians would increase their abilities in visuo-spatial memory. We finally expected that these patterns would be maintained in older adults.

# MATERIALS AND METHODS

## Participants

We recruited three groups of healthy subjects differing only in regular and sustained practice of a specific leisure activity (music; theater, no specific practice for the control group). The dataset was obtained from 146 participants, 50 controls; 50 "musicians" and 46 "actors" matched for age and education. Three participants (1 by group) were excluded because they only partially completed the neuropsychological assessment (**Table 1**). All participants were native French speakers, with normal hearing and normal or corrected to normal vision without any reported neurological or psychiatric conditions, as assessed by a medical interview. None of them presented signs of cognitive impairment (i.e., two or more scores in two or more cognitive domains below two standard deviations of the norms for their age class). All participants provided informed consent before being tested and all procedures were conducted in agreement with the ethical principles of Declaration of Helsinki.

Controls were defined as participants who had practiced any leisure activity regularly (more than 4 h/week) associated with formal lessons (physical activity or drawing lesson for examples) and had never taken any formal music or acting lessons, that could neither play nor read music.

Participants were included in "Musicians" group when they reported current and regular practice at the moment of the study for more than 3 years of musical instrument without interruption, more than 4 h/week, and if they had received formal music training. In addition to this, Musicians had to never have practiced theater. Musicians were recruited from several French conservatories or music schools (no self-educated musicians were included). They played various musical instruments (piano, guitar, trumpet, etc.). To study the influence of instrumental practice and avoid confounding effect between singing and instrumental practices (e.g., Mansens et al., 2018) we excluded participants who previously performed choral singing in all groups.

Participants were included in "Actors" group when they reported current and regular acting practice at the moment of the study for more than 3 years without interruption, more than 4 h/week, and if they had taken formal theater lessons. In addition to this, Actors must had never practiced a musical instrument. All Actors were recruited from theater companies or cultural associations which offered acting lessons. Without being professional, they learned new texts on a regular basis and had regular live performances (two per month in average).

Musicians' and Actors' background information is provided in **Table 1**, which includes the age onset of practice, number of years of practice, weekly practice hours, number of exhibitions by year.

We then reduced our sample to individuals who were 50 years and older to study the difference in cognition between older practitioners (Schneider et al., 2019). This sample was composed of 27 Controls, 24 Musicians and 15 Actors (see **Table 2**).

## Cognitive Functioning

Cognitive functioning was measured using several assessments covering various cognitive domains (**Table 3**) that are clearly impacted by normal aging (Hedden and Gabrieli, 2004; Salthouse, 2010), the earliest and most concerned being processing speed, working memory (maintenance and manipulation of information during a short period of time), spatial ability, reasoning, and episodic memory (declarative long-term contextual remembering of personal events or information). The entire test battery was administered in a single session, which lasted about 90 min, and took place in a quiet room.

### Long-Term Memory

Long-term verbal memory was measured using the 12-word subtest from the Signoret BEM-144 (Signoret, 1991). This verbal memory test consists of learning 12 words during 3 sessions. After every trial, participants are asked to recall as many words as

**TABLE 1** | Demographic data of participants and practice background information for Musicians and Actors.

| | Controls | Musicians | Actors | Stats | *p*-value |
|---|---|---|---|---|---|
| N of subjects | 49 | 49 | 45 | | |
| Gender (F/M) | 30F + 20 M | 24F + 26 M | 30F + 16 M | $X^2 = 3.096$ | 0.213 |
| Age | 47.47 ± 17.78 [18–80] | 47.84 ± 18.3 [20–83] | 41.58 ± 18.13 [18–84] | $F = 1.748$ | 0.178 |
| Years of education | 13.9 ± 2.946 [9–20] | 14.73 ± 2.139 [9–19] | 13.89 ± 2.648 [7–20] | $F = 1.689$ | 0.188 |
| Age onset of practice | n.a | 13.02 ± 12.82 [2–65] | 24.9 ± 15.34 [4–70] | U MW = 506 | **<0.001** |
| Years of practice | n.a | 31.51 ± 16.67 [5–65] | 17.13 ± 10.82 [4–45] | U MW = 539.5 | **<0.001** |
| Weekly practice hours | n.a | 13.03 ± 11.03 [1–49] | 10.18 ± 9.83 [2–40] | U MW = 1.318 | 0.191 |
| Number of exibition/year | n.a | 18.14 ± 24.84 [0–100] | 14.43 ± 22.14 [0–100] | U MW = 0.76 | 0.448 |

**TABLE 2 |** Demographic data of participants of 50 years and older Practice background information for Musicians and Actors of 50 years and older.

| | Controls | Musicians | Actors | Stats | *p*-value |
|---|---|---|---|---|---|
| N of subjects | 27 | 24 | 15 | | |
| Gender (F/M) | 18F + 9 M | 12F + 12 M | 9F + 6 M | $X^2$ = 1.47 | 0.48 |
| Age | 61.1 ± 7.29 [51–80] | 63.9 ± 8.81 [50–83] | 62.6 ± 9.73 [52–84] | *F* = 0.723 | 0.489 |
| Years of education | 14.5 ± 2.65 [9–19] | 14.2 ± 1.99 [9–17] | 12.9 ± 2.45 [10–20] | *F* = 2.31 | 0.108 |
| Age onset of practice | n.a | 18.5 ± 16.6 [2–65] | 37.4 ± 17 [12–70] | U MW = 69.5 | **<0.001** |
| Years of practice | n.a | 43 ± 15 [12–65] | 23.7 ± 13.4 [4–45] | U MW = 60.5 | **<0.001** |
| Weekly practice hours | n.a | 12.8 ± 11.00 [3.5–49] | 10.5 ± 11.8 [3–40] | U MW = 130 | 0.152 |
| Number of exibition/year | n.a | 14.3 ± 18.6 [0–60] | 15.9 ± 20.4 [2–80] | U MW = 142 | 0.27 |

possible. Then participants are distracted by performing a non-verbal task for approximately 7 min. After that, they are asked to recall as many words as possible. We used two scores on the 12-word BEM test: the total score of the three trials to evaluate total learning (BEM Total), and the number of words recalled during the delayed recall (BEM Recall) to assess episodic memory.

Long-term visual memory was evaluated using the Baddeley's Doors test (Baddeley, 1994). This test is a non-verbal recognition test based on colored photographs of doors composed of two parts (A and B). For each part, 12 doors are shown individually for 3 s. Then, the participants are asked to pick out the one out of the four that had been shown before. The score is the number of correct answers of the two parts combined (Doors Total).

The Rey-Osterrieth complex figure (Rey, 1959) was administered and consists of redrawing an abstract geometrical shape from memory that had been copied 3 min earlier (Rey Recall). The maximum final score is 36. This test is classically used for evaluating of visuospatial constructional ability and visual episodic memory.

**TABLE 3 |** Description of the tests and the dependent variables used in the study.

| Cognitive functions assessed | Psychometric tests | Dependent variables (outcomes) |
|---|---|---|
| Long-term verbal memory | BEM-144's 12 words (Signoret, 1991) | Total score of the three trials in learning phase *(BEM Total)* |
| | | Number of word recalled in the delayed recall *(BEM Recall)* |
| Long-term visual memory | Doors test (Baddeley, 1994) | Number of doors recognized *(Doors Total)* |
| | Rey-Osterrieth complex figure (Rey, 1959) | Score of redraw fidelity (number of details, their completeness and location) *(Rey Recall)* |
| Working memory | Forward digit span (Godefroy et al., 2008) | Highest number of digits properly recalled in 2/3 trials *(Digit Span)* |
| Attentional abilities | d2 Test (Brickenkamp, 1981) | Total number of target symbols correctly identified *(GZ-f)* |
| Executive control and verbal abilities | Phonemic Fluency task (Cardebat et al., 1990) | Total number of words in 2 min *(Phonemic Fluency)* |
| | Semantic Fluency task (Cardebat et al., 1990) | Total number of words in 2 min *(Semantic Fluency)* |
| Processing speed | d2 Test (Brickenkamp, 1981) | Number of items processed *(GZ)* |
| | Digit-symbol coding subtest (Wechsler, 2000) | Number of correct associations in 2 min *(Codes)* |
| Non-verbal reasoning | Matrix Reasoning subtests (Wechsler, 2000) | Number of matrices properly completed *(Matrix)* |

## Executive Functioning

The phonological loop of the working memory, which is the ability to retain verbal information for a short time by mean of mental repetition (Baddeley et al., 1998; Baddeley, 2003), was evaluated using the forward digit span (Godefroy et al., 2008). Participants had to immediately recall series of digits in the order they were presented. The score recorded was the size of the forward digit span with 2 successive correct recalls (Digit Span).

We evaluated visual attention using the d2 Test. This test consists of a paper with 14 rows of 47 interspersed "p" and "d" characters. The participant had to cross out as many "d" with two marks above or below them as possible, in any order (target symbols), and had to jump to the next rows every 15 s. The target symbols are relatively similar to the distractors (a "p" with two marks or a "d" with one or three marks). In this study, we used the overall performance score (GZ-f) corresponding to the total number of target symbols correctly identified (Brickenkamp, 1981).

The phonemic fluency task (Cardebat et al., 1990) was used to measure executive functioning. The participant had to name as many words starting with the letter R as possible in 2 min. The score used is the total number of words in 2 min. The semantic fluency was also proposed, in which participants had to name as many fruits as possible in 2 min. Language processing and semantic memory are most the critical components for this task.

Processing speed was measured using the digit-symbol coding subtest from the Weschler Adult Intelligence Scale (WAIS-III, Wechsler, 2000). In this task, each digit (from 1 to 9) was combined with a specific symbol (example 1/— and 9/=) in the upper row. Participants then had 2 min to complete the number maximum of symbols corresponding to the digits presented in the lower rows. The score was the number of correct associations performed in 2 min (Codes). We also used the number of items processed (Gz) in the d2 test (Brickenkamp, 1981) to evaluate the processing speed.

## Reasoning

We administered the Matrix Reasoning subtests of the WAIS-III (Wechsler, 2000) to estimate participants' non-verbal reasoning skills. In this test, participants were presented with an unfinished matrix of drawings, and had to choose the drawing that logically completed the matrix. This task is classically associated with fluid intelligence (Matrix) (Carpenter et al., 1990). We used the number of matrices properly completed as a performance score.

## Procedure and Statistical Analysis

To compare all cognitive variables with each other, scores were standardized. Thus, we transformed all neuropsychological measures into *z*-scores using the mean and standard deviation of the control groups ($n = 49$) as the reference population for each measure due to the lack or poor reliability of French published norms for some assessments. Thus, all variables were on the same scale, with a mean of 0 and a standard deviation of 1 based on the control group. The higher the *z*-score, the better the performance. This allowed comparing every performance on the same normalized scale.

In order to test for group difference among cognitive tests, we performed multivariate analyses of covariance (MANCOVA), with type of practice (Controls; Musicians; Actors) as between-subjects factor, the cognitive test scores (BEM Total, BEM Recall, Doors Total, Rey Recall, Digit Span, Matrix, Codes, d2 GZ, d2 GZ-f, Phonemic Fluency, Semantic Fluency) as dependent variables and age as confounding variable. As we included more than two dependent variables and because they are intercorrelated, we opted for a MANCOVA. This statistical analysis accounts for the relationship between dependent variables (Warne, 2014). To complete the multivariate analyses and examine group differences for each cognitive variable, univariate analyses were performed. A family wise Bonferroni's correction for multiple comparison analyses was carried out 2-by-2 for every significant test.

To further our exploration of the relationship between musical or theatrical practice and cognitive abilities, we performed a second multivariate analysis on cognitive variables restricted to Musicians and Actors and including Age and Years of practice as variables.

In a separate analysis, the same procedure was implemented reducing the sample to people who were 50 years and older to study the effect of expertise on cognition in older adults.

The sample size was based on a power analysis, conducted in G*Power 3.1. Regarding behavioral interactions between age and cognition, assuming an effect size of Cohen's $f = 0.65$ [derived from Carey et al. (2015)], an alpha of 0.05, and three groups, we determined that a total sample size of at least 15 individuals per study would provide 95% power to detect the effects.

All the statistical analyses were performed with STATISTICA software. The partial Eta square ($\eta^2_p$) was utilized to estimate effect size. Results were considered significant at $p < 0.05$.

# RESULTS

## Effect of Expertise on Cognitive Variables in Adults

Results from the multivariate tests on the associations between groups of practice (Controls; Musicians; Actors) and cognitive tests adjusted for age exhibited a significant group effect: Wilks' Lambda [$F_{(22,258)} = 3{,}005$, $p = 0.000015$, $\eta^2_{\text{partial}} = 0.204$]. Results showed a significant effect of groups of practice with higher values for Musicians relative to Controls for Rey Recall ($p = 0.024$); Codes ($p = 0.002$); d2 GZ ($p = 0.023$); d2 GZ-f ($p = 0.023$); Phonemic Fluency ($p = 0.002$), Semantic Fluency ($p = 0.019$);relative to Actors for Matrix ($p = 0.006$); and compared to both Controls and Actors for Span (*respectively* $p = 0.0002$; $p = 0.008$). Results also highlighted significantly higher values for Actors than Controls for BEM Total ($p = 0.002$), and Phonemic Fluency ($p = 0.009$), and compared to both Controls and Musicians for BEM Recall (*respectively* $p = 0.001$; $p = 0.033$) (**Table 4** and **Figure 1**).

**TABLE 4 |** Statistical results of MANCOVA for each cognitive variable.

| | Controls | Musicians | Actors | Statistics | | |
|---|---|---|---|---|---|---|
| | **Mean ± SD** | **Mean ± SD** | **Mean ± SD** | ***F*** | ***p*-value** | ***Post hoc*** |
| BEM Total | 25.673 ± 4.451 | 26.857 ± 4.21 | 28.778 ± 4.617 | **4.450** | **0.0134** | Actors > Controls |
| BEM Recall | 8.449 ± 2.39 | 8.939 ± 2.277 | 10.067 ± 1.876 | **5.256** | **0.006** | Actors > Controls<br>Actors > Musicians |
| Doors Total | 17.673 ± 2.593 | 18.082 ± 2.448 | 18.40 ± 2.934 | 0.632 | 0.533 | |
| Rey Recall | 20.806 ± 4.716 | 23.592 ± 4.745 | 23.067 ± 6.463 | **3.763** | **0.026** | Musicians > Controls |
| Digit Span | 5.592 ± 1.29 | 6.551 ± 1.174 | 5.822 ± 0.960 | **9.449** | **0.0001** | Musicians > Controls<br>Musicians > Actors |
| Matrix | 20.531 ± 2.792 | 21.735 ± 2.564 | 20.289 ± 2.928 | **5.328** | **0.006** | Musicians > Actors |
| Codes | 69.184 ± 17.257 | 78.776 ± 16.37 | 73.222 ± 12.269 | **6.906** | **0.001** | Musicians > Controls |
| d2 GZ | 391.306 ± 72.378 | 431.327 ± 80.379 | 401.556 ± 79.371 | **4.729** | **0.010** | Musicians > Controls |
| d2 GZ-f | 376.878 ± 65.827 | 413.551 ± 77.785 | 382.422 ± 71.237 | **5.285** | **0.006** | Musicians > Controls |
| Phonemic Fluency | 21.408 ± 5.733 | 25.755 ± 5.445 | 25.20 ± 7.191 | **7.619** | **0.0007** | Musicians > Controls<br>Actors > Controls |
| Semantic Fluency | 21.265 ± 4.420 | 23.796 ± 4.509 | 23 ± 4.661 | **3.937** | **0.022** | Musicians > Controls |

*SD, standard deviation. Bold values correspond to significant differences between groups.*

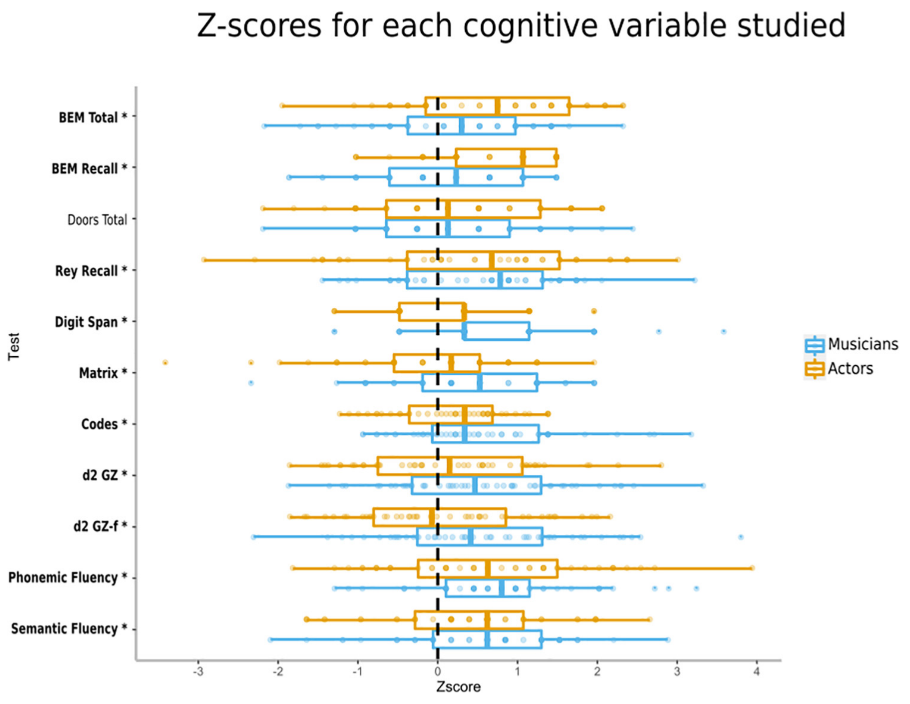



**FIGURE 1 |** Median, 1st and 3rd quartiles, min and max *z*-scores for each cognitive variable studied. Mean and standard deviation of the control group (*n* = 49) serves as reference population for each measure, as such variables were on the same scale with 0 as the mean and 1 as the standard deviation of the control group. The higher the *z*-score, the better the performance. (*) After a test name indicates a significative difference between Controls and Musicians or Actors.

The multivariate analysis exhibited an effect of the Age: Wilks' Lambda [$F_{(11,129)} = 5.286$, $p = 0.000001$], and univariate analyses suggested a significant effect of the Age on BEM Total [$F_{(1,139)} = 12.49$, $p = 0.0005$], BEM Recall [$F_{(1,139)} = 11.32$, $p = 0.001$], Rey Recall [$F_{(1,139)} = 12.319$, $p = 0.0006$], Matrix [$F_{(1,139)} = 13.85$, $p = 0.0003$], Codes [$F_{(1,139)} = 38.56$, $p < 0.001$], d2 GZ [$F_{(1,139)} = 17.85$, $p = 0.0004$], and d2 GZ-f [$F_{(1,139)} = 20.77$, $p = 0.0001$].

## Effect of Years of Practice on Cognitive Variable of Adult Musicians and Actors

The multivariate analysis exhibited a significant effect of groups of practice, Wilks' Lambda [$F_{(11,80)} = 4{,}058$, $p = 0.0001$, $\eta^2_{partial} = 0.358$], a significant effect of Age, Wilks' Lambda [$F_{(11,80)} = 2{,}957$, $p = 0.0025$, $\eta^2_{partial} = 0.289$] and no effect of Years of practice on cognitive variables Wilks' Lambda [$F_{(11,80)} = 1{,}776$, $p = 0.072$, $\eta^2_{partial} = 0.196$]. The univariate analyses on groups of practice confirmed the higher values for Actors relative to Musicians after controlling of Age and Years of practice for BEM Total ($p = 0.039$) and BEM Recall ($p = 0.0046$) and higher values for Musicians relative to Actors for Span ($p = 0.0023$), Matrix ($p = 0.022$), Codes ($p = 0.0109$), d2 GZ ($p = 0.0404$) and d2 GZ-f ($p = 0.0164$).

## Effects of Expertise on Cognitive Variables in Older Adults

In a separate analysis, we reduced our sample to participant who were 50 years and older to study the effects of expertise on cognitive aging. Results from the multivariate tests that studied the associations between groups of practice (Controls; Musicians; Actors) and the cognitive tests adjusted for Age exhibited a significant group effect: Wilks' Lambda [$F_{(22,104)} = 1.732$, $p = 0.0347$, $\eta^2_{partial} = 0.268$]. Results of univariate tests showed a significant effect on groups of practice, with significantly higher values in Musicians relative to Controls for Rey Recall ($p = 0.025$), Digit Span ($p = 0.018$), Codes ($p = 0.026$) and Semantic Fluency ($p = 0.005$). No significantly higher values in Actors relative to Controls and difference between Actors and Musicians (**Table 5** and **Figure 2**). The multivariate an effect of age, Wilks' Lambda [$F_{(11,52)} = 2.869$, $p = 0.005$, $\eta^2 = 0.378$] and univariate analyses suggested a significant effect of age on Doors [$F_{(1,62)} = 12.82$, $p = 0.0007$], Matrix [$F_{(1,62)} = 5.216$, $p = 0.026$], Codes [$F_{(1,62)} = 14.45$, $p = 0.0003$], Phonemic Fluency [$F_{(1,62)} = 4.614$, $p = 0.036$].

## Effect of Years of Practice on Cognitive Variables of Older Musicians and Actors

The multivariate analysis exhibited only a significant effect of Age, Wilks' Lambda [$F_{(11,25)} = 2{,}675$, $p = 0.020$, $\eta^2_{partial} = 0.541$]. No effect of Years of practice on cognitive variables Wilks' Lambda [$F_{(11,25)} = 1.396$, $p = 0.235$, $\eta^2_{partial} = 0.381$] and Groups of practice were observed, Wilks' Lambda [$F_{(11,25)} = 0.614$, $p = 0.219$, $\eta^2_{partial} = 0.387$].

**TABLE 5 |** Statistical results of MANCOVA limited to older adults.

| | Controls | Musicians | Actors | Statistics | | |
|---|---|---|---|---|---|---|
| | Mean ± SD | Mean ± SD | Mean ± SD | *F* | *p*-value | *Post hoc* |
| BEM Total | 24.8 ± 4.03 | 26 ± 4.20 | 26.1 ± 3.81 | 1.039 | 0.359 | |
| BEM Recall | 8 ± 2.43 | 8.46 ± 2.32 | 9 ± 1.69 | 1.065 | 0.351 | |
| Doors Total | 17.6 ± 2.61 | 18 ± 2.51 | 17.9 ± 3.66 | 0.62 | 0.541 | |
| Rey Recall | 18.8 ± 4.51 | 22.6 ± 4.26 | 20.8 ± 7.01 | **4.413** | **0.016** | Musicians > Controls |
| Digit Span | 5.44 ± 1.19 | 6.38 ± 1.28 | 5.87 ± 0.915 | **4.428** | **0.016** | Musicians > Controls |
| Matrix | 19.6 ± 2.49 | 20.6 ± 2.18 | 19.6 ± 3.68 | 1.541 | 1.756 | |
| Codes | 61.3 ± 15.30 | 71.1 ± 13 | 64.3 ± 13.9 | **5.367** | **0.007** | Musicians > Controls |
| d2 GZ | 380 ± 59.60 | 401 ± 79.5 | 366 ± 64.5 | 1.632 | 0.204 | |
| d2 GZ-f | 365 ± 50.70 | 383 ± 76.5 | 347 ± 56.8 | 1.884 | 0.161 | |
| Phonemic Fluency | 22.6 ± 6.23 | 26.3 ± 6.81 | 26.6 ± 8.48 | **3.261** | **0.045** | |
| Semantic Fluency | 20.1 ± 4.67 | 24.3 ± 4.44 | 22.1 ± 4.83 | **5.754** | **0.005** | Musicians > Controls |

*SD, standard deviation. Bold values correspond to significant differences between groups.*

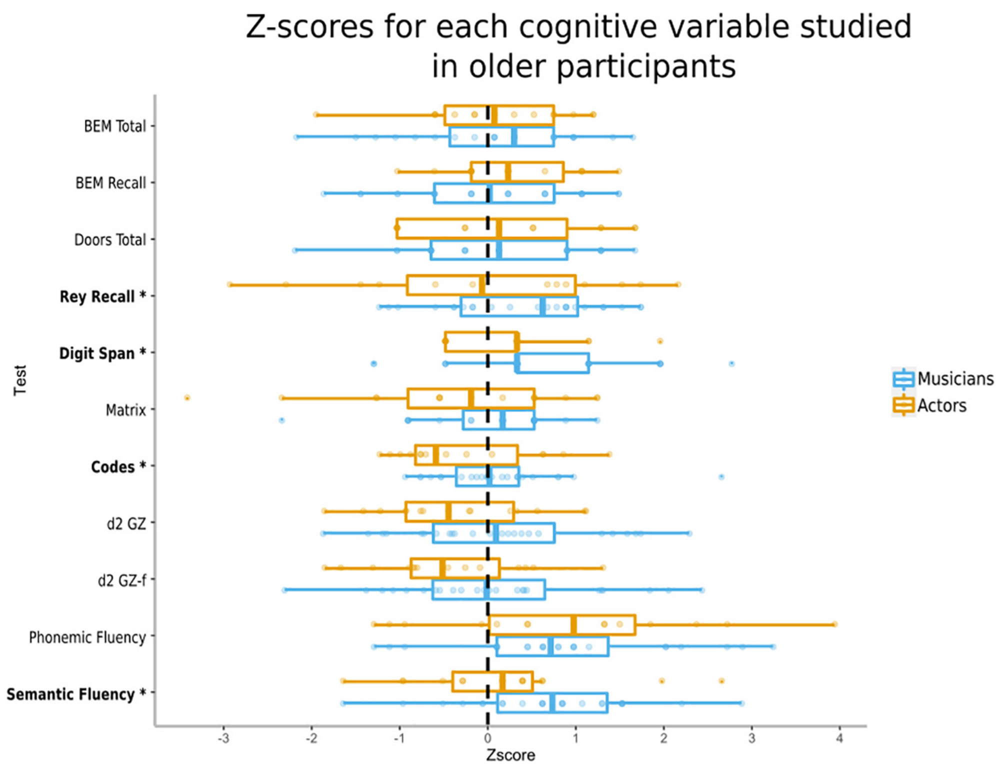



**FIGURE 2 |** Median, 1st and 3rd quartiles, min and max *z*-scores of the older participants for each cognitive variable studied. Mean and standard deviation of the control group (*n* = 49) serves as reference population for each measure, as such variables were on the same scale with 0 as the mean and 1 as the standard deviation of the control group. The higher the *z*-score, the better the performance. (*) after a test name indicates a significative difference between Controls and Musicians or Actors.

## DISCUSSION

The aim of the present study was to investigate possible differences in cognition between different art practitioners (Musicians and Actors), and also people without artistic training. Several studies have demonstrated the benefits of musical practice on cerebral activity and cognition, but comparing this practice to another artistic practice such as theater had yet to be done. Our results are consistent with previous works suggesting that adult musicians outperformed control subjects in standardized cognitive tasks (Fauvel et al., 2013; Schlaug, 2015; Sutcliffe et al., 2020) but a lifespan approach was never adopted in anterior studies. In fact, "previous work has focused on younger musicians or older musicians whereas our sample had a very wide age range (from 18 to 84 years), allowing us to study practitioners" cognitive differences throughout the lifespan. We observed a superiority of musicians in long-term visual-spatial memory, working memory, processing speed, executive functioning and non-verbal reasoning. Nevertheless, Musicians did not seem to outperform Controls on verbal episodic memory and visual memory. This result appears consistent with result of a meta-analysis performed by Talamini et al. (2017) on memory. In fact, these authors suggested a small effect size for long-term memory and a possible domain-specific stimuli effect in favor of musical stimuli (Talamini et al., 2017).

Nevertheless, we also found a difference in the theater group on cognition across the lifespan. In line with results of Noice et al. (2004) obtained after a 4-month theatrical intervention, we observed a better long-term verbal memory and verbal fluency in these subjects, compared to controls and musicians. Our results were observed in a large sample of younger to older adults that presented a sustainable practice and intense training. Actors' better performance in verbal episodic memory is consistent with the abilities developed by them while learning a text and retrieving it during performance. In fact, most actors use mnemotechnical strategies to encode and retrieve their scripts (Noice et al., 2004; Banducci et al., 2017). Strategies were indeed found to improve memory in both young and older adults (e.g., Hinault et al., 2017a,b). However, Actors did not show a difference on executive functioning, working memory and non-verbal reasoning when compared to Control participants. Thus, in future studies it would seem relevant to consider a more detailed evaluation cognition (including strategy use) in order to confirm whether these effects are limited to verbal and memory aspects and do not influence executive processes, or whether self-monitoring abilities are required in theater practice (Nettle, 2006).

This work is the first to statistically compare these two artistic practices, with the same cognitive assessment battery. It highlights for the first time that both musical and theatrical practices could lead to differences in cognition across the lifespan, confirming previous studies on leisure activities, lifestyle and cognition (Hertzog et al., 2008; Reuter-Lorenz and Park, 2014). Furthermore, we observed domain-specific differences, musical practice being associated with better executive functions and reasoning, and theatrical practice with better long-term verbal memory. In fact, Musicians had better performance when compared with Actors on working memory, processing speed, executive functioning, and non-verbal reasoning whereas Actors outperformed Musicians for long-term verbal memory.

In older adults, this pattern seems to be confirmed for musicians, with higher performances on long-term visual-spatial memory, working memory, processing speed and verbal fluency (Hanna-Pladdy and MacKay, 2011; Hanna-Pladdy and Gajewski, 2012; Amer et al., 2013; Fauvel et al., 2014; Mansens et al., 2018; Criscuolo et al., 2019; Ferreri et al., 2019). The EEG study of Moussard et al. (2016) on elderly musicians (currently practicing about 11 h/week) and non-musicians, confirmed a beneficial effect of musical practice on executive control, and highlighted a more anterior distribution of the P3 wave in musicians, suggesting successful functional reorganization in elderly musicians according to the authors. Moreover, longitudinal studies showed that 6 months of piano lessons given to older non-musicians adults could improve working memory and executive functioning (Bugos et al., 2007; Seinfeld et al., 2013). In older actors, Banducci et al. (2017) reported modifications on verbal long term-memory and fluency after a 4-month theatrical intervention in older adults which suggests a cognitive benefit even after a short period of active art practice. While we reported better performances on verbal long-term memory and fluency for Actors compared with Controls, we could not find any significant difference in the elderly for Actors relative to Musicians. These results must to be taken with caution with regard to the small sample size of the older Actors ($n = 15$) and would require further investigation to confirm stronger verbal cognition associated with theatrical practice in aging.

In line with several studies on musicians that suggested no association between practice time and cognitive functions (e.g., Fauvel et al., 2014; Mansens et al., 2018), our results did not reveal any effect of the number of years of practice on assessed cognitive functioning. Thus, having a regular and current practice appears to better explain the cognitive differences we studied rather than years of practice. These results are interesting, as even a short period of practice can lead to an improvement in cognitive performance in adults across the lifespan. There is a growing consensus toward aging brains remaining plastic and consequently involvement in leisure activities such as music or theater remains of significant interest since it is possible to start this type of practice at any age (Noice et al., 2014).

These findings are constrained by several limitations that need to be considered in future research. First, our study is essentially descriptive because of its correlational approach and does not allow us to validate causality. Only future interventional or follow-up studies could confirm these results. Second, we partially evaluated the working memory abilities because working memory updating ability was not assessed, as the digit backward span was not among the cognitive assessment. Future studies could aim at specifying music and theatrical practice differential effects on this cognitive process. Third, musicians and actors differed in their average years of practice, musicians showing a longer practice duration than actors. Although this variable was included as covariate in analyses without significant interactions, it could have explained some of the cognitive differences between our groups. Four, global cognitive functioning was not assessed, but no participant was below the passing score in more than one cognitive measure, in line with preserved overall cognitive performance. Furthermore, although power analyses and previous work support the selected sample size, future studies should investigate cognitive differences between older musicians and actors with a larger sample size.

To conclude, our results suggest that artistic practices can account for different individuals' aging trajectories (Raz and Kennedy, 2009), and that regular artistic practice could promote the constitution of cognitive and cerebral reserve (Stern, 2009; Stern et al., 2018). Therefore, promoting access to artistic practice could help people maintaining or even improving their cognition, besides the obvious and well-documented interest such activities have on socialization (Belgrave, 2011); well-being (Noice et al., 2004; Castora-Binkley et al., 2010) and developing creativity (Salimpoor and Zatorre, 2013; Reynolds et al., 2016). In line with the evidence reviewed by Sutcliffe et al. (2020) on music training and cognition on aging, our study suggests that musical or theatrical practices, even started late in life, could have an effect on cognitive decline. Ferreira et al. (2015) suggested associations between specific activities and the functioning of individual cognitive domains. Results suggest that cognitive training programs could be individually adjusted to observed cognitive deficits following a neuropsychological assessment, without making it a unique criterion for choosing such activity of course. However, previous works on cognitive interventions for aging and dementia showed mixed results (e.g., Alves, 2013). In future clinical studies it would be interesting to determine whether theatrical interventions could improve language and episodic memory processes for people with deficits in theses domains. Conversely, as both previous works and our results highlight the possible positive effect of musical practice on executive functioning (Mansens et al., 2018; Koshimori and Thaut, 2019; Platel and Groussard, 2020), it seems relevant to specify the effect of musical practice interventions for people with deficits in these processes.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

MG and HP conceived the study and coordinated the collect of data. MG and TH performed the statistical analysis. MG wrote the first draft of the manuscript. RC, TH, and HP revised the different version of the manuscript. All the authors approved the submitted version.

## ACKNOWLEDGMENTS

We would like to thank Baptiste Fauvel and the master students of Caen Normandy University for their help in collecting data and their role in this study. Many thanks to Melanie Ambler for proofreading and reviewing English. MG thanks Arthur and Robin Pesnel for their interventions and discussions during writing.

## REFERENCES

Altenmüller, E. (2008). Neurology of musical performance. *Clin. Med*. 8, 410–413. doi: 10.7861/clinmedicine.8-4-410

Alves, J. (2013). Non-pharmacological cognitive intervention for aging and dementia: current perspectives. *World J. Clin. Cases* 1:233. doi: 10.12998/wjcc.v1.i8.233

Amer, T., Kalender, B., Hasher, L., Trehub, S. E., and Wong, Y. (2013). Do older professional musicians have cognitive advantages? *PLoS One* 8:e71630. doi: 10.1371/journal.pone.0071630

Baddeley, A. (1994). The magical number seven: still magic after all these years? *Psychol. Rev*. 101, 353–356. doi: 10.1037/0033-295X.101.2.353

Baddeley, A., Gathercole, S., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychol. Rev*. 105, 158–173. doi: 10.1037/0033-295x.105.1.158

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci*. 4, 829–839. doi: 10.1038/nrn1201

Banducci, S. E., Daugherty, A. M., Biggan, J. R., Cooke, G. E., Voss, M., Noice, T., et al. (2017). Active experiencing training improves episodic memory recall in older adults. *Front. Aging Neurosci*. 9:133. doi: 10.3389/fnagi.2017.00133

Bidelman, G. M., and Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J. Neurosci*. 35, 1240-1249. doi: 10.1523/JNEUROSCI.3292-14.2015

Belgrave, M. (2011). The effect of a music therapy intergenerational program on children and older adults' intergenerational interactions, cross-Age attitudes, and older adults' psychosocial well-being. *J. Music Ther*. 48, 486–508. doi: 10.1093/jmt/48.4.486

Brickenkamp, R. (1981). Test d2 - Aufmerksamkeits-Belastungs-Test. *Handbuch Psychologischer Und Pädagogischer Tests*. Available at: http://library.mpib-berlin.mpg.de/toc/z2008_2431.pdf, 270–273.

Brown, R. M., Zatorre, R. J., and Penhune, V. B. (2015). Expert music performance: cognitive, neural, and developmental bases. *Prog. Brain* Res. 217 57–86. doi: 10.1016/bs.pbr.2014.11.021

Bugos, J. A., Perlstein, W. M., McCrae, C. S., Brophy, T. S., and Bedenbaugh, P. H. (2007). Individualized Piano Instruction enhances executive functioning and working memory in older adults. *Aging Ment. Health* 11, 464–471. doi: 10.1080/13607860601086504

Cardebat, D., Doyon, B., Puel, M., Goulet, P., and Joanette, Y. (1990). Formal and semantic lexical evocation in normal subjects. Performance and dynamics of production as a function of sex, age and educational level. *Acta Neurol. Belgica* 90, 207–217.

Carey, D., Rosen, S., Krishnan, S., Pearce, M. T., Shepherd, A., Aydelott, J., et al. (2015). Generality and specificity in the effects of musical expertise on perception and cognition. *Cognition* 137, 81–105. doi: 10.1016/j.cognition.2014.12.005

Carpenter, P. A., Just, M. A., and Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the raven progressive matrices test. *Psychol. Rev*. 97:404. doi: 10.1037/0033-295x.97.3.404

Castora-Binkley, M., Noelker, L., Prohaska, T., and Satariano, W. (2010). Impact of arts participation on health outcomes for older adults. *J. Aging Human. Arts* 4, 352–367. doi: 10.1080/19325614.2010.533396

Chan, D., Shafto, M., Kievit, R., Matthews, F., Spink, M., Valenzuela, M., et al. (2018). Lifestyle activities in mid-life contribute to cognitive reserve in late-life, independent of education, occupation, and late-life activities. *Neurobiol. Aging* 70, 180-183

Colcombe, S., and Kramer, A. F. (2003). Fitness effects on the cognitive function of older adults. *Psychol. Sci*. 14, 125–130. doi: 10.1111/1467-9280.t01-1-01430

Creech, A., Hallam, S., Varvarigou, M., McQueen, H., and Gaunt, H. (2013). Active music making: a route to enhanced subjective well-being among older people. *Perspect. Public Health* 133, 36–43. doi: 10.1177/1757913912466950

Criscuolo, A., Bonetti, L., Särkämö, T., Kliuchko, M., and Brattico, E. (2019). On the association between musical training, intelligence and executive functions in adulthood. *Front. Psychol*. 10:1704. doi: 10.3389/fpsyg.2019.01704

Degé, F., Kubicek, C., and Schwarzer, G. (2011). Music lessons and intelligence: A relation mediated by executive functions. *Music Percept*. 29, 195–201. doi: 10.1525/mp.2011.29.2.195

Fancourt, D., and Steptoe, A. (2019). The art of life and death: 14 year follow-up analyses of associations between arts engagement and mortality in the english longitudinal study of ageing. *BMJ* 367:l6377. doi: 10.1136/bmj.l6377

Fauvel, B., Groussard, M., Eustache, F., Desgranges, B., and Platel, H. (2013). Neural implementation of musical expertise and cognitive transfers: could they be promising in the framework of normal cognitive aging? *Front. Human Neurosci. 7*:693. doi: 10.3389/fnhum.2013.00693

Fauvel, B., Groussard, M., Mutlu, J., Arenaza-Urquijo, E. M., Eustache, F., Desgranges, B., et al. (2014). Musical practice and cognitive aging: two cross-sectional studies point to phonemic fluency as a potential candidate for a use-dependent adaptation. *Front Aging Neurosci*. 6:227. doi: 10.3389/fnagi.2014.00227

Ferreira, N., Owen, A., Mohan, A., Corbett, A., and Ballard, C. (2015). Associations between cognitively stimulating leisure activities, cognitive function and age-related cognitive decline. *Int. J. Geriatr. Psychiatry* 30, 422–430. doi: 10.1002/gps.4155

Ferreri, L., Moussard, A., Bigand, E., Schellenberg, E. G., Ferreri, L., Moussard, A., et al. (2019). "Music and the Aging Brain," *The Oxford Handbook of Music and the Brain* A. Moussard, L. Cuddy, S. Belleville (London: Elsevier Science), 622–644. doi: 10.1093/oxfordhb/9780198804123.013.25

Godefroy, O., Jeannerod, M., Allain, P., and Le Gall, D. (2008). Frontal lobe, executive functions and cognitive control. *Rev. Neurol*. 164(Suppl. 3), 119–127. doi: 10.1016/S0035-3787(08)73302-2

Godefroy, O., Gibbons, L., Diouf, M., Nyenhuis, D., Roussel, M., Black, S., et al. (2014). Validation of an integrated method for determining cognitive ability: implications for routine assessments and clinical trials. *Cortex* 54, 51–62. doi: 10.1016/j.cortex.2014.01.016

Groussard, M., La Joie, R., Rauchs, G., Landeau, B., Chételat, G., Viader, F., et al. (2010). When music and long-term memory interact: effects of musical expertise on functional and structural plasticity in the Hippocampus. *PLoS One* 5:8. doi: 10.1371/journal.pone.0013225

Groussard, M., Viader, F., Landeau, B., Desgranges, B., Eustache, F., and Platel, H. (2014). The effects of musical practice on structural plasticity: the dynamics of grey matter changes. *Brain Cogn. 90*, 174–180. doi: 10.1016/j.bandc.2014.06.013

Hanna-Pladdy, B., and Gajewski, B. (2012). Recent and past musical activity predicts cognitive aging variability: direct comparison with general lifestyle activities. *Front. Human Neurosci*. 6:198. doi: 10.3389/fnhum.2012.00198

Hanna-Pladdy, B., and MacKay, A. (2011). The relation between instrumental musical activity and cognitive aging. *Neuropsychology* 25, 378–386. doi: 10.1037/a0021895

Hedden, T., and Gabrieli, J. D. E. (2004). Insights into the ageing mind: a view from cognitive neuroscience. *Nat. Rev. Neurosci.* 5, 87–96. doi: 10.1038/nrn1323

Hertzog, C., Kramer, A. F., Wilson, R. S., and Lindenberger, U. (2008). Enrichment effects on adult cognitive development. *Psychol. Sci. Public Int*. 9, 1–65. doi: 10.1111/j.1539-6053.2009.01034.x

Hinault, T., Lemaire, P., and Touron, D. (2017a). Aging effects in sequential modulations of poorer-strategy effects during execution of memory strategies. *Memory*, 25, 176–186. doi: 10.1080/09658211.2016.1146300

Hinault, T., Lemaire, P., and Touron, D. (2017b). Strategy combination during execution of memory strategies in young and older adults. *Memory*, 25, 619–625. doi: 10.1080/09658211.2016.1200626

Hinault, T., and Lemaire, P. (2020). Aging effects on brain and cognition: What do we learn from a strategy perspective? in *The Cambridge Handbook of Cognitive Aging: A Life Course Perspective* eds A. Gutchess, and A. Thomas (Cambridge: Cambridge University Press).

Jaeggi, S. M., Buschkuehl, M., Jonides, J., and Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proc. Natl. Acad. Sci*. 105, 6829–6833. doi: 10.1073/pnas.0801268105

Jung, M. S., Kim, H., Lee, Y., Kim, M., and Chung, E. (2017). Different effects of cognitive and non-exercise physical leisure activities on cognitive function by age in elderly Korean individuals. *Osong Public Health Res. Perspect* 8, 308–317. doi: 10.24171/j.phrp.2017.8.5.04

Kalpouzos, G., Eustache, F., and Desgranges, B. (2008). Réserve cognitive et fonctionnement cérébral au cours du vieillissement normal et de la maladie d'Alzheimer [Cognitive reserve and neural networks in normal aging and Alzheimer's disease]. *Psychol. Neuropsy. Vieillissement* 6, 97–105. doi: 10.1684/pnv.2008.0120

Karbach, J., and Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Dev. Sci*. 12, 978–990. doi: 10.1111/j.1467-7687.2009.00846.x

Koshimori, Y., and Thaut, M. H. (2019). New perspectives on music in rehabilitation of executive and attention functions. *Front. Neurosci*. 13:1245. doi: 10.3389/fnins.2019.01245

Mansens, D., Deeg, D. J. H., and Comijs, H. C. (2018). The association between singing and/or playing a musical instrument and cognitive functions in older adults. *Aging Ment. Health* 22, 964–971. doi: 10.1080/13607863.2017.1328481

Mella, N., Grob, E., Döll, S., Ghisletta, P., and de Ribaupierre, A. (2017). Leisure activities and change in cognitive stability: a multivariate approach. *Brain Sci*. 7, 8–11. doi: 10.3390/brainsci7030027

Moussard, A., Bermudez, P., Alain, C., Tays, W., and Moreno, S. (2016). Life-long music practice and executive control in older adults: an event-related potential study. *Brain Res. 1642*, 146–153. doi: 10.1016/j.brainres.2016.03.028

Nettle, D. (2006). Psychological profiles of professional actors. *Personal. Individ. Diff*. 40, 375–383. doi: 10.1016/j.paid.2005.07.008

Noice, H., and Noice, T. (1999). Long-term retention of theatrical roles. *Memory*, 7, 357–382. doi: 10.1080/096582199387977

Noice, H., Noice, T., and Staines, G. (2004). A short-term intervention to enhance cognitive and affective functioning in older adults. *J. Aging Health* 16, 562–585. doi: 10.1177/0898264304265819

Noice, T., Noice, H., and Kramer, A. F. (2014). Participatory arts for older adults: a review of benefits and challenges. *Gerontologist* 54, 741–753. doi: 10.1093/geront/gnt138

Parbery-Clark, A., Strait, D. L., Hittner, E., and Kraus, N. (2013). Musical training enhances neural processing of binaural sounds. *J. Neurosci*. 33, 16741-16747. doi: 10.1523/JNEUROSCI.5700-12.2013

Platel, H., and Groussard, M., (2020). "Benefits and limits of musical interventions in pathological aging," in *Music and the Aging Brain*, eds L. Cuddy, S. Belleville, and A. Moussard (Cambridge, MA: Academic Press).

Raz, N., and Kennedy, K. M. (2009). "A systems approach to the aging brain: neuroanatomic changes, their modifiers and cognitive correlates," in *Imaging the Aging Brain*, eds W. Jagust, and M. D'Esposito (New York, NY: Oxford University Press), 43–70. doi: 10.1093/acprof:oso/9780195328875.003.0004

Reuter-Lorenz, P. A., and Park, D. C. (2014). How does it STAC Up? revisiting the scaffolding theory of aging and cognition. *Neuropsychol. Rev*. 24, 355–370. doi: 10.1007/s11065-014-9270-9

Rey, A. (1959). Manuel de la figure complexe de Rey. Paris: editions du centre de psychologie appliquée .vel. *Acta Neurol.Belg*. 90, 207–217.

Reynolds, J., Bernard, M., Rezzano, J., and Rickett, M. (2016). Ageing, drama, and creativity: translating research into practice. *Geront. Geriatr. Educ*. 37, 307–328. doi: 10.1080/02701960.2016.1152267

Salimpoor, V. N., and Zatorre, R. J. (2013). Neural interactions that give rise to musical pleasure. *Psychol. Aesthet. Creat. Arts* 7, 62–75. doi: 10.1037/a0031819

Salthouse, T. A. (2010). Selective review of cognitive aging. *J. Int. Neuropsychol. Soc*. 16, 754–760. doi: 10.1017/S1355617710000706

Scarmeas, N., and Stern, Y. (2003). Cognitive reserve and lifestyle NIH public access. *J. Clin. Exp. Neuropsychol.* 25, 625–633.

Schellenberg, E. G., and Weiss, W. M. (2013). 12 - Music and Cognitive Abilities *The Psychology of Music*. Amsterdam: Elsevier, 499–550. doi: 10.1016/B978-0-12-381460-9.00012-2

Schlaug, G. (2015). Musicians and music making as a model for the study of brain plasticity. *Prog. Brain Res*. 217, 37–55. doi: 10.1016/bs.pbr.2014.11.020

Schneider, C. E., Hunter, E. G., and Bardach, S. H. (2019). Potential cognitive benefits from playing music among cognitively intact older adults: a scoping review. *J. Appl. Gerontol*. 38, 1763–1783. doi: 10.1177/0733464817751198

Seinfeld, S., Figueroa, H., Ortiz-Gil, J., and Sanchez-Vives, M. V. (2013). Effects of music learning and piano practice on cognitive function, mood and quality of life in older adults. *Front. Psychol*. 4:810. doi: 10.3389/fpsyg.2013.00810

Signoret, J. (1991). *Batterie d'Efficience Mnésique*. Paris:Elsevier.

Sprague, B. N., Freed, S. A., Webb, C. E., Phillips, C. B., Hyun, J., and Ross, L. A. (2019). The impact of behavioral interventions on cognitive function in healthy older adults: a systematic review. *Ageing Res. Rev*. 52, 32–52. doi: 10.1016/j.arr.2019.04.002

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia* 47, 2015–2028. doi: 10.1016/j.neuropsychologia.2009.03.004

Stern, Y., Arenaza-Urquijo, E. M., Bartrés-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., et al. (2018). Whitepaper: defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimer's Dementia* 14, S1552-S5260 doi: 10.1016/j.jalz.2018.07.219

Sutcliffe, R., Du, K., and Ruffman, T. (2020). Music making and neuropsychological aging: a review. *Neurosci. Biobehav. Rev*. 113, 479-491 doi: 10.1016/j.neubiorev.2020.03.026

Talamini, F., Altoè, G., Carretti, B., and Grassi, M. (2017). Musicians have better memory than nonmusicians: a meta-analysis. *PLoS One*, *12*:e0186773. doi: 10.1371/journal.pone.0186773

Verghese, J., LeValley, A., Derby, C., Kuslansky, G., Katz, M., Hall, C., et al. (2006). Leisure activities and the risk of amnestic mild cognitive impairment in the elderly. *Neurology* 66, 821–827. doi: 10.1212/01.wnl.0000202520.68987.48

Verghese, J., Lipton, R. B., Katz, M. J., Hall, C. B., Derby, C. A., Kuslansky, G., et al. (2003). Leisure activities and the risk of dementia in the elderly. *New Engl. J. Med*. 348, 2508–2516. doi: 10.1056/NEJMoa022252

Wan, C. Y., and Schlaug, G. (2010). Music making as a tool for promoting brain plasticity across the life span. *Neuroscientist* 16, 566–577. doi: 10.1177/1073858410377805

Warne, R. T. (2014). A primer on multivariate analysis of variance (MANOVA) for behavioral scientists. *Pract. Assess. Res. Evaluat. 19*, 1–10.

Wechsler, D. (2000). *Echelle d'Intelligence de Wechsler Pour Adultes, Troisieme édition*. Paris: Centre de Psychologie Appliquée.

Zendel, B. R., West, G. L., Belleville, S., and Peretz, I. (2019). Musical training improves the ability to understand speech-in-noise in older adults. *Neurob. Aging* 81, 102-115. doi: 10.1016/j.neurobiolaging.2019.05.015

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 18 September 2020
doi: 10.3389/fnhum.2020.578112

# Musical Tension Associated With Violations of Hierarchical Structure

*Lijun Sun[1,2], Li Hu[1,2], Guiqin Ren[3] and Yufang Yang[1,2]**

[1]Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, [2]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China, [3]College of Psychology, Liaoning Normal University, Dalian, China

Tension is one of the core principles of emotion evoked by music, linking objective musical events and subjective experience. The present study used continuous behavioral rating and electroencephalography (EEG) to investigate the dynamic process of tension generation and its underlying neurocognitive mechanisms; specifically, tension induced by structural violations at different music hierarchical levels. In the experiment, twenty-four musicians were required to rate felt tension continuously in real-time, while listening to music sequences with either well-formed structure, phrase violations, or period violations. The behavioral data showed that structural violations gave rise to increasing and accumulating tension experience as the music unfolded; tension was increased dramatically by structural violations. Correspondingly, structural violations elicited N5 at GFP peaks, and induced decreasing neural oscillations power in the alpha frequency band (8–13 Hz). Furthermore, compared to phrase violations, period violations elicited larger N5 and induced a longer-lasting decrease of power in the alpha band, suggesting a hierarchical manner of musical processing. These results demonstrate the important role of musical structure in the generation of the experience of tension, providing support to the dynamic view of musical emotion and the hierarchical manner of tension processing.

**Keywords: musical tension, hierarchical structure, global field power, N5, alpha**

OPEN ACCESS

***Edited by:***
*Franco Delogu,*
*Lawrence Technological University,*
*United States*

***Reviewed by:***
*José Lino Oliveira Bueno,*
*University of São Paulo, Brazil*
*Brian A. Coffman,*
*University of New Mexico,*
*United States*

***Correspondence:**
*Yufang Yang*
*yangyf@psych.ac.cn*

***Specialty section:***
*This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience*

***Received:*** *30 June 2020*
***Accepted:*** *27 August 2020*
***Published:*** *18 September 2020*

***Citation:***
*Sun L, Hu L, Ren G and Yang Y (2020) Musical Tension Associated With Violations of Hierarchical Structure. Front. Hum. Neurosci. 14:578112. doi: 10.3389/fnhum.2020.578112*

## INTRODUCTION

Music is universal and essential in human societies, owing to its strong power of emotion (e.g., Koelsch, 2014; Swaminathan and Schellenberg, 2015). However, it remains unclear how the auditory elements in music generate emotion responses (Juslin and Laukka, 2004; Juslin and Västfjäll, 2008; Juslin, 2013). In the literature of music theories and music psychology, tension is considered as the link between musical events and experienced emotions (Lehne et al., 2013; Lehne and Koelsch, 2015b) and so that as an appropriate point of penetration to investigate the dynamics and richness of musical emotion. Musical tension is an affective state that is associated with conflict, dissonance, instability, or uncertainty and create a yearning for resolution (Lehne and Koelsch, 2015b). It is believed that the recurrent alterations between tension and relaxation create a music-generated experience (Lerdahl and Jackendoff, 1983).

As to the relationship between musical structure and tension, presently most studies have primarily focused on the relationship between tonally hierarchical structure and tension, and found that tonal regularity plays a critical role in tension generation; moving away from tonal center would induce tension, whereas returning to the tonal center a sense of resolution (Bigand et al., 1996; Bigand and Parncutt, 1999; Steinbeis et al., 2006; Lerdahl and Krumhansl, 2007). For example, Bigand and his colleagues manipulated the stability of certain chords in music sequences and examined the tension rating of the key chords, suggesting that unstable chords gave rise to higher tension experience than stable chords (Bigand et al., 1996; Bigand and Parncutt, 1999). Steinbeis et al. (2006) constructed harmonic regular and irregular conditions by manipulating the last chord of short and real music pieces. The results showed that structural violations induced a strong sense of tension because the ending of the music did not fulfill the expectation. While these studies shed light on the relationship between tonal structure and the experience of tension, musical stimuli used in these studies were short of large-scale structures.

According to the generative theory of tonal music (GTTM) and tonal tension model (TTM), music expresses complicated meaning and emotion with hierarchical structures. Not only each musical event expresses specific tension and relaxation, but the events are also organized into a hierarchy with a prolongational reduction tree, creating a complex pattern of tension and resolution (Lerdahl and Jackendoff, 1983; Lerdahl and Krumhansl, 2007). In real music pieces, discrete elements are organized at multiple time scales to form structural units such as musical section, phrase, period and movement (Lerdahl and Jackendoff, 1983). Based on different hierarchical structural units, multiple local patterns of tension and relaxation constitute a global pattern of tension and relaxation. That is, small tension arches are always embedded in large tension arches based on the hierarchical structure with different timescales in music. Based on the structural relations in music, several tension arches overlap and interweave to develop the tension and resolution in large-scale structures (Koelsch, 2012).

Music psychology has explored the cognitive and neural bases of tension generation. The model of music tension proposed by Margulis (2005) suggests that listeners make predictions for the upcoming events continuously based on the musical context. The mismatch between the actual stimulus and the prediction in the brain will induce tension experience, such as tonal inconsistency and violations. Recently, a general model of tension and suspense has been proposed by Lehne and Koelsch (2015a) based on the predictive coding theory (Friston and Kiebel, 2009). They suggest that prediction errors enable the brain to update predictions for musical events in the process of music listening. Specifically, when an unpredicted event occurs in music progression, listeners' mental models are not maintained stable, but rather updated based on the current perceptual information. In other words, prediction violations in music could lead to predictive errors and event model resetting, which dynamically engaged working memory and cognitive control processes, such as attention switching (Zacks et al., 2007; Kurby and Zacks, 2008).

In terms of the proposals as to cognitive bases underlying tension generation, we assumed that musical tension could be modulated by the hierarchical levels of musical structure. That is, compared with low-level units, the processing of high-level units would be more difficult and required more cognitive efforts. The difficulties may result from: first, much more information included in musical events at higher levels; second, long-distance dependency involved in integrating multiple small units into a large unit at higher levels. The evidence has been provided by previous studies. It was found that larger closure positive shift (CPS) was elicited by period boundaries than section and phrase boundaries in music, suggesting more retrospective processing of preceding information (Zhang et al., 2016); it was more difficult for a global structure to be integrated into the context than the local structure (Zhang et al., 2018).

How tension experience is modulated by hierarchical structures with different time scales? Whether higher tension would be induced by structural violations at a higher level than that at a lower level? What are the cognitive neural bases underlying the music tension generation? Using continuous behavioral rating and electroencephalography (EEG), the present study explored these questions by investigating the processing of tension experience and neural responses induced by structural violations at different hierarchical levels. In the experiment, four-phrase music sequences were used to organize hierarchical units at different timescales. There were three types of music structures, in which the consistency of the tonality at different hierarchical levels was manipulated, while the musical structure within each phrase was kept the same across conditions. In well-formed sequences, four phrases were organized into a structure with three hierarchical levels and had consistent tonality from the beginning to the end. In sequences with phrase violations and period violations, the tonal consistency was disrupted at either phrase or period boundary respectively.

We hypothesized that tension ratings could be raised by structural violations, and higher tension would be induced by higher-level violations than the lower ones. Furthermore, for the cognitive mechanism underlying tension experience, larger N5 response, a neurophysiological marker of the integrated process of musical structure, would be elicited by sequences with structural violations compared to the well-formed sequences (Poulin-Charronnat et al., 2006; Loui et al., 2007; Miranda and Ullman, 2007; Koelsch and Jentschke, 2010; Sun et al., 2020b). Also, we predicted that the power in the alpha frequency band would decrease in response to structural violations compared with the well-formed sequences since the alpha power was tightly associated with the cognitive resources and attention (Meyer et al., 2013; Sadaghiani and Kleinschmidt, 2016). Finally, we hypothesized that both the N5 component and the power of alpha-band would be influenced by the hierarchical level of structural violations because more cognitive resources were required for working memory updating and attention switching in large structural units than the small ones.

# MATERIALS AND METHODS

## Participants

Twenty-four musicians, highly proficient in Western tonal music ($M_{age}$ = 23.21 years, $SD$ = 2.38, 17 female) participated in the experiment. They had received formal Western instrumental training for an average of 15 years (ranging from 8 to 22 years), playing piano, violin, viola, cello, or tuba. They were all right-handed with normal hearing. None of them reported a history of neural impairment or psychiatric illness. The study was approved by the Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences, and accords with the ethical principles of the Declaration of Helsinki. All participants were provided with informed consent.

## Stimuli

Ten original chorale sequences were composed in 2/4 meter. Four phrases were organized into a well-formed structure with multi-level hierarchies in each original sequence. As illustrated in **Figure 1A** (condition 1), the first two phrases and the last two phrases grouped into two music periods with a consistent tonality relationship. There were no pauses during the whole musical sequence. Musical units were mainly separated by the different durations of notes. The boundaries of phrase and period were signified by half notes, which lasted for 1,200 ms and were longer than other notes. Based on each original sequence, two modified versions (condition 2 and condition 3) were created (see **Figures 1B,C**). The modified versions had structural violations at phrase or period level, while the local structure within each phrase was consistent across conditions. In condition 2, the musical structure was violated at the phrase level by modifying the tonality consistency of each phrase. Thus, the expectation of the listeners cannot be fulfilled at the beginning of each phrase. In condition 3, the musical structure was violated at the period level by modifying the tonality consistency of the two periods. Under such a situation, the expectation of the listeners cannot be fulfilled at the beginning of the second period (that is also the beginning of the third phrase).

The 30 chorale sequences (10 sequences for each condition) were then transposed to three other keys to form a total of 120 sequences. In terms of the sequences, all stimuli were created using the Sibelius 7.5 software (Avid Tech. Incorporated) and were exported in mid format. Using Cubase 5.1 software, the mid files were set at a constant velocity of 100 and were archived with a Yamaha piano timbre at a tempo of 100 beats per minute in.wav format.

## Procedure

There were three versions of sequences: original version with no violations, a modified version with violations at the phrase level, another modified version with violations at the period level. Each version has 40 sequences, resulting in 120 chorale sequences in total. They were presented in a pseudorandom order under two constraints: a given version could not be repeated more than three times in succession, and consecutive sequences were not to originate from the same original chorale sequence.

To reveal the dynamic qualities of the tension experience, participants were required to continuously make real-time judgments of the felt tension using the computer mouse. Tension values were recorded using the Psychpy 1.0. software interface and data were collected at a sampling rate of 20 Hz. Before the beginning of each trial, they will see a sliding bar, raging from 0 to 100, located at the center of the screen. The task for participants was to move the slider up or down through the mouse to indicate the degree of tension they felt while listening to music. Consistent with previous studies (Steinbeis et al., 2006; Lehne et al., 2013), the meaning of tension was not defined by specific descriptions. The initial position of the slider was set at 25 to prevent the rating bar from reaching maximal values throughout the musical progression. The experiment was conducted in an acoustically and electrically shielded room. All stimuli were presented binaurally through Audio Technica CKR30iS headphones. The loudness was adjusted by participants to their comfort levels. Three practice trials were performed before the experimental session to familiarize participants with the stimuli and procedure.

## EEG Recording and Analysis

Using Brain Products (Munich, Germany), EEG data were recorded from 64 Ag/AgCI electrodes mounted at International 10–20 system scalp locations. The data were digitized at a rate of 500 Hz with an additional notched filter at 50 Hz. The FCz was used as an online reference electrode, and an electrode placed between Fz and FPz served as the ground electrode. Blinks and vertical eye movements were recorded with an electrode below the right eye. The impedance of all electrodes was maintained less than 5 kΩ and EEG data were amplified with AC amplifiers.

The raw EEG data were preprocessed with EEGLAB (Delorme and Makeig, 2004) in a MATLAB environment. The continuous data were referenced offline to the algebraic mean of the left and right mastoid electrodes. For the ERPs analysis, the data were filtered offline with a band-pass filter of 1–30 Hz. The high-pass filter of 1 Hz was applied to be consistent with previous literature and assure methodological comparison with other studies (e.g., Hu et al., 2014; Kuhn et al., 2015). Then, the data were segmented into epochs of 22.2 s ranging from 1,000 ms before the onset of the first chord, to 2 s after the offset of each chorale sequence. Next, trials were baseline corrected using the 1,000 ms pre-stimulus interval. The data of each participant were then corrected using an Independent Component Analysis (ICA) algorithm (Makeig et al., 1997; Delorme and Makeig, 2004) implemented in EEGLAB to delete ocular and muscle activity-generated artifacts. After the procedure of ICA components removal, we inspected epochs for each participant visually and rejected the contaminated epochs. Epochs in the same experimental condition were averaged for each participant, yielding three average waveforms in each participant.

Single-subject average waveforms were subsequently calculated to obtain global field power (GFP) values at each instant. At any instant, GFP is calculated as the root of the mean of the squared potential differences between each electrode and the mean of the instantaneous potential across electrodes


**FIGURE 1 |** Samples of the musical sequences used in the study. Condition 1: a sequence with a well-formed hierarchical structure **(A)**. Condition 2: a sequence with phrase violations **(B)**. Condition 3: a sequence with a period violation **(C)**.

(Lehmann and Skrandies, 1980). To determine whether the GFP values elicited by the three conditions differed as a function of musical hierarchical structure, we performed a point-by-point repeated measures analysis of variance (RM ANOVA) test. After the significance level (*p*-value) was corrected for using a false discovery rate (FDR) procedure (Durka et al., 2004), a time course of $p$ values was obtained. In searching for the time window corresponding to $p$ values below 0.05, we decided on the latency of the EEG responses elicited by the three conditions. Then, within the time window identified as being statistically significant, overall ANOVA tests were followed up with further paired comparisons. Finally, given that the strongest filed potentials and highest topographic signal-to-noise ratios were represented by local maxima of the GFP curve (Lehmann and Skrandies, 1980), we calculated scalp topography around 20 ms, on average, from the GFP peaks and compared differences among the three conditions.

For the time-frequency analysis, the preprocessing steps remained the same as for the ERP analysis, other than filtering with a 1–100 Hz band-pass filter. To obtain time-frequency distributions, Fast Fourier Transforms (FFTs) with a fixed 600 ms Hanning window was applied to each epoch for each participant using time steps of 4 ms and frequency steps of 1 Hz. The time-frequency distributions were corrected with a baseline of an interval from 800 to 200 ms pre-stimulus to include subtle stimulus-induced changes in ongoing oscillatory power. Based on previous findings (Knyazev et al., 2008; Parvaz et al., 2012), we selected alpha band frequencies (8–13 Hz) and nine central-frontal electrodes including Fz, F1, F2, FCz, Fc1, Fc2, C1, Cz, C2 electrodes as the regions of interest (ROIs). A point-by-point RM ANOVA test was performed and the significance level ($p$-value) was corrected using a FDR procedure (Durka et al., 2004). The time window of statistical significance was defined as that in which the $p$ values were below 0.05. Next, further paired comparisons were individually performed within each significant time window.

# RESULTS

## Behavioral Results

To minimize the differences in slider moving ranges across participants, data were normalized to a zero mean and unit standard deviation ($z$-score) for each participant. Tension ratings were averaged across participants under each condition at each time point. **Figure 2** shows average tension profiles under the three conditions, indicating dynamic changes in tension throughout the whole musical pieces. In all conditions, within each phrase, tension values always increased from the beginning to the middle, then declined until the end. Across all three conditions, tension values overlapped only in the first phrase and separated in the other three.

The tension values were averaged for each participant in the following five time-windows: 0.5–4.8, 5.3–9.6, 10.1–14.4, 14.9–19.2, and 19.7–21.2 s (corresponding to the time windows of phrases 1–4 and 1.5 s after the end of sequences). In order to reduce the influence of the response delay, data in the time windows of 0–500 ms after phrase onsets were not analyzed. Two-way RM ANOVA taking time window and condition as within-subject factors was conducted. The results showed that the interaction effect between condition and time window ($F_{(8,184)} = 30.07$; $p < 0.001$, partial $\eta^2 = 0.57$), and the main effects of condition ($F_{(2,46)} = 46.70$; $p < 0.001$, partial $\eta^2 = 0.67$) and time window ($F_{(4,92)} = 33.41$; $p < 0.001$, partial $\eta^2 = 0.59$) were significant. Then, we conducted ANOVA test taking condition as within-subject factor in each time window. The results revealed significant main effects for all time windows except for the first phrase (the first phrase: $p = 0.08$; the second phrase: $F_{(2,46)} = 32.22$; $p < 0.001$, partial $\eta^2 = 0.58$; the third phrase: $F_{(2,46)} = 42.76$; $p < 0.001$, partial $\eta^2 = 0.65$; the last phrase: $F_{(2,46)} = 45.62$; $p < 0.001$, partial $\eta^2 = 0.67$; post-ending: $F_{(2,46)} = 38.84$; $p < 0.001$, partial $\eta^2 = 0.63$).

Further paired comparisons among the three conditions showed that for the second phrase, more tension was induced by phrase violations than well-formed structures ($p < 0.001$) and period violations ($p < 0.001$); however, no significant difference of tension was found in conditions between well-formed structures and period violations ($p = 0.24$). For the third phrase, more tension was induced by phrase violations and period violations than well-formed structures ($p$s $< 0.001$), and more tension was induced by period violations than phrase violations ($p = 0.049$). For the last phrase and post-ending, more tension was induced by phrase violations than in the other two conditions ($p$s $< 0.001$), while more tension was induced by period violations than well-formed structures ($p$s $< 0.001$).

## ERP Results

GFP profiles and scalp topography at the instants corresponding to the GFP peaks under different conditions are shown in **Figure 3**. The segregation of GFP profiles under the three conditions is shown around 700 ms after the onsets of two to four phrases.

Point-by-point RM ANOVA test found that the difference across conditions was significant in the time windows of 5,544–5,590 ms (744–790 ms after the onset of the second phrase), 10,342–10,396 ms (742–796 ms after the onset of the third phrase), and 14,960–15,204 ms (560–804 ms after the onset of the last phrase). Given that frontal-distributed negativities were elicited at the instants corresponding to GFP peaks, further paired comparisons were performed using the average amplitudes at frontal electrodes within the time window of 40 ms around each GFP peak. We conducted a one-way ANOVA taking condition as a within-subject factor in each time window. The results revealed significant main effects for all time windows (5,532–5,572 ms: $F_{(2,46)} = 7.03$; $p = 0.002$, partial $\eta^2 = 0.23$; 10,350–10,390: $F_{(2,46)} = 13.43$; $p < 0.001$, partial $\eta^2 = 0.37$; 15,132–15,172: $F_{(2,46)} = 4.73$; $p = 0.014$, partial $\eta^2 = 0.17$). Results of further paired comparisons showed that, for the time window of 5,532–5,572 ms, larger negativities were elicited by phrase violations than in the other two conditions (well-formed structures, $p = 0.001$; period violations, $p = 0.024$), whereas there was no significant difference in conditions between period violations and well-formed structures ($p = 0.437$). For the time window of 10,350–10,390 ms, larger negativities were elicited by phrase and period violations than in well-formed structures (phrase violations, $p = 0.032$; period violations, $p < 0.001$), while larger negativities were elicited by period violations than phrase violations ($p = 0.047$). For the time window of 15,132–15,172 ms, larger negativities were elicited by phrase violations than in the two other

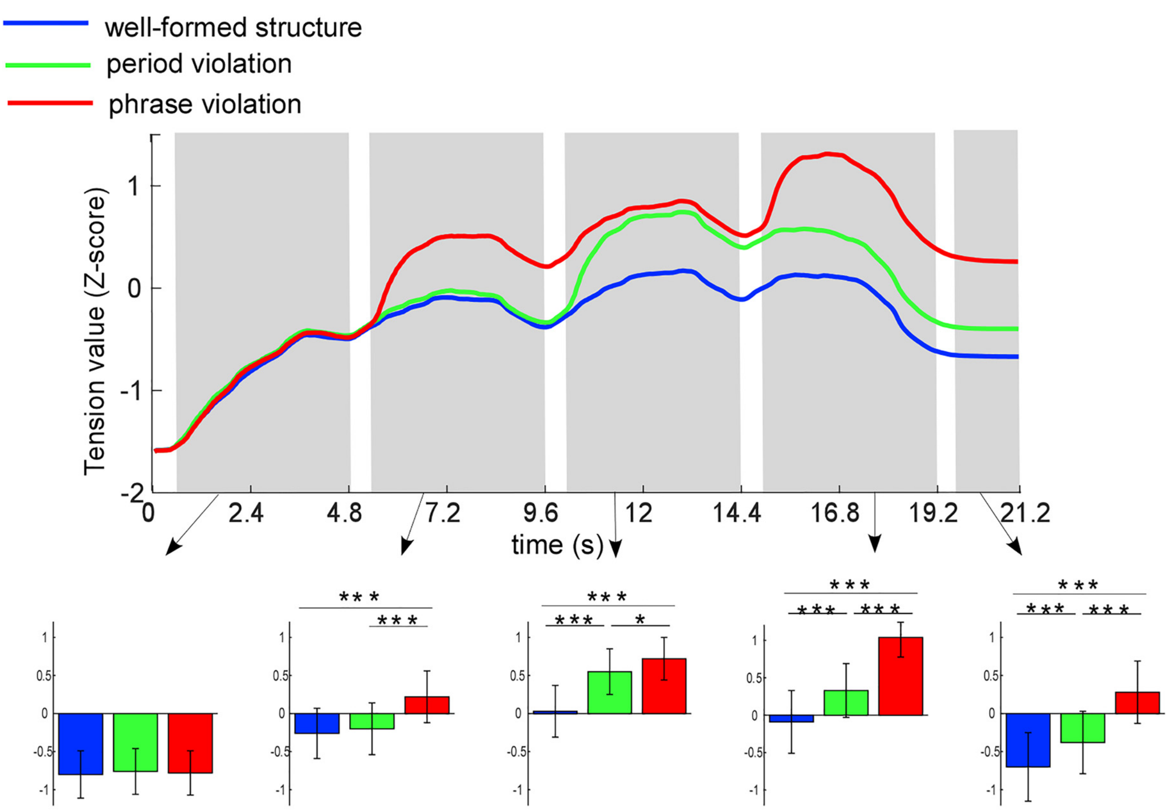



**FIGURE 2 |** Mean ratings (*z*-score) of tension values for the three conditions at each time point. Five time-windows are analyzed: 0.5–4.8, 5.3–9.6, 10.1–14.4, 14.9–19.2, and 19.7–21.2 s, which marked in gray bars. Colored bars represent the mean of the tension values for the three conditions (expressed as mean ± SD). *$p < 0.05$; ***$p < 0.001$.

conditions (well-formed structures, $p = 0.026$; period violations, $p = 0.036$), whereas there was no significant difference in conditions between well-formed structures and period violations ($p = 1.00$).

## Time-Frequency Results

**Figure 4** shows the time-frequency responses under different conditions. Significant time windows under different conditions for alpha frequency band (8–13 Hz) power are represented by gray-shaded areas. Scalp topographies at significant instants under different conditions reveal that the decrease in alpha power induced by phrase violations occurred around 800 ms after the onset of phrases 2–4, while this decrease lasted longer to period violations.

Results of a one-way RM ANOVA revealed significant main effects in the time windows of 5,536–5,808 ms (736–1,008 ms after the onset of the second phrase), 10,284–10,432 ms (684–832 ms after the onset of the third phrase), 11,100–11,332 ms (1,500–1,732 ms after the onset of the third phrase), and 15,300–15,532 ms (900–1,132 ms after the onset of the last phrase). For the time window of 5,536–5,808 ms, further paired comparisons indicated that the alpha band power induced by phrase violations was smaller than in the other two conditions (well-formed structures, $p = 0.048$; period violations, $p = 0.007$), whereas there was no significant difference in conditions between well-formed structures and period violations ($p = 0.29$). For the time window of 10,284–10,432 ms, smaller alpha band power was induced by phrase and period violations than well-formed structures (phrase violations, $p = 0.038$; period violations, $p = 0.034$), whereas there was no significant difference in conditions between phrase violations and period violations ($p = 0.93$). For the time window of 11,100–11,332 ms, smaller alpha band power was induced by period violations than well-formed structures ($p = 0.014$), whereas there was no significant difference between other comparisons ($ps > 0.06$). For the time window of 15,300–15,532 ms, smaller power was induced by phrase violations than in the other two conditions (well-formed structures, $p = 0.050$; period violations, $p = 0.002$), whereas there was no significant difference in conditions between well-formed structures and period violations ($p = 0.51$).

# DISCUSSION

The current experiment examined the effects of long-range hierarchical structure on tension experience and its underlying neurocognitive mechanisms. The behavioral data showed that the violations of hierarchical structure induced high tension experience, which was not completely resolved at structural

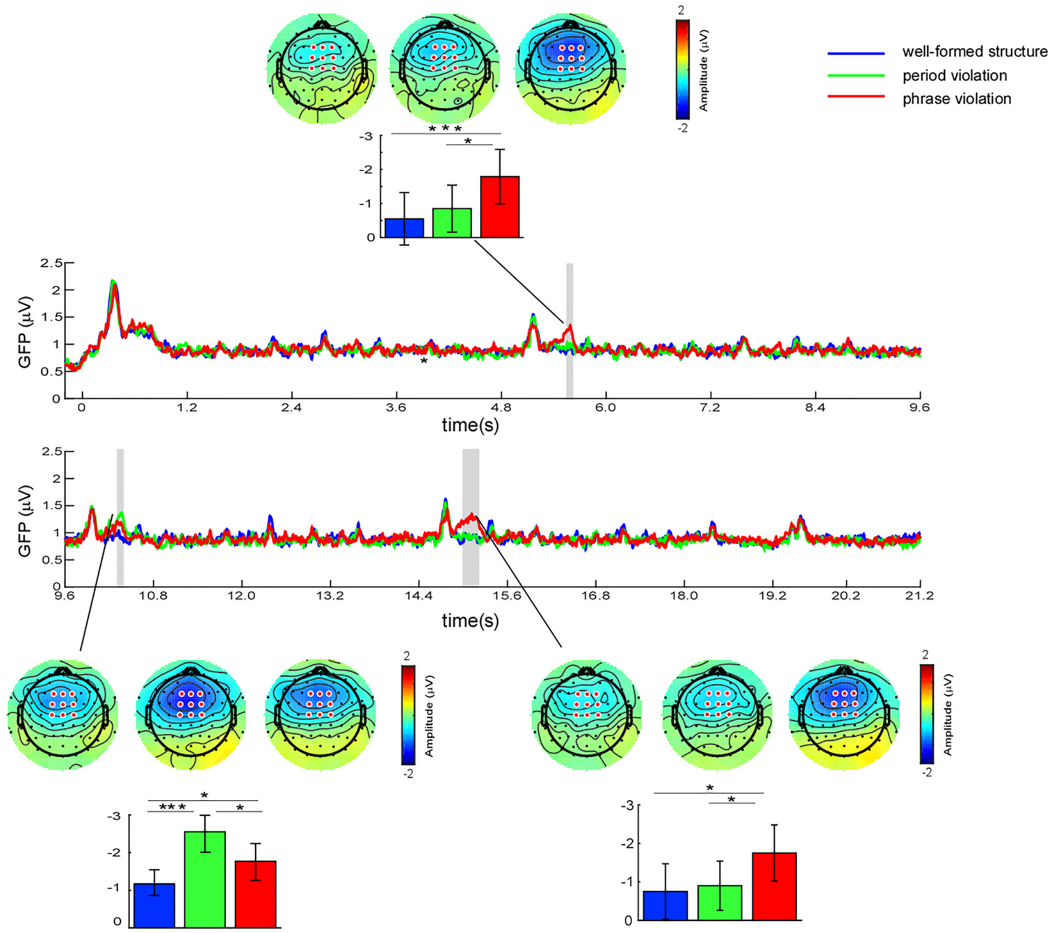



**FIGURE 3 |** Global field power (GFP) profiles for the three conditions over the whole musical sequences. Significant regions are marked in the gray bar [$p < 0.05$, one-way analysis of variance (ANOVA) test, and false discovery rate (FDR)-corrected]. The scalp topographies were obtained around the GFP peaks for different conditions within the time window of 5,532–5,572 ms (732–772 ms after the onset of the second phrase), 10,350–10,390 ms (750–790 ms after the onset of the third phrase), and 15,132–15,172 ms (732–772 ms after the onset of the last phrase). Colored bars represent the mean of the GFP values and the amplitude of the N5 component at central-frontal electrodes for the three conditions (expressed as mean ± SD). $^*p < 0.05$; $^{***}p < 0.001$.

boundaries but accumulated during subsequent musical passages unfolding. The EEG results showed larger frontal-distributed negativities (N5) at GFP peaks, and decreasing power in the alpha band (8–13 Hz) in response to the structural violations. Compared to phrase violations, period violations elicited larger N5 and induced a longer-lasting decrease of power in the alpha band, indicating a hierarchical manner of musical processing. These results are discussed below in more detail.

## Dynamic and Accumulating Tension Experience as Music Unfolding

The behavioral tension ratings demonstrated that the tension induced by music is rarely static, but changing constantly over time, in line with a dynamic view of music emotion (Scherer and Moors, 2019). Previous studies investigating music emotion were mainly focused on a limited set of discrete emotion categories (e.g., pleasant vs. unpleasant) to uncover the specific relationship between some musical features and induced emotion (Menon and Levitin, 2005; Koelsch et al., 2006, 2013; Omigie et al., 2014). The methodology used in these studies cannot show the dynamic processes of musical emotion, therefore is inadequate to characterize the richness of emotion as the music unfolds over time. Our study was concerned with the dynamic process of musical tension as music unfolding in attempting to explore the development of musical tension and resolution.

Three tension curves showed a common trend across experimental conditions, in that musical tension always decreases towards the end of each phrase. The results suggested that musical boundaries were correlated with the resolution, resulting from the dominant or tonic chords in each phrase ending. Based on the relationship between the stability of chords and the strength of the generated tension (Bigand et al., 1996; Steinbeis et al., 2006), the harmonic cadence represented a strong conclusiveness, inducing the sense of resolution (Meyer, 1956; Lerdahl and Jackendoff, 1983).

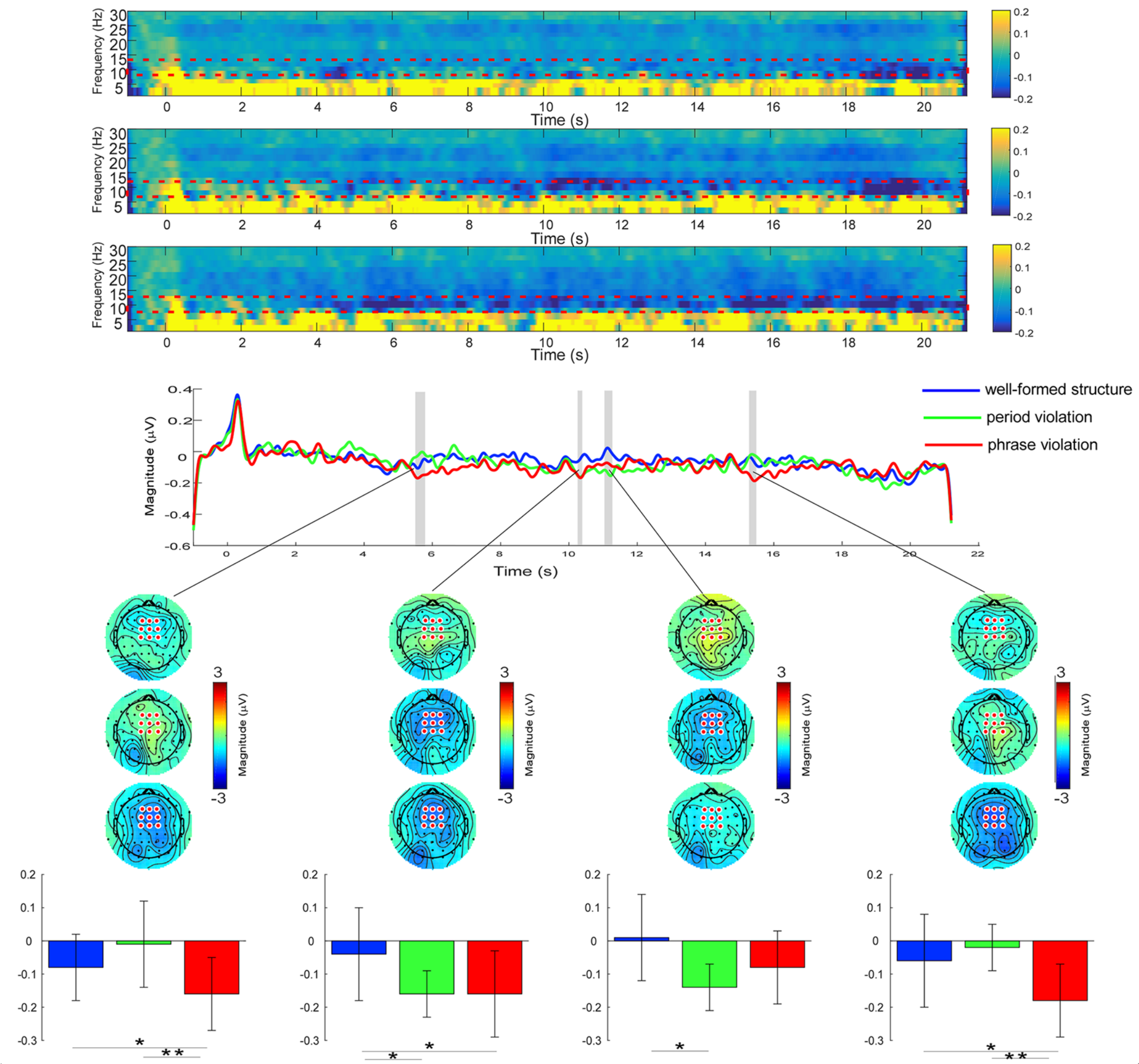



**FIGURE 4 |** *Top*: the time-frequency analysis of electroencephalographic series in the three conditions at the FCz electrode. Alpha band frequencies (8–13 Hz) were selected as the regions of interest (ROI). The color scale represents the average decrease or increase of oscillation power. *Middle*: the alpha power at frontal-central electrodes during the whole music sequences. Significant regions are marked in the gray bar ($p < 0.05$, one-way ANOVA test, and FDR-corrected). *Bottom*: the scalp topography of the alpha band power for the three conditions within the significant time-window of 5,536–5,808 ms (736–1,008 ms after the onset of the second phrase), 10,284–10,432 ms (684–832 ms after the onset of the third phrase), 11,100–11,332 ms (1,500–1,732 ms after the onset of the third phrase), and 15,300–15,532 ms (900–1,132 ms after the onset of the last phrase). Colored bars represent the mean of alpha power in the three conditions for each time window (expressed as mean ± SD). $^{*}p < 0.05$; $^{**}p < 0.05$.

A significant tension difference across conditions was found in the last three phrases. High tension experience was induced by structural violations, which can be explained by the Gestalt principles. Many aspects of music progression abide by Gestalt laws (Narmour, 1992). The sequences with phrase and period violations did not conform to the Gestalt principle of closure because of the tonality inconsistency. The violation of the Gestalt principle unfulfilled the prediction based on musical context, and in turn induced higher tension experience. In our study, the experimental design was careful and able to disentangle the influence of acoustic features and hierarchical structure on the experience of tension. In the first phrase, there was no significant difference in tension experience across conditions, although their acoustic features were different. In contrast, in the second phrase, the tension experience differed between sequences with well-formed structure and phrase violations, despite they had the same acoustic features.

The experience of tension was descendent but neither resolved completely nor immediately towards the end of each phrase but rather accumulated throughout subsequent musical pieces. In sequences with phrase violations, the tension induced by each incongruous tonality accumulated and resulted in higher starting points for the subsequent phrases. A similar situation was found in sequences with period violations. Our previous study investigating the tension experience induced by the nested structures also found the characteristics of tension experience accumulation (Sun et al., 2020a). The results could be explained by the event segmentation theory, which proposes a predictive error-based updating mechanism during the process of event segmentation (Zacks et al., 2007; Kurby and Zacks, 2008). When an unpredicted change occurs, event models are not maintained stable, but rather are updated based on current perceptual information. In our study, structural violations lead to predictive errors and event model resetting, which dynamically engaged working memory and cognitive control processes such as attention switching. This well explains why listeners' felt tension increased at phrase boundaries.

It is worth noting that different patterns of tension and resolution induced by hierarchical structures in the present study and nested structures in the previous study (Sun et al., 2020a). The tension curves showed that more tension arches were induced in hierarchical structures than nested structures, which might due to the stronger dependence of tonality and more close connection in nested structures. However, it also might because only one musical phrase was included in sequences used by our previous study (Sun et al., 2020a), whereas four phrases in the present study, implying the important role of the organization of musical units on tension experience. Future studies should further investigate the difference of tension experience induced by nested structures and hierarchical structures to discriminate tension experience delicately and shed new light on structural processing.

## The Cognitive Neural Mechanism Underlying the Tension Experience

Given our experiment was the first study to investigate the dynamic EEG underlying musical tension experience, we adopted a data-driven method to determine the significant time windows. The topographical approach of GFP does not require any *a priori* hypothesis, such as the prerequisite of electrodes and time intervals, and therefore is amenable to determining significant time points statistically and reliably. GFP in EEG signal reflects stronger global brain activity, its value represents the strength of the electric potential over all EEG electrodes on the scalp, and has been used in previous studies to measure the global brain activity (Hu et al., 2014; Khanna et al., 2015). In our study, to assess the dynamic process evoked by the whole musical pieces, we analyzed epochs as long as 21.2 s. It would be inadequate to use traditional ERP analytical methods with low high pass filters because it was hard to exclude slow-wave drifts due to the long time-span of the musical pieces. GFP is superior in eliminating the influence of a poor signal-to-noise ratio (Milz et al., 2017), therefore suitable for characterizing rapid changes in brain activity.

It was found that frontal-distributed negativities at GFP peaks always occurred around 600–800 ms after the structural violations, the latency and scalp distributions were similar to the N5 component. Previous studies have suggested that the N5 is related to the violation of expectation and the integration of incongruent harmonic chord into musical context (Koelsch, 2005; Miranda and Ullman, 2007; Steinbeis and Koelsch, 2008; Koelsch and Jentschke, 2010). In the time-frequency domain, structural violation induced a decrease in central-frontal alpha power. The decreased power in the alpha band may reflect orienting responses (Klimesch, 1999; Krause, 2006) and attention switching (Meyer et al., 2013; Sadaghiani and Kleinschmidt, 2016). In our experiment, listeners constructed a psychological model to predict the upcoming musical events based on musical contexts. Musical violations led to prediction errors, which required the brain to change the comparatively stable model to reduce prediction errors for the upcoming events. Event model resetting engaged more working memory and attention, the neural bases of which were reflected in the decreased power in the alpha band. Furthermore, more cognitive resources were devoted to the integration of harmonic structures, as indicated by the larger N5 component.

It is worth noting that the N5 effects elicited by period violations were larger than phrase violations because it is more difficult for listeners to integrate the period violations into the musical context (Zhang et al., 2018). Furthermore, longer-lasting decreases in alpha power were induced by period violations (around 800 ms to 1,500 ms) compared with the phrase violations (around 800 ms), suggesting more attentional resources were paid for the high-level violations. The results suggested a hierarchical manner of music processing. In line with previous studies in both music and language domains (Ding et al., 2016; Zhang et al., 2016; Harding et al., 2019), our study indicated that structural processing at different hierarchical levels might engage different neural mechanisms. Furthermore, the results supported the predictive coding theory (Friston, 2005, 2010), pointing to the hierarchical predictive coding process inherent to the auditory system, and active construction while listening to music (Koelsch, 2018).

## The Relationship Between Tension Ratings and EEG Brain Responses

Compared with well-formed sequences, larger N5 and decreased power in the alpha band were shown when listening to the music sequences with structural violations at phrase and period levels. The effects of these distinct brain responses were in parallel with the changes of behavioral tension ratings, in that the close corresponding relationship is indicated by the similar time windows. The modulation of musical structure on tension ratings and brain responses can be explained by the predictive processes. The mismatch between the actual stimulus and the prediction induced tension experience (Meyer, 1956; Margulis, 2005; Huron, 2006), which required listeners to pay more cognitive resources to update working memory, switch attention (Zacks et al., 2007; Kurby and Zacks, 2008) and integrate the violated events into a musical context, as implied by large N5 and decrease power in the alpha band.

Although tension ratings and brain responses induced by music structural violations are closely related, the processing processes they reflect are different. In particular, tension ratings reflected the online tension experience, while the brain responds to the integrated processing of the musical event into the musical context. Thus, the tension rating induced by each structural violation was based on the ending point in the previous phrase. That is, the tension value in one moment represented the tension induced by the present musical event and the accumulated tension from the previous musical context. In contrast, the N5 and the power in the alpha band did not exhibit the accumulation characteristics, but fading away within one phrase. The strength of the brain responses only reflected the cognitive effort to integrate the present musical event into the musical context.

In conclusion, the present study using behavioral rating and EEG investigated the influence of music hierarchical structure on the experience of tension and its underlying neural mechanism. Behavioral results showed that structural violations gave rise to an increasing and accumulating tension experience as indicated by tension curves. In parallel, brain responses including larger N5 effects at GFP peaks and decreased power in alpha frequency band were in response to the structural violations, reflecting more attention and cognitive resources engaged in the integrated processing. Moreover, compared with phrase violations, period violations elicited larger N5 and induced a longer-lasting decrease in alpha power, revealing the hierarchical manner of structure processing in music. The dynamic tension experience revealed by our findings is of significance to emotional regulation. Furthermore, the hierarchical manner of structure processing also reminded musicians to pay more attention to large-scale music structure while listening to music.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

LS and YY proposed the study and designed the experiment. LS and GR conducted the tests. LS and LH contributed to data analysis. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31470985 and 31971034) and the China Postdoctoral Science Foundation (Y9BH032001).

## REFERENCES

Bigand, E., and Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychol. Res.* 62, 237–254. doi: 10.1007/s004260050053

Bigand, E., Parncutt, R., and Lerdahl, F. (1996). Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion and musical training. *Percept. Psychophys.* 58, 124–141.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164. doi: 10.1038/nn.4186

Durka, P. J., Zygierewicz, J., Klekowicz, H., and Ginter, J. (2004). On the statistical significance of event-related EEG desynchronization and synchronization in the time-frequency plane. *IEEE Trans. Biomed. Eng.* 51, 1167–1175. doi: 10.1109/TBME.2004.827341

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11:127. doi: 10.1038/nrn2787

Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1211–1221. doi: 10.1098/rstb.2008.0300

Harding, E., Sammler, D., Henry, M. J., Large, E. W., and Kotz, S. A. (2019). Cortical tracking of rhythm in music and speech. *NeuroImage* 185, 96–101. doi: 10.1016/j.neuroimage.2018.10.037

Hu, L., Valentini, E., Zhang, Z. G., Liang, M., and Iannetti, G. D. (2014). The primary somatosensory cortex contributes to the latest part of the cortical response elicited by nociceptive somatosensory stimuli in humans. *NeuroImage* 84, 383–393. doi: 10.1016/j.neuroimage.2013.08.057

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation.* Cambridge, MA: MIT Press.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Phys. Life Rev.* 10, 235–266. doi: 10.1016/j.plrev.2013.05.008

Juslin, P. N., and Laukka, P. (2004). Expression, perception and induction of musical emotions: a review and a questionnaire study of everyday listening. *J. New Music Res.* 33, 217–238. doi: 10.1080/0929821042000317813

Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 559–575. doi: 10.1017/S0140525X08005293

Khanna, A., Pascual-Leone, A., Michel, C. M., and Farzan, F. (2015). Microstates in resting-state EEG: current status and future directions. *Neurosci. Biobehav. Rev.* 49, 105–113. doi: 10.1016/j.neubiorev.2014.12.010

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Brain Res. Rev.* 29, 169–195. doi: 10.1016/s0165-0173(98)00056-3

Knyazev, G. G., Bocharov, A. V., Levin, E. A., Savostyanov, A. N., and Slobodskoj-Plusnin, J. Y. (2008). Anxiety and oscillatory responses to emotional facial expressions. *Brain Res.* 1227, 174–188. doi: 10.1016/j.brainres.2008.06.108

Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Curr. Opin. Neurobiol.* 15, 207–212. doi: 10.1016/j.conb.2005.03.005

Koelsch, S. (2012). *Brain and Music.* Oxford, UK: John Wiley and Sons.

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666

Koelsch, S. (2018). Investigating the neural encoding of emotion with music. *Neuron* 98, 1075–1079. doi: 10.1016/j.neuron.2018.04.029

Koelsch, S., and Jentschke, S. (2010). Differences in electric brain responses to melodies and chords. *J. Cogn. Neurosci.* 22, 2251–2262. doi: 10.1162/jocn.2009.21338

Koelsch, S., Fritz, T., Cramon, D. Y., Müller, K., and Friederici, A. D. (2006). Investigating emotion with music: an fMRI study. *Hum. Brain Mapp.* 27, 239–250. doi: 10.1002/hbm.20180

Koelsch, S., Skouras, S., Fritz, T., Herrera, P., Bonhage, C., Küssner, M. B., et al. (2013). The roles of superficial amygdala and auditory cortex in music-evoked fear and joy. *NeuroImage* 81, 49–60. doi: 10.1016/j.neuroimage.2013.05.008

Krause, C. M. (2006). Cognition- and memory-telated ERD/ERS responses in the auditory stimulus modality. *Prog. Brain Res.* 159, 197–207. doi: 10.1016/s0079-6123(06)59013-2

Kuhn, A., Brodbeck, V., Tagliazucchi, E., Morzelewski, A., von-Wegner, F., and Laufs, H. (2015). Narcoleptic patients show fragmented EEG-microstructure during early nrem sleep. *Brain Topogr.* 28, 619–635. doi: 10.1007/s10548-014-0387-1

Kurby, C. A., and Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends Cogn. Sci.* 12, 72–79. doi: 10.1016/j.tics.2007.11.004

Lehmann, D., and Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr. Clin. Neurophysiol.* 48, 609–621. doi: 10.1016/0013-4694(80)90419-8

Lehne, M., and Koelsch, S. (2015a). "Tension-resolution patterns as a key element of aesthetic experience: psychological principles and underlying brain mechanisms," in *Art, Aesthetics and the Brain*, eds J. P. Huston, M. Nadal, F. Mora, L. F. Agnati, and C. J. C. Conde (Oxford: Oxford University Press).

Lehne, M., and Koelsch, S. (2015b). Toward a general psychological model of tension and suspense. *Front. Psychol.* 6:79. doi: 10.3389/fpsyg.2015.00079

Lehne, M., Rohrmeier, M., Gollmann, D., and Koelsch, S. (2013). The influence of different structural features on felt musical tension in two piano pieces by mozart and mendelssohn. *Music Percept.* 31, 171–185. doi: 10.1525/mp.2013.31.2.171

Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music.* Cambridge, MA: MIT Press.

Lerdahl, F., and Krumhansl, C. L. (2007). Modeling tonal tension. *Music Percept.* 24, 329–366. doi: 10.1525/mp.2007.24.4.329

Loui, P., Grent, T., Torpey, D., and Woldorff, M. (2007). Effects of attention on the neural processing of harmonic syntax in western music. *Cogn. Brain Res.* 25, 678–687. doi: 10.1016/j.cogbrainres.2005.08.019

Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., and Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci. U S A* 94, 10979–10984. doi: 10.1073/pnas.94.20.10979

Margulis, E. H. (2005). A model of melodic expectation. *Music Percept.* 22, 663–714. doi: 10.1525/mp.2005.22.4.663

Menon, V., and Levitin, D. J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *NeuroImage* 28, 175–184. doi: 10.1016/j.neuroimage.2005.05.053

Meyer, L. B. (1956). *Emotion and Meaning in Music.* Chicago, IL: University of Chicago Press.

Meyer, L., Obleser, J., and Friederici, A. D. (2013). Left parietal a enhancement during working memory-intensive sentence processing. *Cortex* 49, 711–721. doi: 10.1016/j.cortex.2012.03.006

Milz, P., Pascualmarqui, R. D., Achermann, P., Kochi, K., and Faber, P. L. (2017). The EEG microstate topography is predominantly determined by intracortical sources in the alpha band. *NeuroImage* 162, 353–361. doi: 10.1016/j.neuroimage.2017.08.058

Miranda, R. A., and Ullman, M. T. (2007). Double dissociation between rules and memory in music: an event-related potential study. *NeuroImage* 38, 331–345. doi: 10.1016/j.neuroimage.2007.07.034

Narmour, E. (1992). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model.* Chicago, IL: University of Chicago Press.

Omigie, D., Dellacherie, D., Hasboun, D., George, N., Clement, S., Adam, C., et al. (2014). An intracranial EEG study of the neural dynamics of musical valence processing. *Cereb. Cortex* 25, 4038–4047. doi: 10.1093/cercor/bhu118

Parvaz, M. A., Macnamara, A., Goldstein, R. Z., and Hajcak, G. (2012). Event-related induced frontal alpha as a marker of lateral prefrontal cortex activation during cognitive reappraisal. *Cogn. Affect. Behav. Neurosci.* 12, 730–740. doi: 10.3758/s13415-012-0107-9

Poulin-Charronnat, B., Bigand, E., and Koelsch, S. (2006). Processing of musical syntax tonic versus subdominant: an event-related potential study. *J. Cogn. Neurosci.* 18, 1545–1554. doi: 10.1162/jocn.2006.18.9.1545

Sadaghiani, S., and Kleinschmidt, A. (2016). Brain networks and α-oscillations: structural and functional foundations of cognitive control. *Trends Cogn. Sci.* 20, 805–817. doi: 10.1016/j.tics.2016.09.004

Scherer, K. R., and Moors, A. (2019). The emotion process: event appraisal and component differentiation. *Annu. Rev. Psychol.* 70, 719–745. doi: 10.1146/annurev-psych-122216-011854

Steinbeis, N., and Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cereb. Cortex* 18, 1169–1178. doi: 10.1093/cercor/bhm149

Steinbeis, N., Koelsch, S., and Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological and neural responses. *J. Cogn. Neurosci.* 18, 1380–1393. doi: 10.1162/jocn.2006.18.8.1380

Sun, L., Feng, C., and Yang, Y. (2020a). Tension experience induced by nested structures in music. *Front. Hum. Neurosci.* 14:210. doi: 10.3389/fnhum.2020.00210

Sun, L., Thompson, W. F., Liu, F., Zhou, L., and Jiang, C. (2020b). The human brain processes hierarchical strctures of meter and harmony differently: evidence from musicians and nonmusicians. *Psychophysiology* 57:e13598. doi: 10.1111/psyp.13598

Swaminathan, S., and Schellenberg, E. G. (2015). Current emotion research in music psychology. *Emot. Rev.* 7, 189–197. doi: 10.1177/1754073914558282

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., and Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychol. Bull.* 133, 273–293. doi: 10.1037/0033-2909.133.2.273

Zhang, J., Jiang, C., Zhou, L., and Yang, Y. (2016). Perception of hierarchical boundaries in music and its modulation by expertise. *Neuropsychologia* 91, 490–498. doi: 10.1016/j.neuropsychologia.2016.09.013

Zhang, J., Zhou, X., Chang, R., and Yang, Y. (2018). Effects of global and local contexts on chord processing: an ERP study. *Neuropsychologia* 109, 149–154. doi: 10.1016/j.neuropsychologia.2017.12.016

**Conflict of Interest**: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 07 October 2020
doi: 10.3389/fpsyg.2020.02242

# How Soundtracks Shape What We See: Analyzing the Influence of Music on Visual Scenes Through Self-Assessment, Eye Tracking, and Pupillometry

*Alessandro Ansani*[1,2]*, *Marco Marini*[2,3], *Francesca D'Errico*[4] *and Isabella Poggi*[1]

[1] Cosmic Lab, Department of Philosophy, Communication, and Performing Arts, Roma Tre University, Rome, Italy, [2] Department of Psychology, Sapienza University of Rome, Rome, Italy, [3] Institute of Cognitive Sciences and Technologies, Rome, Italy, [4] Department of Education Science, Psychology, Communication Science, University of Bari Aldo Moro, Bari, Italy

OPEN ACCESS

**Edited by:**
Franco Delogu,
Lawrence Technological University,
United States

**Reviewed by:**
Bruno Laeng,
University of Oslo, Norway
Antoine Coutrot,
UMR 6004 Laboratoire des Sciences du Numérique à Nantes (LS2N),
France

***Correspondence:**
Alessandro Ansani
alessandro.ansani@uniroma1.it

**Specialty section:**
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology

**Received:** 28 April 2020
**Accepted:** 10 August 2020
**Published:** 07 October 2020

**Citation:**
Ansani A, Marini M, D'Errico F and Poggi I (2020) How Soundtracks Shape What We See: Analyzing the Influence of Music on Visual Scenes Through Self-Assessment, Eye Tracking, and Pupillometry. Front. Psychol. 11:2242. doi: 10.3389/fpsyg.2020.02242

This article presents two studies that deepen the theme of how soundtracks shape our interpretation of audiovisuals. Embracing a multivariate perspective, Study 1 ($N = 118$) demonstrated, through an online between-subjects experiment, that two different music scores (melancholic vs. anxious) deeply affected the interpretations of an unknown movie scene in terms of empathy felt toward the main character, impressions of his personality, plot anticipations, and perception of the environment of the scene. With the melancholic music, participants felt empathy toward the character, viewing him as more agreeable and introverted, more oriented to memories than to decisions, while perceiving the environment as cozier. An almost opposite pattern emerged with the anxious music. In Study 2 ($N = 92$), we replicated the experiment in our lab but with the addition of eye-tracking and pupillometric measurements. Results of Study 1 were largely replicated; moreover, we proved that the anxious score, by increasing the participants' vigilance and state of alert (wider pupil dilation), favored greater attention to minor details, as in the case of another character who was very hard to be noticed (more time spent on his figure). Results highlight the pervasive nature of the influence of music within the process of interpretation of visual scenes.

**Keywords: soundtrack, film music, audiovisual, interpretation, empathy, eye tracking, pupillometry, environment perception**

## INTRODUCTION

The influence of music on human behavior has been studied since the dawn of time. Although a vast amount of studies analyzed the influence on several kinds of performances, among which physical tasks (Edworthy and Waring, 2006), work performance (Lesiuk, 2005), text and verbal memory (Taylor and Dewhurst, 2017), and learning (Lehmann and Seufert, 2017), the vast majority of the studies, starting from the 1980s, focused on marketing, shopping, and advertising (Bruner, 1990). Nowadays, this tradition continues, although several modifications have been made within the experimental paradigms to involve new contemporary scenarios such as online shopping, website atmospherics, and driving game performance (Brodsky, 2001).

Another flourishing tradition has started to bloom in the last two decades on the psychology of music in gambling environments: numerous researchers have deepened the influence of music on gambling, virtual roulette, ultimatum game, casino environment, and lottery (Dixon et al., 2007).

Concerning aspects in the domain of affect, a lot have been written about the processes through which music is useful for personal enhancement (Brown and Theorell, 2006), being able to express and induce moods and emotions (Västfjäll, 2001). An increasing number of studies measure plausible behavioral changes in dependence on the listening of music pieces evoking or inducing different emotions. Several social- and moral-related domains have been explored: facial emotion recognition (Woloszyn and Ewert, 2012); awareness, acceptance, and recall of unethical messages (Ziv et al., 2012); moral judgment and prosocial behavioral intentions (Ansani et al., 2019; Steffens, 2020); and compliance with requests to harm a third person (Ziv, 2015).

## INDUCTION VS. BACKGROUND MUSIC

Among studies on music influence, we can distinguish two categories, depending on their exploiting the musical stimulus either before the task or during the task. We call the former method *induction* and the latter *background music*. This premise is paramount since the underlying mental processes that preside over these two experimental situations might be substantially different: according to a previous study (Pisciottano, 2019), in the case of induction music, the participant feels the emotion him/herself, while in the background case, s/he attributes the emotion to the scene character.

Our study focuses on soundtracks, that is musical stimuli administered as background. Background music implies a parallel and multimodal processing, which can vary among music/unrelated, music/related, and music/visual tasks. In general, depending on the task, music may have either positive, integrative, or detrimental effects.

### Detrimental Effects: Music as a Source of Distraction

Hearing music while performing an experimental task may be distracting, being a secondary source of information. Indeed, according to Kahneman's (1973) Cognitive-Capacity Model, "there is a general limit on man's capacity to perform mental work. [...] this limited capacity can be allocated with considerable freedom among concurrent activities. [...] the ability to perform several mental activities concurrently depends, at least in part, on the effort which each of these activities demands when performed in isolation. The driver who interrupts a conversation to make a turn is an example" (Kahneman, 1973, p. 9).

Not necessarily does an effortful workload exclude a mood effect—several researchers who use music during the task (e.g., Au et al., 2003) provide accounts in terms of mood, but the parallel elaboration often implies the emergence of other phenomena. In their meta-analysis, Kämpfe et al. (2011) conclude that background music has detrimental effects on several memory-related tasks and produces decreases in reading performance, being a source of distraction during cognitive tasks *per se* (Furnham and Strbac, 2002; Kallinen, 2002; Salvucci et al., 2007) or depending on its tempo (Mentzoni et al., 2014; Nguyen and Grahn, 2017; Israel et al., 2019) or its volume (Noseworthy and Finlay, 2009).

### Integrative Effects

When it comes to music in audiovisuals, things become more complicated. In this case too music can be a distractor, leading, for instance, to a reduced recall of the ads' messages (Fraser and Bradford, 2013). Nevertheless, the effects of music are overall integrative: on the one hand, the human mind expects music to exhibit some sort of synchrony (Rogers and Gibson, 2012) and, most of all, congruity to what is stated and depicted (i.e., visual information) by the main message, whether movies or advertising (Bolivar et al., 1994; Boltz, 2004; North, 2004; Oakes, 2007; Herget et al., 2018), as stated by the Congruence-Association Model by Cohen (2013). On the other hand, through the mood communicated, music can convey semantic and content-related information by activating specific schemas: cognitive structures developed through experience that represent "knowledge about concepts or types of stimuli, their attributes, and the relations among those attributes" (Shevy, 2007, p. 59). Such schemas, in turn, influence the building of a mood-coherent audiovisual narrative.

## SOUNDTRACK AND INTERPRETATION

As already implied by Hoeckner et al. (2011), background music provides an interpretive framework for the audiovisuals (for a more detailed analysis of several cognitive frameworks of soundtracks, see Branigan, 2010); moreover, it can be seen as a second source of emotion besides the film itself (Cohen, 2001): it shapes audience's understanding not only of a character's actions, emotions, and intentions (Marshall and Cohen, 1988; Tan et al., 2007), by framing "visual meanings" (Nelson and Boynton, 1997), but also of characters' moral judgments (Steffens, 2020), general evaluations (Shevy, 2007), and plot anticipations (Bullerjahn and Güldenring, 1994; Vitouch, 2001; Shevy, 2007) or by generating expectations (Killmeier and Christiansen, 2004). This is well-known to any director and soundtrack composer: "the music in a film may be original or not, but what matters most, from a textual and communicative point of view, is the relationship established between the music, and the script, and the photography, and how they all add up and combine with each other, so that viewers can interpret them in a certain way" (Zabalbeascoa, 2008, p. 24). Several scholars claim in fact the existence of proper semiotics of music for film and TV (Tagg, 2013).

Tagg (2006) let their subjects listen to 10 title themes for film or television, asking them "to write down what they thought might be happening on the screen along with each tune. The results were collected and reduced to single concepts;" the authors called these visual–verbal associations (VVAs). Surprisingly enough, they found some of the themes to be strongly connected with male figures and some other with female figures; moreover, masculine themes were associated with concepts like Western, fast, detective, robbery, concrete, business, traffic, shooting, and planning, whereas feminine themes were associated with love, sad, parting, destiny, tragic, death, sentimental, sitting, France, and harmonious. Along the same line, Huron (1989) claims music to be a "very effective non-verbal identifier" and thus useful for targeting certain demographic and social groups as well as determining a character's ethos. Despite such encouraging preliminary results, only a few studies focused on the different interpretations of audiovisuals that music may foster by experimentally manipulating it.

Iwamiya (1997) studied the effect of listening to different music on the impression obtained from landscapes viewed from a car, showing that they were more pleasant when music was played as opposed to silence, and the ratings of pleasantness were highest when relaxing music was on.

Boltz (2001) analyzed the interpretations of three ambiguous clips in positive music, negative music, and no music conditions. A negative rating was connected with extreme violence and death, while the highest rating was given when the interpretation was about very happy outcomes. Coherently with her hypothesis, compared to the no music condition, the interpretations of all three clips were positive in the presence of positive music and negative in the other case. Furthermore, assumptions about the main character's personality were measured: in the positive condition, he was considered as caring, loving, and playful, while in the negative condition, the most significant adjectives were deranged, manipulative, and mysterious.

Ziv and Goshen (2006) obtained the same results with 5- to 6-year-old children. Using the first 21 bars of the melody of Chopin's Mazurka op. 68 n. 2 in A minor as sad music and a modified version of the same piece (transposed in C major and played faster) as happy music, the authors showed that children's interpretations were significantly affected by the background music: sadder in the first case and happier in the second.

Using a more ecological covert design (i.e., participants were presented with an original vs. fake score of the same film sequence), Vitouch (2001) found that "viewers' anticipations about the further development of a sequence are systematically influenced by the underlying film music" (p. 70).

In his fascinating work, Bravo (2013) studied the effect of tonal dissonance on interpretations of the emotional content of an animated short film. He hypothesized that in the same film sequence, different levels of tonal dissonance would elicit different interpretations and expectations about the emotional content of the movie scene. The short film he used as a suitable stimulus to be interpreted was very ambiguous since it did not involve clear facts in its scenes. Bravo created two soundtracks only differing as to their degree of dissonance. Comparing the subjects' interpretations in the consonant vs. dissonant condition, it emerged that in the latter, the main character was judged as more scared, alienated, sadder, less confident, and was thought to be trying to destroy something; the character was also believed to be more sinister than nostalgic and its story more tragic than hopeful.

Finally, Hoeckner et al. (2011) deepened the interesting question about how viewers relate to movie characters in correspondence of different music and how their sense of empathy is shaped by two soundtracks: thriller music and melodrama music. They found that "compared to melodramatic music, thriller music significantly lowered likability and certainty about characters' thoughts," while "melodramatic music increased love attributions and lowered fear attributions." Moreover, for the first time, they introduced the theme of empathy into the debate and, although not directly assessing its level through a specific scale, demonstrated that "musical schemas used in underscoring modulate viewers' theory of mind and emotional contagion in response to screen characters, thus providing antecedents for empathic accuracy and empathic concern."

All of these studies are overviewed in Herget's (2019) comprehensive review on music's potential to convey meaning in film; she concludes her work by underlining some weak points that should be overcome to improve such domain of research:

(1) Research on this issue is sparse. This results in experiments each analyzing a single psychological construct.
(2) Complex psychological constructs such as sympathy and empathy toward media protagonists could and should be investigated through all the available established measuring instruments (Herget, 2019).
(3) There are hardly any ecologically valid investigations carried out in natural contexts such as during a visit to the cinema, a television evening with the whole family, or alone in a young person's room (Bullerjahn, 2005);
(4) Most research designs are too complicated and extensive.
(5) Within-subjects designs risk drawing the participants' attention to the musical manipulation (Tan et al., 2007).

Since we strongly agree with the bulk of Herget's criticism, our aim in this work is to investigate the effects of music in the interpretation of visual scenes by specifically addressing these demands. In our Study 1 below, we intend to:

(1) provide a global view of the influence of background music on scene interpretation by examining various psychological constructs, such as empathy, affective states, and perceived personality traits;
(2) accurately measure such constructs by relying on available established measures and tools;
(3) improve ecological validity by running an online study to be directly done from the participants' homes on laptops and other devices;
(4) reduce the number of factors to have better control and lower the number of experimental subjects required;
(5) plan a between-subjects design to prevent the subjects' awareness of the manipulation.

Moreover, in Study 2, we employ the eye-tracking methodology to investigate the influence of music on a scene interpretation also from a physiological perspective.

## STUDY 1: ONLINE SURVEY

As stated above, the literature on the interpretation of audiovisuals has proved the ability of music to convey meanings through associations (Cohen, 2013) and activation of cognitive schemas (Boltz, 2001). In our study, we consider interpretation in a multidimensional fashion as a global process involving several interconnected cognitive operations: attribution of emotions, personality traits, thoughts or behavioral intentions to the characters on the scene, empathy toward them, and perception of the surrounding environment.

### Research Questions

Our aim is to investigate how in a visual scene the following dependent variables are affected by background music:

(1) empathy toward the character.
(2) affective states attributed to the character.
(3) impressions of the character's personality.
(4) plot anticipation.
(5) environment perception.

### Method: Rationale and Recruitment

We designed a between-subjects experiment ($N$ = 118–44, female; age = 37 ± 11, see **Table 1** for gender and age distribution) in which participants watched a scene (01′ 55″) from an almost unknown short movie (Duras, 1981) (**Figure 1**): an emotionally neutral male character slowly walks toward some large windows in a lonesome building, with the seaside in the distance. He walks, looks outside, stops, and moves out of the frame.

Through Adobe Premiere Pro, we created three versions of the scene—the three experimental conditions—with the video accompanied respectively by a dogged and anxious orchestral piece (*The Isle of the Dead* by S. Rachmaninov), a soft, melancholic jazz solo piano (*Like Someone in Love* by B. Evans), or by ambient sound only. We chose these two pieces based on the findings of Juslin and Laukka (2004), and subsequent studies listed by Cespedes-Guevara and Eerola (2018), concerning several psychoacoustic parameters associated with emotional expression in music. The two pieces both evoke negative feelings but differ in the arousal dimension: Evans track's mellow tone and soft intensity can be associated with delicacy, gracefulness, relaxation, and quietness (Fabian and Schubert, 2003) or with sadness and tenderness (Quinto et al., 2014). On the contrary, Rachmaninov track's large sound level variability, rapid changes in sound level, and ascending pitch could be linked to fear (Juslin and Laukka, 2004), while its increasingly louder intensity could communicate restlessness, agitation, tension (Fabian and Schubert, 2003) or anger, fear (Scherer et al., 2015), and scariness (Eerola et al., 2013).

After viewing the scene, participants were asked how they felt toward the character and what they thought he was feeling, what kind of personality he could have, whether they thought he was remembering or planning instead, and how they perceived the environment in which the scene was set. To avoid sequence effects, the order of questions was randomized for each participant.

Aiming at a better ecological validity, to let people participate in a less detached situation than a lab, we build the procedure on Qualtrics.com. By accessing a single unreusable link,[1] they could run the experiment directly from home on their laptops, smartphones, or tablets; recruited through Amazon Mechanical Turk, they were paid according to the standard American minimum wage: 1$ for a ∼15 min task.

**TABLE 1 |** Gender and age distribution (mean age ± SD).

| | | Controls | Evans | Rachmaninov | Total |
|---|---|---|---|---|---|
| Gender | M | 27 (36 ± 10) | 24 (37 ± 10) | 23 (34 ± 13) | 74 (36 ± 11) |
| | F | 8 (44 ± 11) | 13 (38 ± 13) | 23 (35 ± 12) | 44 (38 ± 12) |
| Total | | 35 (38 ± 10) | 37 (38 ± 11) | 46 (35 ± 12) | 118 (37 ± 11) |

### Measures and Hypotheses

We hypothesize that the narratives (hence, the interpretations) that the participants will build on the scene, influenced by the soundtrack, will be very different among them. We plan to shed light on such differences by taking a fine-grained look into some of the psychological constructs involved. Below, we describe each construct and its related measurement separately, stating our hypotheses at the bottom of each subparagraph.

In a nutshell, here is an example of two different plausible narratives:

- *Evans*: we see a sad man who walks alone in an empty building, he must be an introverted guy, we see that he's watching outside the window, maybe he's thinking about the past, maybe a loved one, the scene is sweet and quite gloomy.
- *Rachmaninov*: we see an ambiguous character walking in an unsettling hall, he shows a solemn gait, something bad is happening; probably he's planning something harmful. I wouldn't trust this man.

#### Empathy Toward the Character

To assess the participants' empathy toward the main character, after comparing various indexes (Neumann et al., 2015), we opted for a 14-item two-factor scale by Batson et al. (1983). The scale involves 14 adjectives that describe affective states of distress (alarmed, grieved, upset, worried, disturbed, perturbed, distressed, troubled) and empathic interest (sympathetic, moved, compassionate, tender, warm, softhearted). The score obtained from the difference between empathic interest and distress should therefore be the most significant assessment of the empathic response (Batson et al., 1983; Leone et al., 2008): higher ratings correspond to higher empathic interest, while lower ratings stand for enhanced distress-like feelings.

*$H_1$: Evoking feelings of delicacy and tenderness, the melancholic track (Evans) will encourage empathy toward the character. On the contrary, the negative feelings evoked by Rachmaninov will dampen empathy.*

[1] An antiballot box stuffing was employed in order to avoid multiple participations from the same device.

**FIGURE 1 |** Illustration of three representative frames of the scene.

### Affective State Attributed to the Character

We administered a classic 10-item Positive and Negative Affect Schedule (I-PANAS-SF) for the emotions attributed to the character. The used version was previously validated by Karim et al. (2011). Moreover, we added the item *wistful*, as we were convinced that it could have been significantly different among the conditions.

> *$H_{2a}$: The dogged and menacing track (Rachmaninov) will lead to attribute a more positive affective state; the character will appear as adamant; on the contrary, Evans track will let the participant attribute the character more negative affective states.*

The reason for such a hypothesis is intuitive: in the first case, music can make one imagine an evil character, possibly determined to do something harmful; while in the second case, music mood will let one picture a depressed/nostalgic character, therefore with a more negative affective state.

> *$H_{2b}$: Evans track will show higher scores in wistfulness as opposed to Rachmaninov's.*

### Impressions of Personality

To measure the participants' personality impressions (Asch, 1946) about the character, we employed a 15-items assessment of the Big Five (Lang et al., 2011) previously validated with satisfying results.

> *$H_3$: In the light of the melancholic track, the character will be seen as more agreeable and open (i.e., very emotional) and less extroverted; on the contrary, in dependence of Rachmaninov's track, the character will be regarded as more neurotic and conscientious (e.g., a lucid criminal) (**Table 2**).*

### Plot Anticipation: Past Perspective vs. Future Perspective

As for the plot anticipation, we simply asked the participants whether they thought that the main character was remembering the past (*past perspective*) or taking a decision (*future perspective*). It was also possible to choose both options. In the first case, several five-point Likert scales were presented about the emotions that the characters could have been feeling in relation to his memory. In the second one, other Likert scales were presented on the nature of such a decision; in particular, we asked whether it could have been a morally good, neutral, or bad action. In the event that both the options (i.e., remembering and taking a decision) were chosen, both the questions on the memory and the decision appeared.

> *$H_4$: When viewed with Evans music in the background, the participants will think about someone who is remembering something nostalgic; with Rachmaninov, he will be seen as a planner of possibly evil deeds.*

### Environment Perception

We were interested in understanding whether a place could be seen as cozier and warmer rather than inhospitable and unpleasant in dependence of different music; therefore, we took inspiration from a study by Yamasaki et al. (2015): they analyzed the impact of music on the impressions of the environment on the three standard dimensions of emotions: activation, valence, and potency. We decided to administer a short list of five bipolar five-step Likert scales by picking only the couples of adjectives that were somehow related to the idea of coziness, so we chose four out of five from those of the valence dimension (we excluded one for reasons of redundancy), and we also added a new couple that we considered crucial: dangerous–safe.

**TABLE 2 |** Impressions of personality (hypotheses).

| Personality trait | Evans | Rachmaninov |
|---|---|---|
| Neuroticism | – | + |
| Agreeableness | + | – |
| Conscientiousness | – | + |
| Extraversion | – | + |
| Openness | + | – |

*$H_5$: The melancholic track will let the environment be perceived as cozier. On the contrary, Rachmaninov will let our participants perceive an unpleasant environment.*

## Preliminary Sample Data Analysis

Every online procedure has the merit of guaranteeing a significant number of participants in a few days; nevertheless, lacking in experimental control, a careful preliminary analysis is necessary. To improve the reliability of our sample, first, we added an attention check question in which a multiple-item Likert scale was presented with an explicit instruction to avoid filling it out; thus, we excluded all those participants who compiled such a scale. Second, we added a time count on the screen containing the video so to exclude all of those participants who had not watched the whole scene.

After such exclusions, our sample decreased from 309 to 118 participants. No further outliers were excluded.

## Results

For all statistical analyses, IBM SPSS 26.0 was used; violin plots were made by means of XLStat 2020 3.1.

### Empathy Toward the Character

From a one-way ANOVA, it emerged that the soundtrack significantly affected the empathy felt toward the main character (scale reliability $\alpha = 0.93$), $F(2,115) = 6.86$, $p = 0.002$, $\eta_p^2 = 0.107$, $(1 - \beta) = 0.92$. The Evans group showed the highest empathy ($M = 1.16$, $SD = 1.75$), followed by the Rachmaninov group ($M = 0.21$, $SD = 0.1.53$) and controls ($M = -0.13$, $SD = 1.29$).

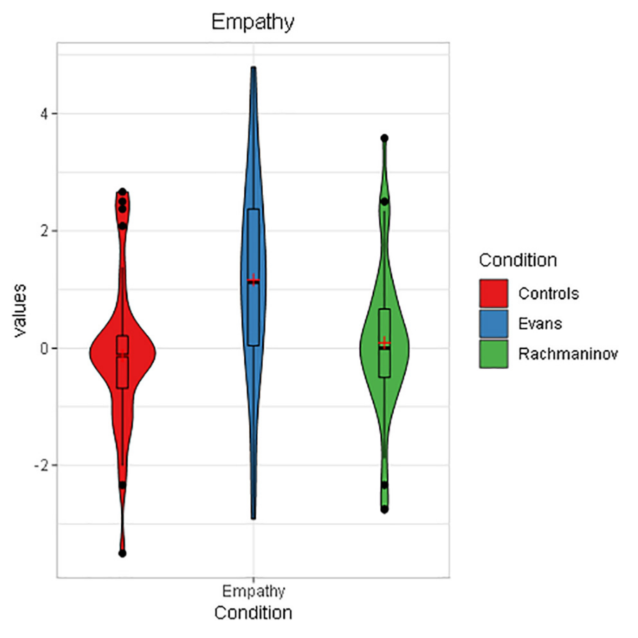



**FIGURE 2 |** Empathy toward the character as a function of condition in Study 1 (violin plot). The boxplots within each violin represent interquartile ranges (IQRs). Red crosses indicate means, black horizontal lines indicate median, and black points are outliers. Participants in Evans condition showed significantly higher Empathy ratings.

Bonferroni corrected *post hoc* revealed that Evans group's means were significantly different from that of the controls ($p = 0.002$) and Rachmaninov group ($p = 0.02$) (**Figure 2**).

We can conclude that $H_1$ was verified even if we did not record the decrease in empathy in the Rachmaninov condition as opposed to controls. This event (i.e., the absence of statistically significant differences between controls and Rachmaninov group) will appear in several analyses throughout the paper; we discuss it thoroughly in Section "General Discussion."

### Affective State Attributed to the Character

The soundtrack did not affect the affective state ($H_{2a}$) attributed to the character (scale reliability $\alpha = 0.85$); the omnibus one-way ANOVA was not significant ($p > 0.05$). Nevertheless, the adjective *wistful* showed significance. In line with $H_{2b}$, the jazz melancholic track (Evans) led to higher scores in wistfulness ($M = 3.22$, $SD = 1.39$), while the Rachmaninov group ($M = 2.43$, $SD = 1.32$) and controls ($M = 2.94$, $SD = 1.30$) showed lower ratings, $F(2,115) = 3.64$, $p = 0.029$, $\eta_p^2 = 0.060$, $(1 - \beta) = 0.66$. In this case, the mean difference of the two soundtracks was significantly different ($p = 0.029$).

### Impressions of Personality

A one-way multivariate ANOVA (MANOVA) was conducted with the five personality traits as dependent variables and the soundtrack as the independent variable. A significant multivariate main effect was found for the soundtrack, $F(10,222) = 3.30$, Wilks' $\Lambda = 0.76$, $p = 0.001$, $\eta_p^2 = 0.13$, $(1 - \beta) = 0.98$.

The soundtrack significantly affected three out of five attributed personality traits (scale reliability $\alpha = 0.73$), as in **Table 3**.

The Evans group showed the highest agreeableness toward the character ($M = 4.51$, $SD = 1.33$), followed by the Rachmaninov group ($M = 3.98$, $SD = 1.08$) and controls ($M = 3.84$, $SD = 1.09$). Subsequent Bonferroni-corrected *post hoc* comparisons revealed that Evans group's means were significantly different from that of the controls only ($p = 0.048$).

The Rachmaninov group, on the contrary, imagined the character as more conscientious ($M = 5.61$, $SD = 1.27$) as opposed to the Evans group ($M = 4.91$, $SD = 1.32$) and controls ($M = 4.82$, $SD = 1.16$). The same *post hoc* showed that the Rachmaninov group significantly differed from both the Evans group ($p = 0.041$) and controls ($p = 0.017$).

As for extraversion, participants in the Evans group registered the lowest score, namely, imagining a shier character ($M = 2.72$, $SD = 1.08$), as opposed to the Rachmaninov group ($M = 3.38$,

**TABLE 3 |** ANOVA—personality traits attributed to the character.

| Personality trait | $F$ | $p$ | $\eta_p^2$ | $(1 - \beta)$ |
|---|---|---|---|---|
| Neuroticism | 0.720 | 0.488 | 0.012 | 0.170 |
| Agreeableness | 3.43 | 0.036 | 0.056 | 0.633 |
| Conscientiousness | 4.99 | 0.008 | 0.080 | 0.803 |
| Extraversion | 3.10 | 0.049 | 0.051 | 0.587 |
| Openness | 0.648 | 0.525 | 0.011 | 0.156 |

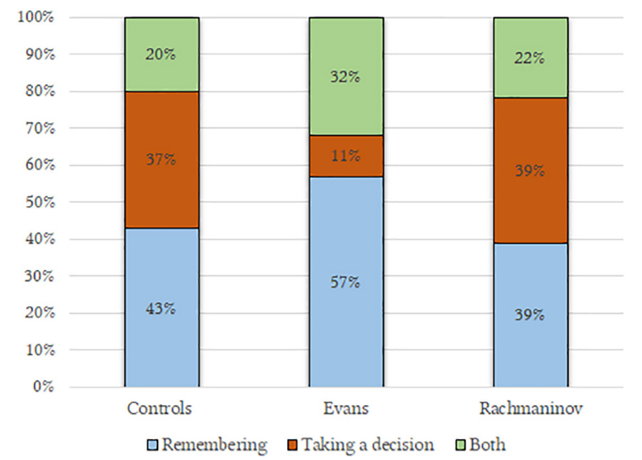



**FIGURE 3 |** Plot anticipations as a function of the condition. Participants in Evans condition imagined the character to be remembering something; on the contrary, Rachmaninov's track led the participants to imagine that the character was taking a decision.

*SD* = 1.29) and controls (*M* = 3.20, *SD* = 1.25). *Post hoc* enlightened a significant mean difference between the two soundtracks only ($p = 0.047$).

We can conclude that $H_3$ has been partially proven for three out of five personality traits.

## Plot Anticipation: Past vs. Future Perspective

As in $H_4$, Evans music made the participants think that the character was above all remembering (57%) (*past perspective*), more than in the Rachmaninov group (39%), while Rachmaninov's music let the character be seen as someone planning something (39%) (*future perspective*), more than in the Evans group (11%), $c^2(4,118) = 9.34$, $p = 0.05$, η condition dependent = 0.046 (**Figure 3**). The controls' pattern is comparable to one of the Rachmaninov group.

### *Emotions related to possible character's memories*

Evans' music led to imagine slightly pleasant memories ($M = 0.10$, $SD = 0.98$) as opposed to negative memories in Rachmaninov's track ($M = -0.70$, $SD = 1.16$) and controls ($M = -0.23$, $SD = 1.11$), $F(2,80) = 4.26$, $p = 0.017$, $\eta_p^2 = 0.09$, $(1 - \beta) = 0.73$. Subsequent Bonferroni-corrected *post hoc* comparisons revealed that the significant mean difference was the one between the two soundtracks ($p = 0.014$) (**Table 8**).

### *Moral nature of the decisions*

In line with our hypotheses, participants who embraced the future perspective were influenced by the soundtrack in foreseeing the character's decisions (*behavioral intentions*), $F(2,61) = 4.87$, $p = 0.011$, $\eta_p^2 = 0.138$, $(1 - \beta) = 0.78$: the Evans group thought it to be a good decision ($M = 1.21$, $SD = 1.04$) as opposed to the controls ($M = 0.25$, $SD = 1.24$) and Rachmaninov group ($M = 0.01$, $SD = 1.35$). The mean difference between the two soundtracks emerged through a Bonferroni-corrected *post hoc* analysis ($p = 0.010$) as the significative one (**Table 8**).

## Environment Perception

Having reached an acceptable scale reliability ($\alpha = 0.71$), to verify $H_5$, we proceeded to a one-way ANOVA that showed a main effect of the condition, $F(2,115) = 4.15$, $p = 0.02$, $\eta_p^2 = 0.067$, $(1 - \beta) = 0.72$. Subsequent Bonferroni-corrected *post hoc* analysis showed the mean difference between the two soundtracks only to be significant ($p = 0.022$). As hypothesized in $H_5$, the Evans group perceived the environment as cozier ($M = 4.34$, $SD = 1.15$) as opposed to the controls ($M = 4.17$, $SD = 1.24$) and Rachmaninov group ($M = 3.67$, $SD = 0.95$) (**Table 8**).

## Gender Differences

In order to search for gender differences, gender was added in every analysis as the factor but showed neither main effects nor interactions.

# STUDY 2: SELF-ASSESSMENT AND EYE TRACKING

Given the very satisfying results of Experiment 1, through which we proved the multifaceted influence of the soundtrack on the interpretation of a scene, we wondered whether this influence could be due to the schemas mentioned above (Boltz, 2001), whose activation could be demonstrated by eye movements; therefore, we planned to replicate the same experiment in our lab but with the addition of eye tracking.

Eye tracking has proved to be a precious technique for the analysis of several domains of cognitive science, among which were visual attention (Wedel and Pieters, 2017), cognitive workload (Kosch et al., 2018), and human interaction (Brône and Oben, 2018).

Concerning our question, we believe that different eye activities could show the activation of different cognitive schemas (Boltz, 2001), therefore configuring the (often unconscious) music perception as a proper top–down process shaping the interpretative process. To the best of our knowledge, after the seminal work of Coutrot et al. (2012), in which the presence vs. absence of sound determined different eye movements, there have been few attempts to measure eye-movement parameters in correspondence of different scenes with music, with no congruent results due to the different experimental designs and the dependent variables analyzed. Auer et al. (2012) managed to find differences in scanpaths and attention (perception of a red X in a clip with different soundtracks); similar results on attention, operationalized through the concept of spatial exploration length, have also been found by Mera and Stumpf (2014). Moreover, in a pilot experiment with a tiny sample, Wallengren and Strukelj (2015) showed an influence of film music on fixation durations in several clips; the same authors ran an improved version of the same experiment (Wallengren and Strukelj, 2018), finding marginally significant differences in the eyeblinks in dependence of different soundtracks without replicating the findings on fixation durations. In particular, they found that eyeblinks increased when film clips and music were congruent.

As for the more general analysis of dynamic stimuli such as movie scenes, not much has been done: Breeden and Hanrahan (2017) released a useful dataset for the analysis of attention in features films; in a psycho-narratology study, Kruger (2012) successfully correlated eye-tracking data with viewer constructions of the narrative of a film. Moreover, by applying a Cognitive Computational Cinematics (CCC) approach to film cognition, in two works, Smith (2013, 2014) managed to confirm filmmakers' intuitions about the influence of motion, feature contrast, and faces on viewer attention, using a combination of eye-tracking and computer vision analyses of video content. Besides, in a recent similar work, Batten and Smith (2018) faced the same theme with a similar procedure, analyzing also gaze similarity values between audio and silent conditions. They did not find an effect of sound on gaze, but the effect of audio consisted in an earlier capture of attention and self-reported higher measures of happiness and excitement.

As a matter of fact, the viewing of a narrative scene is a complex process involving at least two intertwining components, since eye movements can be exogenous or endogenous: in the former case, they are stimulus driven and depend on the visual features of the video (bright objects, camera movements, faces); the latter are, on the contrary, linked to high-level cognitive processes, such as search tasks. We believe music to be involved in this latter case, as it properly shapes the visual scene by providing a frame through which it can be interpreted.

## Research Questions

In this study, both the hypotheses and measurements of the self-report are the same as for Study 1. Yet, we additionally hypothesized that the anxious soundtrack, increasing the participants' arousal, could lead them to pay greater attention to the scene. To test this, a detail of the scene was exploited: a hidden cameraman appearing in the first and the last part of the scene; being set in a darkened building, the scene was overall dark, and this character appeared in the darkest part of the hall, very hard to be seen. Therefore, the only additional question in this second study asked participants whether they had seen another character aside from the main one.

## Method and Recruitment

We recruited participants ($N$ = 92, 63 were female; age = 26 ± 8, see **Table 4** for gender and age distribution) on a voluntary basis among students of Psychology of Communication and Cognitive science. They were symbolically rewarded with 2€ each, all of them had normal or corrected-to-normal vision. The experiment was carried out in our lab at Roma Tre University. The setting consisted of a desk with a computer and an eye-tracking device on. The average distance from the pupil to the screen was 65 cm. The screen resolution was 1,920 × 1,080 (48 cm × 27 cm).

The audio (stereo—192 kHz, 24 bit) was transmitted in the room through two speakers (JBL professional LSR305 First-Generation 5′′ two-way powered studio monitor). To avoid intensity effects, the volumes of the two tracks were normalized using a Loudness, K-weighted, relative to Full Scale (LKFS) (Grimm et al., 2010).

**TABLE 4** | Gender and age distribution (mean age ± SD).

| | | Controls | Evans | Rachmaninov | Total |
|---|---|---|---|---|---|
| Gender | M | 9 (23 ± 3) | 10 (27 ± 10) | 10 (27 ± 7) | 29 (26 ± 8) |
| | F | 21 (30 ± 11) | 21 (24 ± 4) | 21 (24 ± 6) | 63 (26 ± 8) |
| Total | | 30 (28 ± 10) | 31 (25 ± 7) | 31 (25 ± 7) | 92 (26 ± 8) |

The lighting of the room was standardized, with no sunlight entering. To improve ecological validity and to avoid a possible Hawthorne effect, during the entire experimental procedure, the operator was in the adjoining room with no possibility to see the participant. Participants were told to call for help at any time needed.

Before the task, a calibration of the eye tracker was performed for each participant. In this procedure, the participant had to gaze-follow a circle (2 cm diameter) that moved throughout the screen in a Z shape. At the end, the software reported the quality of the calibration. Only subjects with good or excellent quality were tested. When the quality was poor, a new calibration was run until reaching a satisfying quality. Besides, an operator (Author 1 or 2) checked during the whole task whether the gaze paths were reasonable (i.e., absence of any critical artifacts).

In the following, we first overview the results of Study 2 that replicated Study 1—empathy toward the character (*Empathy Toward the Character*), attribution of affective states (*Affective State Attributed to the Character*), impressions of personality (*Impressions of Personality*), plot anticipation (*Plot Anticipation: Past vs. Future Perspective*), and environment perception (*Environment Perception*); then, we detail hypotheses, measurements, and results of the eye-tracking part of Study 2 (*Eye Tracking and Pupillometry: Methodology, Metrics, and Hypotheses*); and finally, we overview the results of the survey and eye-tracking taken together (*Aggregation of Self-Reports and Eye-Tracking Data*).

## Results of the Study 2 Survey

### Empathy Toward the Character

Results on empathy replicated those of Study 1, although the scale reliability lowered until $\alpha = 0.75$. A one-way ANOVA showed a main effect of the condition, $F(2,89) = 7.96$, $p = 0.001$, $\eta_p^2 = 0.15$, $(1 - \beta) = 0.95$. Subsequent Bonferroni-corrected *post hoc* analyses revealed a significant mean difference between Evans ($M = 0.98$, $SD = 1.49$) and Rachmaninov ($M = -0.80$, $SD = 1.77$, $p < 0.001$) and Evans and Controls ($M = -0.17$, $SD = 2.07$, $p = 0.041$) (**Figure 4**). It is worth noting that in this study, the rating of the Rachmaninov group was much lower than in Study 1. We hypothesize that the laboratory environment, as opposed to a domestic setting, could have played a role in intensifying the anxious feeling, especially since the immersive sound quality of the laboratory audio system could have led to perceive better the peculiar low frequencies of this piece, the main carrier of the anxiety evoked.

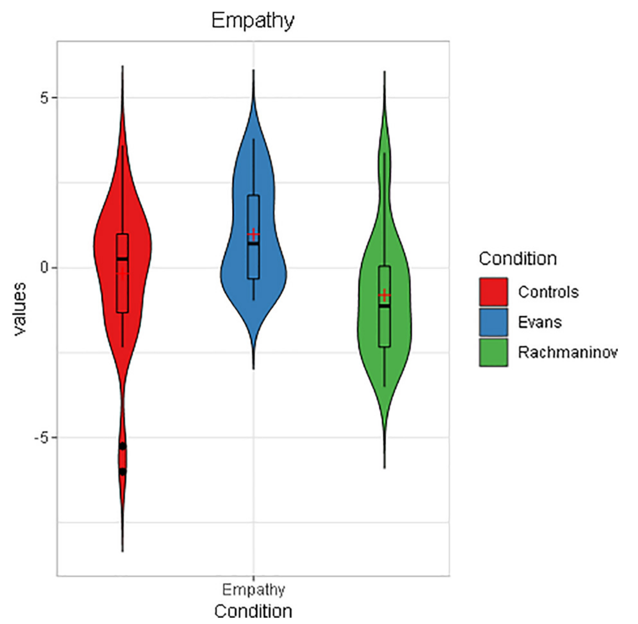



**FIGURE 4 |** Empathy toward the character as a function of condition in Study 2 (violin plot).The boxplots within each violin represent interquartile ranges (IQRs). Red crosses indicate means, black horizontal lines indicate median, and black points are outliers. As in Study 1, participants in Evans condition showed significantly higher empathy ratings.

**TABLE 5 |** ANOVA—personality traits attributed to the character.

| Personality trait | *F* | *p* | $\eta_p^2$ | (1 − β) |
|---|---|---|---|---|
| Neuroticism | 2.37 | 0.09 | 0.05 | 0.469 |
| Agreeableness | 6.54 | <0.001 | 0.16 | 0.965 |
| Conscientiousness | 7.70 | 0.001 | 0.14 | 0.942 |
| Extraversion | 0.84 | 0.316 | 0.02 | 0.250 |
| Openness | 2.31 | 0.089 | 0.05 | 0.487 |

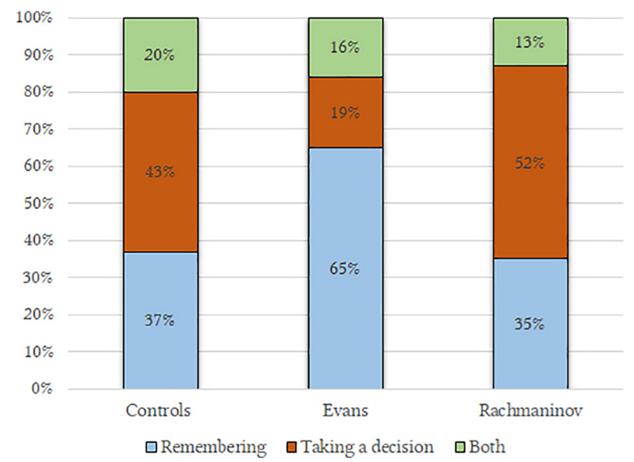



**FIGURE 5 |** Plot anticipations as a function of the condition. Coherently with Study 1, the Evans' track led the participants to imagine a remembering character; in contrast, half of the participants in the Rachmaninov's group imagined a character on the point of taking a decision.

## Affective State Attributed to the Character

As in Study 1, results on the affective states attributed to the character were not significant ($p > 0.05$). We deepen this negative finding in Section "General Discussion."

## Impressions of Personality

As in Study 1, a one-way MANOVA was conducted with personality traits as dependent variables and soundtrack as the factor. A significant multivariate main effect was found for the soundtrack, $F(10,170) = 4.40$, Wilks' $\Lambda = 0.63$, $p < 0.001$, $\eta_p^2 = 0.21$, $(1-\beta) = 0.99$.

In particular, the soundtrack significantly affected two out of five attributed personality traits (scale reliability $\alpha = 0.61$) (**Table 5**).

According to Bonferroni-corrected *post hoc* tests, in correspondence of the Rachmaninov track, the character was seen as less agreeable ($M = 3.08$, $SD = 0.87$) as opposed to the Evans group ($M = 3.95$, $SD = 0.74$, $p = 0.008$) and controls ($M = 3.77$, $SD = 0.98$, $p < 0.001$). Contrarily, in the case of conscientiousness, the Rachmaninov group showed the highest rating ($M = 5.03$, $SD = 0.84$) as opposed to the Evans group ($M = 4.45$, $SD = 0.73$, $p = 0.02$) and controls ($M = 4.18$, $SD = 0.98$, $p = 0.001$) (**Table 7**).

## Plot Anticipation: Past vs. Future Perspective

The distribution is very similar to that of Study 1, although the chi-squared test only grazed the statistical significance, possibly due to the smaller sample, $\chi^2(4,92) = 8.68$, $p = 0.06$, η condition dependent = 0.092 (**Figure 5**).

### *Emotions related to a possible character's memory*

Once again, the same pattern as in Study 1, Evans' music led to imagine more pleasant memories ($M = 0.19$, $SD = 0.1.00$) as opposed to Rachmaninov's track ($M = -0.92$, $SD = 0.86$) and controls ($M = -0.04$, $SD = 0.71$), $F(2,54) = 7.61$, $p = 0.001$, $\eta_p^2 = 0.22$, $(1 - \beta) = 0.93$. Subsequent Bonferroni-corrected *post hoc* comparisons revealed significant mean differences both between the two soundtracks ($p = 0.001$) and between Rachmaninov and controls ($p = 0.022$) (**Table 8**).

### *Moral nature of the decisions*

Concerning the moral nature of the decision, we observed the same pattern of Study 1, namely the Rachmaninov group thought of very negative decisions ($M = -0.70$), followed by controls ($M = -0.39$) and Evans ($M = -0.31$); nevertheless, results did not reach the significance ($p > 0.05$), possibly due to the less numerous sample (**Table 8**).

## Environment Perception

As concerns the scale reliability, we grazed a moderate/acceptable alpha ($\alpha = 0.67$), we replicated Study 1 results. A one-way ANOVA revealed a main effect of the condition on the environment perception, $F(2,89) = 5.02$, $p = 0.009$, $\eta_p^2 = 0.10$, $(1 - \beta) = 0.80$. The mean difference between the two soundtracks was significant ($p$ = 0.007); like in the previous experiment, Evans elicited the best evaluation of the environment ($M$ = 4.21, $SD$ = 0.90), followed by controls ($M$ = 3.74, $SD$ = 1.10) and Rachmaninov ($M$ = 3.45, $SD$ = 0.84) (**Table 8**).

### Gender Differences

As in the previous study, gender was added in every analysis as the factor but showed no main effects nor interactions.

## Eye Tracking and Pupillometry: Methodology, Metrics, and Hypotheses

We used iMotions (iMotions) software for the stimuli presentation, connected to a Tobii X2-30 Compact (screen-based) for eye-tracking recording. Gaze data have been captured at 30 Hz.

In the vast domain of eye tracking, five main metrics may be of use for our aims: time spent, fixations, revisits, dispersion (Coutrot et al., 2012), and pupillometry.

### Time Spent, Fixations, Revisits, and Dispersion

A fixation is a period during which our eyes are locked toward a specific point. Typically, the fixation duration is considered to be 100–300 ms. Fixations are considered a good measure of interest and visual attention in studies dealing with still images; nevertheless, given the dynamicity of our stimuli, and the contradictory evidence in the above-reported studies (*Study 2: Self-Assessment and Eye Tracking*), we preferred to use the mere time spent on a particular area of interest.

Concerning revisits, their number indicates how many times viewers returned their gaze to a particular point. Since one might revisit a spot because s/he considers it to be pleasing or confusing, revisits can be seen as a cue to the visual interest.

For the measurement of gaze points, given that the stimulus was not a still image but a movie scene, we built two moving areas of interest (mAOIs): the first one on the main character's full body (head included) and the second on the almost hidden cameraman.

As for the dispersion, it consists of a metric apt "to estimate the variability of eye positions between observers" (Coutrot et al., 2012, p. 4). As in the here mentioned study, we defined it as follows:

$$D(\mathrm{p}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}$$

To the best of our knowledge, it is the first time that this metric is used with the same visual stimulus accompanied by two different soundtracks and control condition.

We hypothesized that the anxiety communicated by Rachmaninov's track would induce an enhanced state of alert, the feeling that something is about to happen; therefore, the participants in this condition will explore the space in search for possible lurking dangers. On the contrary, Evans' track, characterized by lower arousal and a melancholic mood, might dampen the state of alertness, reducing the space exploration and promoting the focusing on the main character.

*$H_1$: Rachmaninov group will pay more attention to the hidden character (i.e., more time spent on his AOI and more revisits)*

*$H_2$: Rachmaninov group will pay less attention to the main character*

*$H_3$: Rachmaninov group will have higher dispersion ratings*

### Pupillometry

Pupillometry is considered as a reliable response system in psychophysiology. Changes in pupil size can reflect diverse cognitive and emotional states (for reviews, see Sirois and Brisson, 2014; Mathôt, 2018), ranging from arousal, interest, and effort in social decisions (Kret and Sjak-Shie, 2019). Here our focus is on arousal: we believe that the two soundtracks will elicit opposite arousal tendencies, with Rachmaninov being the most and Evans the least arousing.

We measured pupil dilation during the whole visual stimulus. Since Rachmaninov's track induces an anxious mood, we hypothesized a greater pupil dilation in that condition.

*$H_4$: The Rachmaninov group will have greater pupil dilation as opposed to the Evans group.*

### Eye Tracking and Pupillometry Results

First of all, following the criteria of Breeden and Hanrahan (2017), we invalidated all recorded data in which the error rate (eye-tracker error + gaze point off-screen) for a specific participant exceeded a threshold of 10% ($N$ = 3). *Time Spent, Revisits, and Dispersion* and *Pupillometry* report the analyses for time spent and revisits and for pupillometric data, respectively.

#### *Time spent, revisits, and dispersion*

Three one-way ANOVA analyses were conducted to verify $H_1$, with the soundtrack as the factor and time spent and revisits as dependent variables. Both time spent and revisits on the hidden character were significant: we found a main effect of the soundtrack, $F(2,86) = 3.10$, $p = 0.05$, $\eta_p^2 = 0.06$, $(1 - \beta) = 0.58$ and $F(2,86) = 3.00$, $p = 0.05$, $\eta_p^2 = 0.06$, $(1-\beta) = 0.56$. Subsequent least significant difference (LSD) *post hoc* analyses revealed that for time spent (ms), the Rachmaninov group ($M$ = 6,197, 15%) showed higher ratings as opposed to the Evans group ($M$ = 4596, 11%, $p$ = 0.04) and controls ($M$ = 4,486, 11%, $p$ = 0.02). As for revisits, only the mean difference of the two experimental conditions showed significance ($p$ = 0.02) with Rachmaninov group's ratings being $M$ = 11.07 and Evans $M$ = 8.07; that of the controls were $M$ = 8.67. We can conclude that $H_1$—Rachmaninov inducing more attention to the hidden character—was verified.

Nevertheless, these effects could also be due to participants making, on average, more eye movements in the Rachmaninov group and not in particular on the hidden character. To control this, we run the same analysis on an irrelevant portion of the scene (a window), the ANOVA resulted to be not significant. These results were also confirmed by an analysis on time spent and revisits made on the whole scene.

As for $H_2$, concerning the attention paid to the main character, the time spent and revisits on the mAOI were not different in the three conditions.

We then analyzed dispersion[2] to verify $H_3$. An analysis of covariance (ANCOVA) with the frames as a covariate and the condition as the factor revealed a main effect of the condition, $F(2,10157) = 132.76$, $p < 0.001$, $\eta_p^2 = 0.025$, $(-\beta) > 0.99$. In greater detail, our prediction ($H_3$) was not verified since the multiple comparisons *post hoc* analyses with Bonferroni correction showed that the controls ($M = 198.09$, $SD = 54.09$) had significantly higher dispersion as opposed to the Evans group ($M = 180.24$, $SD = 52.79$, $p < 0.001$) and the Rachmaninov group ($M = 179.43$, $SD = 52.90$, $p < 0.001$). Although we were unable to prove $H_3$, what we found (i.e., higher dispersion in the absence of music) is not a new result in the literature (Coutrot et al., 2012), and we claim it to be explicable in terms of attention focusing (see section "General Discussion"). Lastly, an effect of the frames was found, $F(1,10156) = 28.95$, $p < 0.001$, $\eta_p^2 = 0.003$, $(1 - \beta) > 0.99$, in that the dispersion significantly varied over time.

***Pupillometry***

iMotions provides autoscaled normalized data for pupil dilation in which each value is obtained by averaging the pupil size of both eyes. The average pupil size of the first 150 ms from the stimulus onset (fade in from black) was used as a baseline for two reasons: first, dealing with a video stimulus widely changing in luminance, it was not useful to create a blank screen with the average luminance of the stimulus; second, as the music onset in each of the three conditions was after the first 150 ms and the very first screen was exactly the same in terms of luminance, screen used, and room luminosity, the three baselines could by no means differ among each other. However, to be on the safer side, the baseline fragments were not included in any of the analyses.

Following Lemercier et al. (2014), the percentage change in pupil diameter was assessed for each data point by using the following equation:

$$\%\ change = \frac{(Xdata - baseline)}{baseline} \times 100$$

Since the timeframe is coherent among the three conditions, we run a one-way ANOVA with the soundtrack as the factor and the percentage change of pupil diameter as the dependent variable. A main effect of the condition was found, $F(2,5139) = 38.09$, $p < 0.001$, $\eta_p^2 = 0.01$, $(1 - \beta) > 0.99$. Subsequent *post hoc* analyses using Bonferroni correction revealed two differences to be significant ($p < 0.001$ both for Evans/CC and Evans/Rachmaninov) (**Table 6**).

As hypothesized in $H_4$, the anxious soundtrack (Rachmaninov) caused greater pupil dilation ($M = 0.27$, $SD = 0.24$) as opposed to the melancholic one (Evans) ($M = 0.20$, $SD = 0.23$). Nevertheless, quite unexpectedly, the greatest pupil dilation overall was found in the controls ($M = 0.27$, $SD = 0.24$) (**Figure 6**). This finding contradicts previous findings by Wöllner and Hammerschmidt (2018), in which participants' pupil diameters were larger in an audiovisual compared to a visual-only condition but confirms the findings by Batten and Smith (2018). Our result, as well of that of Batten and Smith (2018), might be also due to a strong attentional increase: while the soundtracks somehow help disambiguate the scene and frame it by providing information, the viewing of that dark and quite static scene without any music might produce a sense of hanging, an increase in the arousal and cognitive load observable through the pupil dilation.

Apparently, our data show that in control and Rachmaninov conditions, pupil dilation was higher, indicating more arousal, while in Evans, it was lower, cueing to more relaxation. We can account for such different patterns stating that the total absence of the soundtrack (or the awkward presence of silence) produced a higher state of alert, whereas the romantic/melancholic mood evoked by Evans favored a sweeter and more relaxing view.

Nonetheless, pupillometric data should always be interpreted with the caveat of their intrinsic non-specificity: pupil dilation can be due to an increase in either arousal or cognitive load (Sirois and Brisson, 2014). Unfortunately, both of these features are crucial to the interpretation process. Thereupon, it is not trivial to bear in mind that any observed difference in pupil dilation might heavily depend on the inherent characteristics of a specific visual stimulus.

## Aggregation of Self-Reports and Eye-Tracking Data

### The Hidden Character: Did Participants Really See Him?

Participants in the Evans group answered that they had noticed the hidden character less than the others: only 48% of them responded positively (level of chance), as opposed to the Rachmaninov group (64%) and controls (73%).

Therefore, we wanted to verify through mAOIs whether the participants who said they had seen someone else actually looked at the hidden cameraman more. An alternative explanation could have been that the anxious music would have led to imagine someone else even if they had not seen him. For this purpose, we ran two one-way ANOVA analyses with the answer to the question "Have you seen someone else in the scene?" (i.e., yes/no) as the factor and time spent and revisits as dependent variables. Here, we found a main effect of the yes/no answer both for time spent, $F(1,86) = 20.06$, $p < 0.001$, $\eta_p^2 = 0.18$, $(1 - \beta) = 0.99$, and for revisits, $F(1,86) = 10.70$, $p = 0.002$, $\eta_p^2 = 0.11$, $(1 - \beta) = 0.89$; namely, regardless of the soundtrack, those who answered *yes* actually spent more time ($M_Y = 14.95\%$, $SD = 7.57$; $M_N = 8.38\%$, $SD = 5.00$) and revisits ($M_Y = 10.56$, $SD = 4.96$; $M_N = 7.12$, $SD = 4.59$) on the mAOI of the hidden character.

### Does the Interpretation Change in Dependence on the Time Spent on the Main Character?

Our next question was, how does the scene interpretation change depending on how much the participant's attention is focused on the main character? In order to answer this, we split our sample into three balanced subsamples based on the percentage of time spent on the main character, and we ran

[2]Given that we located each gaze point by means of Cartesian coordinates overlapping the screen resolution ($0 \leq x \leq 1920 \mid 0 \leq y \leq 1080$), dispersion values are given in pixels.

**TABLE 6 |** Bonferroni *post hoc* analysis of percentage change in pupil diameter.

| Condition | Condition | Mean difference | *SE* | Significance |
|---|---|---|---|---|
| CC (ambient sound only) (*M* = 0.27; *SD* = 0.24) | Evans (melancholic jazz) | 0.070 | 0.008 | 0.000 |
| | Rachmaninov (anxious orchestral) | 0.017 | 0.008 | 0.107 |
| Evans (melancholic jazz) (*M* = 0.20; *SD* = 0.23) | CC (ambient sound only) | −0.070 | 0.008 | 0.000 |
| | Rachmaninov (anxious orchestral) | −0.052 | 0.008 | 0.000 |
| Rachmaninov (anxious orchestral) (*M* = 0.25; *SD* = 0.24) | CC (ambient sound only) | −0.017 | 0.69 | 0.107 |
| | Evans (melancholic jazz) | 0.052 | 0.008 | 0.000 |

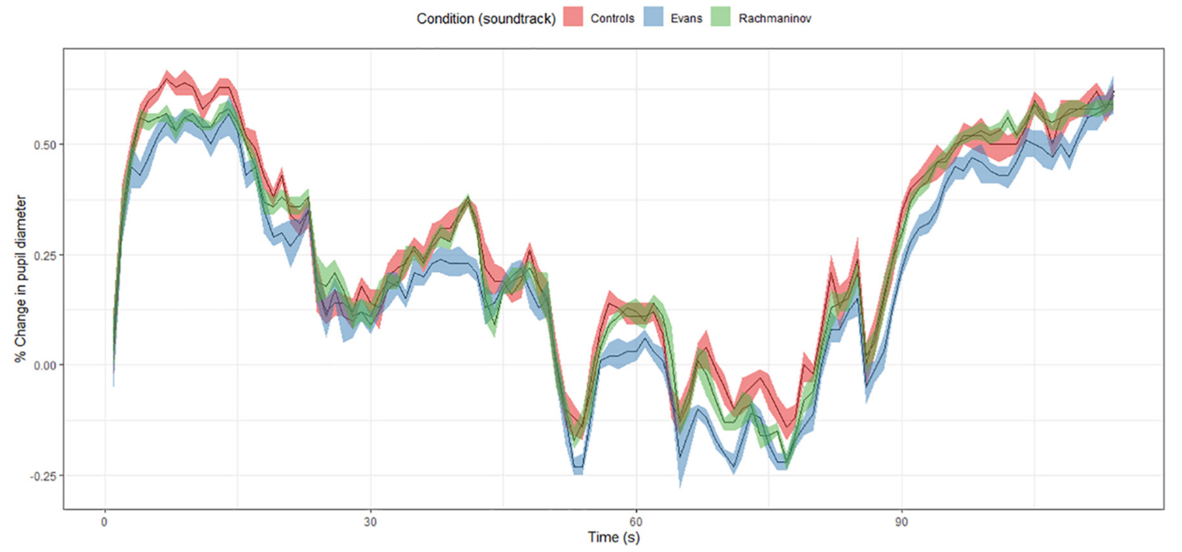



**FIGURE 6 |** Percentage change in pupil diameter as a function of time (s) and condition. Throughout the whole scene, the pupils of the participants in the Evans group had significantly lower dilation as opposed to those in the other conditions.

a two-way ANOVA with factors being the time spent on the mAOI of the main character and the soundtrack and dependent variables empathy toward the character, affective state attributed to him, and impressions of personality. No significant main effect was found, but we found a strong interaction effect for the attributed affective state, $F(4,80) = 3.35$, $p = 0.014$, $\eta_p^2 = 0.144$, $(1 - \beta) = 0.82$; namely, it seems that with the anxious music, the more our participants looked at the main character, the more negative the affective state attributed to him, whereas those who focalized more on the character with the melancholic track tended to attribute him a more positive affective state (**Figure 7**).

## GENERAL DISCUSSION

Our work aimed to show the extent to which a soundtrack affects the interpretation of a short movie scene; we did it by measuring, for the first time to the best of our knowledge, several dependent variables together in a randomized fashion. The results are very satisfying.

The strongest finding concerns empathy. In both studies, we saw a substantial increase in the empathic response toward the main character in correspondence of the sweet jazz soundtrack (Empathy Toward the Character in Study 1: Online Survey and Empathy Toward the Character in Study 2: Self-Assessment and Eye Tracking) (**Figures 2**, **4** and **Table 7**), while in the other conditions, our participants had scores close to 0. As hypothesized by previous work (Pisciottano, 2019), extradiegetic music was unconsciously perceived as an emotionalizing comment on the scene, something speaking for the main character's feelings. Evans' music communicated sweet and melancholic feelings, paving the way to an empathic response; on the contrary, the low empathy scores of controls and Rachmaninov's track suggest higher feelings of distress due to the unpredictable situation in the former case and a state of alert in the latter.

As for the affective state attributed to the main character, we did not find any effect (*Affective State Attributed to the Character* in *Study 1: Online Survey* and *Affective State Attributed to the Character* in *Study 2: Self-Assessment and Eye Tracking*) (**Table 7**): an unexpected result, that music influenced the vast majority of our dependent variables without affecting the emotions attributed to the main character. Ultimately, emotional contents are what music is all about. We shall come back to this issue in *Conclusion*. Nevertheless, we are confident to find such an effect in further studies, possibly by using different scales that can better deepen the role of the most involved emotions.

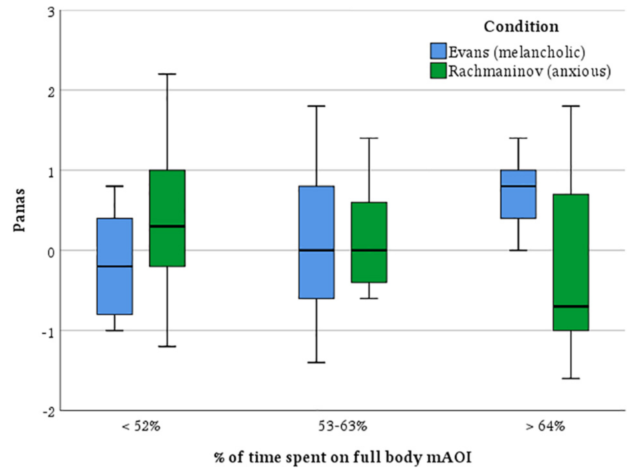



**FIGURE 7 |** Affective state attributed to the main character as a function of the time spent on his moving area of interest (mAOI). The boxplots represent interquartile ranges (IQRs). Black horizontal lines indicate median. As the participants watched the main character to a larger extent, with Rachmaninov's track, they progressively decreased the affective state attributed to him; conversely, with Evans' track, they progressively increased the affective state attributed to the main character.

Concerning the impressions of personality, our findings are mixed but encouraging; on the one hand, we found a multivariate effect both in Study 1 and Study 2 (*Impressions of Personality* in *Study 1: Online Survey* and *Impressions of Personality* in *Study 2: Self-Assessment and Eye Tracking*). As to main effects, things become more varied; we found significance for two personality traits in both studies (i.e., agreeableness and conscientiousness), while extraversion was significant in Study 1, and neuroticism and openness were nearly so in Study 2. It must be stressed that in both studies, the most affected traits were the ones related with relational and moral aspects more than mere psychological features: as for the high agreeableness, this attribution might be considered a precursor to empathy; in fact, in both studies we found a correlation between them, $r(118) = 0.36$, $p < 0.001$ and $r(92) = 0.46$, $p < 0.001$. Overall, the data suggest that the Evans music increased the character's agreeableness, something similar to what was found by Hoeckner et al. (2011), where the likeability of a character increased with melodramatic music as opposed to thriller music or no music. On the contrary, Rachmaninov decreased agreeableness, while in parallel, it increased the character's perceived conscientiousness, possibly due to the sense of austere and serious atmosphere evoked by the lowest notes of double basses and cellos.

The results show that music also influenced plot anticipations, confirming other findings (Bullerjahn and Güldenring, 1994; Vitouch, 2001; Shevy, 2007); in both our studies, Evans led our participants to imagine a character who was remembering more than deciding something i.e., embracing a past perspective. In particular, the memories associated with this music were characterized as slightly pleasant whereas Rachmaninov let our participants think of someone planning morally bad deeds (*Moral Nature of the Decisions*) (**Table 8**).

Concerning environment perception, in both studies, Evans music made it appear (*Environment Perception* in *Study 1: Online Survey* and in *Environment Perception* in *Study 2: Self-Assessment and Eye Tracking*) (**Table 8**) more interesting, pleasant, likable, warmer, and safer, like in similar results already found in very different contexts (Michon and Chebat, 2004; Yamasaki et al., 2015). This result confirms the holistic nature of the interpretation process, showing that music, a human artifact, extends its capacity of shaping not only on the narratives involving human features (emotions, feelings, plans, etc.) but also on inanimate objects and the environment of the scene. It would be very reductive to say that music only expresses feelings: much more than that, it actually shapes our view.

Finally, eye-tracking data of Study 2 provide new physiological insights about scene viewing: music can improve focusing (*Time Spent, Revisits, and Dispersion*), as found by Auer et al. (2012) and Mera and Stumpf (2014), and also influences pupil dilation (*Pupillometry*). Unfortunately, no consistent evidence can be claimed on this matter, since the only similar study (Wöllner and Hammerschmidt, 2018) found opposite results, namely pupil diameters larger in the audiovisual (both with tranquil and dogged music) compared to the visual-only condition. We might explain such conflicting evidence as follows: since two out of the three clips used by those authors depict very dynamic actions (a beating and a chase), participants would not need music to be aroused, and soundtrack worked as an *associative enhancer* (Brown, 2006) in an already-arousing situation. Conversely, in our control condition, silence could have worked as an enhancer of arousal by eliciting a sense of hanging (*Time Spent, Revisits, and Dispersion*), making the pattern of this and other dependent variables often overlap with that of the anxious condition. Furthermore, coherently with Coutrot et al. (2012), our control condition reported the highest ratings in dispersion, confirming the role of music in the attention focusing process throughout a visual scene.

## LIMITATIONS

Despite its satisfying results, this study has some relevant limitation: first of all, the two pieces differ from each other not only in terms of their emotional content but also in their genre. Moreover, one is a solo performance, while the other is fully orchestral. Strictly speaking, we cannot be sure whether such differences could have affected the interpretations; we should have at least created two more conditions: one with a solo/anxious piece and another with an orchestral/melancholic one searching for genre or quantity of instruments effects. Undoubtedly, several other finer-grained manipulations could have been performed; we could have manipulated the mode only of the same melody: major vs. minor, as in Ziv and Goshen (2006), or degree of dissonance, as in Bravo (2013). Nevertheless, as a first attempt, we opted for two tracks that were radically dissimilar so to be able to discover also moderate effects with a relatively small sample.

Another concern lies in the absence of a listening-only condition in which the participants had to imagine a scene

**TABLE 7 |** Results of Study 1 and Study 2: empathy, affective state, and impressions of personality.

| | | | | | Big 5 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *N* | Empathy | Aff. State | Neuroticism | Agreeableness | Conscientiousness | Extraversion | Openness |
| **Study 1 (*N* = 118)** | **Rachmaninov** | 46 | 0.21 | | | 3.98 | 5.61** | 3.38** | |
| | **Evans** | 37 | 1.16** | | | 4.51* | 4.91 | 2.72 | |
| | **Controls** | 35 | −0.13 | NS | NS | 3.84* | 4.82 | 4.82 | NS |
| | **omnibus *p*** | | 0.002 | | | 0.036 | 0.008 | 0.049 | |
| | $\eta_p^2$ | | 0.10 | | | 0.05 | 0.08 | 0.05 | |
| **Study 2 (*N* = 92)** | **Rachmaninov** | 31 | −0.80 | | 4.09 | 3.08** | 5.03** | | 4.16 |
| | **Evans** | 31 | 0.98** | | 3.93 | 3.95 | 4.45 | | 4.69 |
| | **Controls** | 30 | −0.17 | NS | 3.53 | 3.77 | 4.18 | NS | 4.34 |
| | **omnibus *p*** | | 0.001 | | 0.09 | <0.001 | 0.001 | | 0.08 |
| | $\eta_p^2$ | | 0.15 | | 0.05 | 0.16 | 0.14 | | 0.05 |

*Bonferroni correction post hoc analyses. All the results for p < 0.10 are listed. *Only the mean difference between the two with the single asterisk is significant. **All mean differences with the other conditions are significant (p < 0.05).*

**TABLE 8 |** Results of Study 1 and Study 2: plot anticipations and environment perception.

| | | Plot anticipation | | | | |
|---|---|---|---|---|---|---|
| | | Remembering | Taking a decision | Emotions of memories | Moral nature of decisions | Environment perception |
| **Study 1 (*N* = 118)** | **Rachmaninov** | 39% | 39% | −0.70* | 0.01* | 3.67* |
| | **Evans** | 57% | 11% | 0.10* | 1.21* | 4.34* |
| | **Controls** | 43% | 37% | −0.23 | 0.25 | 4.17 |
| | **omnibus *p*** | 0.05 | | 0.017 | 0.011 | 0.025 |
| | $\eta_p^2$ | | | 0.09 | 0.13 | 0.067 |
| **Study 1 (*N* = 92)** | **Rachmaninov** | 35% | 52% | −0.92** | | 3.45* |
| | **Evans** | 65% | 19% | 0.19 | | 4.21* |
| | **Controls** | 37% | 43% | −0.04 | NS | 3.74 |
| | **omnibus *p*** | 0.06 | | 0.001 | | 0.009 |
| | $\eta_p^2$ | | | 0.22 | | 0.10 |

*Bonferroni correction post hoc analyses. All the results for p < 0.10 are listed. *Only the mean difference between the two with the single asterisk is significant. **All mean differences with the other conditions are significant (p < 0.05).*

instead of watching it; this could have been profitable to understand the extent to which the semantic musical information was related to the music itself or to the associative processes with the visual material. We plan to use a similar design in further studies.

Finally, we are aware that these results heavily depend on the semantic content of the visual scene: in the current study, to be sure enough to find noticeable effects of the soundtrack, given the intrinsic primacy of the video contents on the audio ones when making sense of an audiovisual, we opted for a semantically weak scene in which no concrete action was done. We are confident that as the visual content becomes stronger, the power of the music becomes weaker, shifting from a decisive role to an ancillary one. Furthermore, as the visual becomes stronger, we hypothesize that there is an increase in the search for the congruity with the soundtrack, possibly leading to stronger and more coherent contents when satisfied or highly memorable and uncanny scenes when unsatisfied. For instance, who does not remember Hannibal Lecter butchering the guards to the soft (and diegetic) sound of Johann Sebastian Bach's *Aria da Capo*? (BWV 988)? Or Alex DeLarge and his droogs singing *Singin' in the Rain* in the famous raping scene of *A Clockwork Orange*?

## CONCLUSION

Every time we confront a new stimulus—an event, a news, a novel, a piece of music, a poem—we need to interpret it, that is, to make sense of it: to understand it and integrate it into our previous knowledge base by connecting new and old information through inferences and logical relations. To do so, we need to add information to the stimulus, whether on our own or thanks to an "interpreter." A judge in his verdict interprets the rule by shaping it onto a specific transgression; a pianist interprets a Sonata by adding his expressivity; a historian allows us to better interpret the present conflict between two neighboring countries by telling of their history. Information may be added either on a top–down or a bottom–up basis (Shevy, 2007). In the former case, it is provided by our point of view: on a deeper level by our goals or previous knowledge, our personality, and memories; on a more surface level by our contingent affective and cognitive state. Such a point of view superimposes a frame to the stimulus, like sand molds shape the sand or like colored lenses give a color to the world we see. In the bottom–up case, information is drawn from outside: we may better comprehend a poem or a philosophical essay, thanks to a literary critic or an exegetical comment. However, if nobody else hands us such information, given our incessant need, as humans, to search for knowledge everywhere, we can figure out some, filling in the gaps of the unsaid by our inference and imagination. For instance, to understand the plot of a novel or a movie, we need to know the characters' goals and personalities, and if we have no information whatsoever, we try to take advantage of any cue to guess them. More specifically, we need to build narratives that must be coherent among the involved modalities, therefore audio and video in our case, as in Cohen's (2013) Congruence-Association Model.

In our studies on how music affects a visual scene interpretation, this is what seems to happen in our participants. Viewing quite an ambiguous scene and needing to make sense of it, they resort to whatever contextual information, and they profoundly rely on music. If the background music is sweet and melancholic, they either feel (in the emotion-induction hypothesis, Juslin and Laukka, 2004) or cognitively represent (in the representation hypothesis, Brown, 2006) a sad feeling that paves the way for an empathic feeling toward and makes the participants attribute him sad memories. Conversely, if the music is disquieting, attributing this to a sense of alert and evil intentions, they do not feel empathy toward him. In this way, participants can build a coherent and interconnected structure of the scene and expectancies about its plot. This is how we might integrate the results of our studies: in fact, although we have found that multiple variables are affected by the soundtrack, we do not know what are the relationships among them. We plan to investigate their specific intertwining and causal chain in future studies.

In any case, our work provides new evidence of the extent to which music can shape the interpretation of a scene; by measuring several dependent variables entailed by this process (empathy toward a character, attributed affective states and impressions of personality, plot anticipations, environment perception), we came up with a multidimensional perspective of scene interpretation. Robust effects emerged on empathy, plot anticipations, environment perception, and two impressions of personality (agreeableness and conscientiousness) were heavily influenced by music, while the three others (neuroticism, extraversion, openness) less so. Taken together, these results indicate that the soundtrack had a greater impact on some attributed dimensions that are more stable and personal (i.e., impressions of personality and empathic feeling) compared to the contingent affective state; indeed, we may account for this result by recurring to the actor–observer bias (Jones, 1972), according to which people tend to judge their own behavior relying on contingent factors, while the others' actions as caused by dispositional aspects. Surely, further studies need to prove this point, but it looks plausible that when anticipating or judging a character's behavior, viewers would be exposed to such a bias, namely, they tend to overattribute causality to the character's personality (i.e., dispositional factors) while avoiding to explain motivations by situational factors as the contingent affective state.

These studies represent an attempt to fill in the theoretical and methodological gaps reported by Herget's (2019) review. Although our results seem very promising, further investigations are needed. Here, the two musical excerpts differed from one another in terms of several musical features; having found solid effects with such tracks, the further challenge might be to employ subtler manipulations on specific musical cues already proved to be linked to emotional expression (Eerola et al., 2013), such as mode (Ziv and Goshen, 2006), tempo (Fernández-Sotos et al., 2016), dynamics, and timbre (Hailstone et al., 2009; Wu et al., 2014). Another topic to explore is what it means that a music piece *fits* a visual scene; something similar has been proposed for audiovisual advertising (Herget et al., 2018), but one might deepen what are the effects provoked by music that does not fit a scene and whether this can elicit cognitive dissonance.

Furthermore, being the political campaigns advertising conceived as proper audiovisual contents (Killmeier and Christiansen, 2004; Ezell, 2012; Shevy and Hung, 2013), it would be profitable to analyze how audio (speech and background music)-visual fitness can affect the message appeal in terms of credibility and voting intentions.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Commissione Etica of Roma Tre University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

AA conceived and designed the study and wrote the manuscript. AA and MM collected the data, organized the database, and performed the statistical analyses. FD'E supervised the analysis. IP contributed in the conclusion and supervised the whole work. All authors contributed to manuscript revision, read and approved the submitted version.

## FUNDING

This work was partly funded by the Italian MIUR, PRIN "Cultural Heritage Resources Orienting Multimodal Experiences (CHROME)," grant no. 2015WXBPYK, and partly by a scholarship for research on the influence of music issued by the Department of Philosophy, Communication and Performing Arts, Roma Tre University, grant no. 1236-2019.

## ACKNOWLEDGMENTS

AA desires to thank Marina Mishchenko (Hôpital Universitaire Robert-Debré) for her advices on eye-tracking data and Innocenzo De Marco for his insight about smoothening pupillometry graphs and dispersion Python script. All the authors gratefully thank the two reviewers for their significant and profitable comments that improved the quality of the present work.

## REFERENCES

Ansani, A., D'Errico, F., and Poggi, I. (2019). 'You will be judged by the music I hear': a study on the influence of music on moral judgement. *Web Intellig.* 1, 53–62. doi: 10.3233/WEB-190400

Asch, S. E. (1946). Forming impressions of personality. *J. Abnorm. Soc. Psychol.* 41:258.

Au, K., Chan, F., Wang, D., and Vertinsky, I. (2003). Mood in foreign exchange trading: Cognitive processes and performance. *Organ. Behav. Hum. Decis. Process.* 91, 322–338. doi: 10.1016/S0749-5978(02)00510-1

Auer, K., Vitouch, O., Koreimann, S., Pesjak, G., Leitner, G., and Hitz, M. (2012). "When music drives vision: Influences of film music on viewers' eye movements," in *Proceeding of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*, New York, NY.

Batson, C. D., O'Quin, K., Fultz, J., Vanderplas, M., and Isen, A. M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. *J. Person. Soc. Psychol.* 45:706. doi: 10.1037/0022-3514.45.3.706

Batten, J. P., and Smith, T. J. (2018). "Looking at sound: sound design and the audiovisual influences on gaze," in *Seeing into Screens: Eye Tracking the Moving Image*, eds T. Dwyer, C. Perkins, S. Redmond, and J. Sita (London: Bloomsbury Publishing), 85–102.

Bolivar, V. J., Cohen, A. J., and Fentress, J. C. (1994). Semantic and formal congruency in music and motion pictures: effects on the interpretation of visual action. *Psychomusicolo. J. Res. Music Cogn.* 13:28. doi: 10.1037/h0094102

Boltz, M. G. (2001). Musical soundtracks as a schematic influence on the cognitive processing of filmed events. *Music Percept.* 18, 427–454. doi: 10.1525/mp.2001.18.4.427

Boltz, M. G. (2004). The cognitive processing of film and musical soundtracks. *Mem. Cogn.* 32, 1194–1205. doi: 10.3758/BF03196892

Branigan, E. (2010). Soundtrack in mind. *Projections* 4:140. doi: 10.3167/proj.2010.040104

Bravo, F. (2013). "The influence of music on the emotional interpretation of visual contexts," in *From Sounds to Music and Emotions*, Vol. 7900, eds M. Aramaki, M. Barthet, R. Kronland-Martinet, and S. Ystad (Berlin: Springer), 366–377. doi: 10.1007/978-3-642-41248-6_20

Breeden, K., and Hanrahan, P. (2017). Gaze data for the analysis of attention in feature films. *ACM Trans. Appl. Percept.* 14, 1–14. doi: 10.1145/3127588

Brodsky, W. (2001). The effects of music tempo on simulated driving performance and vehicular control. *Transp. Res. Part F: Traffic Psychol. Behav.* 4, 219–241. doi: 10.1016/S1369-8478(01)00025-0

Brône, G., and Oben, B. (eds) (2018). *Eye-Tracking in Interaction: Studies on the Role of Eye Gaze in Dialogue*. Amsterdam: John Benjamins Publishing Company.

Brown, S. (2006). "How does music work? toward a pragmatics of musical communication," in *Music and Manipulation*, eds S. Brown and U. Volgsten (New York, NY: Berghahn Books). doi: 10.1093/acprof:oso/9780198529361.003.0001

Brown, S., and Theorell, T. (2006). "The social uses of background music for personal enhancement," in *Music and Manipulation: On the Social Uses and Social Control of Music*, eds S. Brown and U. Volgsten (New York, NY: Berghahn Books), 126–161.

Bruner, G. C. (1990). Music, mood, and marketing. *J. Market.* 54, 94–104. doi: 10.1177/002224299005400408

Bullerjahn, C. (2005). "Musik und bild [music and image]," in *Musikpsychologie*, eds H. Bruhn, R. Kopiez, and A. C. Lehmann (Das neue Handbuch), 205–222.

Bullerjahn, C., and Güldenring, M. (1994). An empirical investigation of effects of film music using qualitative content analysis. *Psychomusicology* 13, 99–118. doi: 10.1037/h0094100

Cespedes-Guevara, J., and Eerola, T. (2018). Music communicates affects, not basic emotions - a constructionist account of attribution of emotional meanings to music. *Front. Psychol.* 9:215. doi: 10.3389/fpsyg.2018.00215

Cohen, A. J. (2001). "Music as a source of emotion in film," in *Series in Affective Science. Music and Emotion: Theory and Research*, eds P. N. Juslin and J. A. Sloboda (Oxford: Oxford University Press), 249–272.

Cohen, A. J. (2013). "Congruence-association model of music and multimedia: Origin and evolution," in *The psychology of Music in Multimedia*, eds S.-L. Tan, A. J. Cohen, S. D. Lipscomb, and R. A. Kendall (Oxford: Oxford University Press), 17–47. doi: 10.1093/acprof:oso/9780199608157.003.0002

Coutrot, A., Guyader, N., Ionescu, G., and Caplier, A. (2012). Influence of soundtrack on eye movements during video exploration. *J. Eye Mov. Res.* 5, 1–10.

Dixon, L., Trigg, R., and Griffiths, M. (2007). An empirical investigation of music and gambling behaviour. *Intern. Gambl. Stud.* 7, 315–326. doi: 10.1080/14459790701601471

Duras, M. (1981). *L'homme Atlantique*. Available online at: https://youtu.be/AI-N5VKiN3A (accessed Augest 26, 2020).

Edworthy, J., and Waring, H. (2006). The effects of music tempo and loudness level on treadmill exercise. *Ergonomics* 49, 1597–1610. doi: 10.1080/00140130600899104

Eerola, T., Friberg, A., and Bresin, R. (2013). Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Front. Psychol.* 4:487. doi: 10.3389/fpsyg.2013.00487

Ezell, J. E. (2012). *Affective, Attitudinal, and Cognitive Responses to Music in Modern US Political Advertising*. London: University of Alabama Libraries.

Fabian, D., and Schubert, E. (2003). Expressive devices and perceived musical character in 34 performances of variation 7 from bach's goldbergvariations. *Mus. Sci.* 7, 49–71. doi: 10.1177/10298649040070s103

Fernández-Sotos, A., Fernández-Caballero, A., and Latorre, J. M. (2016). Influence of tempo and rhythmic unit in musical emotion regulation. *Front. Comput. Neurosci.* 10:80. doi: 10.3389/fncom.2016.00080

Fraser, C., and Bradford, J. A. (2013). Music to your brain: background music changes are processed first, reducing ad message recall: MUSIC TO YOUR BRAIN. *Psychol. Market.* 30, 62–75. doi: 10.1002/mar.20580

Furnham, A., and Strbac, L. (2002). Music is as distracting as noise: the differential distraction of background music and noise on the cognitive test performance of introverts and extraverts. *Ergonomics* 45, 203–217. doi: 10.1080/00140130210121932

Grimm, E. M., Van Everdingen, R., and Schöpping, M. (2010). Toward a recommendation for a European standard of peak and LKFS loudness levels. *SMPTE Mot. Imag. J.* 119, 28–34. doi: 10.5594/j11396

Hailstone, J. C., Omar, R., Henley, S. M. D., Frost, C., Kenward, M. G., and Warren, J. D. (2009). It's not what you play, it's how you play it: timbre affects perception of emotion in music. *Q. J. Exper. Psychol.* 62, 2141–2155. doi: 10.1080/17470210902765957

Herget, A.-K. (2019). On music's potential to convey meaning in film: a systematic review of empirical evidence. *Psychol. Music* 2019:030573561983501. doi: 10.1177/0305735619835019

Herget, A.-K., Schramm, H., and Breves, P. (2018). Development and testing of an instrument to determine musical Fit in audio-visual advertising. *Mus. Sci.* 22, 362–376. doi: 10.1177/1029864917731127

Hoeckner, B., Wyatt, E. W., Decety, J., and Nusbaum, H. (2011). Film music influences how viewers relate to movie characters. *Psychol. Aesthet. Creat. Arts* 5, 146–153. doi: 10.1037/a0021544

Huron, D. (1989). Music in advertising: an analytic paradigm. *Mus. Q.* 73, 557–574. doi: 10.1093/mq/73.4.557

iMotions. *iMotions Biometric Research Platform*. Available at: https://imotions.com/ (accessed October 1, 2020).

Israel, A., Lahav, E., and Ziv, N. (2019). Stop the music?: the effect of music on risky financial decisions: an experimental study. *SSRN Electron. J.* 42:151. doi: 10.2139/ssrn.3318151

Iwamiya, S. (1997). Interaction between auditory and visual processing in car audio: Simulation experiment using video reproduction. *Appl. Hum. Sci.* 16, 115–119. doi: 10.2114/jpa.16.115

Jones, E. E. (1972). *Attribution: Perceiving the Causes of Behavior*. Morristown, NJ: General Learning Press.

Juslin, P. N., and Laukka, P. (2004). Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. *J. New Music Res.* 33, 217–238. doi: 10.1080/0929821042000317813

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Princeton University.

Kallinen, K. (2002). Reading news from a pocket computer in a distracting environment: effects of the tempo of background music. *Comput. Hum. Behav.* 18, 537–551. doi: 10.1016/S0747-5632(02)00005-5

Kämpfe, J., Sedlmeier, P., and Renkewitz, F. (2011). The impact of background music on adult listeners: a meta-analysis. *Psychol. Music* 39, 424–448. doi: 10.1177/0305735610376261

Karim, J., Weisz, R., and Rehman, S. U. (2011). International positive and negative affect schedule short-form (I-PANAS-SF): testing for factorial invariance across cultures. *Proc. Soc. Behav. Sci.* 15, 2016–2022. doi: 10.1016/j.sbspro.2011.04.046

Killmeier, M., and Christiansen, P. (2004). Wolves at the door: musical persuasion in a 2004 Bush-Cheney campaign advertisement. *MedieKultur J. Med. Commun. Res.* 27:21. doi: 10.7146/mediekultur.v27i50.2857

Kosch, T., Hassib, M., Woźniak, P. W., Buschek, D., and Alt, F. (2018). "Your eyes tell: Leveraging smooth pursuit for assessing cognitive workload," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Amsterdam.

Kret, M. E., and Sjak-Shie, E. E. (2019). Preprocessing pupil size data: guidelines and code. *Behav. Res. Methods* 51, 1336–1342. doi: 10.3758/s13428-018-1075-y

Kruger, J.-L. (2012). Making meaning in AVT: eye tracking and viewer construction of narrative. *Perspectives* 20, 67–86. doi: 10.1080/0907676X.2011.632688

Lang, F. R., John, D., Lüdtke, O., Schupp, J., and Wagner, G. G. (2011). Short assessment of the big five: robust across survey methods except telephone interviewing. *Behav. Res. Methods* 43, 548–567. doi: 10.3758/s13428-011-0066-z

Lehmann, J. A., and Seufert, T. (2017). The influence of background music on learning in the light of different theoretical perspectives and the role of working memory capacity. *Front. Psychol.* 8:1902. doi: 10.3389/fncom.2016.01902

Lemercier, A., Guillot, G., Courcoux, P., Garrel, C., Baccino, T., and Schlich, P. (2014). "Pupillometry of taste: Methodological guide-from acquisition to data processing-and toolbox for MATLAB," in *Quantitative Methods for Psychology*, Vol. 10, (University of Ottawa; School of Psychology), 179–195.

Leone, G., D'Errico, F., Serino, C., and Marzano, M. (2008). *Empatia e Costi Psicologici Dell'aiuto', La Mente Del Cuore, Le Emozioni Nel Lavoro Nella Scuola Nella Vita*. Roma: Armando.

Lesiuk, T. (2005). The effect of music listening on work performance. *Psychol. Music* 33, 173–191. doi: 10.1177/0305735605050650

Marshall, S. K., and Cohen, A. J. (1988). Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Percept.* 6, 95–112. doi: 10.2307/40285417

Mathôt, S. (2018). Pupillometry: psychology, physiology, and function. *J. Cogn.* 1:16. doi: 10.5334/joc.18

Mentzoni, R. A., Laberg, J. C., Brunborg, G. S., Molde, H., and Pallesen, S. (2014). Type of musical soundtrack affects behavior in gambling. *J. Behav. Addict.* 3, 102–106. doi: 10.1556/JBA.3.2014.006

Mera, M., and Stumpf, S. (2014). Eye-tracking film music. *Music Mov. Image* 7:3. doi: 10.5406/musimoviimag.7.3.0003

Michon, R., and Chebat, J.-C. (2004). The interaction effect of background music and ambient scent on the perception of service quality. *J. Bus. Res.* 34, 191–196.

Nelson, J. S., and Boynton, G. R. (1997). *Video Rhetorics: Televised Advertising in American Politics*, Vol. 1. Champaign, IL: University of Illinois Press.

Neumann, D. L., Chan, R. C. K., Boyle, G. J., Wang, Y., and Rae, W. H. (2015). "Measures of Empathy," in *Measures of Personality and Social Psychological Constructs*, eds Y. Wang and M. Davidson (Amsterdam: Elsevier), 257–289. doi: 10.1016/B978-0-12-386915-9.00010-3

Nguyen, T., and Grahn, J. A. (2017). Mind your music: the effects of music-induced mood and arousal across different memory tasks. *Psychomusicology* 27, 81–94. doi: 10.1037/pmu0000178

North, A. C. (2004). Uses of music in everyday life. *Music Percept.* 22, 41–77. doi: 10.1525/mp.2004.22.1.41

Noseworthy, T. J., and Finlay, K. (2009). A comparison of ambient casino sound and music: Effects on dissociation and on perceptions of elapsed time while playing slot machines. *J. Gambl. Stud.* 25, 331–342. doi: 10.1007/s10899-009-9136-x

Oakes, S. (2007). Evaluating empirical research into music in advertising: a congruity perspective. *J. Advert. Res.* 47, 38–50. doi: 10.2501/s0021849907070055

Pisciottano, F. (2019). *Effetti Della Musica Sull'interpretazione di Stimoli Visivi*. Roma: Roma Tre University.

Quinto, L., Thompson, W. F., and Taylor, A. (2014). The contributions of compositional structure and performance expression to the communication of emotion in music. *Psychol. Music* 42, 503–524. doi: 10.1177/0305735613482023

Rogers, A., and Gibson, I. (2012). "Audiovisual coalescence: the effect of synchrony in film score," in *Proceedings of the 9th International Symposium on Computer Modeling and Retrieval*, London.

Salvucci, D. D., Markley, D., Zuber, M., and Brumby, D. P. (2007). "iPod distraction: effects of portable music-player use on driver performance," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, CA.

Scherer, K. R., Sundberg, J., Tamarit, L., and Salomão, G. L. (2015). Comparing the acoustic expression of emotion in the speaking and the singing voice. *Comput. Speech Lang.* 29, 218–235. doi: 10.1016/j.csl.2013.10.002

Shevy, M. (2007). The mood of rock music affects evaluation of video elements differing in valence and dominance. *Psychomusicology* 19, 57–78. doi: 10.1037/h0094034

Shevy, M., and Hung, K. (2013). "Music in television advertising and other persuasive media," in *The Psychology of Music in Multimedia*, eds S. L. Tan, A. J. Cohen, S. Lipscomb, and R. A. Kendall (Oxford: Oxford University Press), 311–334.

Sirois, S., and Brisson, J. (2014). Pupillometry. *Wiley Interdiscipl. Rev.* 5, 679–692. doi: 10.1002/wcs.1323

Smith, T. J. (2013). "Watching you watch movies: using eye tracking to inform cognitive film theory," in *Psychocinematics: Exploring Cognition at the Movies*, ed. A. P. Shimamura (New York, NY: Oxford University Press), 165–191. doi: 10.1093/acprof:oso/9780199862139.003.0009

Smith, T. J. (2014). "Audiovisual correspondences in sergei eisenstein's alexander nevsky: a case study in viewer attention," in *Cognitive Media Theory*, eds P. Taberham and T. Nanicelli (New York, NY: Taylor & Francis Group), 85–106.

Steffens, J. (2020). The influence of film music on moral judgments of movie scenes and felt emotions. *Psychol. Music* 48, 3–17. doi: 10.1177/0305735618779443

Tagg, P. (2006). Music, moving images, semiotics, and the democratic right to know. *Music Manipul. Soc. Uses Soc. Contr. Music* 21, 163–186.

Tagg, P. (2013). *Music's Meanings: A Modern Musicology for Non-Musos*. New York, NY: Mass Media Music Scholar's Press.

Tan, S.-L., Spackman, M. P., and Bezdek, M. A. (2007). Viewers' interpretations of film characters' emotions: effects of presenting film music before or after a character is shown. *Music Percept.* 25, 135–152. doi: 10.1525/mp.2007.25.2.135

Taylor, A. C., and Dewhurst, S. A. (2017). Investigating the influence of music training on verbal memory. *Psychol. Music* 45, 814–820. doi: 10.1177/0305735617690246

Västfjäll, D. (2001). Emotion induction through music: a review of the musical mood induction procedure. *Mus. Sci.* 5, 173–211. doi: 10.1177/10298649020050S107

Vitouch, O. (2001). When your ear sets the stage: musical context effects in film perception. *Psychol. Music* 29, 70–83. doi: 10.1177/0305735601291005

Wallengren, A.-K., and Strukelj, A. (2015). Film music and visual attention: a pilot experiment using eye-tracking. *Music Mov. Image* 8, 69–80. doi: 10.5406/musimoviimag.8.2.0069

Wallengren, A.-K., and Strukelj, A. (2018). "Into the film with music: measuring eyeblinks to explore the role of film music in emotional arousal and narrative transportation," in *Seeing into Screens: Eye Tracking and the Moving Image*, eds T. Dwyer, C. Perkins, S. Redmond, and J. Sita (Bloomsbury: Bloomsbury Academic), 65–84.

Wedel, M., and Pieters, R. (2017). "A review of eye-tracking research in marketing," in *Review of Marketing Research*, ed. N. K. Malhotra (London: Routledge), 123–147. doi: 10.4324/9781351550932-5

Wöllner, C., and Hammerschmidt, D. (2018). The impact of music and stretched time on pupillary responses and eye movements in slow-motion film scenes. *J. Eye Mov. Res.* 11:jemr.11.2.10. doi: 10.16910/jemr.11.2.10

Woloszyn, M. R., and Ewert, L. (2012). Memory for facial expression is influenced by the background music playing during study. *Adv. Cogn. Psychol.* 8:226. doi: 10.5709/acp-0118-9

Wu, B., Horner, A., and Lee, C. (2014). The correspondence of music emotion and timbre in sustained musical instrument sounds. *J. Audio Eng. Soc.* 62, 663–675. doi: 10.17743/jaes.2014.0037

Yamasaki, T., Yamada, K., and Laukka, P. (2015). Viewing the world through the prism of music: Effects of music on perceptions of the environment. *Psychol. Music* 43, 61–74. doi: 10.1177/0305735613493954

Zabalbeascoa, P. (2008). The nature of the audiovisual text and its parameters. *Didact. Audiovis. Transl.* 7, 21–37. doi: 10.1075/btl.77.05zab

Ziv, N. (2015). Music and compliance: can good music make us do bad things? *Psychol. Music* 14:55.

Ziv, N., and Goshen, M. (2006). The effect of 'sad' and 'happy' background music on the interpretation of a story in 5 to 6-year-old children. *Br. J. Music Educ.* 23:303. doi: 10.1017/S0265051706007078

Ziv, N., Hoftman, M., and Geyer, M. (2012). Music and moral judgment: the effect of background music on the evaluation of ads promoting unethical behavior. *Psychol. Music* 40, 738–760. doi: 10.1177/0305735611406579

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 08 October 2020
doi: 10.3389/fnins.2020.554731

# Neural Dynamics of Improved Bimodal Attention and Working Memory in Musically Trained Children

*Leonie Kausel[1,2]*, Francisco Zamorano[2,3,4], Pablo Billeke[2], Mary E. Sutherland[1], Josefina Larrain-Valenzuela[2], Ximena Stecher[3,4,5], Gottfried Schlaug[6] and Francisco Aboitiz[1]*

[1] Centro Interdisciplinario de Neurociencias, Pontificia Universidad Católica de Chile, Santiago, Chile, [2] Laboratorio de Neurociencia Social y Neuromodulación, Centro de Investigación en Complejidad Social (CICS), Universidad del Desarrollo, Santiago, Chile, [3] Unidad de Imágenes Cuantitativas Avanzadas, Clínica Alemana, Universidad del Desarrollo, Santiago, Chile, [4] Departamento de Imágenes, Clínica Alemana, Universidad del Desarrollo, Santiago, Chile, [5] Neuroradiology, Radiology Department, Clinica Alemana de Santiago, Santiago, Chile, [6] Department of Neurology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, United States

OPEN ACCESS

**Edited by:**
*Elvira Brattico,*
*Aarhus University, Denmark*

**Reviewed by:**
*Peter Schneider,*
*Heidelberg University, Germany*
*Jochen Kaiser,*
*Goethe University Frankfurt, Germany*

***Correspondence:**
*Leonie Kausel*
*lmkausel@uc.cl;*
*leoniekausel@gmail.com*

**Specialty section:**
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

**Received:** *22 April 2020*
**Accepted:** *04 September 2020*
**Published:** *08 October 2020*

**Citation:**
*Kausel L, Zamorano F, Billeke P, Sutherland ME, Larrain-Valenzuela J, Stecher X, Schlaug G and Aboitiz F (2020) Neural Dynamics of Improved Bimodal Attention and Working Memory in Musically Trained Children. Front. Neurosci. 14:554731. doi: 10.3389/fnins.2020.554731*

Attention and working memory (WM) are core components of executive functions, and they can be enhanced by training. One activity that has shown to improve executive functions is musical training, but the brain networks underlying these improvements are not well known. We aimed to identify, using functional MRI (fMRI), these networks in children who regularly learn and play a musical instrument. Girls and boys aged 10–13 with and without musical training completed an attention and WM task while their brain activity was measured with fMRI. Participants were presented with a pair of bimodal stimuli (auditory and visual) and were asked to pay attention only to the auditory, only to the visual, or to both at the same time. The stimuli were afterward tested with a memory task in order to confirm attention allocation. Both groups had higher accuracy on items that they were instructed to attend, but musicians had an overall better performance on both memory tasks across attention conditions. In line with this, musicians showed higher activation than controls in cognitive control regions such as the fronto-parietal control network during all encoding phases. In addition, facilitated encoding of auditory stimuli in musicians was positively correlated with years of training and higher activity in the left inferior frontal gyrus and the left supramarginal gyrus, structures that support the phonological loop. Taken together, our results elucidate the neural dynamics that underlie improved bimodal attention and WM of musically trained children and contribute new knowledge to this model of brain plasticity.

**Keywords: attention, working memory, fronto-parietal control network, phonological loop, musical training**

## INTRODUCTION

Executive functions, which include goal-directed attention and working-memory capacity, allow us to regulate, control, and manage our thoughts, emotions, and decision making (Aboitiz and Cosmelli, 2009). Attention allows us to select the stimuli that are relevant for us at each moment, and working memory (WM) allows us to keep the information in an accessible state for a short time. Attention and WM are closely related, because paying attention to certain information makes it easier to remember (Chun and Turk-Browne, 2007; Fougnie, 2008). Switching attention from one task to another also requires cognitive flexibility, an ability that helps to adjust one's behavior according to a changing environment (Monsell, 2003; Armbruster et al., 2012), and which is a core component of executive functions (Diamond, 2013). Greater cognitive flexibility is associated with favorable outcomes throughout lifespan, such as higher resilience, improved reading abilities in childhood, higher creativity, and a better quality of life (Dajani and Uddin, 2015). These skills, used every day to interact with our world (Hinton et al., 2012), develop during childhood and adolescence and can be improved by training (Diamond, 2013).

One activity that has been proposed to improve executive functions is playing a musical instrument (Miendlarzewska and Trost, 2014). Playing a musical instrument is a very challenging activity that puts high demands on motor and multisensory skills and is usually begun at an early age. Musicians have to master independent motor control for each hand, listen to what they play, react to what they hear, and also pay attention to other players (when playing in an ensemble). Score reading, which implies transforming visual symbols into auditory patterns by means of playing the instrument, is also part of most musical trainings. It has been shown that musical training produces structural and functional changes in the brain. As such, it has been proposed as a model for the study of brain plasticity (Schlaug, 2015).

Research has shown that adult musicians outperform their untrained peers on tasks assessing cognitive flexibility, WM, and verbal fluency (Zuk et al., 2014). Increased WM capacity has also been found in musically trained children and adolescents, with improved visuo-spatial and verbal WM and improved processing speed and reasoning (Bergman Nutley et al., 2014). In this latter study, researchers also found that changes in WM were proportional to the weekly hours spent on music practice. Still, results that show that better performance of musically trained children on visual WM tasks has not always been consistent (Talamini et al., 2017). Ho et al. (2003) found that children with musical training had better verbal memory, but not visual memory. Importantly, the improvements in verbal memory were maintained in those students who had begun or continued musical training after a year. Also, there is some behavioral evidence from young adults that musical training enhances task switching and dual task performance (e.g., Moradzadeh et al., 2015), which are tasks that require high performance of executive functions. Furthermore, it has been shown that musically trained young adults have higher efficiency of the executive attention network, which is involved in top-down attentional control (Medina and Barraza, 2019). It has also been shown that musical training has a positive impact on children who have auditory- and attention-related developmental disorders [attention deficit hyperactivity disorder (ADHD) or dyslexia], improving their neural efficiency of auditory cortex and promoting intrahemispheric synchronization (Seither-Preisler et al., 2014; Serrallach et al., 2016). Even though it has been shown that musically trained children have enhanced executive functions, the neural dynamics underlying these improvements are not well known.

Recent brain imaging studies have shown that musically trained children have higher activation of the bilateral supplementary motor area (SMA), the inferior frontal gyrus (IFG), the anterior cingulate cortex and the insula in a visual Stroop task (Sachs et al., 2017), and the pre-SMA/SMA and the right ventrolateral prefrontal cortex in a set-shifting task (Zuk et al., 2014). A study in young adult musicians showed that they had higher activation than a control group in cognitive control-related areas such as the bilateral posterior dorsal prefrontal cortex and the anterior cingulate gyrus when solving a musical-sound WM task (Pallesen et al., 2010). These studies suggest that it is plausible that musical training could influence the neural networks that underlie better performance of executive functions in musically trained children.

In order to better determine the neural dynamics involved in performance of musically trained children in executive function tasks, particularly in a bimodal context, our study sought to determine the neural correlates that underlie bimodal auditory/visual attention and WM in musically trained children. We hypothesized that playing a musical instrument improves these functions and that the neural networks underlying these skills would be boosted in children who regularly learn and play a musical instrument.

In the present study, we used functional magnetic resonance imaging (fMRI) to investigate the influence of musical training on the neural correlates that underlie bimodal attention and WM in musically trained children. To achieve our goal, we adapted and implemented the bimodal attention task of Johnson and Zatorre (2006). Participants were presented with a simultaneous pair of bimodal stimuli (auditory and visual) and were asked to pay attention only to the auditory, only to the visual (selective attention), or to both at the same time (divided attention). Both stimuli were afterward tested with a memory task in order to confirm attention allocation. By combining behavioral measures and brain activity recordings, we were able to determine the neural dynamics underlying the improved performance of musically trained children on our task.

## PARTICIPANTS AND METHODS

### Participants

Forty healthy, right-handed, Spanish-speaking children aged 10–13, with normal hearing and normal or corrected-to-normal vision, participated in our study. Written informed consent was obtained from all children and their parents for a protocol approved by the ethics committee of the Pontificia Universidad Católica de Chile. Participants completed the Wechsler Intellectual Scale for Children (WISC III) (Wechsler, 1991) validated for Chilean population (Ramírez and Rosas, 2007) and answered the Spanish version of the standardized Montreal Music History Questionnaire (Coffey et al., 2011), which inquired about their personal experience in music listening and performing. In a second session, participants solved the bimodal selective and divided attention task while their brain activity was being measured with fMRI. Participants received monetary compensation for travel costs.

Twenty musically trained participants were recruited from different youth orchestras in Santiago, Chile. Inclusion criteria encompassed playing a melodic instrument, having at least 2 years of instrumental lessons, practicing at least 2 h/week, and regularly playing in an orchestra or an ensemble. Six children played wind instruments (three clarinets, one traverse flute, one horn, and one saxophone), and 14 played string instruments (12 violins, one viola, and one cello). Age of onset of musical training was 9.1 ± 1.6 years (range from 6 to 11), average musical training was 3.7 ± 1.3 years (range from 2 to 6 years), intensity of practice over the last year was 9.2 ± 5.3 h/week (range from 2 to 21), and all participants had studied music continuously since the onset of training. All children were trained based on more non-aural strategies and had individual or small group (two to three participants) instrumental lessons and also played in an orchestra, having rehearsals at least once a week over the last year. Twenty control children were recruited from public schools in Santiago and had no additional musical training than the one provided in school curricula. In contrast to musically trained children, control children all declared to be unable to read or write musical scores.

Importantly, groups were matched for gender, age, intelligence coefficient (WISCIII), and socioeconomic status (educational level of both parents) (**Table 1**). For parental education, the highest, successfully completed education level of the parents was re-coded into a measure reflecting level of education, ranging from 1 (incomplete middle school education) to 10 (complete PhD). The average of both parents was used (Liberatos et al., 1988). The guardian of one musically trained child did not provide father's education, and the guardian of one control child did not provide parental education.

Five participants were excluded because of excessive movement during scanning. Finally, 18 musically trained children (10 female, mean age = 12.2 ± 0.8 years) and 17 non-musically trained children (11 female, mean age = 12.2 ± 0.8 years) were included in the analysis (**Table 1**). **Table 2** shows the musical training details of the musically trained children who were included in the analysis.

**TABLE 1 |** General demographics of the study population.

| | Musically trained children | Control children | |
|---|---|---|---|
| **n** | 18 | 17 | |
| Females | 10 | 11 | |
| | **Mean ± *SD*** | **Mean ± *SD*** | ***t*-value (*p*-value)** |
| Age (years) | 12.2 ± 0.8 | 12.2 ± 0.8 | −0.08 (0.53) |
| IQ | 109.5 ± 10.3 | 105.9 ± 11.1 | 0.96 (0.17) |
| Parental education | 3.8 ± 1.7 | 4.1 ± 1.6 | −0.65 (0.74) |

*There were no significant differences between groups for age, IQ, and parental education. IQ, intelligence coefficient.*

**TABLE 2 |** Characteristics of musical training in musically trained children.

| **Musically trained children (*n* = 18)** | | |
|---|---|---|
| **Group characteristics** | **Mean ± *SD*** | **Range** |
| Age at onset of musical training (years) | 9.1 ± 1.6 | 6–11 |
| Intensity of practice over the last year (hours/week) | 9.2 ± 5.3 | 2–21 |
| Duration of musical training (years) | 3.7 ± 1.3 | 2–6 |
| **Type of musical instrument** | **Number of children** | |
| Strings | 13 | |
| Woodwinds | 3 | |
| Brass | 2 | |

## Experimental Paradigm and Stimuli

### Experimental Task

The bimodal (auditory/visual) attention task that was used was adapted from Johnson and Zatorre (2006). In particular, we adapted the length of the stimuli by making them shorter (4 s) and adding the memory retrieval task after each stimulus pair. Participants solved this task while their brain activity was measured with fMRI.

### Stimuli

Auditory (melodies) and visual (figures) stimuli were 4 s long. We included a defined feature to the stimuli—a chord in the melody and a line of a different color in the figure—in order to help children to direct their attention to only one modality during the selective attention conditions. They were asked to report the chord by button press during the encoding phase of the auditory selective attention condition (ASA) and the red line during the visual selective attention condition (VSA). Melodies were in major tonalities and comprised pitches drawn from the Western musical scale centered around the mid-range of the piano from F3 (175 Hz) to G6 (784 Hz), with quarter and eighth notes. They were all in wav format and were presented in a piano timbre. All melodies contained one chord, which had to be reported by button press during the ASA. The melodies were presented binaurally at a comfortable listening level for each subject through MR-compatible sound transmission headphones (Resonance Technology Inc.)[1]. Figures consisted of equally long nine black lines and one red line, which had to be reported by button press during VSA. In order to "draw" each figure on a white background, individual shapes had the same starting point and new lines were presented sequentially aligned either horizontally or vertically every 300 ms. An abstract shape formed by 10 consecutively incorporated lines was completed after 3,000 ms and remained in view for 1,000 ms. The MRI head coil had a mirror attached, so that participants could see the screen where visual stimuli were displayed. A total of 160 melodies and figures were created. When presented simultaneously the auditory and visual stimuli, started and stopped at exactly the same time, but the individual elements of the two stimuli

[1] http://www.mrivideo.com/

never synchronized. Stimuli were presented using Presentation Software (Neurobehavioral Systems).

### Procedure

Each trial of our task had two parts, the encoding phase and the memory retrieval tasks (**Figure 1A**). The encoding phase started with an instruction to pay attention to either (or both) the melody or figure and then presented a pair of stimuli, which included an evolving abstract figure (visual) and a melody (auditory). Both stimuli lasted 4 s and started and stopped at exactly the same time, but the individual elements of the two stimuli never synchronized. Where attention was directed to was given by attention instruction (**Figure 1C**) and defined the auditory selective (ASA), visual selective (VSA), and divided attention (DA) conditions, respectively. We also included one condition in which children were instructed to passively observe the stimuli. This was the passive condition (P), and these trials did not include the memory retrieval tasks. The same/different memory retrieval tasks for both the auditory and visual stimuli followed each active attention encoding phase with a delay of 1,600 ms. These tasks allowed us to evaluate attention allocation (**Figure 1B**). Participants had 2.5 s to respond and had to report their answer via button press. Children did not receive any specific instruction on how to press the button (e.g., "respond as fast as possible"). During training outside of the scanner, we explained to the children that since we were interested in studying attention allocation, the most important thing during the experiment was that they followed the attention instruction of each attention condition. Better performance on the memory task for the attended stimuli was expected. Accuracy (correct responses) and reaction time of correct responses on the retrieval memory tasks were our behavioral outcome measures.

The trials in the same attention condition were presented as a block (**Figure 1D**). The order of attention condition blocks was randomized across participants. All conditions included unique stimuli, and stimulus pairs presented during encoding phases were defined randomly for each subject. The task had a duration of 16.05 min and included 36 trials, nine trials for each attention condition. All participants completed the task while being scanned in the MRI machine.

## Data Acquisition

Images were acquired at the Radiology Department of the Clínica Alemana de Santiago with a 3T Siemens Skyra scanner and a 20-channel head coil. Participants were prepared for the MRI and were instructed to relax and keep still during image acquisition. For each subject, a 3D structural T1-weighted scan [voxel size, $1 \times 1 \times 1$ mm; slices per slab, 176; field of view (FoV), 256 mm; repetition time (TR) = 2.53 s; echo time (TE) = 2.19 ms], phase and magnitude field maps (voxel size, $2.7 \times 2.7 \times 2.3$ mm; slices, 72; FoV, 208; TR = 731 ms; TE1 = 4.92 ms; TE2 = 7.38 ms) and a functional T2*-weighted gradient echo planar imaging scan (voxel size, $3 \times 3 \times 3$ mm; slices, 38; FoV, 220; TR = 2.21 s; TE = 30 ms) were acquired.

During functional T2*-weighted gradient echo planar imaging, our bimodal attention task was presented using Presentation Software (Neurobehavioral Systems). Auditory stimuli were presented over MRI-compatible headphones (Resonance Technology Inc.)[2], and visual stimuli were presented on a screen located in the MRI room at the same viewing distance for all subjects. The coil had a mirror attached, so that participants could see the screen where visual stimuli were displayed. Answers were given via button press on a keypad.

## Data Analysis

### Behavioral Data

Behavioral data were studied using RStudio (R Version 3.1.2). Accuracy and reaction time for memory tasks were analyzed with a $2 \times 3 \times 2$ mixed analysis of variance (ANOVA) to compare the main effects and interactions of group (between-subject factor: musicians, controls), attention condition (within-subject factor: ASA, VSA, DA), and retrieval memory task [within-subject factor: visual memory task (VMT), auditory memory task (AMT)]. Whenever the assumption of sphericity was violated, the Greenhouse–Geisser correction for epsilon was applied. Interaction effects were further assessed with pairwise *t*-tests. Bonferroni correction for multiple comparisons was applied where necessary. Alpha level of 0.05 was used for all statistical tests.

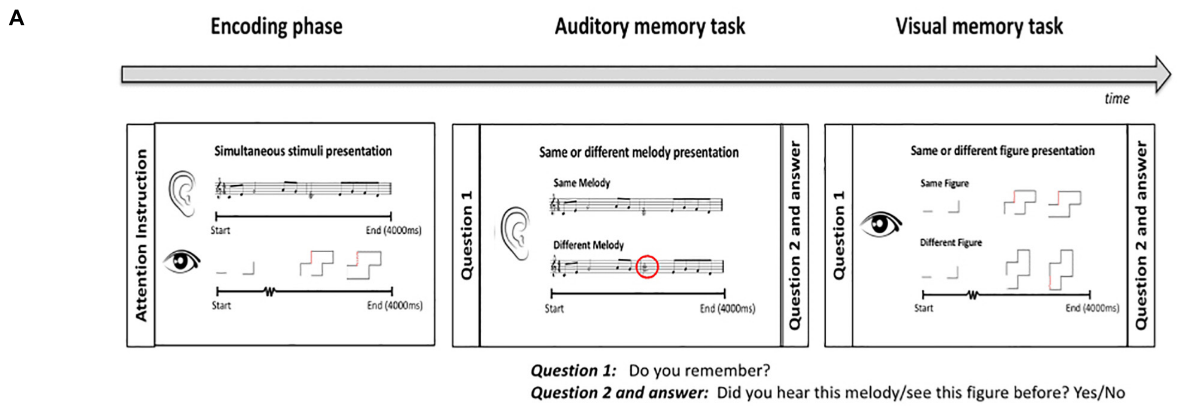
### Functional Magnetic Resonance Imaging

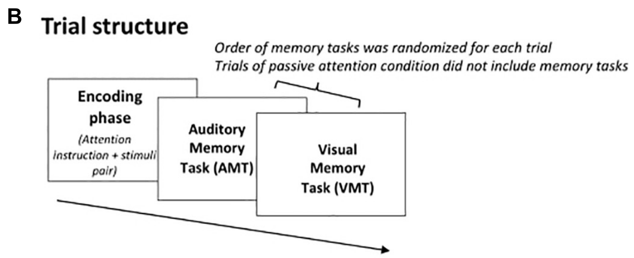
fMRI data were analyzed using FMRIB Software Library (FSL, version 5.0.10)[3] (Smith et al., 2004; Jenkinson et al., 2012). Data preprocessing involved the following steps: motion correction including field map unwarping (MCFLIRT), slice timing corrections, brain extraction (BET), spatial smoothing with a 6 mm full width at half maximum (FWHM) Gaussian kernel, and high-pass temporal filtering using Gaussian-weighted least-squares straight line fitting with sigma = 100.0 s, and pre-whitening. The blood oxygenation level-dependent (BOLD) response was modeled using a separate explanatory variable (EV) for the encoding phase of each attention condition (ASA, VSA, DA, and P). The design was convolved with a double gamma hemodynamic response function and temporal derivatives of each EV time course, and motion correction parameters were included as additional nuisance regressors. Estimated beta maps for contrasts were normalized to MNI152 standard space using linear transformations (FLIRT) in two stages. First, functional images were aligned with the subjects' high-resolution T1 using boundary-based registration (BBR). Then the T1 was registered to the standard Montreal Neurological Institute (MNI) atlas with a 12-degree-of-freedom affine transformation. Finally, these transformations were then applied to the functional data. Second-level activation maps were calculated with FSL using mixed-effect model (FLAME1 + 2). All reported results are based on an initial uncorrected voxel-level threshold of $z > 3.1$ and cluster inference using a familywise error-corrected threshold of $p < 0.05$, according to new MRI analysis guidelines (Eklund et al., 2016; Nichols et al., 2017).

In the first-level analysis, we modeled the encoding phases for each subject. The Hillyard principle (Hillyard et al., 1973)

[2] http://www.mrivideo.com/

[3] https://fsl.fmrib.ox.ac.uk/fsl




C

| Attention condition | Visually presented instruction | Instruction meaning |
|---|---|---|
| Auditory Selective Attention (ASA) | MELODY | Pay attention to the melody |
| Visual Selective Attention (VSA) | FIGURE | Pay attention to the figure |
| Divided Attention (DA) | MELODY AND FIGURE | Pay attention to the melody and the figure |
| Passive (P) | PASSIVE | Listen and look passively |

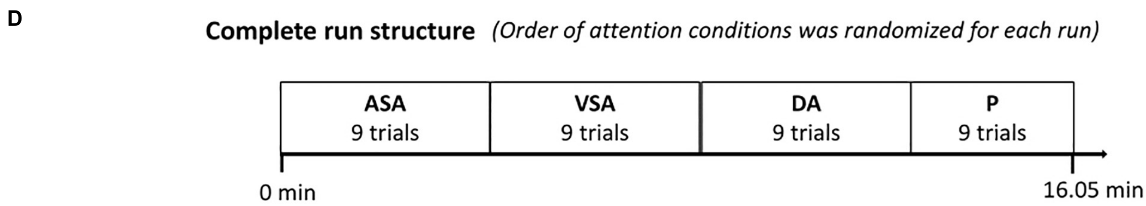



**FIGURE 1 |** Experimental protocol. **(A)** Scheme of the structure of the task and examples of the stimuli presented during the encoding phase and the retrieval memory tasks. **(B)** Trial structure. ASA, VSA, and DA included encoding phase and retrieval memory tasks. Order of memory tasks was randomized for each trial. P only included encoding phase. **(C)** The attention condition of each trial was given by the instruction presented at the beginning of the encoding phase. **(D)** Complete run structure. Order of attention conditions was randomized for each run. All participants completed one run of the task in the MRI scanner. ASA, auditory selective attention condition; VSA, visual selective attention condition; DA, divided attention condition; P, passive condition; VMT, visual memory task; AMT, auditory memory task.

states that in order to assess the effects of directed attention, responses should be compared with the same physical stimuli while holding overall arousal level and task demands constant, such that all that differs is the focus of directed attention. See **Figure 1C** for instructions given in each condition. In short, children were instructed to pay attention to the figure (VSA), the melody (ASA), both the figure and the melody (DA) or to listen and look passively at the presented stimuli (P). We modeled the attention component of the encoding phases of the active attention conditions by subtracting the passive condition from the encoding phase of the other three attention conditions, resulting in the contrasts [ASA > P], [VSA > P], and [DA > P]. Note that the arousal level and task demand could be different between the active and passive conditions, because the passive condition was not followed by memory tasks. Nevertheless, this design was chosen considering that there is a tradeoff between arousal level and focusing attention, following previous literature (Johnson and Zatorre, 2006).

Then we carried out three second-level-analysis models. In the first one, we explored for differences between groups in encoding phases. In the second one, we added a regressor of the accuracy in the AMT of VSA trials to the [VSA > P] contrast, in order to further investigate the three-way interaction effect that we found in the behavioral analysis. In this model, we also explored for differences between groups. In a third model, we added a regressor for the time of musical training (years) to the [ASA > P] contrast in the musician group, in order to disentangle if the results obtained with the anterior model would be explained by a facilitation in the encoding of the auditory stimulus. Finally, we performed a conjunction analysis to determine the overlaps between (1) the contrasts that showed differences between musicians and controls in the encoding phases determined with the first model and (2) the results determined with the second and third models.

Activation maps selected for figures were overlaid on a high-resolution brain image in MRIcroGL or FSLeyes for visualization. Activation locations were confirmed using the Harvard-Oxford Cortical Structure Atlas. Data are presented following the radiological convention (L, left; R, right), and coordinates are in MNI space.

# RESULTS

## Behavioral Results

Accuracy and reaction times of correct responses for both groups for each memory task across attention conditions are shown in **Figure 2** and **Table 3**, respectively. Results were analyzed with a 2 × 3 × 2 mixed ANOVA with group, attention condition, and retrieval memory task as factors.

Our behavioral results showed an interaction effect among attention condition and memory task [$F(3.1, 102.3) = 11.3$, $p[GG] = 0.0007$, $\eta^2 = 0.09$], which indicated that in both groups, attention condition significantly modulated the correct responses for memory tasks, with attended stimuli being better remembered than unattended ones. In other words, both groups remembered melodies better in auditory selective and divided attention conditions, whereas both groups remembered figures better in visual selective and divided attention conditions. This same modulation was found for adults in Johnson and Zatorre (2006), from where we adapted our task. Our results for correct responses to memory tasks also showed a significant main effect of group [$F(1, 33) = 6.1$, $p = 0.019$, $\eta^2 = 0.05$]. Overall, musically trained children had a better performance on memory tasks than control children independent of attention condition (Mus: mean = 7.09, *SD* = 1.69; Cont: mean = 6.28, *SD* = 2.3). Finally, the three-way interaction that we found between group, attention condition, and memory task [$F(3.1, 102.3) = 3.8$, $p[GG] = 0.047$, $\eta^2 = 0.03$] was given by the correct responses to the AMT in the VSA ($t = 3.0226$, $df = 28.005$, uncorrected $p = 0.00531$; Bonferroni-corrected $p = 0.032$). On average, musicians had 6.2 (*SD* = 1.6) correct responses as opposed to controls, who had an average of 4.1 (*SD* = 2.4) correct responses (**Figure 2**).

Behavioral results for reaction time of correct responses showed no significant main or interaction effects. There was no main effect of group [$F(1, 33) = 0.02$, $p = 0.88$] or interaction effects between group, attention condition, and memory task [$F(2, 66) = 0.3$, $p = 0.74$]. Overall mean reaction time was 809 ms (*SD* = 284 ms) (**Table 3**).

## Functional Magnetic Resonance Results

In order to determine the neural activity underlying attention during the encoding phases, we modeled our contrasts by

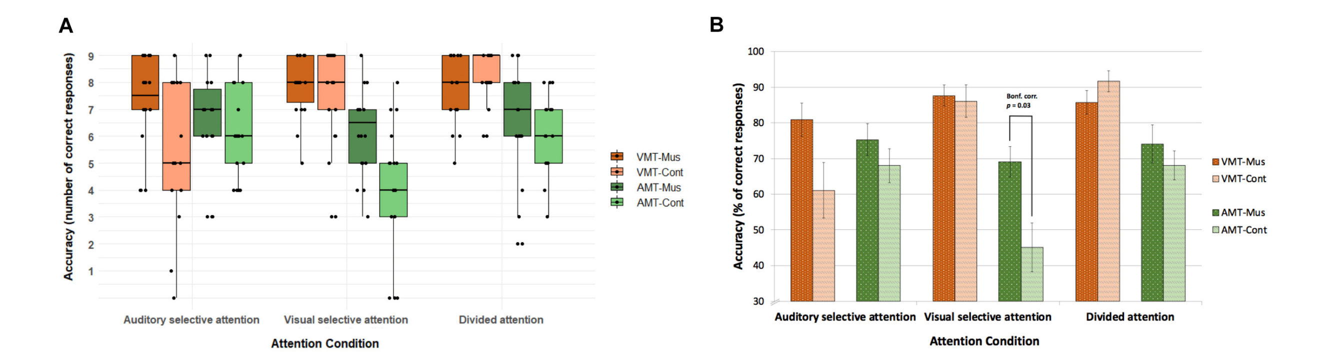



**FIGURE 2 |** Accuracy on the retrieval memory tasks in the three active attention conditions. **(A)** Accuracy with individual data points for number of correct responses. **(B)** Accuracy in percentage of correct responses. Overall musically trained children had more correct responses than control children [$F(1, 33) = 6.1$, $p = 0.019$). The three-way interaction effect between group, attention condition, and memory task [$F(3.1, 102.3) = 3.8$, $p[GG] = 0.047$) was given by the correct responses to the AMT in the visual selective attention condition trials (Bonf. corr. $p = 0.03$). Error bars in **(B)** indicate standard error of the mean. VMT, visual memory task; AMT, auditory memory task; Mus, musically trained children; Cont, control children.

**TABLE 3 |** Reaction times for correct responses of the retrieval memory tasks after each attention condition.

| Attention condition before MT | VMT | | AMT | |
|---|---|---|---|---|
| | Musicians RT ± *SEM* (ms) | Controls RT ± *SEM* (ms) | Musicians RT ± *SEM* (ms) | Controls RT ± *SEM* (ms) |
| ASA | 808.6 ± 55.9 | 830.2 ± 118.3 | 835.8 ± 74.1 | 856.6 ± 74.0 |
| VSA | 772.1 ± 49.5 | 839.9 ± 54.3 | 788.3 ± 66.8 | 750.5 ± 123.8 |
| DA | 810.8 ± 52.7 | 777.1 ± 61.4 | 755.7 ± 47.4 | 765.4 ± 67.7 |

*There were no significant differences between reaction times. ASA, auditory selective attention; VSA, visual selective attention; DA, divided attention; VMT, visual memory task; AMT, auditory memory task.*

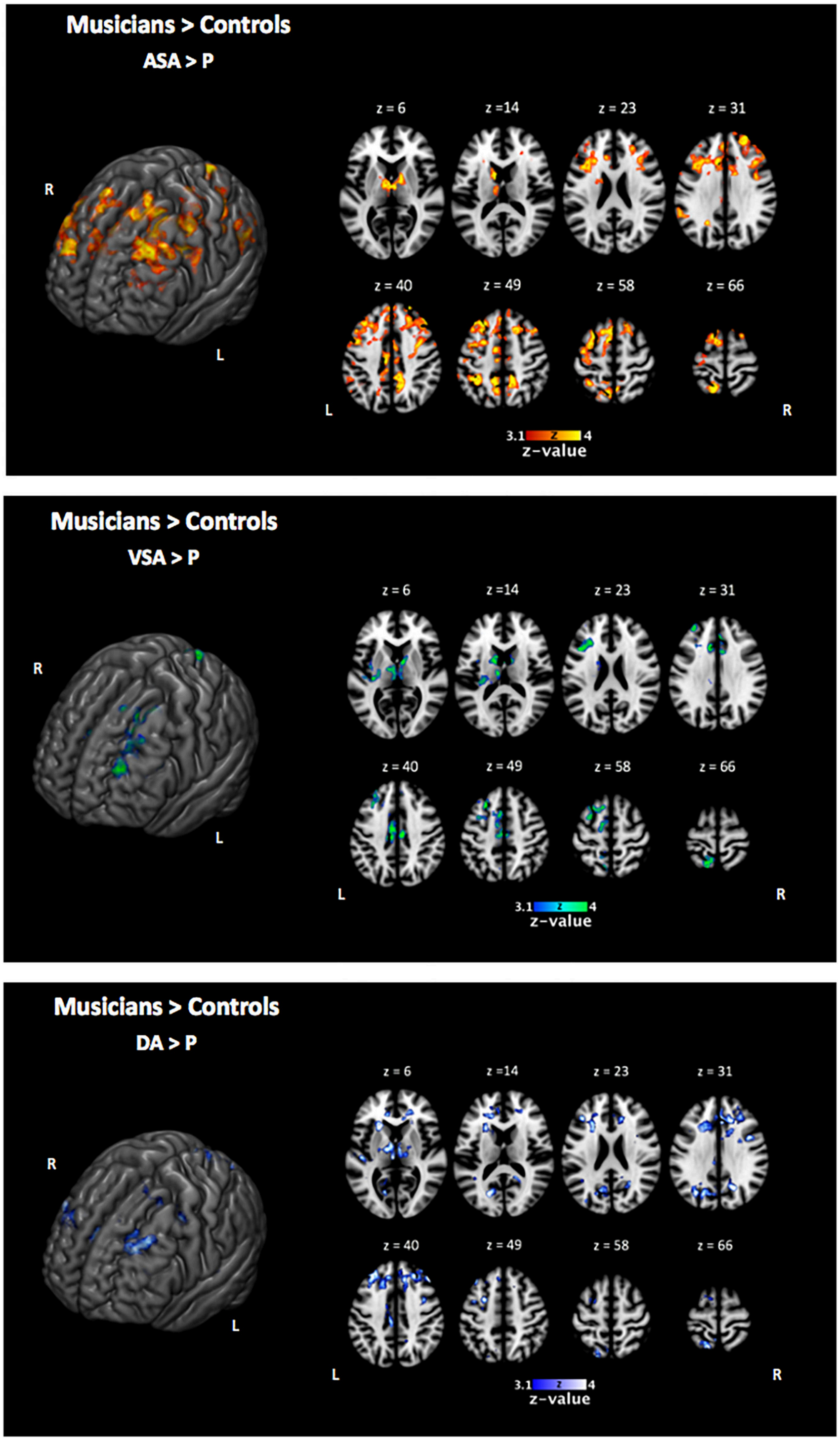



**FIGURE 3 |** Two-sample comparison of musically trained children over control children during encoding phase [ASA > P], [VSA > P], and [DA > P] contrasts (corrected $p < 0.05$). ASA, auditory selective attention condition; VSA, visual selective attention condition; DA, divided attention condition; P, passive condition.

subtracting the passive condition from the other three attention conditions. We expected that the main effect would be given by differences during the encoding phase, due to the role of attention on selecting the items that will be encoded in memory. The memory tasks were used to test the encoding process. We also explored if there were any differences during memory tasks, but we did not find any at our threshold levels (corrected $p < 0.05$).

**TABLE 4 |** Peaks of activity of group differences (Mus > Cont children) for the encoding phase contrasts [ASA > P], [VSA > P], and [DA > P].

| Area | x | y | z | Z-score | Cluster size | Corrected *p*-value |
|---|---|---|---|---|---|---|
| **Mus > Cont ASA > P** | | | | | | |
| Right precuneus | 12 | −42 | 47 | 5.19 | 83,684 | 2.32E–39 |
| Right cuneal cortex | 12 | −82 | 35 | 5.07 | 83,684 | 2.32E–39 |
| Left precentral gyrus | −41 | −7 | 55 | 4.47 | 83,684 | 2.32E–39 |
| Right cingulate gyrus, anterior division | 4 | −0.4 | 37 | 4.04 | 83,684 | 2.32E–39 |
| Left middle frontal gyrus/superior frontal gyrus | −26 | 24 | 48 | 3.9 | 83,684 | 2.32E–39 |
| Left superior frontal gyrus | −2 | 17 | 58 | 3.76 | 83,684 | 2.32E–39 |
| Right middle frontal gyrus/superior frontal gyrus | 21 | 24 | 48 | 3.6 | 83,684 | 2.32E–39 |
| Left superior parietal lobe | −34 | −47 | 59 | 3.5 | 83,684 | 2.32E–39 |
| Left cingulate gyrus, anterior division | −2 | 0.3 | 36 | 3.35 | 83,684 | 2.32E–39 |
| Left caudate | −10 | 5 | 14 | 5.08 | 5,731 | 3.34E–06 |
| Left thalamus | −3 | −7 | 5 | 5 | 5,731 | 3.34E–06 |
| Right thalamus | 4 | −11 | 4 | 4.28 | 5,731 | 3.34E–06 |
| Left supramarginal gyrus | −58 | −42 | 36 | 4.6 | 2,111 | 0.00579 |
| **Mus > Cont VSA > P** | | | | | | |
| Right cingulate gyrus, posterior division | 6 | −18 | 45 | 4.94 | 15,899 | 2.69E–13 |
| Left cingulate gyrus, posterior division | −7 | −20 | 45 | 4.85 | 15,899 | 2.69E–13 |
| Left paracingulate gyrus | −3 | 15 | 50 | 4.82 | 15,899 | 2.69E–13 |
| Left superior frontal gyrus | −24 | 19 | 54 | 4.03 | 15,899 | 2.69E–13 |
| Left middle frontal gyrus | −29 | 34 | 42 | 5.58 | 15,899 | 2.69E–13 |
| Left caudate | −7 | 3 | 14 | 4.79 | 8,859 | 7.63E–09 |
| Left thalamus | −7 | −11 | 2 | 4.02 | 8,859 | 7.63E–09 |
| Right thalamus | 7 | −12 | 2 | 3.48 | 8,859 | 7.63E–09 |
| Left posterior insular cortex/planum polare | −39 | −19 | −4 | 5.16 | 5,502 | 2.62E–06 |
| Left posterior insular cortex | −37 | −17 | −1 | 4.81 | 5,502 | 2.62E–06 |
| Left medial insular cortex | −36 | −5 | −1 | 4.38 | 5,502 | 2.62E–06 |
| Left putamen/pallidum | −25 | −14 | 4 | 4.35 | 5,502 | 2.62E–06 |
| Left cingulate gyrus, anterior division | −8 | 21 | 29 | 4.61 | 2,650 | 0.0011 |
| Left lateral occipital cortex, superior division | −12 | −60 | 67 | 4.53 | 2,405 | 0.00201 |
| Left superior parietal lobe | −14 | −55 | 69 | 4.39 | 2,405 | 0.00201 |
| **Mus > Cont DA > P** | | | | | | |
| Left middle frontal gyrus | −29 | 35 | 41 | 4.79 | 28,912 | 6.89E–19 |
| Left paracingulate gyrus | 2 | 35 | 33 | 4.77 | 28,912 | 6.89E–19 |
| Left putamen | −22 | 21 | 5 | 4.74 | 28,912 | 6.89E–19 |
| Right middle frontal gyrus | 41 | 32 | 38 | 3.67 | 28,912 | 6.89E–19 |
| Left cingulate gyrus, anterior division | −8 | 20 | 31 | 3.45 | 28,912 | 6.89E–19 |
| Cuneal cortex | 0 | −76 | 28 | 4.73 | 11,088 | 9.34E–10 |
| Left precuneus | −13 | −68 | 25 | 4.7 | 11,088 | 9.34E–10 |
| Left planum polare | −42 | −23 | −2 | 4.53 | 7,241 | 2.98E–07 |
| Left parahippocampal gyrus, posterior division | −31 | −24 | −22 | 4.46 | 7,241 | 2.98E–07 |
| Left thalamus | −14 | −11 | 2 | 4.42 | 7,241 | 2.98E–07 |
| Left pallidum | −26 | −18 | −1 | 4.42 | 7,241 | 2.98E–07 |
| Right thalamus | 12 | −19 | 7 | 3.61 | 7,241 | 2.98E–07 |
| Left lateral occipital cortex, superior division | −14 | −61 | 65 | 4.63 | 2,757 | 0.00139 |
| Left medial postcentral gyrus | −7 | −48 | 73 | 4.47 | 2,757 | 0.00139 |
| Left superior parietal lobe | −13 | −55 | 71 | 3.37 | 2,757 | 0.00139 |
| Left middle frontal gyrus/superior frontal gyrus | −25 | 4 | 52 | 4.4 | 2,055 | 0.00724 |
| Left cingulate gyrus, posterior division | −5 | −29 | 36 | 4.05 | 1,612 | 0.0225 |
| Brain stem | 5 | −29 | −13 | 4.51 | 1,388 | 0.0413 |

*Mus, musically trained children; Cont, control children; ASA, auditory selective attention condition; VSA, visual selective attention condition; DA, divided attention condition; P, passive condition.*

In line with our behavioral results that showed an overall better performance of musically trained children across attention conditions and memory tasks, whole-brain analyses of encoding phase contrasts ([ASA > P], [VSA > P], and [DA > P]) showed a significantly greater activation for musically trained children as compared with control children (corrected $p < 0.05$)

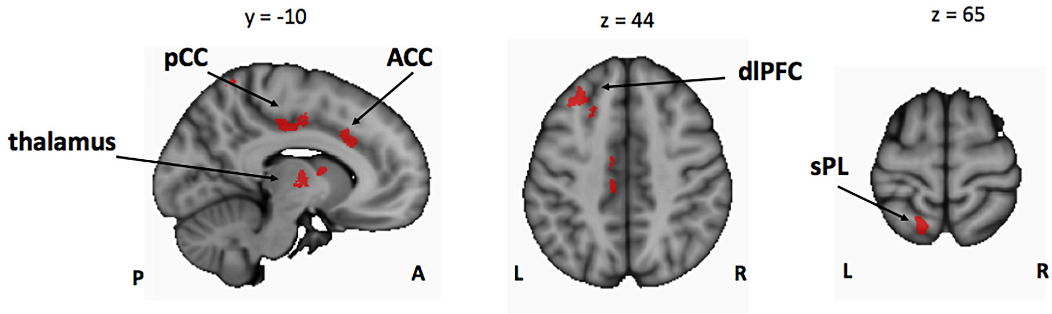



**FIGURE 4 |** Conjunction analysis between the Mus > Cont contrasts of the encoding phase [ASA > P], [VSA > P], and [DA > P] contrasts. Mus, musically trained children; Cont, control children; ASA, auditory selective attention; VSA, visual selective attention; DA, divided attention; P, passive; pCC, posterior cingulate cortex; ACC, anterior cingulate cortex; dlPFC, dorsolateral prefrontal cortex; sPL, superior parietal lobe.

**TABLE 5 |** Peaks of activity of (1) group differences (Mus > Cont children) for the correlation between correct responses on AMT and the contrasts [VSA > P] and (2) musicians for the correlation between years of training and the [ASA > P] contrast.

| Area | x | y | z | Z-score | Cluster size | Corrected *p*-value |
|---|---|---|---|---|---|---|
| **Mus > Cont for correlation between correct responses on AMT and VSA > P** | | | | | | |
| Right supplementary motor cortex | 4 | −9 | 55 | 5.18 | 5,567 | 1.19E–06 |
| Left cingulate gyrus, anterior division | 0 | −9 | 36 | 4.92 | 5,567 | 1.19E–06 |
| Left supramarginal gyrus | −55 | −39 | 51 | 5.22 | 3,228 | 0.000183 |
| Left frontal pole | −40 | 57 | 7 | 5.35 | 1,752 | 0.00849 |
| Inferior frontal gyrus, pars opercularis | −59 | 10 | 15 | 5.1 | 1,305 | 0.0327 |
| **Musicians for correlation between years of training and ASA > P** | | | | | | |
| Left middle frontal gyrus | −42 | 10 | 45 | 5.06 | 5,054 | 3.81E–06 |
| Left inferior frontal gyrus, pars triangularis | −57 | 26 | 14 | 4.65 | 5,054 | 3.81E–06 |
| Left middle frontal gyrus | −44 | 11 | 45 | 4.65 | 5,054 | 3.81E–06 |
| Left inferior frontal gyrus, pars triangularis/pars opercularis | −52 | 21 | 14 | 4.6 | 5,054 | 3.81E–06 |
| Right cingulate gyrus, posterior division | 8 | −28 | 37 | 5 | 3,216 | 0.000208 |
| Left precentral gyrus | −15 | −32 | 45 | 4.91 | 3,216 | 0.000208 |
| Left supramarginal gyrus/postcentral gyrus | −53 | −30 | 51 | 4.63 | 1,927 | 0.00556 |
| Right anterior insula | 38 | 13 | −7 | 4.75 | 1,310 | 0.0343 |
| Right frontal orbital cortex | 35 | 21 | −20 | 4.5 | 1,310 | 0.0343 |
| Right supramarginal gyrus/angular gyrus | 44 | −47 | 64 | 4.98 | 1,278 | 0.0379 |
| Right supramarginal gyrus | 52 | −36 | 52 | 4.33 | 1,278 | 0.0379 |

*Mus, musically trained children; Cont, control children; ASA, auditory selective attention condition; VSA, visual selective attention condition; DA, divided attention condition; P, passive condition; AMT, auditory memory task.*

in areas related to attentional control (**Figure 3** and **Table 4**). Musically trained children showed higher activation in regions including bilateral dorsolateral prefrontal cortex (dlPFC), medial premotor area, right dorsal precentral gyrus (pre-CG), left supramarginal gyrus (SMG), bilateral posterior division of the cingulate cortex (PCC), and bilateral thalamus for the [ASA > P] contrast; left dlPFC, left superior parietal lobe (sPL), bilateral PCC and bilateral thalamus for the [VSA > P] contrast; and bilateral dlPFC, left sPL, left anterior division of the cingulate cortex (ACC), left PCC, and bilateral thalamus for the [DA > P] contrast. The conjunction analysis for these three Musicians > Controls contrasts showed an overlap in the left dlPFC, the left sPL, the ACC, the PCC, and the thalamus (**Figure 4**). The opposite comparison of control children over musically trained children resulted in no activation at our threshold level (corrected $p < 0.05$).

Taken together, our results suggest that the overall better performance of the musically trained children in our bimodal attention task seems to be driven by higher activation of attention control related brain areas from the fronto-parietal control network (e.g., dlPFC, sPL, and ACC) during encoding phase in musically trained children as compared with controls in all active attention conditions (**Figure 4**).

In order to investigate the three-term interaction effect found in our behavioral analysis of accuracy, we correlated the activation in the encoding phase of the [VSA > P] contrast with the correct responses of each subject to the AMT in the VSA condition. When comparing the results among groups, we found a higher activation of the left SMG, the SMA, the ACC, the left superior IFG, and the left frontal pole in musically trained children as compared with controls (**Figure 5** and **Table 5**). This effect could be due to a general facilitation for the encoding of

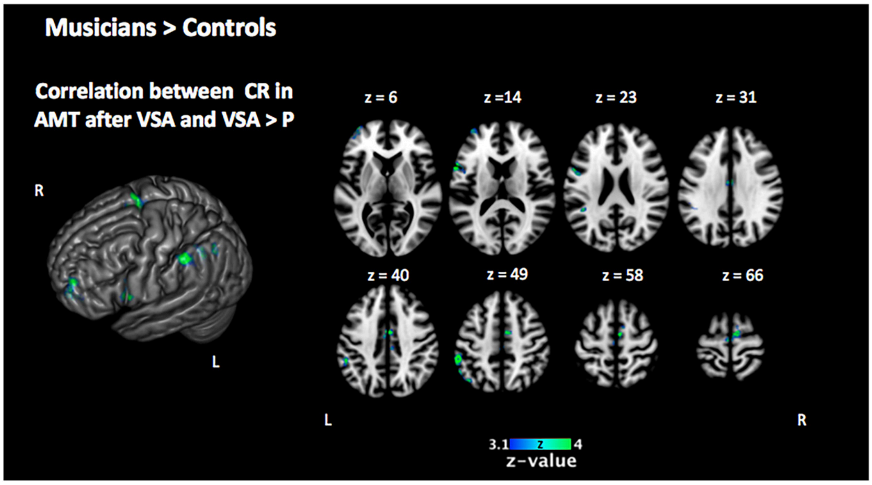



**FIGURE 5 |** Two-sample comparison of Mus > Cont for the correlation between the activation in the encoding phase [VSA > P] contrast with the correct responses of each subject in the auditory memory tasks of the VSA condition trials (corrected $p < 0.05$). Mus, musically trained children; Cont, control children; VSA, visual selective attention; P, passive; CR, correct responses; AMT, auditory memory task.

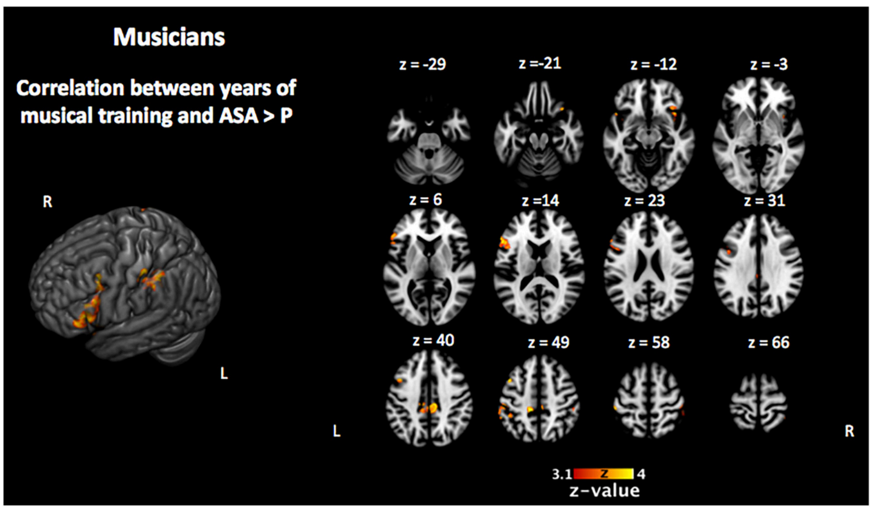



**FIGURE 6 |** Activation that correlated with the years of musical training in the musically trained group during the encoding phase of the [ASA > P] contrast (corrected $p < 0.05$). ASA, auditory selective attention; P, passive.

auditory stimuli in the musically trained children. If the latter was true, a similar modulation should be found when musically trained children pay attention to auditory stimuli (such as in our ASA condition), and this modulation should correlate with the years of musical training. In order to test this, we made a conjunction analysis between the previous result (**Figure 5**) and the activity that was found when correlating the years of musical training with the activity during the encoding phase of the [ASA > P] contrast in the musically trained group (**Figure 6** and **Table 5**). We found two points of overlap: the left IFG (specifically the pars opercularis) and the left SMG (**Figure 7**). To be more certain that these results are specific to auditory processing, we did also explore if there was a correlation between the [ASA > P] contrast and performance on the visual task in the ASA condition, and we did not find any significant activations at our threshold levels (corrected $p < 0.05$).

Taken together, our results suggest that musical training facilitates the encoding of auditory stimuli and that this facilitation relies on the left IFG and the left SMG in musically trained children.

## DISCUSSION

Our study investigated the neural dynamics that underlie the improved performance of children who play a musical instrument on a bimodal (auditory/visual) attention and WM task. We found that two mechanisms seem to contribute to

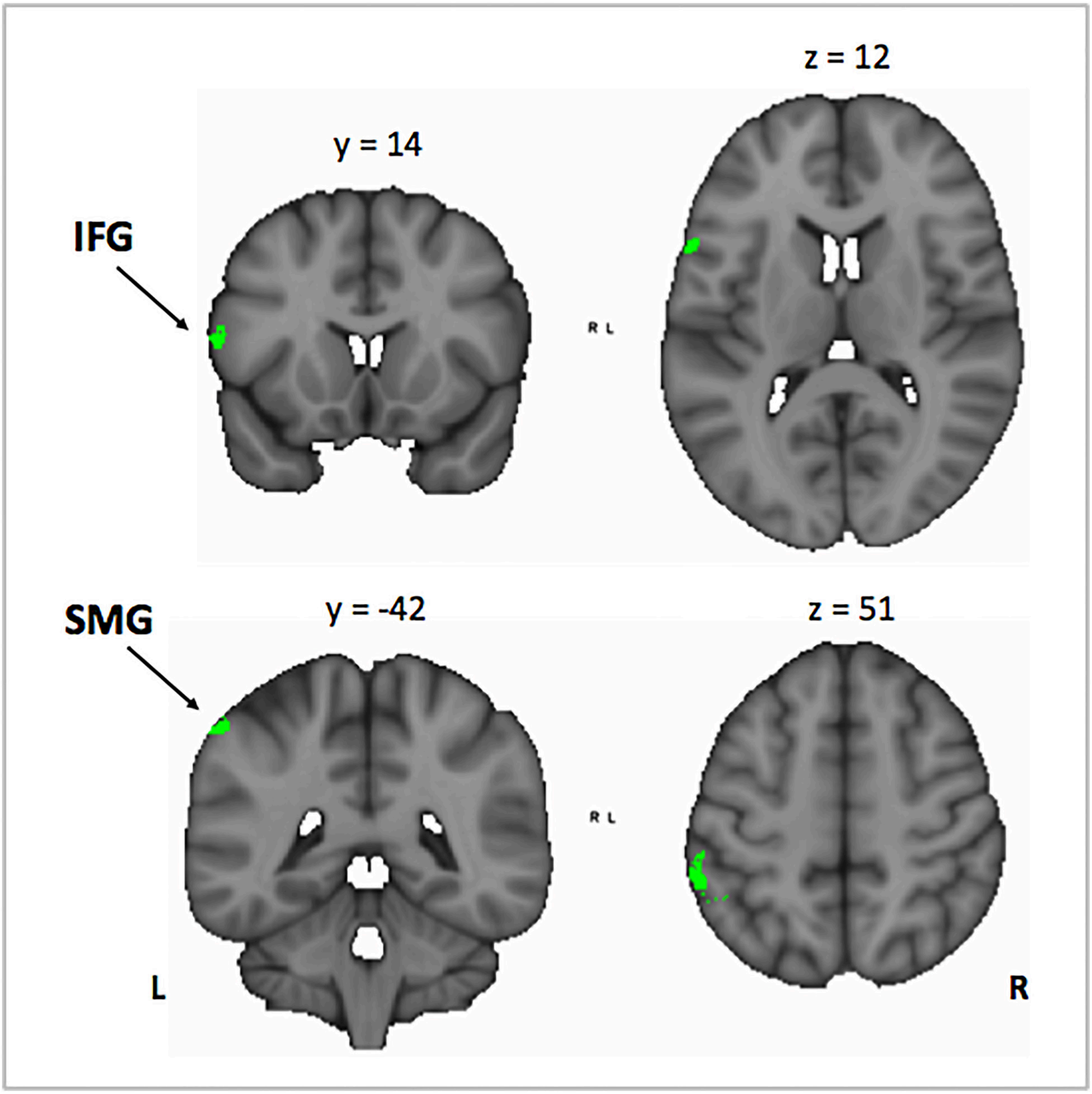



**FIGURE 7 |** Conjunction analysis between the Mus > Cont contrast of the correlation between the activation in the encoding phase [VSA > P] contrast with the correct responses of each subject in the auditory memory tasks of the VSA condition trials, and the correlation between the years of musical training in the musically trained group in the encoding phase [ASA > P] contrast. Mus, musically trained children; Cont, control children; VSA, visual selective attention; ASA, auditory selective attention; P, passive. IFG, inferior frontal gyrus; SMG, supramarginal gyrus.

this improvement. On the one hand, musically trained children show a higher activation of a more domain-general mechanism, the fronto-parietal control network during encoding phases of all attention conditions. On the other hand, they showed a higher activation of a more domain-specific mechanism of auditory encoding, which includes the left IFG and the left SMG, which are structures that support the phonological loop. These results contribute new knowledge that allows us to better understand how the developing brain is influenced by the achievement of this complex ability. Longitudinal studies on groups paired for general demographics before training suggest that musical training has a "nurture" effect on development and brain plasticity (Schlaug et al., 2005; Kraus et al., 2014; Putkinen et al., 2015; Habibi et al., 2017). Our cross-sectional design does not allow us to address whether there were differences in attention and WM prior to musical training. However, our groups were matched on gender, age, IQ, and socioeconomic status, which allowed us to evaluate the relation between musical training and attention and memory without these confounding factors. Also, our control group was a passive control group that did not engage in another type of activity that also requires self-control, concentration, and regular training. It is though important to mention that longitudinal studies that have included active control groups such as Moreno et al. (2009) and Habibi et al. (2018) have found that musical training has an impact on auditory processing skills and also executive functions such as inhibitory control.

Our behavioral results showed that musically trained children had an overall better performance on both memory retrieval tasks than had control children. These results are in line with other studies that have found that musicians perform better on both AMT and VMT (Rodrigues et al., 2013; Slevc et al., 2016; Talamini et al., 2016). Several studies have proposed that this improvement in visual attention and memory skills is due specifically to music reading and playing in an orchestra (Land and Furneaux, 1997; Rodrigues et al., 2007, 2013). Since all of our musically trained participants read music and played in an ensemble, it is plausible that these specific facets of their musical training may have contributed to the overall better performance of the musically trained children in both memory tasks.

In line with the above-presented behavioral results, our functional brain imaging results showed that musically trained children had higher activation in attentional control related brain areas (e.g., dlPFC and superior PL) (Corbetta and Shulman, 2002) in the encoding phase contrasts of all three active attention conditions (auditory selective [ASA > P], visual selective [VSA > P], and divided attention [DA > P]) than had control children. Musically trained children also showed significantly higher activation of the anterior cingulate cortex (ACC) and the thalamus in the encoding phase contrasts of all active attention conditions than did controls. The ACC is involved with monitoring demands for executive control (Mansouri et al., 2017), and its activity is also associated with the fronto-parietal control network (Power et al., 2011). The thalamus is important for sensory processing and integration (McCormick and Bal, 1994; Cappe et al., 2009), language processing (Crosson, 2013), and memory functions (Kopelman, 2015), and it participates in distributed cognitive control (Halassa and Kastner, 2017). A recent study also showed that the functional connectivity of the thalamocortical network is reorganized in musicians (Tanaka and Kirino, 2017). This latter study showed that auditory areas are more strongly connected with the left thalamus in musicians as compared with controls. Our results also expand on the results obtained by Pallesen et al. (2010) with young adult musicians, who showed that the cognitive control network was enhanced during auditory WM in musicians. Taken together, our results suggest that playing a musical instrument boosts cognitive control networks such as the fronto-parietal attention network and the thalamus during the encoding phase and that this subserves the improved memory capacity for auditory and visual stimuli shown by musically trained children in our task.

Another of our behavioral results for accuracy showed a three-way interaction between the factors group, attention condition, and memory task. This interaction was given by the group differences in the correct responses to the AMT after the VSA, with the musically trained children showing significantly better performance than control children. This result shows that even though participants were instructed to pay attention only to the visual stimuli in this condition, musically trained children were still able to encode and remember the auditory stimuli that were presented during the encoding phase far better than control children. When we correlated the brain activity in the encoding phase of VSA (contrast [VSA > P]) with the correct responses on the AMT for this attention condition, we found significant differences among groups, with musicians showing higher activation in the left SMG, among others.

We hypothesized that this behavioral effect could be due to a facilitated encoding of auditory stimuli in the musically trained group. We reasoned that if this was true, a similar modulation should be found when musically trained children pay attention to auditory stimuli (such as in our ASA) and that this modulation should correlate with the years of musical training. In order to test this, we checked if there was an overlap between the previous result and the activity that correlated with the years of musical training in the musically trained group during the encoding phase of the ASA condition ([ASA > P] contrast). We found two points of overlap: the left IFG (specifically the pars opercularis) and the left SMG.

Both these areas are multimodal association areas known to support the phonological loop. The phonological loop is part of the WM system involved in auditory processing; in particular, it is thought to be implicated in establishing auditory–motor connections (Baddeley, 2003; Schulze and Koelsch, 2012). Importantly, it has been shown that these areas are core structures involved in both tonal and verbal auditory WM (Koelsch et al., 2009). In fact, it has been observed that these areas are more activated in musicians than non-musicians during tonal WM tasks (Schulze et al., 2011).

In general, the IFG, specifically Broca's area, is known to be part of the language network and is involved in the perception and vocalizations of speech (Aboitiz, 2017). Notably, this area is also important for the recognition of musical auditory patterns (Chiang et al., 2018). It has also been suggested that the IFG contributes to memory formation (Tang et al., 2018). On the other hand, it has been shown that musicianship seems to have an impact on the structure of the left IFG. Abdul-Kareem et al. (2011) found that increased gray matter volume of left pars opercularis in male orchestral musicians correlated positively with years of musical performance. One could speculate then that our functional finding that subserved auditory memory encoding in our musically trained group could eventually lead to increased gray matter in this area.

On the other hand, the left SMG has been shown to participate in pitch memory (Gaab et al., 2003; Ellis et al., 2013). Notably, research using transcranial direct current stimulation (tDCS) (Vines et al., 2006; Schaal et al., 2017) and transcranial magnetic stimulation (TMS) (Schaal et al., 2015) have implied that the left SMG is causally involved with pitch memory processing. Another recent study that included cross-sectional and longitudinal data also showed that the left SMG is involved in music processing in musically trained children and adults (Ellis et al., 2013). Participants in Ellis et al. (2013) solved the same/different melodic and rhythmic discrimination task. Similar as in our results, they found that activation in the left SMG was related to cumulative hours of musical practice in both tasks for children and adults. Our results suggest that musical training facilitated the encoding of auditory stimuli in the musically trained children and that this facilitation relied on the left IFG and the left SMG. This results could also help to interpret the positive impact that musical interventions have on children with dyslexia (Habib et al., 2016), and this and the abovementioned results also support the overlap and attention conditions of the OPERA hypothesis proposed by Patel for the benefit of musical training on the neural encoding of speech (Patel, 2011).

Interestingly, the IFG and SMG, which we found to be involved specifically in auditory encoding in our musically trained group, have also been shown to be involved in visual stimuli processing. For example, Broca's area in the IFG has been shown to be activated to a greater extent by visually presented sentences when compared with spoken sentences (Carpentier et al., 2001), and the SMG has been causally involved in visual word recognition (Stoeckel et al., 2009). These results suggest that it is plausible that the increased functioning of these areas in musicians could eventually impact their visual processing. Effectively, it has also been found that Broca's area supports enhanced visuospatial cognition in professional orchestral musicians (Sluming et al., 2007).

Our results do not support the neural efficiency hypothesis, which states that subjects with better performance show lower brain activation than individuals with lower performance when working on the same cognitive tasks (Dunst et al., 2014). Evidence has suggested that this phenomenon also seems to be a function of the amount and quality of learning; this means that the specialization of functioning is reached over time (Neubauer and Fink, 2009). It is probable that in the case of this study, the children are still in the "training phase" of the functions, and that is why we see a higher functioning of the networks. This would have to be tested in other experiments.

Taken together, our results describing the neural dynamics the underlie the improved performance of musically trained children in our attention task suggest that musical training improves the allocation of attentional resources by increasing the functioning of the fronto-parietal control network and facilitating the encoding of auditory stimuli. This latter benefit is due to the years of training and depends on the function of left IFG and left SMG, structures that also support the phonological loop. Our results could be relevant for educational policies, and they also suggest that musical training could be used as a non-pharmacological intervention strategy for children with attentional problems in order to improve their overall functioning in daily life.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Medical Faculty of the Pontificia Universidad Católica de Chile. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

LK, FZ, MES, and FA designed the experiment. LK and MES created the stimuli. LK and PB programmed the experiment. LK and JL-V recruited the subjects. JL-V and XS evaluated the subjects. LK and FZ conducted the experiments. LK analyzed the data. LK, FZ, PB, GS, and FA discussed the results. All authors provided revisions and critical feedback on the final draft of the manuscript.

## FUNDING

This research was supported by CONICYT Chile (Regular FONDECYT 1160258 to FA, Regular FONDECYT 1181295 to PB, Regular FONDECYT 1190513 to FZ, Postdoc FONDECYT 3150235 to MES, and Postdoc FONDECYT 3190914 to LK).

## ACKNOWLEDGMENTS

We would like to thank all the participants and their families for making this research possible. We would also like to thank Natalie Julian for her help with the figure of the behavioral results and Gudrun Kausel and Sebastián Bascuñán Hughes for their critical revision of the first version of the manuscript.

## REFERENCES

Abdul-Kareem, I. A., Stancak, A., Parkes, L. M., and Sluming, V. (2011). Increased gray matter volume of left pars opercularis in male orchestral musicians correlate positively with years of musical performance. *J. Magn. Reson. Imaging* 33, 24–32. doi: 10.1002/jmri.22391

Aboitiz, F. (2017). *A Brain for Speech: A View From Evolutionary Neuroanatomy*. London: Palgrave Macmillan UK, doi: 10.1057/978-1-137-54060-7

Aboitiz, F., and Cosmelli, D. (2009). "From attention to goal-directed behavior," in *Neurodynamical, Methodological and Clinical Trends*, Vol. 53, eds F. Aboitiz and D. Cosmelli (Berlin: Springer Berlin Heidelberg).

Armbruster, D. J. N., Ueltzhöffer, K., Basten, U., and Fiebach, C. J. (2012). Prefrontal cortical mechanisms underlying individual differences in cognitive flexibility and stability. *J. Cogn. Neurosci.* 24, 2385–2399. doi: 10.1162/jocn_a_00286

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839. doi: 10.1038/nrn1201

Bergman Nutley, S., Darki, F., and Klingberg, T. (2014). Music practice is associated with development of working memory during childhood and adolescence. *Front. Hum. Neurosci.* 7:926. doi: 10.3389/fnhum.2013.00926

Cappe, C., Rouiller, E. M., and Barone, P. (2009). Multisensory anatomical pathways. *Hear. Res.* 258, 28–36. doi: 10.1016/j.heares.2009.04.017

Carpentier, A., Pugh, K. R., Westerveld, M., Studholme, C., Skrinjar, O., Thompson, J. L., et al. (2001). Functional MRI of language processing: depedence on input modality and temporal lobe epilepsy. *Epilepsia* 42, 1241–1254.

Chiang, J. N., Rosenberg, M. H., Bufford, C. A., Stephens, D., Lysy, A., and Monti, M. M. (2018). The language of music: common neural codes for structured sequences in music and natural language. *Brain Lang.* 185, 30–37. doi: 10.1016/j.bandl.2018.07.003

Chun, M. M., and Turk-Browne, N. B. (2007). Interactions between attention and memory. *Curr. Opin. Neurobiol.* 17, 177–184. doi: 10.1016/j.conb.2007.03.005

Coffey, E. B. J. H., Scala, S. C., and Zatorre, R. J. (2011). "Montreal music history questionnaire: a tool for the assessment of music-related experience in music cognition research," in *Paper Presented at the The Neurosciences and Music IV: Learning and Memory*, Edinburgh.

Corbetta, M., and Shulman, G. L. (2002). Control of Goal-directed and Stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 215–229. doi: 10.1038/nrn755

Crosson, B. (2013). Thalamic mechanisms in language: a reconsideration based on recent findings and concepts. *Brain Lang.* 126, 73–88. doi: 10.1016/j.bandl.2012.06.011

Dajani, D. R., and Uddin, L. Q. (2015). Demystifying cognitive flexibility: implications for clinical and developmental neuroscience. *Trends Neurosci.* 38, 571–578. doi: 10.1016/j.tins.2015.07.003

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

Dunst, B., Benedek, M., Jauk, E., Bergner, S., Koschutnig, K., Sommer, M., et al. (2014). Neural efficiency as a function of task demands. *Intelligence* 42, 22–30. doi: 10.1016/j.intell.2013.09.005

Eklund, A., Nichols, T. E., and Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natil. Acade. Sci. U.S.A.* 113, 7900–7905. doi: 10.1073/pnas.1602413113

Ellis, R. J., Bruijn, B., Norton, A. C., Winner, E., and Schlaug, G. (2013). Training-mediated leftward asymmetries during music processing: a cross-sectional and longitudinal fMRI analysis. *NeuroImage* 75, 97–107. doi: 10.1016/j.neuroimage.2013.02.045

Fougnie, D. (2008). "The relationship between attention and working memory," in *New Research on Short-Term Memory*, ed. N. B. Johansen (New York, NY: Nova Science Publishers, Inc), 1–45. doi: 10.3389/conf.fnhum.2011.207.00576

Gaab, N., Gaser, C., Zaehle, T., Jancke, L., and Schlaug, G. (2003). Functional anatomy of pitch memory—an fMRI study with sparse temporal sampling. *NeuroImage* 19, 1417–1426. doi: 10.1016/S1053-8119(03)00224-6

Habib, M., Lardy, C., Desiles, T., Commeiras, C., Chobert, J., and Besson, M. (2016). Music and dyslexia: a new musical training method to improve reading and related disorders. *Front. Psychol.* 7:26. doi: 10.3389/fpsyg.2016.00026

Habibi, A., Damasio, A., Ilari, B., Elliott Sachs, M., and Damasio, H. (2018). Music training and child development: a review of recent findings from a longitudinal study. *Ann. N. Y. Acad. Sci.* 1423, 73–81. doi: 10.1111/nyas.13606

Habibi, A., Damasio, A., Ilari, B., Veiga, R., Joshi, A. A., Leahy, R. M., et al. (2017). Childhood music training induces change in micro and macroscopic brain structure: results from a longitudinal study. *Cereb. Cortex* 28, 4336–4347. doi: 10.1093/cercor/bhx286

Halassa, M. M., and Kastner, S. (2017). Thalamic functions in distributed cognitive control. *Nat. Neurosci.* 20, 1669–1679. doi: 10.1038/s41593-017-0020-1

Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science* 182, 177–180. doi: 10.1126/science.182.4108.177

Hinton, C., Fischer, K. W., and Glennon, C. (2012). *Mind, Brain, and Education – How The Brain Learns. Mind, Brain and Education: The students at The Center Series*. Available online at: https://studentsatthecenterhub.org/

Ho, Y.-C., Cheung, M.-C., and Chan, A. S. (2003). Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology* 17, 439–450. doi: 10.1037/0894-4105.17.3.439

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *NeuroImage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Johnson, J. A., and Zatorre, R. J. (2006). Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. *NeuroImage* 31, 1673–1681. doi: 10.1016/j.neuroimage.2006.02.026

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., and Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an fMRI study. *Hum. Brain Mapp.* 30, 859–873. doi: 10.1002/hbm.20550

Kopelman, M. D. (2015). What does a comparison of the alcoholic Korsakoff syndrome and thalamic infarction tell us about thalamic amnesia? *Neurosci. Biobehav. Rev.* 54, 46–56. doi: 10.1016/j.neubiorev.2014.08.014

Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., et al. (2014). Music enrichment programs improve the neural encoding of speech in at-risk children. *J. Neurosci.* 34, 11913–11918. doi: 10.1523/JNEUROSCI.1881-14.2014

Land, M. F., and Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philos. Trans. R. Soc. B Biol. Sci.* 352, 1231–1239. doi: 10.1098/rstb.1997.0105

Liberatos, P., Link, B. G., and Kelsey, J. L. (1988). The measurement of social class in epidemiology. *Epidemiol. Rev.* 10, 87–121.

Mansouri, F. A., Egner, T., and Buckley, M. J. (2017). Monitoring demands for executive control: shared functions between human and nonhuman primates. *Trends Neurosci.* 40, 15–27. doi: 10.1016/j.tins.2016.11.001

McCormick, D. A., and Bal, T. (1994). Sensory gating mechanisms of the thalamus. *Curr. Opin. Neurobiol.* 4, 550–556. doi: 10.1016/0959-4388(94)90056-6

Medina, D., and Barraza, P. (2019). Efficiency of attentional networks in musicians and non-musicians. *Heliyon* 5:e01315. doi: 10.1016/j.heliyon.2019.e01315

Miendlarzewska, E. A., and Trost, W. J. (2014). How musical training affects cognitive development: rhythm, reward and other modulating variables. *Front. Neurosci.* 7:279. doi: 10.3389/fnins.2013.00279

Monsell, S. (2003). Task switching. *Trends Cogn. Sci.* 7, 134–140. doi: 10.1016/S1364-6613(03)00028-7

Moradzadeh, L., Blumenthal, G., and Wiseheart, M. (2015). Musical training, bilingualism, and executive function: a closer look at task switching and dual-task performance. *Cogn. Sci.* 39, 992–1020. doi: 10.1111/cogs.12183

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex* 19, 712–723. doi: 10.1093/cercor/bhn120

Neubauer, A. C., and Fink, A. (2009). Intelligence and neural efficiency. *Neurosci. Biobehav. Rev.* 33, 1004–1023. doi: 10.1016/j.neubiorev.2009.04.001

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., et al. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nat. Neurosci.* 20, 299–303. doi: 10.1038/nn.4500

Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., et al. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PLoS One* 5:e11120. doi: 10.1371/journal.pone.0011120

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychol.* 2:142. doi: 10.3389/fpsyg.2011.00142

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., et al. (2011). Functional network organization of the human brain. *Neuron* 72, 665–678. doi: 10.1016/j.neuron.2011.09.006

Putkinen, V., Tervaniemi, M., Saarikivi, K., and Huotilainen, M. (2015). Promises of formal and informal musical activities in advancing neurocognitive development throughout childhood. *Ann. N. Y. Acad. Sci.* 1337, 153–162. doi: 10.1111/nyas.12656

Ramírez, V., and Rosas, R. (2007). Estandarización del WISC-III en chile: descripción del test, estructura factorial y consistencia interna de las escalas. *Psykhe* 16, 91–109. doi: 10.4067/S0718-22282007000100008

Rodrigues, A. C., Guerra, L., and Loureiro, M. (2007). "Visual attention in musicians and non-musicians: a comparative study," in *Proceedings of the 3rd International Conference on Interdisciplinary Musicology*, Tallinn.

Rodrigues, A. C., Loureiro, M. A., and Caramelli, P. (2013). Long-term musical training may improve different forms of visual attention ability. *Brain Cogn.* 82, 229–235. doi: 10.1016/j.bandc.2013.04.009

Sachs, M., Kaplan, J., Der Sarkissian, A., and Habibi, A. (2017). Increased engagement of the cognitive control network associated with music training in children during an fMRI stroop task. *PLoS One* 12:e0187254. doi: 10.1371/journal.pone.0187254

Schaal, N. K., Pollok, B., and Banissy, M. J. (2017). Hemispheric differences between left and right supramarginal gyrus for pitch and rhythm memory. *Sci. Rep.* 7:42456. doi: 10.1038/srep42456

Schaal, N. K., Williamson, V. J., Kelly, M., Muggleton, N. G., Pollok, B., Krause, V., et al. (2015). A causal involvement of the left supramarginal gyrus during the retention of musical pitches. *Cortex* 64, 310–317. doi: 10.1016/j.cortex.2014.11.011

Schlaug, G., Norton, A., Overy, K., and Winner, E. (2005). Effects of music training on the child's brain and cognitive development. *Ann. N. Y. Acad. Sci.* 1060, 219–230. doi: 10.1196/annals.1360.015

Schlaug, G. (2015). "Musicians and music making as a model for the study of brain plasticity," in *Music, Neurology, and Neuroscience: Evolution, the Musical Brain, Medical Conditions, and Therapies*, eds E. Altenmüller, S. Finger, and F. Boller, 1st Edn., Vol. 217 (Amsterdam: Progress in Brain Research, Elsevier B.V.), 37–55. doi: 10.1016/bs.pbr.2014.11.020

Schulze, K., and Koelsch, S. (2012). Working memory for speech and music. *Ann. N. Y. Acad. Sci.* 1252, 229–236. doi: 10.1111/j.1749-6632.2012.06447.x

Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., and Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Hum. Brain Mapp.* 32, 771–783. doi: 10.1002/hbm.21060

Seither-Preisler, A., Parncutt, R., and Schneider, P. (2014). Size and synchronization of auditory cortex promotes musical, literacy, and attentional skills in children. *J. Neurosci.* 34, 10937–10949. doi: 10.1523/jneurosci.5315-13.2014

Serrallach, B., Groß, C., Bernhofs, V., Engelmann, D., Benner, J., Gündert, N., et al. (2016). Neural biomarkers for dyslexia, ADHD, and ADD in the auditory cortex of children. *Front. Neurosci.* 10:324. doi: 10.3389/fnins.2016.00324

Slevc, L. R., Davey, N. S., Buschkuehl, M., and Jaeggi, S. M. (2016). Tuning the mind: exploring the connections between musical ability and executive functions. *Cognition* 152, 199–211. doi: 10.1016/j.cognition.2016.03.017

Sluming, V., Brooks, J., Howard, M., Downes, J. J., and Roberts, N. (2007). Broca's area supports enhanced visuospatial cognition in orchestral musicians. *J. Neurosci.* 27, 3799–3806. doi: 10.1523/JNEUROSCI.0147-07.2007

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23(Suppl. 1), S208–S219. doi: 10.1016/j.neuroimage.2004.07.051

Stoeckel, C., Gough, P. M., Watkins, K. E., and Devlin, J. T. (2009). Supramarginal gyrus involvement in visual word recognition. *Cortex* 45, 1091–1096. doi: 10.1016/j.cortex.2008.12.004

Talamini, F., Altoè, G., Carretti, B., and Grassi, M. (2017). Musicians have better memory than nonmusicians: a meta-analysis. *PLoS One* 12:e0186773. doi: 10.1371/journal.pone.0186773

Talamini, F., Carretti, B., and Grassi, M. (2016). The working memory of musicians and nonmusicians. *Music Percept.*34, 183–191. doi: 10.1525/mp.2016.34.2.183

Tanaka, S., and Kirino, E. (2017). Reorganization of the thalamocortical network in musicians. *Brain Res.* 1664, 48–54. doi: 10.1016/j.brainres.2017.03.027

Tang, L., Shafer, A. T., and Ofen, N. (2018). Prefrontal cortex contributions to the development of memory formation. *Cereb. Cortex* 28, 3295–3308. doi: 10.1093/cercor/bhx200

Vines, B. W., Schnider, N. M., and Schlaug, G. (2006). Testing for causality with transcranial direct current stimulation: pitch memory and the left supramarginal gyrus. *NeuroReport* 17, 1047–1050. doi: 10.1097/01.wnr.0000223396.05070.a2

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children*, 3rd Edn. San Antonio, TX: The Psychological Corporation.

Zuk, J., Benjamin, C., Kenyon, A., and Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLoS One* 9:e99868. doi: 10.1371/journal.pone.0099868

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 08 October 2020
doi: 10.3389/fnhum.2020.535775

# Pitch Processing Can Indicate Cognitive Alterations in Chronic Liver Disease: An fNIRS Study

*Geonsang Jo[1], Young-Min Kim[2,3], Dae Won Jun[4*†] and Eunju Jeong[3,5*†]*

*[1] Daehong Communications Inc, Seoul, South Korea, [2] Graduate School of Technology and Innovation Management, Hanyang University, Seoul, South Korea, [3] College of Interdisciplinary Industrial Studies, Hanyang University, Seoul, South Korea, [4] Department of Internal Medicine, College of Medicine, Hanyang University, Seoul, South Korea, [5] Department of Music and Science for Clinical Practice, Hanyang University, Seoul, South Korea*

Early detection and evaluation of cognitive alteration in chronic liver disease is important for predicting the subsequent development of hepatic encephalopathy. While visuomotor tasks have been rigorously employed for cognitive evaluation in chronic liver disease, there is a paucity of auditory processing task. Here we focused on auditory perception and examined behavioral and haemodynamic responses to a melodic contour identification task (CIT) to compare cognitive abilities in patients with chronic liver disease (CLD, $N = 30$) and healthy controls ($N = 25$). Further, we used support vector machines to examine the optimal combination of channels of functional near-infrared spectroscopy that can classify cognitive alterations in CLD. Behavioral findings showed that CIT performance was significantly worse in the patient group and CIT significantly correlated with neurocognitive evaluation (i.e., number connection test, digit span test). The findings indicated that CIT can measure auditory cognitive capacity and its difference existing between patient group and healthy controls. Additionally, optimal subsets classified the 16-dimensional haemodynamic data with 78.35% classification accuracy, yielding markers of cognitive alterations in the prefrontal regions (CH6, CH7, CH10, CH13, CH14, and CH16). The results confirmed the potential use of behavioral as well as haemodynamic responses to music perception as an alternative or supplementary method for evaluating cognitive alterations in chronic liver disease.

**Keywords: chronic liver disease (CLD), cognitive alteration, nonverbal auditory perception/music perception, melodic contour identification, haemodynamic response, functional near-infrared spectroscopy (fNIRS)**

OPEN ACCESS

***Edited by:***
*Cunmei Jiang,*
*Shanghai Normal University, China*

***Reviewed by:***
*Yun Jiao,*
*Southeast University, China*
*Marta Llansola,*
*Principe Felipe Research Center (CIPF), Spain*

****Correspondence:***
*Dae Won Jun*
*noshin@hanyang.ac.kr*
*Eunju Jeong*
*ejeong@hanyang.ac.kr*

*† These authors have contributed equally to this work*

***Specialty section:***
*This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience*

***Received:*** *17 February 2020*
***Accepted:*** *26 August 2020*
***Published:*** *08 October 2020*

***Citation:***
*Jo G, Kim Y-M, Jun DW and Jeong E (2020) Pitch Processing Can Indicate Cognitive Alterations in Chronic Liver Disease: An fNIRS Study. Front. Hum. Neurosci. 14:535775. doi: 10.3389/fnhum.2020.535775*

## INTRODUCTION

Chronic liver disease (CLD) is a progressive destruction of liver functions over a period more than 6 months leading to fibrosis and cirrhosis (Sharma and Nagalli, 2020). Depending on the severity of impairment, patients with CLD show the limited mental capacity, changes in psychomotor functions and/or hepatic coma (Bernthal et al., 1987; Brodersen et al., 2014; Filipović et al., 2018). CLD-related neurophysiological and psychometric dysfunctions vary, ranging from psychomotor speed to executive functioning (Ortiz et al., 2005; Zhan and Stremmel, 2012) and get worsened as approach to overt hepatic encephalopathy (HE) (Butterworth, 2000; Sánchez-Carrión et al., 2008). Patients with HE present with overt clinical symptoms, such as disorientation, and consciousness disorders, which contributes to an increased risk of death in cirrhotic patients (Bustamante et al., 1999; Weissenborn et al., 2001a, 2005; Bajaj et al., 2007a; Prasad et al., 2007; Stewart et al., 2007).

Early detection of cognitive alteration in chronic liver disease is, thus, critical to predict the subsequent development of HE (Krieger et al., 1996; Romero-Gómez et al., 2001; Chen et al., 2013).

The current cognitive evaluation in chronic liver disease do not cater for cognitive functions in diverse sensory modalities. For example, the West Haven scale is a subjective and semi-quantitative clinical scale that classifies mental state changes (Conn et al., 1977; Groeneweg et al., 1998; Hartmann et al., 2000; Ferenci et al., 2002; Amodio et al., 2004). In addition, there is a battery of psychometric tests that aims to detect neurocognitive impairments, such as neuropsychological and perceptual motor dysfunction (Weissenborn et al., 2005), named the psychometric hepatic encephalopathy score (PHES). The PHES consists of five subtests, including the A and B number connection tests (NCT-A, NCT-B), the line tracing test (LTT), the serial dotting test (SDT), and the digit symbol test (DST) (Weissenborn et al., 2001b; Ferenci et al., 2002; Bajaj et al., 2009), a primarily paper-and-pencil test. The stimuli in these neurocognitive tasks are limited to visual perception and visuomotor agility and rarely provide information about cognitive alteration in auditory modality. Clinicians and researchers have emphasized to a novel method that employs another type of stimuli and tasks to detect cognitive alterations, and that is also time- and cost-effective (Bajaj et al., 2007b, 2008; Romero-Gómez et al., 2007; Sharma et al., 2013; Kircheis et al., 2014; Gupta et al., 2015), which is a central aim of this study.

In this study, we focused on the potential of auditory perception as a new evaluation task (Mehndiratta et al., 1990; Saxena et al., 2001). For instance, Mehndiratta et al. (1990) employed different modality tasks to detect hepatic encephalopathy (HE), auditory task as measured by brain stem auditory evoked potentials were found to be the most sensitive to indicate HE compared to visual and somatosensory evoked potentials. Event-related potentials (ERPs) using an auditory oddball test have been shown to have significantly delayed P300 components in patients with minimal hepatic encephalopathy (MHE, the earliest form of HE characterized by neurocognitive impairment; Stinton and Jayakumar, 2013) and in patients with cirrhosis compared to healthy adults (Ciećko-Michalska et al., 2006; Teodoro et al., 2008). In a similar vein, Moon et al. showed cirrhotic patients had longer latencies for N100, P200, N200, and P300 than healthy adults in the auditory oddball test (Moon et al., 2014). In particular, the N200 latency, a negative peak related to mismatch detection and executive cognitive control function (Folstein and Van Petten, 2008), was significantly prolonged in cirrhotic patients than in healthy adults. The authors suggested that both P300 and N200 were delayed due to a slowness in intracerebral nerve conduction and that the two EEG components can be considered to be the first signs of cerebral deterioration in HE (Moon et al., 2014).

Rather than the auditory oddball test, in which target and non-target stimuli are presented consecutively, this study employed melodic contour identification task (CIT) that is designed to measure the selectivity of auditory attention, which is the core characteristics of auditory information processing. Our previous studies confirmed the validity of using melodic CIT to measure the various types of attention in moderate-to-severe traumatic brain injury patients. Jeong (2013) validated that the melodic contour stimuli could distinguish the different types and capacities of auditory attention existing in the various age groups and that it is a valid and reliable test for auditory cognition. They also that melodic CITs can measure attentional and cognitive dysfunctions existing clinical populations (Jeong and Lesiuk, 2011). More recently, the updated and computerized version of the test was scaled up, showing that the different CITs could distinguish different cognitive loads (Jeong et al., 2018).

We also examined $HbO_2$ (oxygenated hemoglobin) in the frontal lobe, which is known as an indicator of cognitive alterations in patients with chronic liver disease (Mendonça et al., 2013). Keiding and Pavese (2013) revealed that the cerebral oedema yields dysfunction in haemodynamic responses and, thus, it is the main cause of cognitive alterations in HE. Macias-Rodriguez et al. (2016) supported the idea that the severity of HE contributes to haemodynamic alteration, mainly caused by damage to the vascular system and dysfunction of auto-regulatory vascular responses. The prefrontal areas modulate diverse cognitive systems, such as attention, working memory, decision-making, and problem solving (Amodio et al., 2004; Felipo et al., 2012; Jao et al., 2015) and to receive projections from almost all processing levels in the superior temporal gyrus (Poremba et al., 2004; Kusmierek and Rauschecker, 2009; Kikuchi et al., 2010). With regards to oxygen changes, to a lesser extent, prefrontal activity has been examined with patients with MHE. To our best knowledge, there existed a single study examining oxygen consumption changes during cognitive performance at the prefrontal cortex using fNIRS. Nakanishi et al. (2014) compared regional $HbO_2$ in cirrhotic patients without MHE and those with MHE during a word fluency task. Their findings showed a significant difference between groups in $HbO_2$ changes over time. That is, $HbO_2$ in the MHE group were gradually increased throughout tasks, while the non-MHE group showed recurrent patterns of abrupt increase and decrease in concentrations. Also, the increase in $HbO_2$ upon stimulation was significantly delayed in the MHE group than in non-MHE group (i.e., 5 s after stimulus presentation). These findings were suggestive of $HbO_2$ obtained from the frontal area can reflect cognitive alterations along with progress of liver disease.

Thus, the main purpose of this study was to examine the potential of the melodic CIT as a test for cognitive alterations following chronic liver disease. We measured behavioral performance using three subtests of the CIT (focused, selective, and alternating listening, respectively) and examined the criterion validity of CIT, in correlation with standard neurocognitive tests. We also examined whether changes in $HbO_2$ can be indicative of cognitive alterations in chronic liver diseases, and whether the regions of the prefrontal cortex, known to modulate attention and cognition, can be specified to characterize the alterations in chronic liver disease. For this analysis, we employed a support vector machine (SVM) combined with an fNIRS to classify cognitive alterations following liver diseases, a technique which has been increasingly used to classify clinical and healthy populations. Monden et al. (2015) previously measured $HbO_2$ during a go/no-go task and performed classification using a SVM algorithm. The findings of that algorithm yielded a 90% accuracy, indicating the potential of SVM as a tool for classifying and, thus, diagnosing children with attention deficit hyperactivity disorder (ADHD). In addition, Ichikawa et al. (2014) employed an exhaustive search method combined with SVM that explored all possible combinations of the fNIRS channels to classify children with ADHD and autism spectrum disorder (ASD).

# METHODS

## Participants

A total of 55 participants were included: 30 patients with chronic liver diseases (Male = 19, Female = 11), and 25 healthy controls (Male = 8, Female = 17) were matched by age and the level of education. Participants who had a <3 months of regular involvement in musical activities and/or professional training and who had a minimal ability to understand the spoken instruction were eligible to participate in the study. The diagnosis of liver disease was based on either a liver biopsy or the presence of portal hypertension and markers of hepatocyte synthetic dysfunction. We included patients who have been diagnosed as liver disease for more than 6 months. Participants with any of the following conditions were excluded: diabetes mellitus, systemic arterial hypertension, metabolic liver disease (hemochromatosis and Wilson's), personal history of stroke or cancer, use of neuropsychiatric drugs, neuropsychiatric disorders, current alcohol intake or smoking, rotating shift work, or acute inflammatory responses of infectious origin. **Table 1** presents the demographics of the two groups.

**TABLE 1 |** Participant demographics.

| | CL (*N* = 30) | | HC (*N* = 25) | | *p*-value |
|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | |
| Age (years) | 55.80 | 6.91 | 54.64 ± | 5.94 | 0.205 |
| Education (years) | 10.80 | 3.67 | 13.48 ± | 3.10 | 0.394 |

*CL, Patients with chronic liver disease; HC, Healthy control group.*
*N = 55 for this correlation analysis.*

## Musical Stimuli and Tasks

Three melodic contours (ascending, stationary, descending) were adopted from Jeong and Ryu (2016). Melodic contours are a series of tones moving in different directions (i.e., ascending, descending, and stationary). Two different types of contours were combined consecutively to yield six test items (i.e., ascending and descending, ascending and stationary, stationary and ascending, stationary and descending, descending and ascending, descending and stationary). The presentation time of each test item was 5,250 ms, including two contours (2,250 ms for each) and inter-contour interval (750 ms). **Figure 1** shows examples of the pitch contours used in the study.

The six test items were modulated in five different keys (G# to C major) and were presented using three instrument timbres (i.e., piano, flute, string), yielding a total of 90 test items. We selected the instruments based on a previous study that classified various musical instruments according to their spectral features of timbre, such as the harmonic structure, inharmonicity, and harmonic energy skewness (Agostini et al., 2003). The melodic contours were generated by a digital audio workstation (Logic Pro X, Apple Inc., Cupertino, CA, USA) with amplitude normalization. The experimental test was developed as a computerized version, using Visual Studio (Microsoft, Washington, DC, USA).

Contour identification tasks (CITs). The computerized version of the CIT was designed to measure different types of auditory attention and the associated cognitive load changes (**Table 2**). The task stimuli and task structures were adopted from previous studies (Jeong, 2013; Jeong and Ryu, 2016; Jeong et al., 2018) and modified for the current purpose. In CIT1, two consecutive contour directions were presented as target contour with environmental sounds, including traffic, raining, twittering, ticktack, bustling, laughing, gabbling, applause, crying, and jeering sounds. They were randomly presented against target contours (i.e., selective contour identification against environmental noise). In CIT2, participants were presented with target melodic contours against target-like distractors (i.e., another melodic contour played by different instrument timbres) and were asked to identify a target contour presented in a predetermined instrument timbre. In CIT3, two melodic contours were presented, while participants were asked to shift their attention from one to another target contour and identify the direction of contours. For both CIT2 and CIT3, a


**FIGURE 1 |** Sample melodic contours. **(A)** Ascending and stationary contours, **(B)** ascending and descending contours.

visual cue (e.g., a picture of an instrument) was shown on the computer screen to inform about the instrument timbre of target contour. For all CITs, the direction of melodic contours and instrument timbres were randomly selected.

CIT1 had no visual cues, however, in CIT2, a picture of an instrument that plays target contours was presented prior to presenting the item to inform which contour the participants selectively listened to. In CIT3, outlined boxes were additionally used to guide at which contours the participants selectively listened to and shifted from one to another instrument (see **Figure 2**). For example, the first outlined box appeared in the upper or lower line with the first set of contours, and the second box appeared with the second set of contours.

## Haemodynamic Measurements

Oxygenated hemoglobin was measured to evaluate cognitive activation in participants and the loads imposed by the given tasks (Peck et al., 2013; Ogawa et al., 2014; Yasumura et al., 2014). For this evaluation, we employed an fNIRS and a non-invasive to monitor cortical tissue oxygenation (oxygenated hemoglobin, $HbO_2$; deoxygenated hemoglobin, HHb) during cognitive, motor, and sensory stimulation. We used a 16-channel Spectratech OEG-16 (Shimadzu Co. Ltd., Kyoto, Japan) for the measurements (**Figure 3**). Task-related haemodynamic changes in $HbO_2$ were recorded in 16 channels with a sampling rate of 0.65 s. In addition to the fNIRS data, we collected behavioral data, including task performance accuracy and reaction time.

## Experimental Procedure

This study was approved by the Institutional Review Board of Hanyang Medical Centre (HYUH-2013-08-017-002). Healthy adults voluntarily participated and were recruited *via* physical and online advertisements and patients with CLD were recruited *via* Hanyang Medical Center. All participants gave written informed consent in accordance with the Declaration of Helsinki. All other experimental methods were performed in accordance with the relevant guidelines and regulations. Prior to the experiment, two medical doctors with 17 and 20 years of experience, respectively, met with the participants to determine the participation eligibility and administered neurocognitive tests.

After taking a 30-min break, the participants wore a band-type fNIRS containing an array of 12 probes on their forehead. Participants had a short rest period and then the pre-stimulus baseline data were obtained for 20 s while they fixated their eyes on the center of the monitor. A 20-s baseline was also obtained during inter-task rest periods and post-task period. Once the baseline data were obtained, the participants were presented with instruction and examples of melodic contours. Each of the three CITs started with brief instruction in terms of the task characteristics given in each CIT and how to respond to test items. Participants were also instructed to identify the directions of the target contours by clicking the arrow corresponding to the contour direction as accurate and immediate as possible. The contours were delivered *via* a headphone with controlled volume, while the visual cues specifying the target musical instrument were presented to the participants on a monitor. The

**TABLE 2 |** Structure of the CITs.

| | Target | Distractor | Given task | Cognitive load |
|---|---|---|---|---|
| CIT1 | Melodic contour | Environmental sounds | Selective identification-Basic | Low |
| CIT2 | Melodic contour | Target-like contours | Selective identification-Advanced | ↓ |
| CIT3 | Melodic contour | Target-like contours | Alternating | High |

*CIT, contour identification task.*

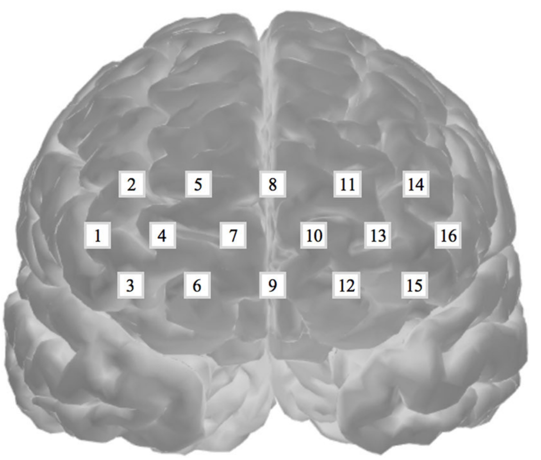



**FIGURE 3 |** Spectratech OEG-16. The locations of the 16 fNIRS channels along the frontal cortex. The center of the measurement unit was placed on the frontopolar (Fp) region according to the international 10–20 system.

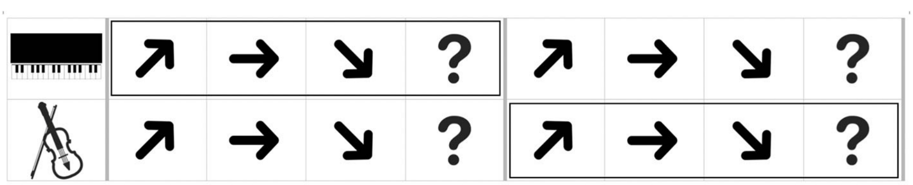



**FIGURE 2 |** An example of an answer page given in a monitor with the musical stimuli (CIT3). The boxes were presented prior to presentation of each contour.

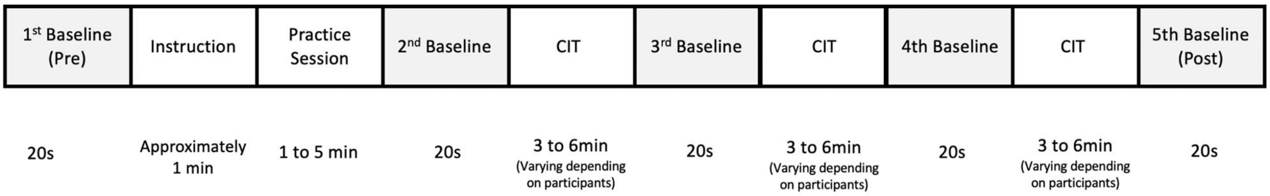



**FIGURE 4 |** The flow of the experiment.

participants underwent a practice session to become familiar with the direction identification task. When their accuracy was over 80%, the main experimental session was administered. A total of 18 test items were presented in each CIT (a blocked design) and the order of CIT was randomized across participants (**Figure 4**). The CITs took about 20 min to complete. The experiment was performed in a sound-proof room, in which light and temperature were controlled.

## Signal Pre-processing

The fNIRS raw data were collected throughout the experiment and were converted into concentration changes of hemoglobin using the modified Beer–Lambert law (Baker et al., 2014). Subsequently, a zero-phase low- and high-pass filter with a cut off frequency from 0.01 to 0.09 Hz was applied to remove any noise from heartbeat pulsations and longitudinal signal drifts (Morren et al., 2004; Akgül et al., 2005; Bauernfeind et al., 2011). In this study, we employed an $HbO_2$ index that was based on previous studies, which reported that $HbO_2$ shows better classification performance than other measures (Li et al., 2015), especially for conditions with a high-dimensional feature and low sample sizes (Mourao-Miranda et al., 2005; Yoon et al., 2007). The obtained $HbO_2$ values were standardized by subtracting the mean of 20-s pre-stimulus baselines in order to compare directly across participants and channels (Herff et al., 2013). Subsequently, the mean $HbO_2$ during each CIT (CIT1, CIT2, CIT3) was calculated for each of the 16 channels. Since we found that the fNIRS response peak was delayed by a few seconds compared to the stimulus onset (Cui et al., 2010; Ichikawa et al., 2014), we eliminated the $HbO_2$ from the first 3–10 s for the statistical analysis. Finally, we obtained 165 data points, including 90 from CL (including 30 from the MHE subgroup) and 75 from HC. We used the data as inputs and the diagnosis (HC, CL, MHE) of the groups as outputs.

## Classification Using Support Vector Machines (SVMs)

In this study, we employed a linear SVM model with a repeated k-fold **cross** validation to solve classification problems between HC and patients with liver disease. K-fold cross validation method is one of the split sample methods that randomly divide data into $k$ subsets, then one of the $k$ subsets are used as the test set and the other k-1 subsets are used as training set. This process is iterated $k$ times so that each subset is used as testing set then results are averaged (Kim, 2009; Mishra and Sahu, 2011). We adopted 5-fold cross validation with repeated 20 times to accurately estimate the generalized performance of the classification (Refaeilzadeh et al., 2009). Then, we evaluated classification performance using the MCC and the bACC, as described by previous studies (Sakiyama et al., 2008; Jiao and Du, 2016). The studies recommended the use of different evaluation methods to confirm high classification performance in practical application (Jiao and Du, 2016).

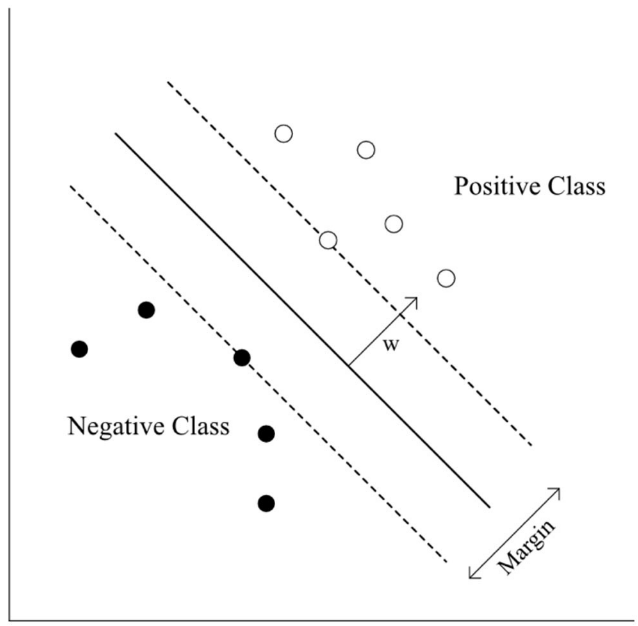



**FIGURE 5 |** SVM classification.

The standardized mean of $HbO_2$ was trained in three different ways, including (1) eight channels from the right hemisphere (CH1–CH8), (2) eight channels from the left hemisphere (CH9–CH16), and (3) 16 channels from both hemispheres (CH1–CH16). The standardized $HbO_2$ data from the eight channels from both hemispheres and the 16 channels have an 8-dimensional ($2^8$- 1 = 255 subsets) or 16-dimensional ($2^{16}$- 1 = 65,565 subsets) feature vector, respectively. The obtained $HbO_2$ data were trained and classified to diagnose groups. SVMs have been applied previously for classification problems in various domains, and have yielded high generalization capability (Bennett and Campbell, 2000; Lotte et al., 2007). An SVM learns the relationship between the input and output, (i.e. classes) from the given set of data. In the feature vector space (**Figure 5**), the SVM algorithm creates a hyperplane that separates the input data into two classes with a maximum margin.

The following equation describes the hyperplane of SVM.

$$\mathbf{w}^{T}\mathbf{x} + b = 0, \quad (1)$$

Where **w** is the normal vector to the hyperplane, the two classes are represented as follows:

$$\mathbf{w}^{T}\mathbf{x}_i + b \geq 1 \text{ for } y_i = 1, \quad (2)$$
$$\mathbf{w}^{T}\mathbf{x}_i + b \leq -1 \text{ for } y_i = -1, \quad (3)$$
$$\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in}), y_i \subset \{1, -1\},$$

where $\mathbf{x}_i$ is the $i^{th}$ example of the training data and $y_i$ is its label.

Once we trained the SVM model, the weight vector **w** is known. To test the model, we put a test sample into the left side of the **Equation (1)**. If the value was more than 0, the sample was classified as positive. If the value was <0, the sample was classified as negative.

## Evaluation

To evaluate the trained model, we first computed the confusion matrix shown in **Table 3**. A confusion matrix describes the classification result of the test samples by counting the number of true positives, true negatives, false positives, and false negatives. If both actual and predicted classes of a sample are positives, the sample is a true positive; similarly, the sample is a true negative when both classes are negatives. If the actual class of a sample is positive but the predicted class of the sample is negative, then the sample is a false negative. If the actual class of a sample is negative but the predicted class is positive, the sample is a false positive.

From the confusion matrix, the basic performance measures were calculated according to the following equations:

$$\text{Sensitivity}: TP/(TP + FN) \quad (1)$$
$$\text{Specificity}: TN/(TN + FP) \quad (2)$$
$$\text{Accuracy}: (TP + TN)/(TP + FP + FN + TN) \quad (3)$$

Sensitivity is the probability to correctly predict the samples that are actually classed as positive. It is an indicator that assess the ability of a model to identify positive samples. Specificity also assess a model's ability to correctly to correctly predict the samples whose class is actually negative. Accuracy is the ability of a model to correctly identify classes for all samples. However, accuracy could lead to evaluation errors when the data is imbalanced. If one class is a majority class, the real performance of the other class cannot be reflected accurately. Moreover, sensitivity and specificity are not proper measures to study the balanced performance. For a balanced evaluation, the following measure was used.

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (4)$$
$$bACC = (\text{Sensitivity} + \text{Specificity})/2 \quad (5)$$

We assessed a modified classification performance measure using a MCC, which is used widely in biomedical research (Van't Veer et al., 2002; Boughorbel et al., 2017). MCC considers all of the confusion matrix categories (true positives, true negatives, false positives, false negatives), making it suitable for providing a more balanced value when the sample numbers between groups were imbalanced (Brodersen et al., 2010; Powers, 2011). Because of a difference in the number of samples in our patients (CL = 30, and HC = 25), we adopted the MCC to classify performance measures. MCC values range between −1 and +1. If an MCC coefficient is +1, it means the classifier can perfectly predict the class of the data. An MCC coefficient = 0 means that it is not different from a random prediction, and a −1 coefficient means it totally mispredicts the class. Also, we considered the bACC, which is the average of sensitivity and specificity since it is also one of the ways to solve an imbalanced dataset (Brodersen et al., 2010; Powers, 2011), for our accuracy measurements. All statistical procedures were performed using R.

**TABLE 3** | Confusion matrix.

| | | Predicted Class | |
|---|---|---|---|
| | | **Positive** | **Negative** |
| Actual Class | Positive | True Positive (TP) | False Negative (FN) |
| | Negative | False Positive (FP) | True Negative (TN) |

# RESULTS

We analyzed behavioral responses to CITs and performed classifications between patients with chronic liver disease (CL) and healthy controls (HC) using $HbO_2$. $HbO_2$ was obtained from 16 fNIRS channels located in the prefrontal cortex. We employed SVM with the exhaustive search method and grouped three data sets, including (1) 8-dimension dataset obtained from the right hemisphere (CH1–CH8), (2) 8-dimension dataset obtained from the left hemisphere (CH9–CH16), and (3) 16-dimension dataset obtained from the bilateral hemispheres (CH1–CH16). For each of these three datasets, we trained the SVM and evaluated its classification accuracy using Matthew correlation coefficient (MCC) and the balanced accuracy (bACC).

## Behavioral Responses

**Table 4** shows the mean accuracy and reaction time for the CL and HC groups on the CIT. The mean accuracy in the HC group was the highest for CIT1(0.75), and was followed by the CIT2 score (0.48). Accuracy in the HC was lowest (0.46) when an attention shift was required between two concurrent melodic contours during CIT3. A similar trend was found across the CITs for the CL group (decreasing from 0.62 to 0.32). However, accuracy declined considerably between both the CIT1 and CIT2 and between the CIT2 and CIT3, while the HC group showed an obvious decrease only between CIT1 and CIT2. Reaction time

**TABLE 4 |** Statistical difference of CIT performance between groups.

| | Accuracy | | | | Reaction time | | | |
|---|---|---|---|---|---|---|---|---|
| | CL ($N = 30$) | | HC ($N = 25$) | | CL ($N = 30$) | | HC ($N = 25$) | |
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| CIT1 | 0.62[a] | 0.30 | 0.75[d] | 0.28 | 8576.44[f1] | 816.34 | 5799.34 | 894.26[h1] |
| CIT2 | 0.45[b] | 0.23 | 0.48[e1] | 0.25 | 9595.72[f2] | 1154.55 | 6695.61 | 1264.74[h2] |
| CIT3 | 0.32[c] | 0.21 | 0.46[e2] | 0.26 | 11139.09[g] | 1490.87 | 6926.38 | 1633.17[h3] |

*CL, Patients with chronic liver disease; HC, Healthy control group; M, Mean; SD, Standard Deviation.*
*Different superscript letters in the same column indicate statistical significance.*
*$p < 0.001$ for the pairs of a and b, b and c, c and a, d and $e^1$, d and $e^2$.*
*$p < 0.05$ for the pairs of f1 and g, $f^2$ and g. The pairs of $e^1$ and $e^2$, $f^1$ and $f^2$, $h^1$, $h^2$, and $h^3$ did not show any statistical significance ($p > 0.05$).*

**TABLE 5 |** Correlation between CIT and neurocognitive evaluation tests.

| | | CL | | | | HC | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | NCT-A | NCT-B | DST-Forward | DST-Backward | NCT-A | NCT-B | DST-Forward | DST-Backward |
| Accuracy | CIT1 | −0.673** | −0.546** | 0.32 | 0.325 | −0.3 | −0.470* | 0.560** | 0.558** |
| | CIT2 | −0.565** | −0.520** | 0.38 | 0.33 | −0.438* | −0.429* | 0.466** | 0.398* |
| | CIT3 | −0.593** | −0.459* | 0.415* | 0.367 | −0.418* | −0.473** | 0.521** | 0.414* |
| | Total | −0.691** | −0.573** | 0.416* | 0.383 | −0.409* | −0.499** | 0.566** | 0.509** |
| Reaction time | CIT1 | 0.753** | 0.567** | −0.248 | −0.332 | 0.752** | 0.759** | −0.431* | −0.304 |
| | CIT2 | 0.670** | 0.576** | −0.289 | −0.231 | 0.626** | 0.649** | −0.380* | −0.24 |
| | CIT3 | 0.692** | 0.566** | −0.386 | −0.182 | 0.721** | 0.714** | −0.467** | −0.319 |
| | Total | 0.752** | 0.602** | −0.311 | −0.279 | 0.723** | 0.732** | −0.435* | −0.295 |

*NCT, Number Connection Test (Time to completion, sec); DST, Digit Span Test (Number of correct answers).*
*$N = 55$ for this correlation analysis.*
*$^{**}p < 0.01$, $^{*}p < 0.05$.*

showed a similar trend between HC and CL groups, and was the shortest in the CIT1 and the longest in the CIT3.

A two-way mixed ANOVA [i.e., Group (CL, HC) × Task (CIT1, CIT2, CIT3)] was performed to validate accuracy and reaction time. For accuracy, there was a significant main effect of the Task [$F_{(2,\ 110)} = 65.566$, $p < 0.001$], indicating that CIT performance for both groups worsened significantly as the CITs became more difficult. The pairwise *post hoc* comparison using the Bonferroni correction revealed that CIT1 was significantly better than CIT2 ($p < 0.001$), and that CIT2 was significantly better than CIT3 ($p < 0.01$). We also found a significant interaction between the Group and Task [$F_{(2,\ 110)} = 3.114$, $p < 0.05$]. Interestingly, the pairwise *post-hoc* comparison revealed that the CL group performed significantly better in the CIT2 than in the CIT3 ($p < 0.001$), while the HC group showed similar performances in the CIT2 and CIT3 ($p > 0.05$). There was a significant main effect of the Task [$F_{(2,\ 110)} = 6.126$, $p < 0.01$] on the response time. The pairwise *post hoc* comparison using the Bonferroni correction revealed that only the CL group performed significantly worse in the CIT3 than CIT1 and CIT2 ($p < 0.05$, respectively). **Table 4** present the statistical difference of behavioral performance between groups. Our behavioral findings collectively indicated that patients with chronic liver disease showed significantly worse auditory perception and cognition, specifically between CIT2 and CIT3. Additionally, we performed a correlation analysis to examine the criterion validity of the CITs. **Table 5** presents the correlations between the CITs and neurocognitive evaluations.

## Classification of Hemodynamic Responses

Further, we classified the differences in haemodynamic responses to CIT1 and CIT2 between HC and CL in combination with an SVM and an exhausted feature selection method. We decided to exclude CIT3 since our behavioral findings clearly showed that CIT3 was too difficult for CL group and, thus, haemodynamic response to CIT3 might not indicate appropriate cognitive load in this group. With the analysis, we found a total of 15 subsets that classified the data more accurately into two groups among the 255 subtests (five each for the right, left, and bilateral hemispheres, respectively). **Table 6** presents the classification performance for each of the 15 subsets. The classification performance was higher for the right hemisphere (MCC = 0.451) than the left hemisphere (MCC = 0.317). The compute score of sensitivity and specificity was also higher for the right hemisphere (bACC = 69.75%) than in the left (bACC = 64.50%) hemisphere. The classification performance was best with CH6, CH7, CH10, CH13, CH14, and CH16 (MCC = 0.577, bACC = 78.35%), indicating that the inclusion of $HbO_2$ obtained from the bilateral hemisphere yielded better classification performance than values obtained from the unilateral hemispheres. After an additional *t*-test, we

**TABLE 6 |** Classification accuracy between HC and CL data sets.

| Hemisphere | Best subset | MCC | bACC |
|---|---|---|---|
| Right | 2,7 | 0.451 | 69.75% |
| | 3,4,7 | 0.435 | 71.05% |
| | 1,2,4,7,8 | 0.366 | 69.40% |
| | 3,4,6,7 | 0.364 | 68.75% |
| | 2,4,6,7 | 0.363 | 62.25% |
| Left | 9,13,16 | 0.317 | 64.50% |
| | 9,13 | 0.309 | 63.65% |
| | 9,13,15 | 0.296 | 62.65% |
| | 9,12,13,14,15 | 0.274 | 60.80% |
| | 9,10,13 | 0.245 | 61.50% |
| Bilateral | 6,7,10,13,14,16 | 0.577 | 78.35% |
| | 6,7,10,12,13,14,16 | 0.575 | 78.00% |
| | 1,2,6,7,12,13,16 | 0.571 | 77.45% |
| | 4,6,7,9,13,14,16 | 0.560 | 77.45% |
| | 4,6,7,8,13,14,16 | 0.548 | 76.70% |

*N = 255 subtests for the right and left hemispheres, N = 65565 subtests for the bilateral hemisphere. Best subsets indicated a group of the channels that yielded optimum classification accuracy. MCC, Matthew Correlation Coefficient; bACC, balanced Accuracy.*

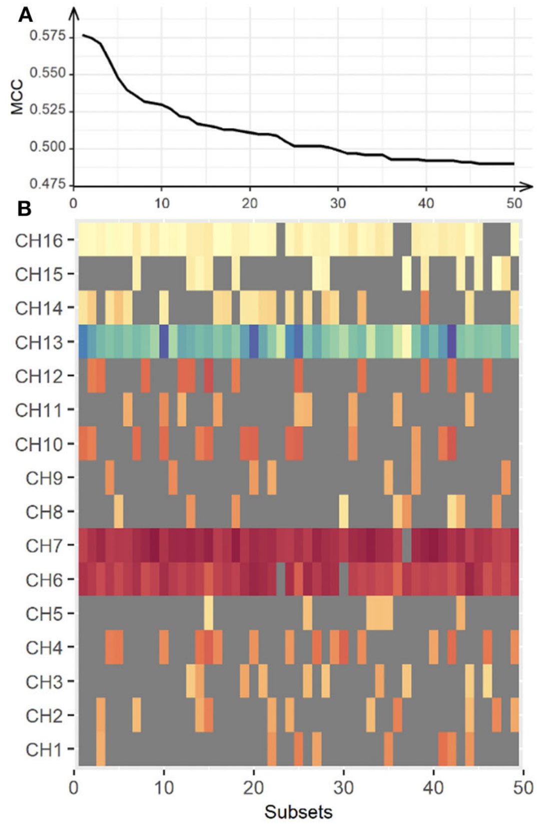



**FIGURE 6 |** Classification performance for the bilateral hemisphere between CL and HC. The upper panel **(A)** represents classification performance for the best 50 out of 65,565 subsets. The lower panel **(B)** shows the weight vector of each channel matched in the subsets. MCC, Matthew Correlation Coefficient. Red indicates higher $HbO_2$ in the CL, while blue indicates higher $HbO_2$ in the HC.

found that classification performance of the bilateral subset was significantly higher than that of the right and left hemisphere subsets ($p < 0.001$, respectively).

Further, we created a matrix plot using 50 subsets and constructed features that correspond to the classification performance of each subtest. **Figure 6** shows that CH6, CH7, CH10, CH13, CH14, and CH16 are effective features. This indicated that $HbO_2$ in CH6, CH7, and CH10 were higher in the CL than in the HC, and that the $HbO_2$ in CH13, CH14, and CH16 were lower in the CL than in the HC. Further, our findings suggest that the frontal areas of each hemisphere were intercommunicating, so it was necessary to include subsets from both hemispheres to yield better classification performance.

## DISCUSSION

In this study, we examined the potential of melodic CITs to evaluate cognitive alterations in chronic liver disease. Our behavioral findings indicated that CITs can differentiate changes in auditory perception and cognition in patients with chronic liver disease from healthy controls. Correlation analysis using standard neurocognitive evaluations revealed that CITs had good criterion validity and potential for measuring cognitive alterations that occur in chronic liver disease. Then, we applied SVMs with 5-fold cross validation to the haemodynamic responses obtained during the CIT performance to classify cognitive alterations in patients with chronic liver disease. We exhaustively searched all subsets of the measurement channels and evaluated each classification performance by repeating the 5-fold cross validation method 20 times. Our findings yielded an optimal subset for the classification of the haemodynamic data with 78.35% accuracy. Our results indicated that the subsets obtained bilaterally can better classify the differences that exist between HC and patients with CL. Also, we found channel features that could specify between groups. Three channels (CH13, CH14, CH16) in the left dorsolateral prefrontal cortex (DLPFC), one channel (CH6) in the right orbitofrontal cortex (OFC), and two channels (CH7, CH10) in the right frontopolar area (FP) were important for the classification of CL from HC. **Figure 7** summarizes the channel specific findings and clearly shows haemodynamic difference existing between groups.

From the behavioral findings, we found poorer overall CIT performance in the CL than in the HC group, and the difference was more obvious between CIT2 and CIT3. The findings indicated that overall auditory cognition deteriorated in the CL group. Given that CIT1 and CIT2 involve auditory selective attention and that CIT3 involves auditory alternating attention, the cognitive threshold of patients with CLD might be auditory alternating attention (or cognitive flexibility). The current findings resembled previous findings where selective attention (Felipo et al., 2012) and cognitive flexibility (Yang et al., 2018) deteriorated in patients with liver-related diseases. The

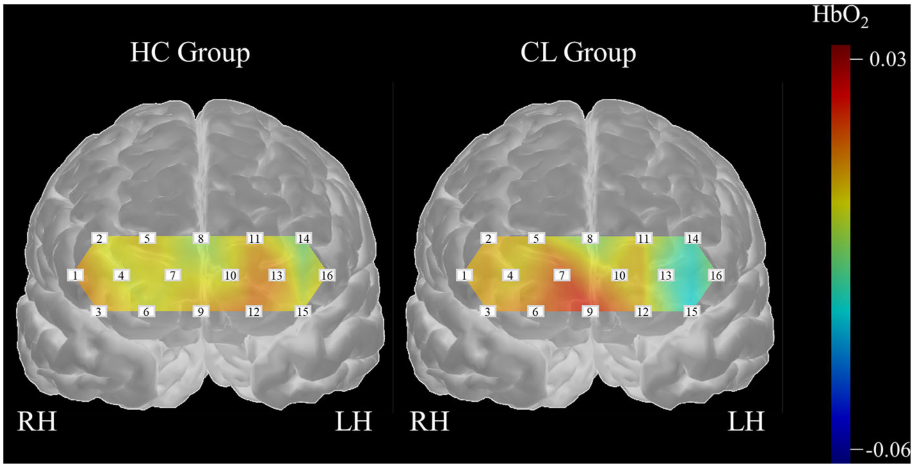



**FIGURE 7 |** Differences of haemodynamic responses between the HC and CL groups.

tasks employed in the previous studies are visual-oriented, so our findings revealed that the deterioration of selective attention and of cognitive flexibility exists similarly in an auditory modality.

Additional correlation analyses indicated that CIT can measure modality-general cognitive abilities and evaluate auditory cognitive deficits in chronic liver disease. In the CL and HC groups, the NCT-A and NCT-B highly correlated with the accuracy and reaction time of CITs. Correlations with DST were specific to groups. DST-Forward significantly correlated with accuracy and reaction time in the HC group, but this was not significant for the CL group (except for the accuracy of CIT3 and CIT total). DST-Backward significantly correlated with accuracy only in the HC group—not for the CL group. DST measures auditory working memory (DST-Forward) and auditory executive function (DST-Backward; Soltani et al., 2018). The current findings might indicate that the HC group utilized auditory working memory and executive function appropriately while performing CITs. The CL group, on the other hand, utilized these functions limitedly during CIT performance. The findings from correlation analyses further suggest that dysfunctions in higher auditory cognition (i.e., processing sequential auditory information) might be one of the characteristics of auditory cognitive deficits in the CL group.

From the classification of haemodynamic responses, we found that the channel features in this study were intriguing. Firstly, $HbO_2$ values obtained from the left DLPFC (CH13, CH14, CH16) in the CL group were lower as compared to the HC group, indicating that patient group consumed less cognitive resources during the CIT performances than healthy adult group. DLPFC regions modulate higher order cognitive systems, such as working memory, decision-making, and problem solving (Amodio et al., 2004; Felipo et al., 2012; Jao et al., 2015). Also, this region is progressively impaired as liver disease advances (Chen et al., 2014), which can be a result of decreased blood flow and reduced glucose uptake activity in the frontal region (Lockwood et al., 1991, 2002). Ni et al. reported that patients with liver disease showed decreased functional coherence in the bilateral prefrontal cortex as the disease progressed (Felipo et al., 2012; Ni et al., 2012; Zhang et al., 2012). Taken together, deactivation in the left DLPFC observed in patients with chronic liver disease seemed to bring poor CIT performance, which can be indicative of cognitive alterations and inattention. Current inactivity of $HbO_2$ in the DLPFC, thus, can be a useful feature to classify patients with chronic liver disease from healthy adults.

In addition, $HbO_2$ values obtained from the right FP (CH7, CH10) were greater in the CL group as compared to HC group. The right frontopolar area, which corresponds to BA10, is involved in a variety of cognitive performances, ranging from simple to highly complex tasks (Burgess et al., 2007; Turner et al., 2008). This area receives projections from almost all processing levels in the superior temporal gyrus (Poremba et al., 2004; Kusmierek and Rauschecker, 2009; Kikuchi et al., 2010), and is known for its engagement in abstract representations of auditory information in organized thought (Medalla and Barbas, 2014). Further, lower activations in DLPFC and greater activation in FP in CL than HC group together indicated that healthy adults seemed to assign their cognitive resources on a higher-order cognitive function while patients with CL placed them in a more general attention and auditory cognition. The current findings possibly suggested a compensatory mechanism that reflects the recruitment and reallocation of cognitive resources due to the liver disease (Qi et al., 2012). Lastly, $HbO_2$ values obtained from the CH6 were greater in the CL than HC group. CH6 receives signals from the FP (BA10), as well as the OFC (BA11). Note that BA11 is known for its role in emotional behaviors, especially in evaluating the emotional valence of external stimuli (Rolls, 2004; Powell et al., 2017). Huang et al. (2016) for example reported that the level of BA11 activation is indicative of an individuals' aesthetic experience. Greater activation CH6 in the CL than in the HC were, thus, possibly due to our participants' aversion or appetite for the auditory stimuli given in CITs rather than features of specific alterations that could be used to diagnose cirrhotic chronic liver disease.

## LIMITATION

The results of this study provide new information regarding the use multiple prefrontal area channels in the diagnosis of cognitive alterations in chronic liver disease. Our study preliminarily utilized auditory/music processing as an evaluation task in chronic liver disease, but it includes a small sample size ($N = 55$) and a skewed female-to-male ratio. Some previous studies have reported gender difference issues in patients with chronic liver disease (Tsai et al., 2015; Barreira et al., 2019). This issue remains controversial, and findings differ depending on the subtypes of cognitive functions and the types of tasks. In this study, the comparison of cognitive functions between males and females was limited by the insufficient sample size. In future upscaled studies, we will directly address this issue by controlling the sample size and gender ratio to investigate the possible influence of gender difference on cognitive functions in chronic liver disease. It is alsonecessary to confirm that auditory attention is affected in chronic liver disease and that it has potential as a biomarker for MHE detection.

Also, there is little doubt that the important aspects of attention and cognition are associated with other brain regions. As the full associations between attention and cognition were not comprehensively covered due to the physical limitations of our fNIRS device, the caveats of our study should be considered when inferring its relationship with overall activations in the cortex. Emergent advances in fNIRS technology that provide more channels should make it possible to cover more regions of the brain in future studies to examine compensatory mechanisms for cognitive alterations in chronic liver disease.

## CONCLUSIONS

In this study, we applied an SVM model with an exhaustive method, as suggested by Ichikawa et al. (2014), to classify the haemodynamic responses during auditory perception. This method was performed successfully and yielded chronic liver disease-specific channel features. Given that the majority of assessment stimuli are visual, these findings implicated the importance of auditory processing in evaluating cognitive alterations in chronic liver disease (Mehndiratta et al., 1990; Sawhney et al., 1997; Saxena et al., 2001; Moon et al., 2012). Second, by virtue of recent brain imaging technology, such as the fNIRS, changes in oxygenated hemoglobin in the prefrontal areas were examined in a time and cost-effective manner. There were several channels that differentiated group-specific cognitive alterations that were reflected in the auditory/music perception. The left DLPFC and frontopolar areas played a task-specific role, while the right DLPFC played a modality- and stimulus-specific role in classification. Lastly, SVMs combined with an exhaustive search method was effective in classifying multivariate haemodynamic data since it allowed us to extrapolate an optimized combination from all possible combinations. Auditory/music perception tasks identified cognitive alterations in chronic liver diseases, which is frequently observed but not yet clearly explained; thus, this method combined with CITs could potentially serve as a supplementary evaluation of cognitive functions in the early detection of HE.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of Hanyang Medical Centre (HYUH-2013-08-017-002). The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

GJ and EJ conceived and designed the experiments. DJ recruited participants and arranged individual experiments. EJ performed the experiments and edited the manuscript. GJ, Y-MK, DJ, and EJ analyzed, interpreted the data, and wrote the manuscript. GJ and EJ prepared tables and figures. All authors reviewed the manuscript.

## FUNDING

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A5A2A03034582).

## REFERENCES

Agostini, G., Longari, M., and Pollastri, E. (2003). Musical instrument timbres classification with spectral features. *Eurasip J. Appl. Signal Proces.* 1:943279. doi: 10.1155/S1110865703210118

Akgül, C. B., Sankur, B., and Akin, A. (2005). Spectral analysis of event-related hemodynamic responses in functional near infrared spectroscopy. *J. Comput. Neurosci.* 18, 67–83. doi: 10.1007/s10827-005-5478-2

Amodio, P., Montagnese, S., Gatta, A., and Morgan, M. Y. (2004). Characteristics of minimal hepatic encephalopathy. *Metab. Brain Dis.* 19, 253–267. doi: 10.1023/B:MEBR.0000043975.01841.de

Bajaj, J. S., Hafeezullah, M., Franco, J., Varma, R. R., Hoffmann, R. G., Knox, J. F., et al. (2008). Inhibitory control test for the diagnosis of minimal hepatic encephalopathy. *Gastroenterology* 135, 1591–1600 e1. doi: 10.1053/j.gastro.2008.07.021

Bajaj, J. S., Hafeezullah, M., Hoffmann, R. G., and Saeian, K. (2007a). Minimal hepatic encephalopathy: a vehicle for accidents and traffic violations. *Am. J. Gastroenterol.* 102, 1903–1909. doi: 10.1111/j.1572-0241.2007.01424.x

Bajaj, J. S., Saeian, K., Verber, M. D., Hischke, D., Hoffmann, R. G., Franco, J., et al. (2007b). Inhibitory control test is a simple method to diagnose minimal hepatic encephalopathy and predict development of overt hepatic encephalopathy. *Am. J. Gastroenterol.* 102, 754–760. doi: 10.1111/j.1572-0241.2007.01048.x

Bajaj, J. S., Wade, J. B., and Sanyal, A. J. (2009). Spectrum of neurocognitive impairment in cirrhosis: implications for the assessment of hepatic encephalopathy. *Hepatology* 50, 2014–2021. doi: 10.1002/hep.23216

Baker, W. B., Parthasarathy, A. B., Busch, D. R., Mesquita, R. C., Greenberg, J. H., and Yodh, A. G. (2014). Modified Beer-Lambert law for blood flow. *Biomed. Opt. Express* 5:4053. doi: 10.1364/BOE.5.004053

Barreira, D. P., Marinho, R. T., Bicho, M., Flores, I., Fialho, R., and Ouakinin, S. (2019). Hepatitis C pretreatment profile and gender differences: cognition and disease severity effects. *Front. Psychol.* 10, 1–9. doi: 10.3389/fpsyg.2019.02317

Bauernfeind, G., Scherer, R., Pfurtscheller, G., and Neuper, C. (2011). Single-trial classification of antagonistic oxyhemoglobin responses during mental arithmetic. *Med. Biol. Eng. Comput.* 49:979. doi: 10.1007/s11517-011-0792-5

Bennett, K. P., and Campbell, C. (2000). Support vector machines: hype or hallelujah? *Acm Sigkdd Explor. Newsl.* 2, 1–13. doi: 10.1145/380995.380999

Bernthal, P., Hays, A., Tarter, R. E., Van Thiel, D., Lecky, J., and Hegedus, A. (1987). Cerebral CT scan abnormalities in cholestatic and hepatocellular disease and their relationship to neuropsychologic test performance. *Hepatology* 7, 107–114, doi: 10.1002/hep.1840070122

Boughorbel, S., Jarray, F., and El-Anbari, M. (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS ONE* 12:e0177678. doi: 10.1371/journal.pone.0177678

Brodersen, C., Koen, E., Ponte, A., Sánchez, S., Segal, E., Chiapella, A., et al. (2014). Cognitive function in patients with alcoholic and nonalcoholic chronic liver disease. *J. Neuropsychiatry Clin. Neurosci.* 26, 241–248, doi: 10.1176/appi.neuropsych.12040091

Brodersen, K. H., Ong, C. S., Stephan, K. E., and Buhmann, J. M. (2010). "The balanced accuracy and its posterior distribution," in *Pattern recognition (ICPR), 2010 20th international conference on* (IEEE) (Istanbul), 3121–3124. doi: 10.1109/ICPR.2010.764

Burgess, P. W., Gilbert, S. J., and Dumontheil, I. (2007). Function and localization within rostral prefrontal cortex (area 10). *Philos. Trans. R. Soc. London B Biol. Sci.* 362, 887–899. doi: 10.1098/rstb.2007.2095

Bustamante, J., Rimola, A., Ventura, P. J., Navasa, M., Cirera, I., Reggiardo, V., et al. (1999). Prognostic significance of hepatic encephalopathy in patients with cirrhosis. *J. Hepatol.* 30, 890–895, doi: 10.1016/S0168-8278(99)80144-5

Butterworth, R. F. (2000). Complications of cirrhosis III. Hepatic encephalopathy. *J. Hepatol.* 32(Supp.1), 171–180. doi: 10.1016/S0168-8278(00)80424-9

Chen, H.-J., Jiao, Y., Zhu, X.-Q., Zhang, H.-Y., Liu, J.-C., Wen, S., et al. (2013). Brain dysfunction primarily related to previous overt hepatic encephalopathy compared with minimal hepatic encephalopathy: resting-state functional MR imaging demonstration. *Radiology* 266, 261–270. doi: 10.1148/radiol.12120026

Chen, H.-J., Wang, Y., Zhu, X.-Q., Li, P.-C., and Teng, G.-J. (2014). Classification of cirrhotic patients with or without minimal hepatic encephalopathy and healthy subjects using resting-state attention-related network analysis. *PLoS ONE* 9:e89684. doi: 10.1371/journal.pone.0089684

Ciećko-Michalska, I., Senderecka, M., Szewczyk, J., Panasiuk, A., Słowik, A., Wyczesany, M., et al. (2006). Event-related cerebral potentials for the diagnosis of subclinical hepatic encephalopathy in patients with liver cirrhosis. *Adv. Med. Sci.* 51, 273–277.

Conn, H. O., Leevy, C. M., Vlahcevic, Z. R., Rodgers, J. B., Maddrey, W. C., Seeff, L., et al. (1977). Comparison of lactulose and neomycin in the treatment of chronic portal-systemic encephalopathy: a double blind controlled trial. *Gastroenterology* 72, 573–583. doi: 10.1016/S0016-5085(77)80135-2

Cui, X., Bray, S., and Reiss, A. L. (2010). Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage* 49, 3039–3046. doi: 10.1016/j.neuroimage.2009.11.050

Felipo, V., Ordoño, J. F., Urios, A., El Mlili, N., Giménez-Garz,ó, C., Aguado, C., et al. (2012). Patients with minimal hepatic encephalopathy show impaired mismatch negativity correlating with reduced performance in attention tests. *Hepatology* 55, 530–539. doi: 10.1002/hep.24704

Ferenci, P., Lockwood, A., Mullen, K., Tarter, R., Weissenborn, K., and Blei, A. T. (2002). Hepatic encephalopathy—definition, nomenclature, diagnosis, and quantification. *Hepatology* 35, 716–721. doi: 10.1053/jhep.2002.31250

Filipović, B., Marković, O., Urić, V., and Filipović, B. (2018). Cognitive changes and brain volume reduction in patients with nonalcoholic fatty liver disease. *Can. J. Gastroenterol. Hepatol.* 2018:9638797. doi: 10.1155/2018/9638797

Folstein, J. R., and Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 45, 152–170. doi: 10.1111/j.1469-8986.2007.00602.x

Groeneweg, M., Quero, J. C., De Bruijn, I., Hartmann, I. J. C., Essink-Bot, M. L., Hop, W. C. J., et al. (1998). Subclinical hepatic encephalopathy impairs daily functioning. *Hepatology* 28, 45–49. doi: 10.1002/hep.510280108

Gupta, D., Ingle, M., Shah, K., Phadke, A., and Sawant, P. (2015). Prospective comparative study of inhibitory control test and psychometric hepatic encephalopathy score for diagnosis and prognosis of minimal hepatic encephalopathy in cirrhotic patients in the Indian subcontinent. *J. Dig. Dis.* 16, 400–407. doi: 10.1111/1751-2980.12248

Hartmann, I. J. C., Groeneweg, M., Quero, J. C., Beijeman, S. J., De Man, R. A., Hop, W. C. J., et al. (2000). The prognostic significance of subclinical hepatic encephalopathy. *Am. J. Gastroenterol.* 95, 2029–2034. doi: 10.1111/j.1572-0241.2000.02265.x

Herff, C., Heger, D., Putze, F., Hennrich, J., Fortmann, O., and Schultz, T. (2013). "Classification of mental tasks in the prefrontal cortex using fNIRS," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE* (Osaka: IEEE), 2160–2163. doi: 10.1109/EMBC.2013.6609962

Huang, P., Huang, H., Luo, Q., and Mo, L. (2016). The difference between aesthetic appreciation of artistic and popular music: evidence from an fMRI study. *PLoS ONE* 11:e0165377. doi: 10.1371/journal.pone.0165377

Ichikawa, H., Kitazono, J., Nagata, K., Manda, A., Shimamura, K., Sakuta, R., et al. (2014). Novel method to classify hemodynamic response obtained using multi-channel fNIRS measurements into two groups: exploring the combinations of channels. *Front. Hum. Neurosci.* 8:480. doi: 10.3389/fnhum.2014.00480

Jao, T., Schröter, M., Chen, C.-L., Cheng, Y.-F., Lo, C.-Y. Z., Chou, K.-H., et al. (2015). Functional brain network changes associated with clinical and biochemical measures of the severity of hepatic encephalopathy. *Neuroimage* 122, 332–344. doi: 10.1016/j.neuroimage.2015.07.068

Jeong, E. (2013). Psychometric validation of a music-based attention assessment: revised for patients with traumatic brain injury. *J. Music Ther.* 50, 66–92. doi: 10.1093/jmt/50.2.66

Jeong, E., and Lesiuk, T. L. (2011). Development and preliminary evaluation of a music-based attention assessment for patients with traumatic brain injury. *J. Music Ther.* 48, 551–572. doi: 10.1093/jmt/48.4.551

Jeong, E., and Ryu, H. (2016). Melodic contour identification reflects the cognitive threshold of aging. *Front. Aging Neurosci.* 8:134. doi: 10.3389/fnagi.2016.00134

Jeong, E., Ryu, H., Jo, G., and Kim, J. (2018). Cognitive load changes during music listening and its implication in earcon design in public environments: an fNIRS study. *Int. J. Environ. Res. Public Health* 15:2075. doi: 10.3390/ijerph15102075

Jiao, Y., and Du, P. (2016). Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quant. Biol.* 4, 320–330. doi: 10.1007/s40484-016-0081-2

Keiding, S., and Pavese, N. (2013). Brain metabolism in patients with hepatic encephalopathy studied by PET and MR. *Arch. Biochem. Biophys.* 536, 131–142. doi: 10.1016/j.abb.2013.05.006

Kikuchi, Y., Horwitz, B., and Mishkin, M. (2010). Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *J. Neurosci.* 30, 13021–13030. doi: 10.1523/JNEUROSCI.2267-10.2010

Kim, J.-H. (2009). Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap. *Comput. Stat. Data Anal.* 53, 3735–3745. doi: 10.1016/j.csda.2009.04.009

Kircheis, G., Hilger, N., and Häussinger, D. (2014). Value of critical flicker frequency and psychometric hepatic encephalopathy score in diagnosis of low-grade hepatic encephalopathy. *Gastroenterology* 146, 961–969 e11. doi: 10.1053/j.gastro.2013.12.026

Krieger, S., Jauss, M., Jansen, O., Theilmann, L., Geissler, M., and Krieger, D. (1996). Neuropsychiatric profile and hyperintense globus pallidus on T1-weighted magnetic resonance images in liver cirrhosis. *Gastroenterology* 111, 147–155. doi: 10.1053/gast.1996.v111.pm8698193

Kusmierek, P., and Rauschecker, J. P. (2009). Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J. Neurophysiol.* 102, 1606–1622. doi: 10.1152/jn.00167.2009

Li, Z., Wang, Y., Quan, W., Wu, T., and Lv, B. (2015). Evaluation of different classification methods for the diagnosis of schizophrenia based on functional near-infrared spectroscopy. *J. Neurosci. Methods* 241, 101–110. doi: 10.1016/j.jneumeth.2014.12.020

Lockwood, A. H., Weissenborn, K., Bokemeyer, M., Tietge, U., and Burchert, W. (2002). Correlations between cerebral glucose metabolism and neuropsychological test performance in nonalcoholic cirrhotics. *Metab. Brain Dis.* 17, 29–40. doi: 10.1023/A:1014000313824

Lockwood, A. H., Yap, E. W. H., Rhoades, H. M., and Wong, W.-H. (1991). Altered cerebral blood flow and glucose metabolism in patients with liver disease and minimal encephalopathy. *J. Cereb. Blood Flow Metab.* 11, 331–336. doi: 10.1038/jcbfm.1991.66

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *J. Neural Eng.* 4:R1. doi: 10.1088/1741-2560/4/2/R01

Macias-Rodriguez, R. U., Ilarraza-Lomelí, H., Ruiz-Margáin, A., Ponce-de-León-Rosales, S., Vargas-Vorácková, F., García-Flores, O., et al. (2016). Changes in hepatic venous pressure gradient induced by physical exercise in cirrhosis: results of a pilot randomized open clinical trial. *Clin. Transl. Gastroenterol.* 7:e180. doi: 10.1038/ctg.2016.38

Medalla, M., and Barbas, H. (2014). Specialized prefrontal "auditory fields": organization of primate prefrontal-temporal pathways. *Front. Neurosci.* 8:77. doi: 10.3389/fnins.2014.00077

Mehndiratta, M. M., Sood, G. K., Sarin, S. K., and Gupta, M. (1990). Comparative evaluation of visual, somatosensory, and auditory evoked potentials in the detection of subclinical hepatic encephalopathy in patients with nonalcoholic cirrhosis. *Am. J. Gastroenterol.* 85, 799–803.

Mendonça, E. B. S., Muniz, L. F., de Carvalho Leal, M., and da Silva Diniz, A. (2013). Applicability of the P300 frequency pattern test to assess auditory processing. *Braz. J. Otorhinolaryngol.* 79, 512–521. doi: 10.5935/1808-8694.20130091

Mishra, D., and Sahu, B. (2011). Feature selection for cancer classification: a signal-to-noise ratio approach. *Int. J. Sci. Eng. Res.* 2, 1–7.

Monden, Y., Dan, I., Nagashima, M., Dan, H., Uga, M., Ikeda, T., et al. (2015). Individual classification of ADHD children by right prefrontal hemodynamic responses during a go/no-go task as assessed by fNIRS. *NeuroImage Clin.* 9, 1–12. doi: 10.1016/j.nicl.2015.06.011

Moon, J. H., Jun, D. W., Yum, M. K., Lee, K. N., Lee, H. L., Lee, O. Y., et al. (2014). Prolonged N200 is the early neurophysiologic change in the patient with minimal hepatic encephalopathy. *Scand. J. Gastroenterol.* 49, 604–610. doi: 10.3109/00365521.2013.878382

Moon, S.-S., Lee, Y.-S., and Kim, S. W. (2012). Association of nonalcoholic fatty liver disease with low bone mass in postmenopausal women. *Endocrine* 42, 423–429. doi: 10.1007/s12020-012-9639-6

Morren, G., Wolf, M., Lemmerling, P., Wolf, U., Choi, J. H., Gratton, E., et al. (2004). Detection of fast neuronal signals in the motor cortex from functional near infrared spectroscopy measurements using independent component analysis. *Med. Biol. Eng. Comput.* 42, 92–99. doi: 10.1007/BF02351016

Mourao-Miranda, J., Bokde, A. L. W., Born, C., Hampel, H., and Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. *Neuroimage* 28, 980–995. doi: 10.1016/j.neuroimage.2005.06.070

Nakanishi, H., Kurosaki, M., Nakanishi, K., Tsuchiya, K., Noda, T., Tamaki, N., et al. (2014). Impaired brain activity in cirrhotic patients with minimal hepatic encephalopathy: Evaluation by near-infrared spectroscopy. *Hepatol. Res.* 44, 319–326. doi: 10.1111/hepr.12127

Ni, L., Qi, R., Zhang, L. J., Zhong, J., Zheng, G., Zhang, Z., et al. (2012). Altered regional homogeneity in the development of minimal hepatic encephalopathy: a resting-state functional MRI study. *PLoS ONE* 7:e42016. doi: 10.1371/journal.pone.0042016

Ogawa, Y., Kotani, K., and Jimbo, Y. (2014). Relationship between working memory performance and neural activation measured using near-infrared spectroscopy. *Brain Behav.* 4, 544–551. doi: 10.1002/brb3.238

Ortiz, M., Jacas, C., and Córdoba, J. (2005). Minimal hepatic encephalopathy: diagnosis, clinical significance and recommendations. *J. Hepatol.* 42(Supp.1), S45–S53, doi: 10.1016/j.jhep.2004.11.028

Peck, E. M. M., Yuksel, B. F., Ottley, A., Jacob, R. J. K., and Chang, R. (2013). "Using fNIRS brain sensing to evaluate information visualization interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (ACM)* (Paris), 473–482. doi: 10.1145/2470654.2470723

Poremba, A., Malloy, M., Saunders, R. C., Carson, R. E., Herscovitch, P., and Mishkin, M. (2004). Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature* 427, 448–451. doi: 10.1038/nature02268

Powell, J. L., Grossi, D., Corcoran, R., Gobet, F., and Garcia-Finana, M. (2017). The neural correlates of theory of mind and their role during empathy and the game of chess: a functional magnetic resonance imaging study. *Neuroscience* 355, 149–160. doi: 10.1016/j.neuroscience.2017.04.042

Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* 2, 37–63.

Prasad, S., Dhiman, R. K., Duseja, A., Chawla, Y. K., Sharma, A., and Agarwal, R. (2007). Lactulose improves cognitive functions and health-related quality of life in patients with cirrhosis who have minimal hepatic encephalopathy. *Hepatology* 45, 549–559. doi: 10.1002/hep.21533

Qi, R., Zhang, L. J., Xu, Q., Zhong, J., Wu, S., Zhang, Z., et al. (2012). Selective impairments of resting-state networks in minimal hepatic encephalopathy. *PLoS ONE* 7:e37400. doi: 10.1371/journal.pone.0037400

Refaeilzadeh, P., Tang, L., and Liu, H. (2009). "Cross-validation," in *Encyclopedia of Database Systems*, eds L. Liu and M. T. Özsu (New York, NY: Springer), 532–538. doi: 10.1007/978-0-387-39940-9_565

Rolls, E. T. (2004). The functions of the orbitofrontal cortex. *Brain Cogn.* 55, 11–29. doi: 10.1016/S0278-2626(03)00277-X

Romero-Gómez, M., Boza, F., Garcia-Valdecasas, M. S., Garcia, E., and Aguilar-Reina, J. (2001). Subclinical hepatic encephalopathy predicts the development of overt hepatic encephalopathy. *Am. J. Gastroenterol.* 96, 2718–2723. doi: 10.1111/j.1572-0241.2001.04130.x

Romero-Gómez, M., Córdoba, J., Jover, R., Del Olmo, J. A., Ramírez, M., Rey, R., et al. (2007). Value of the critical flicker frequency in patients with minimal hepatic encephalopathy. *Hepatology* 45, 879–885. doi: 10.1002/hep.21586

Sakiyama, Y., Yuki, H., Moriya, T., Hattori, K., Suzuki, M., Shimada, K., et al. (2008). Predicting human liver microsomal stability with machine learning techniques. *J. Mol. Graph. Model.* 26, 907–915. doi: 10.1016/j.jmgm.2007.06.005

Sánchez-Carrión, R., Gómez, P. V., Junqu,é, C., Fernández-Espejo, D., Falcon, C., Bargall,ó, N., et al. (2008). Frontal hypoactivation on functional magnetic resonance imaging in working memory after severe diffuse traumatic brain injury. *J. Neurotrauma* 25, 479–494. doi: 10.1089/neu.2007.0417

Sawhney, I. M. S., Verma, P. K., Dhiman, R. K., Chopra, J. S., Sharma, A., Chawla, Y. K., et al. (1997). Visual and auditory evoked responses in acute severe hepatitis. *J. Gastroenterol. Hepatol.* 12, 554–559. doi: 10.1111/j.1440-1746.1997.tb00484.x

Saxena, N., Bhatia, M., Joshi, Y. K., Garg, P. K., and Tandon, R. K. (2001). Auditory P300 event-related potentials and number connection test for evaluation of subclinical hepatic encephalopathy in patients with cirrhosis of the liver: a follow-up study. *J. Gastroenterol. Hepatol.* 16, 322–327. doi: 10.1046/j.1440-1746.2001.02388.x

Sharma, A., and Nagalli, S. (2020). *Chronic Liver Disease*. Treasure Island, FL: StatPearls Publishing. Available online at: http://europepmc.org/books/NBK554597 (accessed July 05, 2020).

Sharma, P., Kumar, A., Singh, S., Tyagi, P., and Kumar, A. (2013). Inhibitory control test, critical flicker frequency, and psychometric tests in the diagnosis of minimal hepatic encephalopathy in cirrhosis. *Saudi J. Gastroenterol. Off. J. Saudi Gastroenterol. Assoc.* 19, 40–44. doi: 10.4103/1319-3767.105924

Soltani, M., Fatemeh Emami Dehcheshmeh, S., Moradi, N., Hajiyakhchali, A., Majdinasab, N., Mahmood Latifi, S., et al. (2018). Comparing executive functions in bilinguals and monolinguals suffering from relapsing- remitting multiple sclerosis. *J. Mod. Rehabil. Sclerosis.*12, 133–139.

Stewart, C. A., Malinchoc, M., Kim, W. R., and Kamath, P. S. (2007). Hepatic encephalopathy as a predictor of survival in patients with end-stage liver disease. *Liver Transplant.* 13, 1366–1371. doi: 10.1002/lt.21129

Stinton, L. M., and Jayakumar, S. (2013). Minimal hepatic encephalopathy. *Can. J. Gastroenterol.* 27, 572–574. doi: 10.1155/2013/547670

Teodoro, V., Bragagnolo Jr, M., Lucchesi, L., Kondo, M., and Tufik, S. (2008). Evaluation of the event-related potential (ERP-P300) in patients with hepatic cirrhosis without encephalopathy. *Arq. Gastroenterol.* 45, 82–86. doi: 10.1590/S0004-28032008000100015

Tsai, C. F., Chu, C. J., Huang, Y. H., Wang, Y. P., Liu, P. Y., Lin, H. C., et al. (2015). Detecting minimal hepatic encephalopathy in an endemic country for hepatitis B: the role of psychometrics and serum IL-6. *PLoS ONE* 10, 1–11. doi: 10.1371/journal.pone.0128437

Turner, M. S., Simons, J. S., Gilbert, S. J., Frith, C. D., and Burgess, P. W. (2008). Distinct roles for lateral and medial rostral prefrontal cortex in source monitoring of perceived and imagined events. *Neuropsychologia* 46, 1442–1453. doi: 10.1016/j.neuropsychologia.2007.12.029

Van't Veer, L. J., Dai, H., Van De Vijver, M. J., He, Y. D., Hart, A. A. M., Mao, M., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530–536. doi: 10.1038/415530a

Weissenborn, K., Ennen, J. C., Schomerus, H., Rückert, N., and Hecker, H. (2001a). Neuropsychological characterization of hepatic encephalopathy. *J. Hepatol.* 34, 768–773. doi: 10.1016/S0168-8278(01)00026-5

Weissenborn, K., Giewekemeyer, K., Heidenreich, S., Bokemeyer, M., Berding, G., and Ahl, B. (2005). Attention, memory, and cognitive function in hepatic encephalopathy. *Metab. Brain Dis.* 20, 359–367. doi: 10.1007/s11011-005-7919-z

Weissenborn, K., Heidenreich, S., Ennen, J., Hecker, H. (2001b). Attention deficits in minimal hepatic encephalopathy. *Metab. Brain. Dis.* 16, 13–19. doi: 10.1023/A:1011654210096

Yang, Z. T., Chen, H. J., Chen, Q. F., and Lin, H. (2018). Disrupted brain intrinsic networks and executive dysfunction in cirrhotic patients without overt hepatic encephalopathy. *Front. Neurol.* 9:14. doi: 10.3389/fneur.2018.00014

Yasumura, A., Inagaki, M., and Hiraki, K. (2014). Relationship between neural activity and executive function: an NIRS study. *ISRN Neurosci.* 2014:734952. doi: 10.1155/2014/734952

Yoon, U., Lee, J.-M., Im, K., Shin, Y.-W., Cho, B. H., Kim, I. Y., et al. (2007). Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. *Neuroimage* 34, 1405–1415. doi: 10.1016/j.neuroimage.2006.11.021

Zhan, T., and Stremmel, W. (2012). The diagnosis and treatment of minimal hepatic encephalopathy. *Dtsch. Arztebl. Int.* 109, 180–187. doi: 10.3238/arztebl.2012.0180

Zhang, L., Qi, R., Wu, S., Zhong, J., Zhong, Y., Zhang, Z., et al. (2012). Brain default-mode network abnormalities in hepatic encephalopathy: a resting-state functional MRI study. *Hum. Brain Mapp.* 33, 1384–1392. doi: 10.1002/hbm.21295

**Conflict of Interest:** GJ was employed by the company Daehong Communications Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 15 October 2020
doi: 10.3389/fnhum.2020.576888

# Mental Effort When Playing, Listening, and Imagining Music in One Pianist's Eyes and Brain

*Tor Endestad[1,2,3], Rolf Inge Godøy[2], Markus Handal Sneve[1], Thomas Hagen[1], Agata Bochynska[1,4] and Bruno Laeng[1,2]**

[1]Department of Psychology, University of Oslo, Oslo, Norway, [2]RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway, [3]Helgelandssykehuset, Mosjøen, Norway, [4]Department of Psychology, New York University, New York, NY, United States

We investigated "musical effort" with an internationally renowned, classical, pianist while playing, listening, and imagining music. We used pupillometry as an objective measure of mental effort and fMRI as an exploratory method of effort with the same musical pieces. We also compared a group of non-professional pianists and non-musicians by the use of pupillometry and a small group of non-musicians with fMRI. This combined approach of psychophysiology and neuroimaging revealed the cognitive work during different musical activities. We found that pupil diameters were largest when "playing" (regardless of whether there was sound produced or not) compared to conditions with no movement (i.e., "listening" and "imagery"). We found positive correlations between pupil diameters of the professional pianist during different conditions with the same piano piece (i.e., normal playing, silenced playing, listen, imagining), which might indicate similar degrees of load on cognitive resources as well as an intimate link between the motor imagery of sound-producing body motions and gestures. We also confirmed that musical imagery had a strong commonality with music listening in both pianists and musically naïve individuals. Neuroimaging provided evidence for a relationship between noradrenergic (NE) activity and mental workload or attentional intensity within the domain of music cognition. We found effort related activity in the superior part of the locus coeruleus (LC) and, similarly to the pupil, the listening and imagery engaged less the LC–NE network than the motor condition. The pianists attended more intensively to the most difficult piece than the non-musicians since they showed larger pupils for the most difficult piece. Non-musicians were the most engaged by the music listening task, suggesting that the amount of attention allocated for the same task may follow a hierarchy of expertise demanding less attentional effort in expert or performers than in novices. In the professional pianist, we found only weak evidence for a commonality between subjective effort (as rated measure-by-measure) and the objective effort gauged with pupil diameter during listening. We suggest that psychophysiological methods like pupillometry can index mental effort in a manner that is not available to subjective awareness or introspection.

**Keywords: imagery, music, audiation, audio-visual integration, pupillometry, fMRI**

OPEN ACCESS

***Edited by:***
*Marta Olivetti Belardinelli, Sapienza University of Rome, Italy*

***Reviewed by:***
*Enrica Laura Santarcangelo, University of Pisa, Italy*
*Chun Meng, University of Electronic Science and Technology of China, China*

****Correspondence:***
*Bruno Laeng*
*bruno.laeng@psykologi.uio.no*

***Specialty section:***
*This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience*

***Received:*** *27 June 2020*
***Accepted:*** *07 September 2020*
***Published:*** *15 October 2020*

***Citation:***
*Endestad T, Godøy RI, Sneve MH, Hagen T, Bochynska A and Laeng B (2020) Mental Effort When Playing, Listening, and Imagining Music in One Pianist's Eyes and Brain. Front. Hum. Neurosci. 14:576888. doi: 10.3389/fnhum.2020.576888*

# INTRODUCTION

## Musical Effort

The piano is not an easy instrument to master. It requires a lifetime of extended practice and early (possibly during childhood) onset of training to achieve a high level of performance. Well-known musical pieces in either classical music or jazz are technically particularly challenging, and they constitute a paradigmatic example of intense practice of cognitive executive functions (Montero, 2016). There is the need of dividing attention over the control of complex, synchronized, sequences of finger positions and movements (Mikumo, 1994) and paying attention to the obtained auditory stimuli. Such a fine and precise motor control of hands and fingers require cognitive and motor control, guiding constant adjustments of bodily actions for both the execution and preparation of the following movements during the different passages of a musical piece, also depending on the level of cognitive and motoric demands in performing the movements. Hence, we assume that playing the piano—or any musical instrument professionally—requires mental resources or "mental effort," as Kahneman (1973) labeled it originally. Such a type of neurocognitive effort is distinguishable from physical effort, though it has clear analogies with it.

Surprisingly, research on "musical effort" or the broad process of allocation of attentional resources during music performance or listening (Keller, 2001; Shenhav et al., 2017) has been so far an understudied aspect of music cognition. In cognitive psychology, pupillometry or the measurement of pupil dilations during a task has been considered the best psychophysiological measure (Kahneman et al., 1969) of the "intensive aspect of attention" or cognitive workload (e.g., Hess and Polt, 1964; Beatty, 1982; Just and Carpenter, 1993). Changes in pupil diameter are not simply evoked by changes in light stimulating the eye, and changes related to mental processing, albeit tiny compared to those provoked by light, are separable. Most importantly, these pupillary changes provide a reliable, "honest" (i.e., difficult to affect voluntarily; see Laeng and Sulutvedt, 2014), and valid measure of the overall aggregate of resource demand and capacity utilization by the brain (Just et al., 2003). A wide variety of pupillometry studies confirm a tight relationship between pupillary dilation and the allocation of attention and load on cognitive resources (for reviews see Laeng et al., 2012; Laeng and Alnæs, 2019). Recently, neurophysiological studies with monkeys (e.g., Joshi et al., 2016) and neuroimaging in humans have indicated the involvement of the noradrenergic (NE) brainstem structure called the locus coeruleus (LC) in the control of pupil size during cognitive work (e.g., Alnæs et al., 2014; Murphy et al., 2014; Mäki-Marttunen et al., 2019). Both pupil diameter and the LC have been known as a physiological measure and structure involved in both cognitive and affective "arousal" (e.g., Nunnally et al., 1967; Libby et al., 1973; Nassar et al., 2012; Sara and Bouret, 2012).

Notably, even just music listening triggers high arousal (e.g., Gingras et al., 2015; Laeng et al., 2016; Weiss et al., 2016; Bowling et al., 2019). To our knowledge, no studies have applied eye pupillometry to the study of "musical effort" during the performance. Instead, there have been a few pupillometry studies on mental effort during the perceptual processing of auditory stimuli like speech (e.g., Zekveld et al., 2014) and music (e.g., Kang and Wheatley, 2015; O'Shea and Moran, 2016; Liao et al., 2018). Moreover, it remains unclear whether the cognitive mechanisms engaged during music listening are equally engaged during musical auditory images (e.g., when a musician "plays in her head" a musical piece) than with music perception or when listening to the same musical piece played on a real instrument. Research within cognitive psychology had mainly focused on visual imagery (e.g., Kosslyn, 1980, 1994), which has overall confirmed a strong overlap of the cognitive mechanisms engaged by both imagery and perception.

## Musical Imagery

Musical imagery consists of actively evoking and maintaining sound "images" of music in our minds, like mental imagery in other specific modalities (Godøy and Jørgensen, 2001). However, musical imagery would seem to be a process at the very heart of not only music-making but also listening, as an ongoing anticipatory activity (e.g., Janata, 2001; Leaver et al., 2009; Gracyk, 2019). An interest in the topic has been rekindled within the last 20 years by several neuroimaging studies investigating the effects of music perception and imagery in the brains of both musicians and non-musicians (e.g., Zatorre, 1999; Lotze et al., 2003; Herholz et al., 2008). Recently, most references to musical imagery have focused on "involuntary musical imagery" (IMI, sometimes called "earworms," "sticky music," "catchy tunes," or simply "hooks"; e.g., Sacks, 2007; Farrugia et al., 2015; Williams, 2015; Moseley et al., 2018). However, in this article, we will focus on the "volitional" or active type of imagery of musical sounds and report a psychophysiological plus neuroimaging study on imagery, listening, and performance by a professional, internationally renowned, pianist.

### Musical Imagery as a Multimodal Experience

Importantly, the phenomenology of music-making and listening is not just related to sounds or finger movements but it is a multisensory or multimodal experience (Vines et al., 2006; Godøy, 2010b; Zimmerman and Lahav, 2012; Fine et al., 2015). The musical imagery of musicians seems particularly embedded in complex action plans, co-articulated executive programs, and predictions of the effects of actions both within the body and in the environment (e.g., Reybrouck, 2001; Wolpert and Flanagan, 2001). Musicians seem capable of constructing "sonic images" when playing, based on timbral and timing features (Wöllner and Williamon, 2007), allowing them to anticipate sonic actions and even perform without auditory feedback (e.g., with sound is switched off). However, also non-musicians appear to have an intuitive and coarsely correct understanding of the visual features linked to playing a particular musical instrument and of its related specific gestures or "gestural affordances" (Godøy and Leman, 2010; Godøy, 2010a); as also testified by the amusing ability of non-musicians in "playing" virtual instruments (e.g., "air-guitar"; see Godøy, 2001; Godøy et al., 2006). Just the sight of the musicians' gestures (with no sound) can influence our understanding and evaluation of music (e.g., Platz and

Kopiez, 2012; Tsay, 2013) and it also triggers activity in the auditory cortex of the observers' brain (Haslinger et al., 2005), probably indicating that the sight of music-making evokes in turn spontaneous musical (auditory) imagery.

Although we can in principle study in isolation each modality-specific type, given the underlying modularity of the sensory and motor system, there is likely to be a great interconnectedness between processes and structures of the brain for musical imagery. The current picture is that the auditory type of imagery is active together with the motor imagery of the sound-producing body motions. Moreover, the motor imagery seems imbued with somesthetic images, characteristic of the kinesthetic feedback that such bodily motions would typically produce (Betts, 1909; Hubbard and Stoeckig, 1988; Reisberg, 1992; Hodges, 2009; Hubbard, 2010, 2013, 2017, 2018; Vlek et al., 2011; Vuvan and Schmuckler, 2011). Indeed, perception, in general, is "active" and makes use of the motor system to achieve perceptual categorization or, at least, to facilitate it (e.g., Liberman et al., 1967; Liberman and Mattingly, 1985; Smith et al., 1995; Gallese and Metzinger, 2003; Gallese and Lakoff, 2005; Galantucci et al., 2006; Glenberg and Gallese, 2012). Embodied cognition accounts give prominence to action and behavior for perceptual processes of all kinds (e.g., Varela et al., 1991; Gallagher, 2005), including especially music (e.g., Cox, 2016). These accounts have influenced music psychology (Leman, 2008; Peñalba, 2011; Schiavio et al., 2014; Korsakova-Kreyn, 2018; Bailes, 2019). We believe that studying volitional musical imagery is important because it could be the gateway to the more systematic exploitation of musical imagery in practical tasks such as composition, improvisation, and performance (Christensen, 2019).

### Musical "Audiation"

A remarkable phenomenon of musical imagery is the so-called notational "audiation" often reported by professional musicians; that is, the simple act of reading a musical score evokes auditory imagery of the music in reading's real-time (e.g., Schürmann et al., 2002; Battisti, 2007; Brodksy et al., 2003, 2008). Remarkably, the electrophysiological activity from the brain of trained musicians during note reading or the actual perception of notes is undistinguishable (Simoens and Tervaniemi, 2013). The ability to "hear" with the "mind"'s ear' or "thinking in sounds" (Combarieu, 1907) and "replay" virtual music with the "inner voice," could assist the making of creative compositions. A well-known case of the power of musical imagery in creating music is Ludwig van Beethoven who composed many of is most praised compositions (e.g., the last piano sonatas and string quartets, the Missa Solemnis, and the Ninth Symphony) while he was practically deaf due to an inner ear problem. Presumably, he was perfectly able to compose music because he could "hear the music in his head" (Jourdain, 1997; Zatorre and Halpern, 2005) despite his spared auditory cortex was unable to be stimulated by the actual sounds of musical instruments.

Hearing music in the head appears to be a ubiquitous experience (Cotter, 2019) but the role that mental control or effort plays in these experiences has not been addressed in current research in any thorough manner. That is, both professional musicians and music writers, as well as musically naïve individuals, have the power to start, stop, shape, and maintain in their head musical images unfolding in time (e.g., Zatorre, 1999; Janata, 2001; Cotter, 2019). However, all these mental "actions" are a form of cognitive work and require focused attention; they are likely to draw on mental resources and engage brain systems the control cognitive arousal (Alnæs et al., 2014). Indeed, musical "mental practice" (Coffman, 1990) is well known as an effective tool for enhancing memorization of music and refining performance (e.g., Driskell et al., 1994; Halpern et al., 2004; Highben and Palmer, 2004; Holmes, 2005; Lotze and Halsband, 2006; Cahn, 2008; Gregg et al., 2008; Keller, 2012; Halpern and Overy, 2019). Deliberately imagining music, both the sound-producing actions (e.g., finger movements in pianists) and the resulting musical sounds (Davidson-Kelly et al., 2015), seems almost as common as actually listening to music (respectively 32 and 44% of times when querying musicians; Bailes, 2006, 2015). According to Fine et al. (2015), about 70% of the classical music performers they surveyed state that they use regularly mental practice, and among these about 90% also experience musical imagery in the form of "audiation" (i.e., "hearing music in the head" while reading a score away from the instrument; Bishop et al., 2013). A well-known case of extensive use of the mental practice is that of the classical pianist Glenn Gould who mentally rehearsed a performance, without touching the piano for prolonged periods, leaving the actual testing of the "mechanics" of the finger movements on the piano to just the last period of preparation (Mesaros, 2008). In a seminal study, Repp (1999) compared the timing profiles of six pianists during a live performance and the "imagined performance" and found that the temporal fluctuations occurred in a very similar manner (i.e., they were positively correlated). Other experiments in "mental chronometry" of imagined music have yielded similar results (Wöllner and Williamon, 2007; Clark and Williamon, 2011; Clark et al., 2012).

## The Present Study

Because a piano piece can present cognitive challenges that rapidly vary during the music's stream, we expect that changes in pupillary diameter in the eyes of the pianist mirror the level of required effort as the music unfolds. We expect that the complete trace of pupillary changes will provide a continuous physiological measure of changes in control processes (executive and attentional) occurring in the brain of a high-level professional pianist as she focusses on the piece over time. We also expect that, to some extent, independent listeners would react to the perceived effort inherent in the music, likely more in musicians than in non-musicians. Hence, we also monitored the eye pupil in a "control" group of pianists and non-musicians, while they listened to the same piano renditions of the musical pieces by our professional pianist.

Importantly, the pupillometry method is currently considered not only a reliable gauge of cognitive workload but also as a window into the activity of the brain's NE arousal system and the involvement of brainstem structures like the LC, involved in the NE control of pupil size (e.g., in other mammals;

Joshi et al., 2016). Very few human neuroimaging studies have explored the role of the brainstem's NE structures and pupillary activity (Alnæs et al., 2014; Murphy et al., 2014; Mäki-Marttunen et al., 2019) and none have related the role of these brainstem's NE structures to music listening performance, and imagery. Hence, we further explored music listening, playing, and imagery, in the same professional pianist, during these tasks with functional MRI, seeking converging evidence to the engagement of mental effort by brainstem structures. Since musical imagery is a quintessential multisensory experience (auditory, kinesthetic, and visuospatial at least), playing music in one's head or imagery should reveal the strongest relationship with actual piano playing, since all of the components of music-making that takes place in the mind (brain) would be re-instated, despite the lack of behavioral enactment. Importantly, we expect to reveal converging evidence between the findings obtained with pupillometry (Experiment 1) and those obtained with neuroimaging (Experiment 2). Specifically, since we posit that the pupil index effort-related cognitive arousal, we expect that pupillary changes across conditions should also appear as changes in activity in the LC or the NE center of the brainstem (Alnæs et al., 2014). Moreover, processes that are similar in terms of pupillary activity and highly correlated should also appear similar in terms of cortical activity.

In sum, we hope to offer initial answers about some fundamental questions in music cognition: (a) Can we measure musical effort through the eye pupil similarly to reading out the cognitive workload and arousal in other domains? (b) Does musical imagery engage sensory and motor areas of the brain concerning the mental effort required by the complexity of the structure and execution of the imagined music? (c) Is musical imagery more similar to music listening in non-musicians in terms of neural networks than in musicians where it engages more motor aspects (i.e., it is multimodal)? (d) Can we in general deduce the degree of functional similarity between playing, imagining, and listening by comparing the activity in the NE system (as indexed by the pupil) and/or the overlap between sensory and motor neural networks in the whole brain? and (e) Specifically, can we reveal "audiation" in an expert musician by the similarity in which the pupil changes during musical listening and imagery?

### Pupillometry of Playing, Listening, and Imagining Music

Pupillometry measures "objectively" mental effort, but an experienced musician may be able to estimate "subjectively" how effortful a moment can be during the piano execution. Hence, we introduce a distinction between "subjective effort" and "objective effort". With the former, we mean what a participant judges to be the processing load based on her private experience, even when estimated *via* ordinal scales (e.g., the NASA-TLX; e.g., Chaffin, 2009). Instead, physiologically driven changes in pupil diameter are an objective measure of mental effort (Kahneman, 1973) since—differently from verbal reports and ratings—they are not under volitional control (Loewenfeld, 1999; Laeng and Sulutvedt, 2014). Given our assumption that musical imagery is intimately linked with the motor imagery of sound-producing body motions, one straightforward expectation is that there will be a close affinity in load on cognitive resources between standard playing and silent playing. Both involve the planning and execution of complex coordinated body motions and both result in actual movement, muscular deployment, and related metabolic expenditure. This would also be consistent with much evidence from pupillometry research indicating that a simple, single, keypress (used in many paradigms to indicate the detection of a target stimulus or a discriminatory choice) results in measurable dilation of the eye pupil (e.g., Simpson, 1969; Simpson and Climan, 1969). However, overt behaviors and muscles' contractions are not necessary for pupillary dilations to occur and there is overwhelming evidence in the literature that dilations can index the presence and degree of internal (cognitive or affective) processes without the need for overt responses (e.g., Kahneman and Beatty, 1966; Laeng et al., 2016).

In the present study, we opted for correlational analyses of the measure-by-measure changes in pupil diameter across conditions within each musical piece. We expect that the pupil sizes will co-vary positively across the different phases of a musical piece as a support to the hypothesis of functional overlap. Similarly, there should be a chronometric correspondence between the time required to perform a musical piece and that required for its execution, but this does not imply that the absolute times cannot differ, even when there is strong functional overlap. Based on the theory and findings of the original "mental scanning" experiments (e.g., Kosslyn et al., 1978; Borst and Kosslyn, 2008), the time to scan increasing distances increases at comparable rates in perception and visual imagery. However, when generating mental images from long-term memory, participants could scan more slowly in the mental image condition. Hence, we expect the time course of imaging to be slower for musical imagery than in listening or playing. This seems also to be the case for the time taken to imagine an action and its actual execution in athletes (e.g., Reed, 2002).

Previous studies with the pupillometry method have shown its ability to measure the moment-by-moment changes in cognitive and affective arousal during music listening (e.g., Laeng et al., 2016; Bowling et al., 2019). To our knowledge, there are no previous pupillometry studies during a musical performance. The present use of pupillometry as a gauge of mental work in music seems novel and potentially fruitful.

## EXPERIMENT 1A: PUPILLOMETRY

If similar processes and mechanisms underlie perception and imagery, executed and imagined movements, then we reason that this functional overlap should be reflected in the level of allocation of cognitive resources and, in turn, in changes in the pupil diameter. Moreover, the degree of effort required during a musical piece may be visible not only in the eyes of a performer but also in those of the listeners. The experience could modulate this response, being stronger in musicians with experience with the same instrument than in the non-musicians. Hence, in this study, we explore how the pianist's eye pupil can index effectively the level of cognitive resources required in a performance by using differentially challenging musical piano pieces and annotating the momentary difficulty, at different points in time during a musical performance. Also, we monitor the pupils of listeners, either pianists or non-musicians, while listening to the pianistic performances of the professional musician.

We expect positive correlations between pupil diameter changes of the professional pianist during different "executions" (i.e., normal, silenced, listened, and imagined) of the same piano piece, which would indicate the presence of similar task demands and load on resources and the engagement of similar cognitive and execution mechanisms. Conversely, a lack of correlation between pupil diameter changes in the different "executions" of the same piano piece (i.e., normal, silenced, listened, and imagined) might indicate the absence of common mechanisms but the strength of the correlations should give a hint to the degree of functional overlap or equivalence.

# Method

We use the pupillometry method based on infrared eye-tracking, which allows the precise tracking of the size of the pupil in both eyes simultaneously in a non-invasive manner and with no restrictions on eye movements, at a sampling rate equal or superior to standard film or television (e.g., PAL is 50 Hz).

## Participants

This first experiment involved the single case of a professional pianist (PP hereafter) who is a music teacher in Oslo, Norway, and an internationally renowned performer (e.g., at Carnegie Hall), with expertise in 19th century piano music and technique. PP is a 41 years old female and she has played the piano since the age of six. Based on her responses to The Goldsmiths Musical Sophistication Index (v1.0), she has a top score of 36 in musical training and a General Sophistication score = 88. PP plays five or more hours per day on her primary instrument (piano), but she can play four other additional instruments. She is highly active in perfecting her technique and has lectured and published on piano techniques from previous centuries. Hence, PP qualifies as exceptional and a true expert in music performance according to the criteria of Montero (2016) and Høffding (2019) and most definitions of expertise regardless of the domain (e.g., for elite sports; Swann et al., 2015). The Department of Psychology's IRB at the University of Oslo approved the present study (Reference number: 3568281) and PP as well as the other participants (in Experiments 1B and 2B) received a consent form before testing and were treated according to the Declaration of Helsinki.

## Musical Stimuli

PP provided us with three piano pieces of different technical and interpretive difficulty: an easy piece, a middle-level one, and an advanced one. She chose two compositions of Edvard Grieg: Wächterlied, from Lyrical Pieces, Op.12 No.3, and Holberg Suite, Op. 40, Praeludium (Allegro vivace); these are considered as an easy and a difficult piano piece, respectively. The medium difficulty composition was Robert Schumann's Träumerei (from Kinderszenen or "Scenes from Childhood," Op. 15). Each piano piece was performed twice on an electronic keyboard set to "grand piano," once normally, "standard playing" (hereafter) and once with no sound (by turning off the audio). Also, PP listened to her normal renditions of each piece (as registered in a MIDI file) as well as in a condition where there was neither auditory nor kinesthetic feedback from finger movements, i.e., the "imagery" (hereafter) condition. In all the four conditions, we showed the first two pages of the musical score on the computer screen of the eye-tracking device.

## Apparatus

We used a Yamaha electronic or digital piano (P-140) set to "grand piano." The P-140 keyboard has 88 graded-hammer keys with sounds based on Yamaha's AWM sampling technology, with the convincing similarity of sound to a real acoustic instrument. We interfaced with a MIDI Unit (MOTU UltraLight Hybrid MK3) which recorded the two performances (with sound and without), plus it allowed the playback of the performance with sound during the listening-only condition. We positioned the piano keyboard on an adjustable desk at a comfortable height for the pianist. We positioned the eye-tracking computer's screen on the same adjustable desk, behind and above the keyboard for optimal visibility. We attached the infrared camera of the ET unit to the lower edge of the computer screen, a flat DELL LCD monitor, with a screen resolution of $1,680 \times 1,050$. We presented the first two pages of each of the three pieces' musical scores, after digitalizing them at a high-resolution, so that the pianist could read the score at a comfortable distance of about 50 cm. To facilitate playing and making the session more natural we did not use a chinrest, which is not problematic for the SMI eye-tracking equipment since it automatically corrects for changes in head position and rotation. Hence, in such testing conditions, it is possible to obtain reliable gaze data and mapped pupil diameter (in mm) that are free of artifacts due to head movements or changes in distance between the eyes and the screen.

A R.E.D. 250 SMI infrared eye and iView X Hi-Speed Software (SMI; Berlin, Germany) recorded eye positions at a sampling rate of 250 Hz. The RED can operate at a distance of 0.5–1.5 m. This device has two sources of infrared light from an infrared light-sensitive video camera, placed under the monitor frame. According to SMI, the RED system can detect changes as small as 0.004 mm. During the experiment, PP looked directly into the screen to the musical scores. We used BeGaze software from SMI to extract the gaze and pupil data, and Microsoft® Excel, JASP, and Statview software for the statistical analyses.

## Procedure

A 4-point calibration procedure preceded each experimental session. The pianist always looked to the music score while playing as well as in the separate conditions, where PP imagined performing the same pieces "in her head." Also, the pianist listened to her own playing of the same piano pieces (as recorded by MIDI from the keyboard and played back with headphones). One condition consisted of playing the piece while looking at the score on the screen with the muted electronic keyboard so that the hands/fingers' movements produced no sound. At the end of all experimental sessions, the pianist rated levels of technical difficulty or expressivity and harmonic intensity on the musical score (by use of a 7-step Likert scale), measure-by-measure (see **Figure 1** for an example).


**FIGURE 1 |** The score of Grieg's Wächterlied for piano, with annotations by the professional pianist of its technical difficulty (measure by measure on a 7-step Likert scale).

To estimate subjective effort, at a later date after the pupillometry session, we asked the professional pianist to estimate with a 7-step rating scale each piece, measure by measure (roughly following the musical metric framework; see Keller, 2001), along with three different parameters of "difficulty". One type we label "technical" (i.e., of the motor-related challenges of playing the notes as indicated in the score); then, the "expressive" (i.e., expressivity-related difficulties, like shaping the performance as intended); and finally, the "harmonic tension" (i.e., subjectively experienced harmonic tension and release).

## Results and Discussion

We first computed descriptive statistics for the ratings given to each piece, measure by measure (see **Figure 2**). The pianist (PP) judged on Likert 7-step scales the technical difficulty, the expressivity, and the harmonic tension. Three separate ANOVAs with each of three ratings as dependent variables showed that all pieces differed from one another. Specifically: (a) technical difficulty, $F_{(2)} = 194.1$, $p < 0.001$, $\eta^2 = 0.86$ (*post hoc* tests: $2.8 < t < 18.6$; $0.02 < p < 0.001$); (b) expression, $F_{(2)} = 45.5$, $p < 0.001$, $\eta^2 = 0.59$ (*post hoc* tests: $4.5 < t < 9.5$; all $p < 0.001$); (c) harmonic tension, $F_{(2)} = 38.3$, $p < 0.001$, $\eta^2 = 0.55$ (*post hoc* tests: $2.9 < t < 8.7$; $0.01 < p < 0.001$). These mean ratings essentially confirm PP's selection of three pieces in terms of pianistic challenges or performance demands, since consistently she rated the Holberg Suite highest on all three measures, Wächterlied was rated the lowest, while Träumerei was placed in between.

Interestingly, a multiple regression analysis with "technical difficulty" as the dependent variable and "expression" and "harmonic tension" as the independent variables revealed a strong positive relationship between the three measures, $F_{(2)} = 85.5$, $p < 0.0001$, $r = 0.855$. At the same time, these measures tap, as intended, on different aspects of

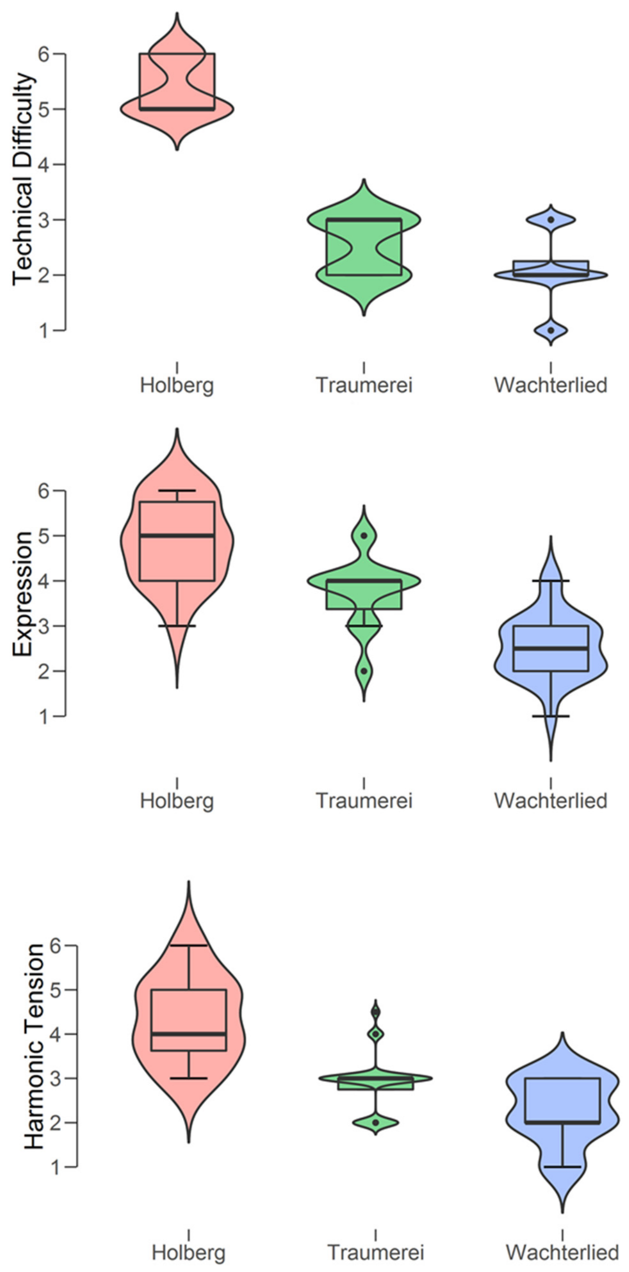



**FIGURE 2** | Violin boxplots of the ratings by the Professional Pianist of the three types of difficulty (from top to bottom: technical, expressive, and harmonic tension) for each of the three piano pieces (Holberg Suite, Träumerei, and Wächterlied).

the subjective effort in performance. Specifically, "technical difficulty" appeared to be more closely related to "harmonic tension" (Regression Coefficient = 0.865, $t = 6.2$, $p < 0.0001$) than to "expression" (Regression Coefficient = 0.280, $t = 2.1$, $p = 0.04$).

Our measure of objective (mental) effort was the pupil diameter during each condition. Since the pianist was looking at the same score in all conditions, we assume that the light stimulation to her eyes remained constant across musical measures and across conditions. In **Figure 5** we show the waveforms of PP's pupil diameter (as color lines) along time (in seconds) for each of the four conditions and for the three pieces (split in panels). These waveforms reveal several interesting aspects. First, the pupillary waveforms when "playing" (i.e., moving the fingers "with" or "without" sound) are consistently above the other two conditions where there is no movement (i.e., "listening" and "imagery").

Second, it is clear that while the duration of each piece when playing either with or without sound differs only a few seconds from one another (and from listening), the imagery condition was—as seen in previous studies (e.g., Janata and Paroo, 2006)—longer than the other conditions. Traäumerei was imagined for about 40 s longer (i.e., a 40% lengthening) than when listening or performing. Similarly, Wächterlied was imagined for about 22 s longer (i.e., a 16% lengthening) than when listening or performing. PP performed the Holberg Suite at a faster rate than the other two pieces (in about 35 s) but, when imagined, its length stretched of about 5 s (i.e., a 14% lengthening).

We also computed descriptive statistics for mean pupil responses. Variations in pupil diameter approximate a normally distributed parameter (e.g., Mathôt et al., 2018). Moreover, *F*-tests remain robust also when data are not entirely normal (Blanca et al., 2017). Hence, we applied a repeated-measures ANOVA with the mean pupil diameter within each measure as the random factor and Conditions (Playing with sound; Playing with no sound; Listening; Imagery) as the within factor. This analysis revealed a significant effect of Conditions, $F_{(3)} = 320.6$, $p < 0.001$, $\eta^2 = 0.83$. *Post hoc* (Bonferroni) tests confirmed that the four conditions differed significantly from one another ($p < 0.001$; Cohen's $d$ range: 0.4–2.68).

As visible in **Figure 4**, the most effortful condition—according to the pupil diameter—was "Playing with no sound," followed by "Playing with sound". The condition of "Listening" and "Imagery" were clearly less effortful (Cohen's $d > 2$) compared to the previous "motoric" conditions. Looking at the score and imagining the music was the least demanding of all conditions.

Also, we run an ANOVA on the mean pupil diameters for the three musical pieces. This analysis revealed a significant effect of the factor of Musical Piece, $F_{(2)} = 18.9$, $p < 0.001$, $\eta^2 = 0.38$. *Post hoc* (Bonferroni) tests showed that both the Holberg Suite (mean pupil diameter = 3.432; SD = 0.13) and Träumerei (mean pupil diameter = 3.429; SD = 0.07) differed significantly ($p < 0.001$) from Wächterlied (mean pupil diameter = 3.296; SD = 0.05). However, they did not differ from each other (see **Figure 5** illustrating in boxplots the pupil diameter for each piano piece and each condition in separate panels).

A multiple regression analysis explored the relationship between the four conditions. One multiple regression used "imagery" as the dependent variable and the other three conditions as independent variables, which revealed a highly significant relationship and a moderate positive relationship, $F_{(3)} = 17.5$, $p < 0.0001$, $r = 0.68$. Specifically, "imagery" was highly significantly related to "listening" (Regression Coefficient = 0.65, $t = 4.85$, $p < 0.0001$). Imagery was also significantly related to "Playing with no sound" (Regression Coefficient = 0.224, $t = 2.1$, $p = 0.04$), but failed to

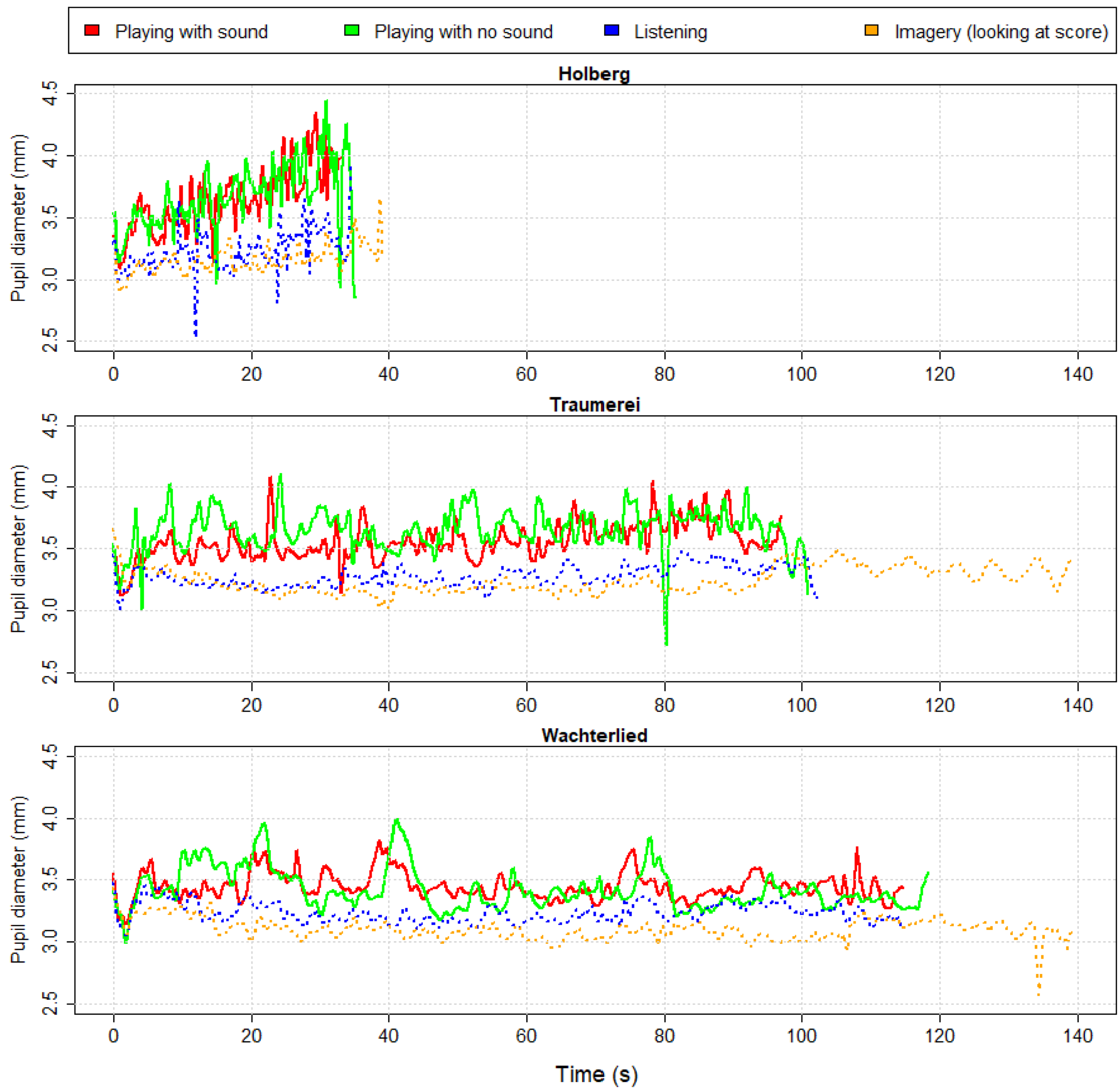



**FIGURE 3 |** Pupillary waveforms of professional pianist (PP) during each of the three piano excerpts (Holberg Suite, Traäumerei and Wächterlied) and for each of the four conditions ("Playing with sound"—red line, or "Playing with no sound"—green line, "listening"—blue dotted line, and "imagery"—orange dotted line).

reach significance with the standard performance condition or "Playing with sound" (Regression Coefficient = 0.12, $t = 1.05$, $p = 0.30$).

**Figure 6** illustrates in detail how the mean pupil diameters (in each musical measure or bar) are related to each other in each condition. Specifically, we subdivided the pupil time series by the number of bars in the score, to standardize the pupil data between conditions differing in length. The results confirmed our expectation, based on the idea that each of the conditions would draw resources or demand mental effort, that the pupil changed similarly during the same moments (or "chunks," i.e., measures) of a musical piece. Indeed, the two motoric conditions (Playing with sound and Playing with no sound), both requiring actual finger movements, showed the strongest relationship (**Figure 6**, top left panel), $F_{(1,64)} = 160.6$, $p < 0.0001$, $r = 0.85$. Most interestingly, the second strongest relationship was between "listening" and "imagery" (i.e., the two conditions without explicit motoric involvement), $F_{(1,64)} = 39.1$, $p < 0.0001$, $r = 0.62$. This positive relationship might be attributed to hearing the music, not only when listening, but also "in the mind"s ear' when imagining.

The two next strongest relationships (displayed in the bottom panels of **Figure 6**) showed only moderate correlations, though both statistically significant. "Playing with sound" was positively related to "Listening" (bottom left panel), $F_{(1,64)} = 20.9$, $p < 0.0001$, $r = 0.49$. Remarkably, "Playing with no sound" (bottom right panel) was positively related to "Imagery," $F_{(1,64)} = 21.6$, $p < 0.0001$, $r = 0.50$. All other simple regression

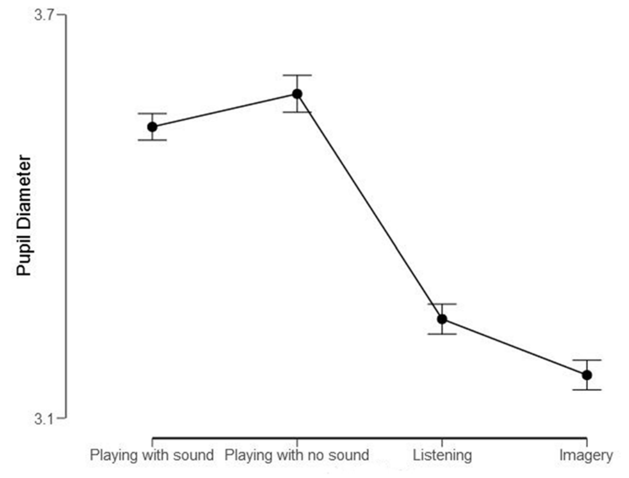



**FIGURE 4 |** Mean pupil diameters (in mm) of the Professional Pianist for the four conditions. Bars represent 95% confidence intervals.

had still positive correlations, but with Spearman's coefficients *r* below 0.5, accounting for less than a quarter of the variance and are not displayed here.

In sum, our objective measure of effort, the pupil size, showed that pupil diameters were largest (**Figure 4**) when "playing" (regardless whether there was sound produced or not) and above the conditions where there is no movement (i.e., "listening" and "imagery"). This suggests that programming and executing motion is considerably more demanding than situations in which action is not required. This could also be partly because mental and physical efforts are meshed during the action and both could show up in the measure of attentive workload.

Although "musical effort" in general has been an understudied aspect of music cognition (see Keller, 2001), a study by O'Shea and Moran (2016, in Study II) examined explicitly how the eye pupil adjusted when pianists performed a piece but also while they simply imagined the performance. They reported no difference between their musicians' pupils during the actual performance and its imagery, which seems consistent with an equivalent deployment of arousal for the same processes in the two conditions. However, they based such a conclusion on the results of an analysis of variance that barely missed the 0.5% cut-off ($p = 0.053$), which might not constitute conclusive evidence for no difference.

We confirmed that subjective effort is related but not identical to the objective effort, since the ordering of the two differed slightly (**Figures 2**, **5**), especially for "listening" and "imagery."

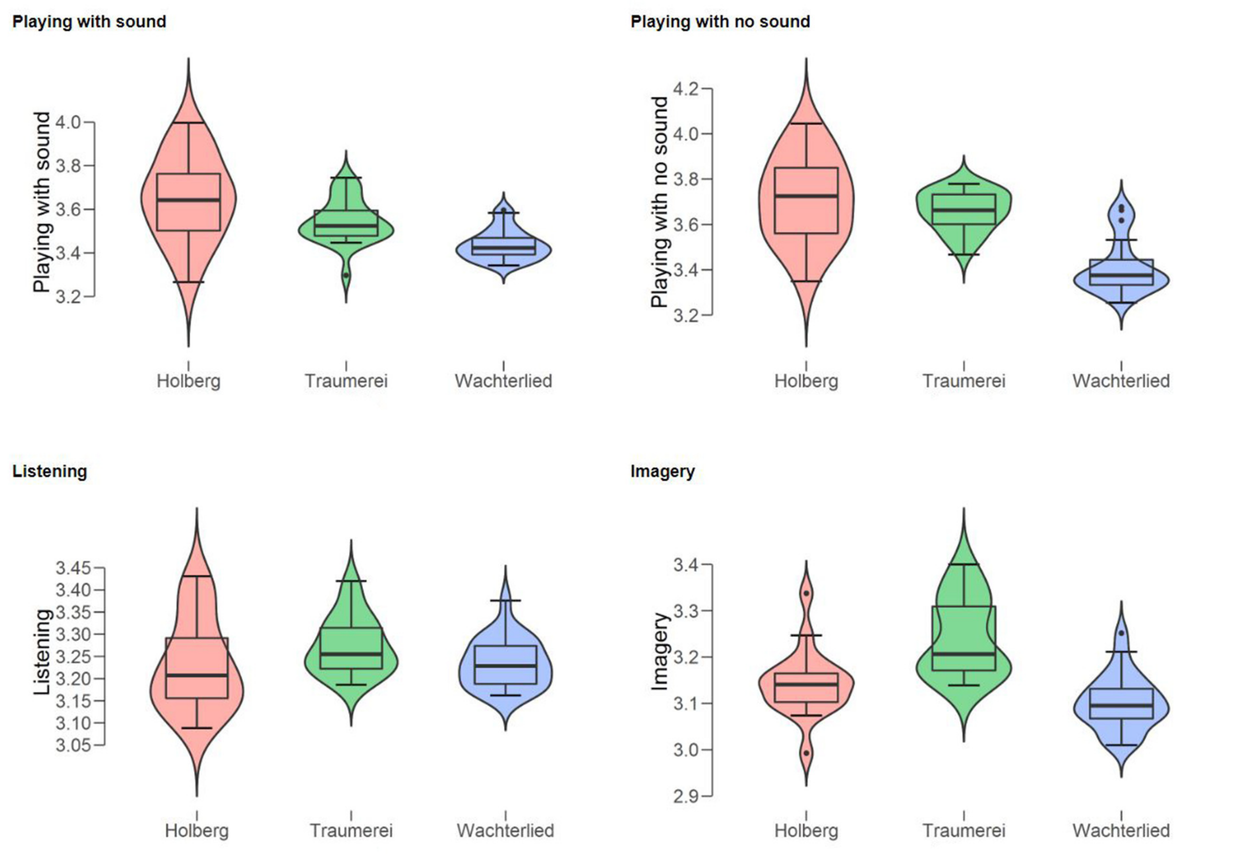



**FIGURE 5 |** Violin boxplots of the mean pupil diameters (in mm) of the Professional Pianist for the three piano pieces (Holberg Suite, Träumerei, and Wächterlied) split by the four conditions (from top left panel: Playing with sound; Playing with no sound; Listening; and Imagery).

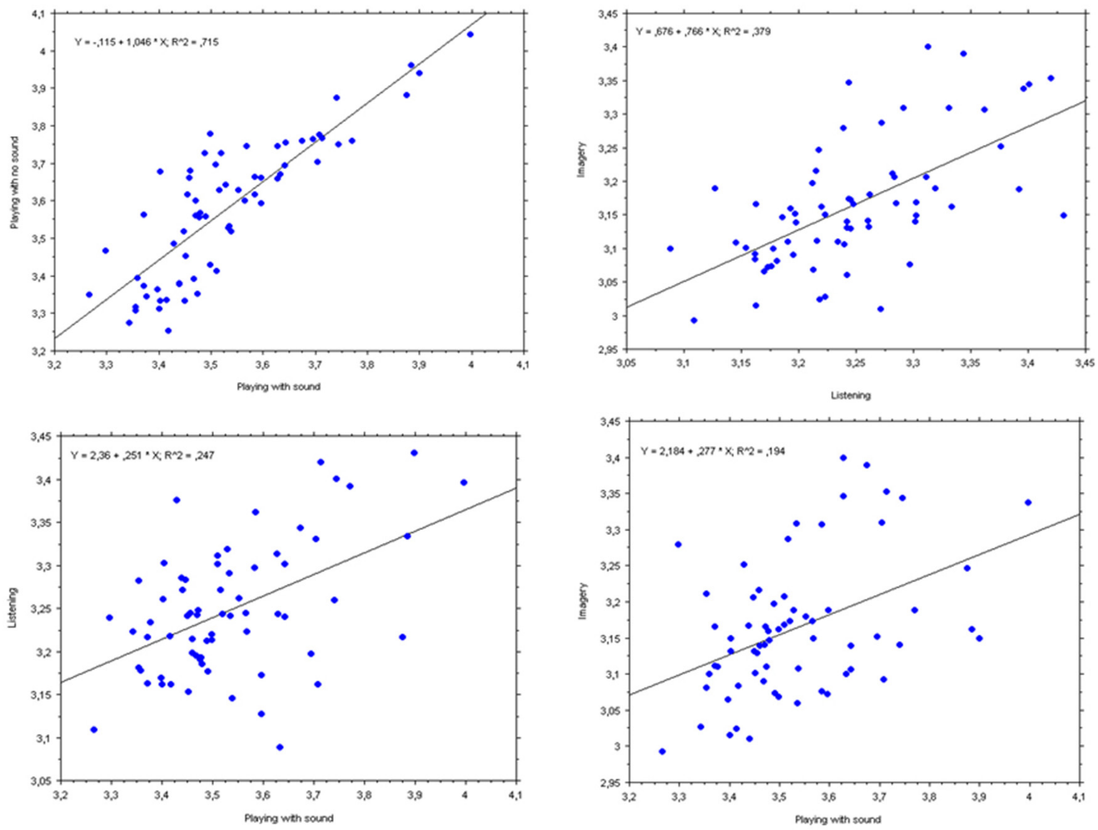



**FIGURE 6** | Bivariate plots and simple regression lines for PP's mean pupil diameters (in mm) for each measure of all three pieces together. Top left panel: dependent variable = "Playing with no sound"; independent variable = "Playing with sound"; Top right panel: dependent variable = "Imagery"; independent variable = "Listening"; Bottom left panel: dependent variable = "Listening"; independent variable = "Playing with sound"; Bottom right panel: dependent variable = "Imagery"; independent variable = "Playing with sound".

Most interestingly, changes in the mental effort as the musical piece evolved (measure-by-measure) were all positively related (**Figure 6**) and their relationship was strong for the two playing conditions and moderate for the two non-playing conditions, revealing that regardless music is performed or only listened (with the physical ear or mind's ear) the pupils' diameter co-vary at each point in time. This can be interpreted as good evidence that all conditions share the same cognitive mechanisms and a similar workload along with the musical piece.

Finally, we replicated the finding that imagery can stretch the timeline. It is possible that when playing music in one's head, there is the luxury to pause or dwell on a particular moment. Because imagery involves additional mechanisms of generation and maintenance as well executive processes (Glover et al., 2020), this could lengthen processing, without apparently increasing the cognitive workload compared to actually listening (**Figure 4**).

## EXPERIMENT 1B

In the following experiment, we sought to provide evidence that also listeners would react to the perceived "musical effort" inherent in the music and likely more in musicians than in non-musicians, while they listened to the piano renditions by our professional pianist. The spontaneous pupillary responses can be compared for similarity across groups and with those of the performer.

### Method

We used the same pupillometry method of Experiment 1A.

#### Participants

We recruited 20 participants (12 females) as volunteers for the listening group (mean age: 28.15 years, range: 19–65). Of these, 10 were pianists (mean age: 26.5 years, 7 females) and 10 were non-musicians (mean age: 29.8 years, 5 females), all based on self-reports. All participants read and signed informed consent.

#### Apparatus

We used the same R.E.D. 250 SMI infrared eye tracker in the same laboratory room used in Experiment 1A. However, we recorded eye positions at a sampling rate of 60 Hz, which is a sufficient sampling rate for pupil measurements (Laeng and Alnæs, 2019).

### Procedure

A four-point calibration procedure preceded each experimental session. Participants looked at all times to a blank gray screen with a circle at the center (5 cm in diameter) while the music was playing as well as in a "baseline" recording, in silence right before the music, of 1 s. Participants seated in front of a monitor with their head supported on a chinrest and listened to recordings of the different pieces performed by PP (Holberg Suite, Träumerei, and Wächterlied). There were two different versions of each piece, one where PP listened to the sound produced by the piano ("Sound" or normal condition) and one where PP did not get to listen to the sound produced ("No sound" or silent condition). All participants received the following instruction: "Please look at the circle in the middle of the screen while listening to the music. Keep your eyes open during the experiment (you can blink as normal)."

## Results and Discussion

Since half of the participants were not musicians, we collected only the measure of objective (mental) effort, i.e., pupil diameter during listening. In **Figure 7** we show the waveforms of the Pianists' and Non- musicians' pupil diameters (split between the left side and right side panels respectively). The color lines show the pupil change along time (in seconds) for the two listening conditions and the three pieces (split in panels vertically).

Fidler and Loftus (2009; see also Loftus and Masson, 1994) have argued that graphs with (appropriate) error bars can replace significance tests. Hence, the plotted average pupil changes (in **Figure 7**) and their 95% confidence intervals reveal a tendency for larger pupil diameters when listening to the normal performance of PP, with sound, compared to her silent performance. This is particularly clear for the pianist group when listening to Träumerei and for the non-musicians group when listening to the Holberg Suite, which are the two most challenging piano pieces, whereas the easier Wächterlied shows no remarkable separation between the pupil waveforms. One interpretation is that the normal performance (where the pianist hears herself) is a more engaging rendition of the pieces than the silent one and therefore captures more attention from the listeners. A visual comparison of **Figures 3**, **7** (showing PP's pupillary waveforms and, in particular, the blue lines of the "Listening" to the normal Playing) reveals that PP's pupil maintained either a constant size (Traäumerei and Wächterlied) or increased over time (Holberg Suite). In contrast, both groups of control participants showed a reduction of their pupil diameters over time.

**Figure 8** illustrates pupil responses during listening only to the normal playing of each piece, which reveals clear differences in pupil response between the two control groups and of both with respect to PP. Specifically, while the non-musicians' average pupil was largest when listening to any of the three piano pieces, PP's pupils showed the smallest responses. **Figure 9** shows the baseline pupil measurements of the two control groups during a baseline pupil measurement at silent rest.

However, of special interest for the present investigation is to what extent the pupil responses, despite the strong difference in absolute response (probably reflecting how arousing was the music listening for the controls and the pianist), changed similarly with the unfolding of the music. Hence, we run multiple regression analyses using the change in average pupils as they occurred measure-by-measure, with PP's average pupils as the dependent variable and Non-musicians' and Pianists' average pupils as the independent variables. The m-regression for the listening of the excerpt from the Holberg Suite showed no significant effects; $R = 0.21$, $F_{(2, 17)} = 0.342$, $p = 0.72$ (Regression Coefficients: Pianists = 0.16; non-musicians = 0.04). In contrast, the m-regression for the listening of Träumerei showed a significant effect, due to a significant relationship between pupil responses of the pianists and PP across measures; $R = 0.52$, $F_{(2, 23)} = 3.82$, $p = 0.03$ (Regression Coefficients: Pianists = 0.27; non-musicians = 0.01). The m-regression for the listening of the Wächterlied showed no significant effects; $R = 0.30$, $F_{(2, 23)} = 1.05$, $p = 0.37$ (Regression Coefficients: Pianists = 0.13; non-musicians = 0.10).

Also, we run separate multiple regression analyses using the change in average pupils as they occurred measure-by-measure, with PP's and each of the control groups' average pupils as the independent variables and as the dependent variable PP's ratings (also measure-by-measure) in the three types of subjective effort) "technical difficulty," "expression", and "harmonic tension").

The m-regressions for "technical difficulty," during listening of the excerpt from the Holberg Suite, showed a significant effect, $F_{(3,17)} = 3.74$, $p = 0.03$. This was due to a positive relationship of subjective effort with the pupil responses of the Non-musicians ($p = 0.02$) but not with those of Pianists ($p = 0.27$) or PP ($p = 0.95$); $R = 0.67$ (Regression Coefficients: Non-musicians = 0.98; Pianists = 0.69; *PP* = 0.03). In contrast, the m-regression for "technical difficulty" and Träumerei showed no significant effects; $R = 0.21$, $F_{(3,23)} = 0.31$, $p = 0.81$ (Regression Coefficients: Non-musicians = 0.46; Pianists = 0.08; *PP* = 0.004). Similarly, there were no effects for Wächterlied; $R = 0.29$, $F_{(3,23)} = 0.59$, $p = 0.63$ (Regression Coefficients: Non-musicians = 0.92; Pianists = 0.62; *PP* = 0.008).

The m-regressions for "expression," during listening of the excerpt from the Holberg Suite, showed no significant effects; $R = 0.51$, $F_{(3,17)} = 1.6$, $p = 0.23$ (Regression Coefficients: Non-musicians = 1.2; Pianists = 0.55; *PP* = 1.5). In contrast, the m-regression for "technical difficulty" and Träumerei showed significant effects for both Non-musicians ($p = 0.009$) and *PP* ($p = 0.004$); $R = 0.70$, $F_{(3,23)} = 6.6$, $p = 0.003$ (Regression Coefficients: Non-musicians = 1.75; Pianists = 0.02; *PP* = 5.9). However, there were no effects for Wächterlied; $R = 0.41$, $F_{(3,23)} = 1.3$, $p = 0.29$ (Regression Coefficients: Non-musicians = 0.16; Pianists = 1.6; *PP* = 3.4).

The m-regressions for "harmonic tension," during listening of the excerpt from the Holberg Suite, showed no significant effects; $R = 0.57$, $F_{(3,17)} = 2.2$, $p = 0.13$ (Regression Coefficients: Non-musicians = 1.4; Pianists = 1.6; *PP* = 0.14). Similarly, the m-regression for Träumerei showed no significant effects; $R = 0.30$, $F_{(3,23)} = 0.66$, $p = 0.59$ (Regression Coefficients: Non-musicians = 0.37; Pianists = 0.78; *PP* = 2.6). Finally, there were no effects for Wächterlied; $R = 0.49$, $F_{(3,23)} = 2.0$, $p = 0.13$

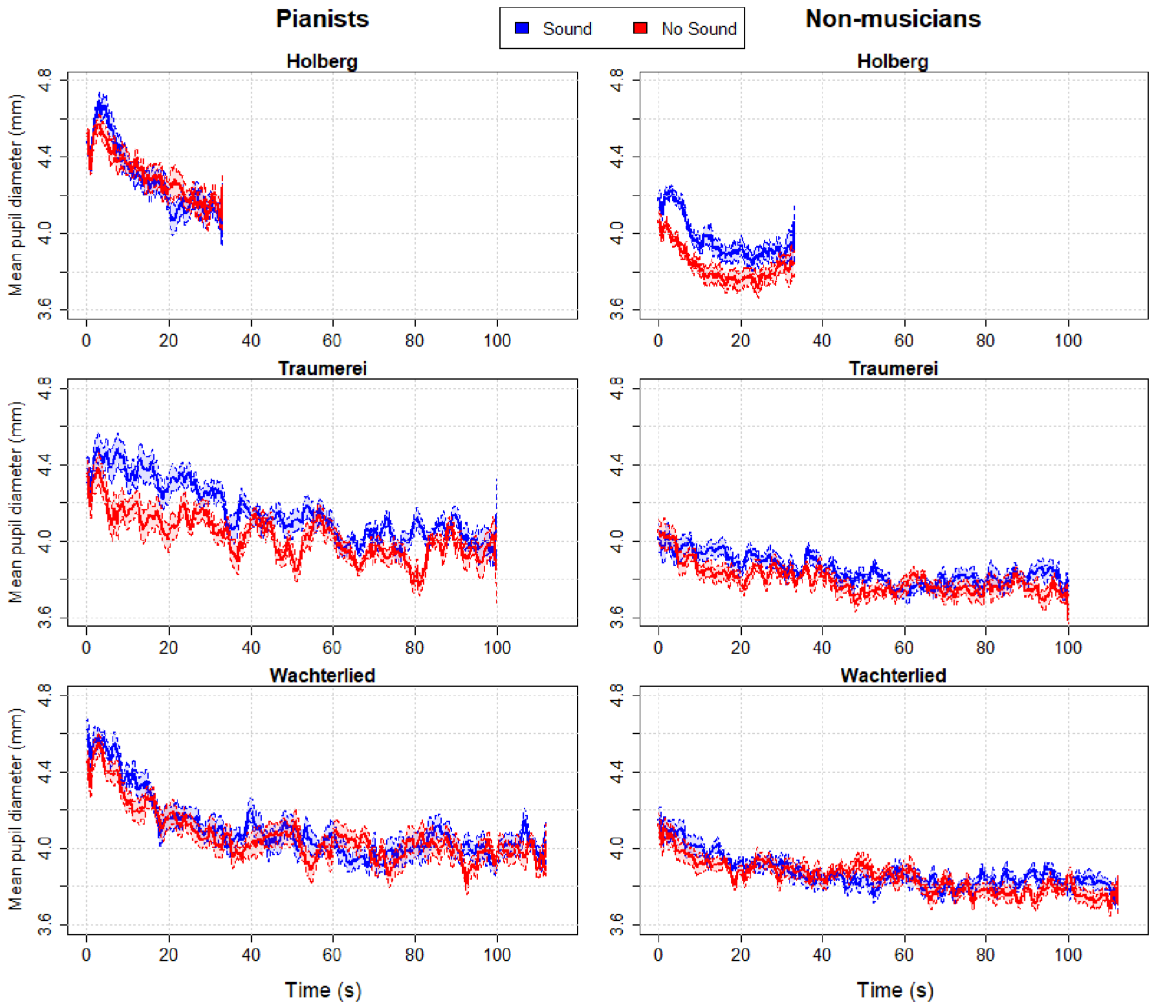



**FIGURE 7 |** Pupillary waveforms of control participants after averaging the time series of pianists and non-musicians of the three piano excerpts. The two waveforms in each panel illustrate pupillary responses to listening to the two versions of each musical piece, one in which PP played with sound—i.e., standard playing condition (in blue)—and the other while listening to PP playing the same piece but with no sound or silent keys (in red). Patches along the curves indicate between-subjects (95%) confidence intervals.

(Regression Coefficients: Non-musicians = 0.89; Pianists = 0.55; *PP* = 2.69).

In sum, pianists listening to the performance would seem to attend more intensively to the most difficult piece than non-musicians, showing that their expertise with the same instrument could effectively engage their cognitive and perceptual system. Since pianists showed larger pupils than non-musicians only for the most difficult piece, it is unlikely that this difference was due to generally larger pupils within one group. Additionally, both the pianists and non-musicians seemed to be engaged more (**Figure 7**) by the standard performance (with sound) than the silent performance (which could have sounded less optimal).

Most interestingly, pupil sizes were smaller, when listening to all three pieces, for the professional pianist and largest for the non-musicians. Although this could be due to spurious pupil size differences across the groups and PP, we find this unlikely and we would like to suggest that these results might be indicative of expertise so that the amount of attention allocated for the same task is always lower in the expert or better performers than in novices.

Finally, there was weak evidence that, during listening, subjective effort measures (ratings) were related to the objective effort (pupil diameter), at least according to the timeline based on measures' subdivisions. The only exceptions were the ability of the pupils of Non-musicians to predict "technical difficulty" for the Holberg Suite excerpt, and of the pupils of both Non-musicians and PP to predict "expression" for "Träumerei."

## EXPERIMENT 2: fMRI

Several neuroimaging studies have specifically addressed music and imagery (e.g., Zatorre et al., 1996, 2007; Halpern and Zatorre, 1999; Satoh et al., 2001; Meister et al., 2004; Kraemer et al., 2005; Cebrian and Janata, 2010; Farrugia et al., 2015;

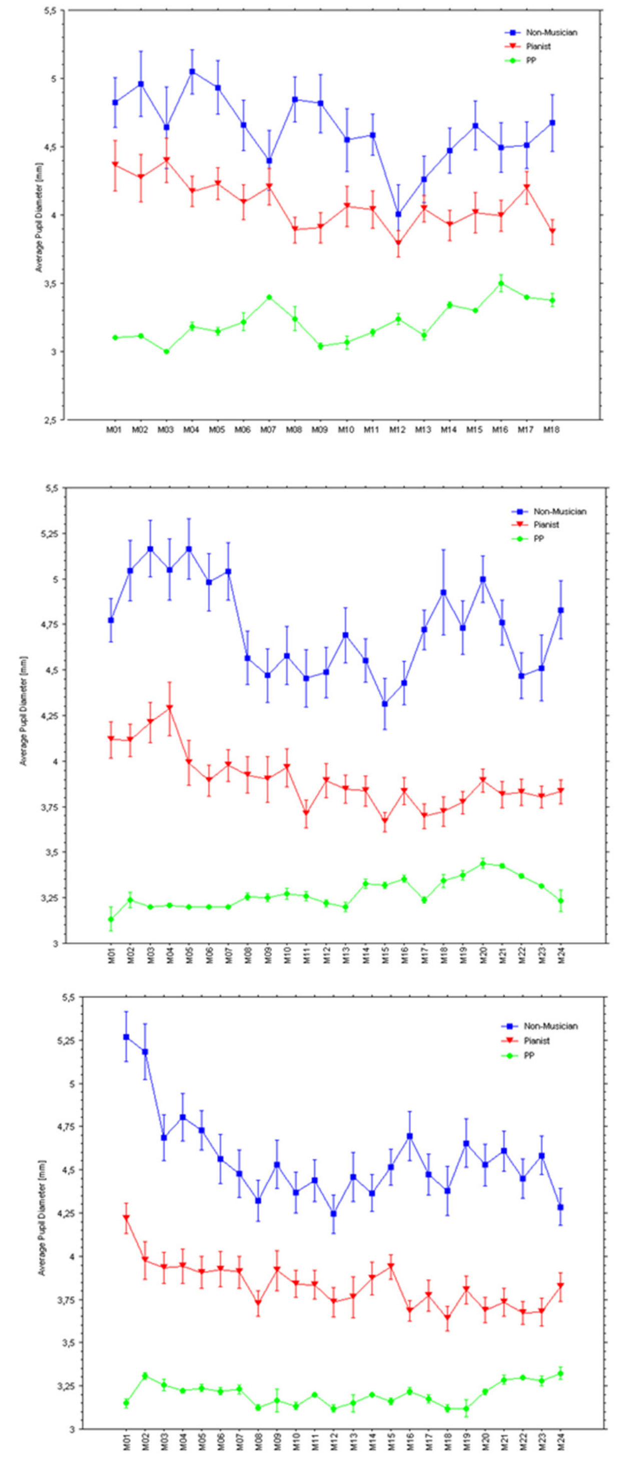



**FIGURE 8 |** Average pupil diameters in mm (bars are SEs) within each measure while listening to PP's playing with a sound each of the three pieces (Holberg Suite: top panel; Träumerei: middle panel; Wächterlied: bottom panel).

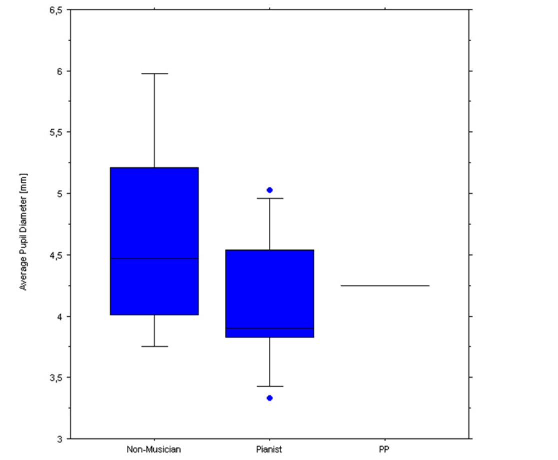



**FIGURE 9 |** Boxplots of pupil diameters of the Non-musicians (left) and Pianists (middle) groups. The horizontal line to the right indicates PP's baseline average pupil.

Lu et al., 2017). In general, they reported activity within cortical structures, including the human motor or pre-motor cortex during imagery. For example, a study by Meister et al. (2004) showed that imagery and playing shared many sensory cortical areas, while bilateral primary motor areas were active only during playing. A few recent fMRI studies have specifically investigated piano players (e.g., Harris and de Jong, 2015), showing activations in the auditory and premotor cortex (PMC) during a motor imagery task compared to simply listening. Many studies converge in revealing that motor regions of the brain are involved in musical imagery and that imagery-like processes are involved in musical perception. However, to our knowledge, no studies have investigated the relationship between playing and imagining playing music when different levels of cognitive effort are involved and/or have specifically investigated the role of the subcortical region beside the cerebellum.

At least two single-case fMRI studies have previously investigated music listening, playing, and imaging in internationally renowned musicians, one with the popular artist Sting (Levitin and Grafton, 2016) and the other with the classical pianist Christopher Seed (Jäncke et al., 2006). The former confirmed substantial overlap of brain regions activated by listening and imagining, while the latter focused specifically on the ability of the classical pianist to play "left-handed" on a mirror keyboard compared "right-handed" on the standard keyboard.

Importantly, besides a study on jazz improvisation (Limb and Braun, 2008), no neuroimaging studies have explicitly examined mental effort in music-making. Most importantly, neuroimaging studies have not examined whether there is an activity within the LC in the brain's brainstem while playing, listening, or imagining playing, that is in the area of the brain that has directly related to mental effort in humans (Alnæs et al., 2014; Mäki-Marttunen et al., 2019). Because in the previous experiment, we applied the method of pupillometry as an objective measurement of mental effort during music listening, performance, and imagery, we have the opportunity to compare brain activity in the different conditions between PP and a control group, in the light of the results of the pupillometry experiment.

## Experiment 2A

We invited PP to participate in an fMRI experiment to assess the whole brain's activity, including subcortical structures while listening, playing, and imagining piano pieces with a simple and complex level of cognitive and technical effort. Specifically, we used the same two Grieg's pieces of Experiment 1, considered easy vs. difficult piano pieces, this also supported by PP's subjective ratings (**Figure 2**) and—most importantly—by the pupil results (**Figure 5**) as shown earlier in Experiment 1. For practical reasons, we omitted during scanning the "playing without sound" condition, focusing on four conditions: listening, imagining the sounds, imagining playing, and actually "playing." To enable a realistic "piano playing" condition, we set up a scanner-adapted piano keyboard, with sound played back through noise-canceling headphones. Our expectations derive from the literature that has revealed substantial overlap between cortical regions during imagery, listening, and playing, even at the single case level (Levitin and Grafton, 2016). Also, based on the results of Experiment 1, we expected to find activity in the brainstem's LC that would differentiate the two levels of effort (easy vs. difficult). We also expected that brain activity would overlap more for listening and imagery than between these two and playing since PP's pupil dilated the most during playing (**Figure 4**) and it was less active and in a similar way during the two other conditions (**Figures 4**, **6**).

## Method

We used the initial part of the Wächterlied, from Lyrical Pieces, Op.12 No.3, and the Holberg Suite, Op. 40, Praeludium (Allegro vivace). In this and the following MRI experiments, adapted versions of subparts of the music sheets were presented on a computer screen, visible in the scanner, representing 15 s of the pieces. The pieces were also adapted to make it possible to play them on a small piano keyboard (with two octaves) in the scanner.

### Participants

The target participant was the single case of PP who returned for testing with MRI about 1 year following the pupillometry experiment.

### Apparatus

Scanning was performed with a Philips Achieva 3 Tesla MR scanner (Philips Medical Systems, Best, The Netherlands), equipped with an eight-channel Philips SENSE head coil. Functional data were collected using a BOLD-sensitive T2* weighted echo-planar imaging sequence [40 slices, no gap; repetition time (TR), 2.5 s; echo time (TE) = 30 ms; flip-angle = 80°; voxel size = 3 × 3 × 3; a field of view (FOV) = 240 × 240 mm; interleaved acquisition]. The slices were oriented to cover the whole cortex, cerebellum, and the brainstem's pons. To avoid T1 saturation effects, five dummy scans were collected at the start of each fMRI run. Each run produced 340 volumes for each session. Anatomical T1-weighted images consisting of 184 sagittally-oriented slices were obtained using a turbo field echo pulse sequence (TR = 6.7 ms; TE = 3.1 ms; flip angle = 8°; voxel size = 1 × 1 × 1 mm; FOV = 256 × 256 mm). Also, to identify the LC, 39 transversally oriented slices of a high-resolution T1-weighted turbo spin-echo sequence were collected (TR = 600 ms; TE = 14 ms; flip angle = 90; voxel size = 0.4 × 0.5 × 3 mm; FOV = 220 × 178).

A polyphonic keyboard adapted and tested for a 3T MRI scanner (Jensen et al., 2017) was used for the experiments. It uses 25 full-size, keys covering two full octaves and it is designed ergonomically for the MRI scanner. The keyboard rests on the participant's legs so that all the keys are reachable by moving the forearms within the MRI scanner. The keyboard was attached by a MIDI cable to Novation NIO 2/4 USB audio interface and then connected, *via* USB cable, to the Windows laptop, using Reaper software for generating audio. The output from Novation audio interface is in turn connected to the Eurorack UB 1002 Audio mixer. The windows computer runs the experiment on E-prime generating audio. The output audio from the E-prime computer is delivered into another input in UB 1002 mixer. To reduce the impact of the scanner's noise, the audio mixer provides the ability to play the mono and stereo audio. The output from the Audio mixer is delivered through active noise-canceling headphones (OptoACTIVE).

### Procedure

The experiment was designed after Experiment 1 but optimized for fMRI by making two separate sessions for the "imagining" vs. "listening" and "playing" vs. "listening" conditions. We adopted a block design structure with an equal length of 15 s for "Listening, Imagining (playing)" and "Playing." These conditions were pseudorandomized, and a rest period of equal length introduced between the active blocks. The music sheets were presented using E-PRIME3 on a calibrated MRI compatible LCD screen (NNL LCD Monitor, Nordic Neurolab, Bergen, Norway) placed behind the scanner bore. In the "Listening and Imagining" session, PP was instructed either to listen to a music clip while viewing the music sheet or to imagine playing the same piece without moving fingers or hands. The "Listening and Playing" session was identical to the previous one, except that PP was instructed to play the music from the score sheet. In the "Imagining the sounds" condition, the participant was instructed to imagine hearing the melody in her head.

### Analysis

Functional data were transferred to 4D nifti and motion-corrected using SPM 12 (Ashburner and Friston, 2005). For the whole-brain analysis, we normalized the anatomical images to the MNI template using the unified segmentation and normalization algorithm implemented in SPM12 (Ashburner and Friston, 2005). The resulting transformation parameters were then applied to the functional images. Images were smoothed with a Gaussian kernel of 8 mm FWHM. A general linear model/GLM) was estimated for each voxel with a canonical hemodynamic response function (HRF) for each condition of interest. The statistical parametric maps were thresholded at $P$ values below $P < 0.05$ (FWE corrected) at cluster level for $t = 5.27$ for the cortical and $t = 5.43$ for the conjunction and clusters (extend threshold 55 voxels) derived from the fixed effects analysis.

For the brainstem analysis, the unsmoothed images were first analyzed in participant space using a general linear model

(GLM) with six HRF-convolved regressors representing blocks of activation during the three conditions (Listening, Imagining Playing, and Playing) separated over the two musical pieces. To obtain precision in localizing LC we first resampled the high-resolution anatomical scan to the Spatially Unbiased Infratentorial Template (SUIT; Diedrichsen, 2007), covering the cerebellum and brainstem. Next, β-images resulting from the GLM were resampled into the same template space at a resolution of 2 × 2 × 2 mm. Finally, we applied a probabilistic mask of LC (Tona et al., 2017) to the unsmoothed *t*-map for all the conditions to identify significant voxels associated with LC. Correspondence between the LC mask and anatomical LC was ensured through visual inspection of the relevant region in the high-resolution anatomical scan. While all LC-specific analyses were performed on unsmoothed data, the SUIT-normalized contrast images were also smoothed with a 3D Gaussian kernel (4 mm FWHM) for illustration purposes.

## Results and Discussion

Our first aim was a whole-brain approach to imagery and effort. Specifically, we subtracted the activations related to the "simple" piano piece from the "complex" piano piece for all the conditions (listening, imagining playing, and playing; see **Figure 10A**). Thus, the T2*-weighted scans were not optimized for the identification of functional activity within such a tiny structure as the LC (see Turker et al., 2019 for a review on issues with identifying LC with fMRI). Still, by applying the LC mask and a small volume correction analysis approach, we found one significant voxel in the left and two in the right superior LC

**TABLE 1** | Brain areas activated for professional pianist (PP) in the three tasks.

| | | Location | BA | MNA Coordinates | | | kE |
|---|---|---|---|---|---|---|---|
| | | | | *x* | *y* | *z* | |
| **Contrast of playing the complex piano piece compared to the simple piece** | | | | | | | |
| Left | Frontal | Postcentral | BA: 2, 3, 4, 5, 7 | −28 | −24 | 54 | 937 |
| | | Precuneus | | −26 | −24 | 66 | 909 |
| | | Supplementary Motor Area | BA: 6 | −8 | −14 | 50 | 143 |
| | | Superior Frontal | | −26 | −8 | 52 | 143 |
| | Limbic | Uncus | BA: 20, 35 | −36 | 14 | −38 | 134 |
| | | Parahippocampal | BA: 38 | −22 | −2 | −28 | 100 |
| | | Hippocampus | BA: 21 | −30 | 4 | −34 | 19 |
| | Cerebellum | Cerebelum_Crus1, 2, Pyramis | | 12 | −82 | −30 | 38 |
| | LC | Superior | | −2 | −36 | −14 | 1* |
| Right | Temporal | Mid Temporal | | 42 | −34 | 2 | 96 |
| | | Superior Temporal | | 48 | −46 | −4 | 38 |
| | | Inferior Temporal | | 50 | −32 | −2 | 12 |
| | Limbic | Mid Cingulum | BA: 23 | 0 | −22 | 28 | 79 |
| | Cerebellum | Cerebelum_4_5_6 | | 4 | −64 | −20 | 250 |
| | | Vermis_4, 5, 6, 8 | | 8 | −52 | −12 | 283 |
| | LC | Superior | | 4 | −36 | −14 | 2* |
| **Contrast imagining the complex compared to the simple piano piece** | | | | | | | |
| | | | NS | | | | |
| **Contrast listening the complex compared to the simple piano piece** | | | | | | | |
| | | | NS | | | | |
| **Conjunction of playing and imagining playing** | | | | | | | |
| Left | Frontal | Supplementary Motor Area | BA: 6 | −4 | −4 | 58 | 832 |
| | | Inferior Frontal Gyrus | BA: 44 | −52 | 8 | 20 | 165 |
| | | Precentral Gyrus | BA: 6 | −48 | −6 | 48 | 240 |
| Right | Occipital | Inferior occipital | BA: 18, 37 | 50 | −66 | −16 | 411 |
| | Cerebellum | Crus 1 | | 52 | −68 | −28 | 362 |
| **Contrast of imagining compared to playing** | | | | | | | |
| | | | NS | | | | |
| **Conjunction of imagining playing and listening** | | | | | | | |
| | | | NS | | | | |
| **Contrast of imagining compared to listening** | | | | | | | |
| Left | Frontal | Supplementary Motor Area | BA: 6 | −4 | −4 | 58 | 744 |
| | | Frontal Superior | BA: 6 | −26 | −6 | 62 | 149 |
| | | Rolandic Operandis | | −56 | 6 | 6 | 87 |
| | | Inferior Frontal Gyrus | | −52 | 6 | 20 | 161 |
| | | Insula | | −38 | 6 | 2 | 220 |
| | Occipital | Fusiform Gyrus | | −38 | −76 | −20 | 47 |
| | Cerebellum | Crus_1, 6 | | 52 | −68 | −28 | 368 |
| Right | Frontal | Supplementary Motor Area | BA: 8 | 3 | 14 | 44 | 492 |
| | Occipital | Middle Occiptal Gyrus | BA: 37 | 50 | −66 | −16 | 41 |
| | | Inferior Occipital | BA: 18 | 40 | −80 | −16 | 101 |

**The LC analysis was done as a small volume corrected analysis with peak activation at threshold of F = 0.001, K = 1.*

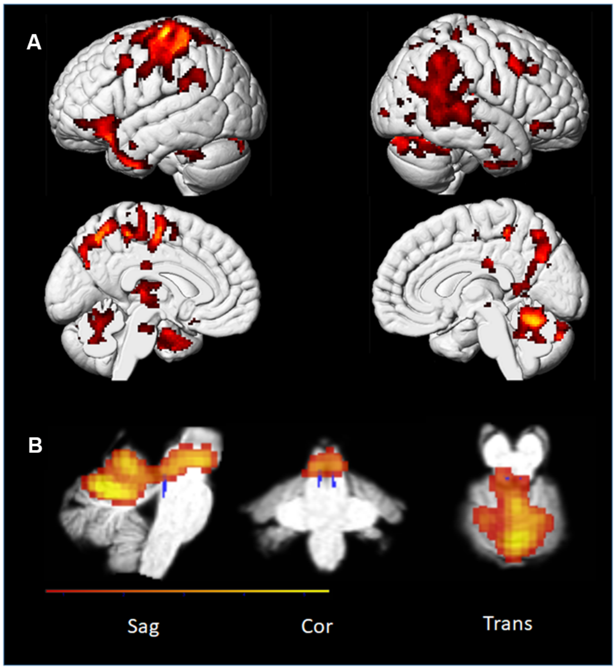



**FIGURE 10 |** PP's brain: *t*-maps representing areas showing activation when playing the difficult music piece contrasted to the simple. Maps are corrected for multiple comparisons using FWE (Cortical: $p < 0.05$, $k = 55$; and $p < 0.05$, $K = 8$ for the brainstem using the SUIT mask). **(A)** *t*-map projected on an MNI template cortical mesh. **(B)** Parametric map from the subcortical specific analysis (on a cropped version of a normalized image of PP). The locus coeruleus (LC) mask from Keren et al. (2009) is shown in blue. Brain stem activations are displayed on the SUIT template (Diedrichsen, 2007) using MRIcron (Rorden et al., 2007).

when comparing the two "playing" conditions (see **Table 1**). This fits with the visual inspection of the brainstem using the SUIT mask (**Figure 10B**). Even though these results must be regarded with caution, we suggest that our findings imply a role for LC related to cognitive load. For the other conditions, we did not find any significant voxels in this area. These results are consistent with the pupil findings that also showed clear differences in pupil size between the difficult Praeludium of the Holberg Suite, compared to Wächterlied and, in particular, during the "Playing" conditions (**Figure 5**).

Interestingly, we found no significant difference between the levels of difficulty for the Listening and Imagining Playing conditions in cortical activity, which underlines the fact that—at least at the single case level—monitoring brainstem's activity may best relate to mental effort than the general cortical activity.

In the whole-brain analyses, the two levels of effort were analyzed together to increase statistical power. We found several cortical areas that were more strongly activated during the Playing condition compared to the others. **Table 1** lists increased cortical activity in several areas associated with monitoring, motor planning, and motor execution. We found no significant result in the conjunction between Imagining Playing and Listening conditions suggesting that they imply different networks. Again, the pupil also differentiated Playing from Listening and Imagery (**Figure 4**) while there was a moderate relationship (**Figure 6**) between Listening and Imagery, suggesting some degree of overlap of the cognitive mechanisms and/or their degree of involvement in these conditions. We also found several areas where Imagining Playing provided more activations than the Listening condition. At the same time, we found that "Imagining" and "Listening" shared significant activations in the auditory cortex (see **Figure 11A**).

In summary, we did find consistently with the previous pupil findings that a complex piece of music contrasted to a simple piece lead to subcortical activations at the brain stem level. At the same time imagining playing shared cortical areas with actual playing (**Figure 11** and **Table 1**; conjunction analyses) which was also consistent with several previous neuroimaging evidence on musical imagery (e.g., Zatorre et al., 2007; Zhang et al., 2017). Also, we confirmed that in our musician's brain simply imagining a melody shared cortical areas with listening to the same melody.

## EXPERIMENT 2B

In the following MRI experiment, we looked at the brain activity in PP and a group of non-musicians when listening or imagining a well-known melody ("Happy Birthday"). In this MRI experiment, we necessarily excluded a condition with piano playing since that would not possible for such a non-musician group. In the "Imagine the melody" experiment, the participants listened to a simple well-known melody in blocks of 10 s and were asked to imagine listening to the same melody also in 10-s blocks. The conditions where pseudorandomized and a 10 s rest period were introduced between each block.

### Method

#### Participants

PP and eight control participants (five females; mean age = 26.2 years; range: 23–42). The control participants declared to be non-musicians; nevertheless, they all completed The Goldsmiths Musical Sophistication Index (v1.0), indicating an average score of 4.5 in musical training (it was 36 for PP) and a General Sophistication score = 48.8. The apparatus was the same as in Experiment 2A. The same prepossessing steps applied to the data. For the control participants, we derived the functional data based on the thresholded ($P < 0.05$, FWE corrected, $t = 5.13$) fixed effects data and extended the cluster threshold to 20 voxels.

#### Apparatus and Analysis

These were identical to the previous MRI experiment.

### Results and Discussion

We found similar activations of "Imagining a melody" and "Listening to the melody" in PP's auditory cortex, as confirmed by a conjunction model (**Figures 11A, 12D**). This is consistent, even at the single case level, with the widespread idea that imagery uses the same neural substrate of perception (Kosslyn, 1980, 1994; Zatorre et al., 1996; Halpern and Zatorre, 1999;

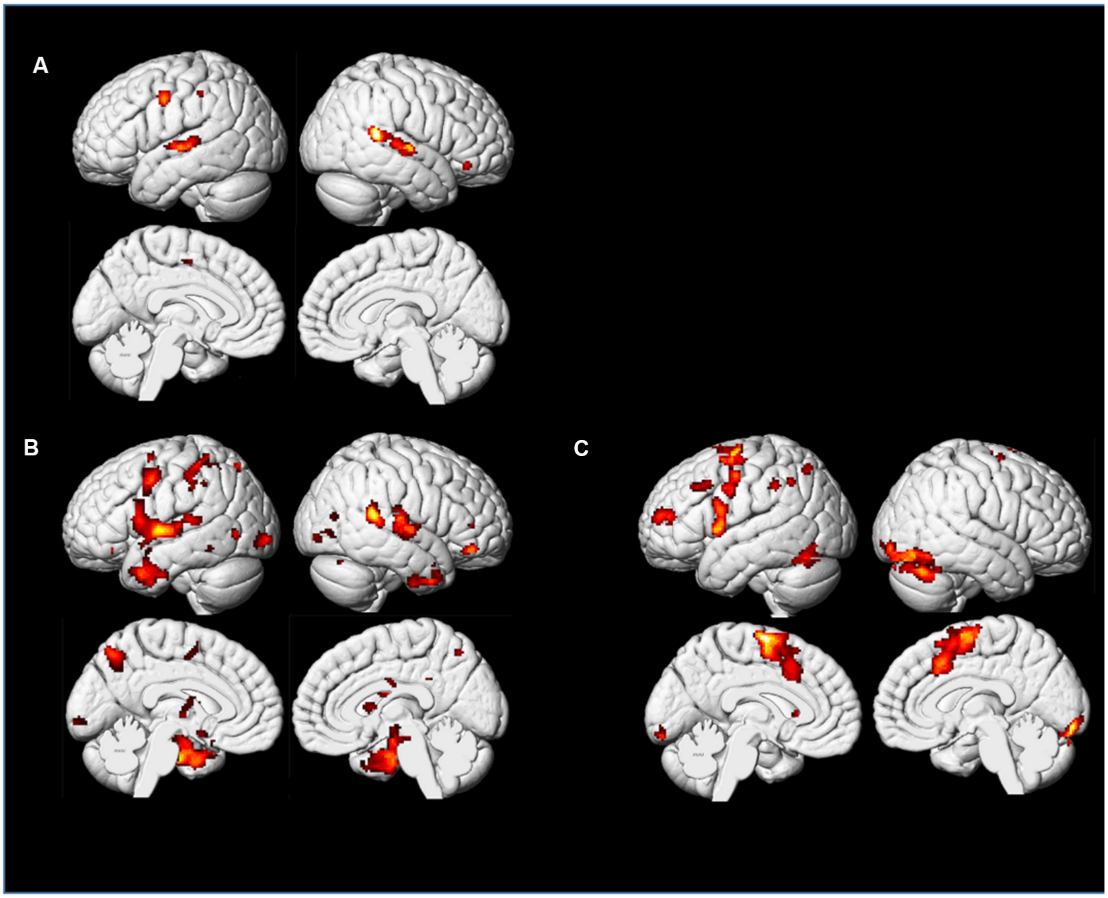



**FIGURE 11 |** PP's brain: *t*-maps representing PP's cortical activations for Playing, Imagining Playing, and Imagining Listening. The maps were corrected for multiple comparisons using FWE ($p < 0.05$, $k = 22$). **(A)** Conjunction analysis of Imagining a melody and Listening to the same melody. **(B)** Conjunction analysis of Imagining Playing and actual Playing of the same piece. **(C)** The contrast between Imagining Playing a piece of music vs. Imagining Listening to it.

Pearson et al., 2015). Interestingly, we found overlapping activations in the auditory cortex for non-musicians during the "Listening" condition, but we also found extended occipital activations (**Figures 12A–C**), within visual areas of the cortex, that were not revealed for PP. Although both PP and her controls were looking at the music score of the melody, there was likely less need of attending to this visual information for the professional pianist than for the non-musicians (see Schön et al., 2002). Moreover, as seen in **Figures 11C**, **12D**, activations were in general lower for PP in the same areas than in the non-musicians. Lower neural activity in experts may seem paradoxical but it may be a hallmark of expertise (in musicians: Jäncke et al., 2000; Krings et al., 2000; Koeneke et al., 2004; but also in sport athletes: Naito and Hirose, 2014). That is, long-term training sharpens the relevant neural networks and dampens or filters irrelevant or noisy activity (Milton et al., 2007), so that the network becomes more efficient and uses lower activity or fewer dedicated units for its operation. Interestingly, the findings of Experiment 1B, where the pupil size when listening had an inverse relationship to expertise (**Figure 8**), seem consistent with the idea that experts require lower levels of effort, perhaps because the relevant neural network has become more efficient, than the less experienced or the novices.

## DISCUSSION

The main goal of the present study was to examine "musical effort" or the cognitive workload that the act of imagining music, playing it, or listening to it, imposes on the mind or brain of a professional pianist as well as other individuals with different musical expertise. We used a multi-pronged approach, by use of both methods of pupillometry and magnetic resonance imaging, comparing these results between a professional pianist and groups of non-professional pianists or non-musicians. Although the present neuropsychological study is certainly limited, we believe that we succeeded in offering some initial but promising results about a few fundamental questions. First, we can measure musical effort through the eye pupil similarly to reading out cognitive workload and arousal in other domains. Second, musical imagery engages sensory and motor areas of the brain concerning the mental effort required by the

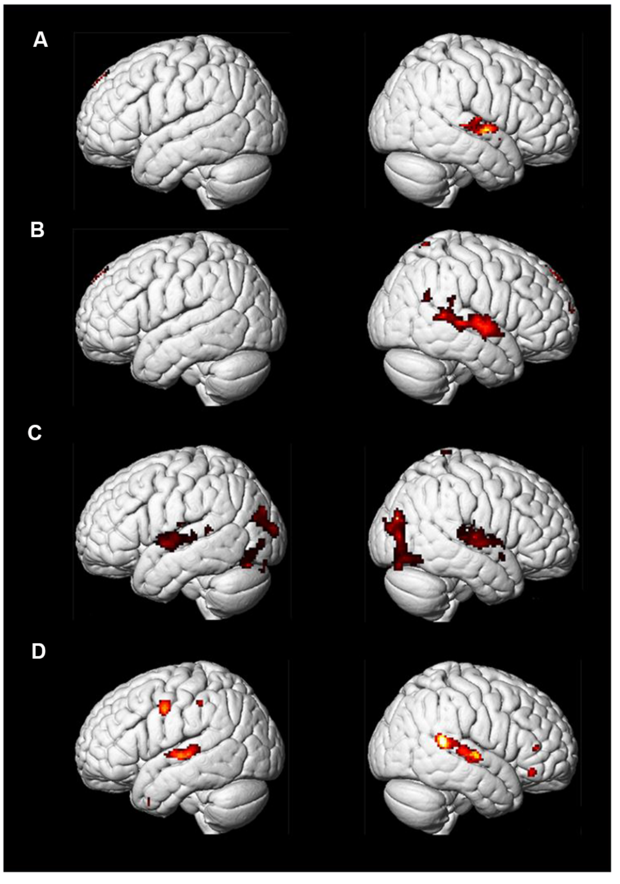



**FIGURE 12** | Control participants and PP: *t*-maps representing areas showing activations in the Imagining and Listening a well-known melody ("Happy Birthday"): Panel **(A)** shows the conjunction of Imagining and Listening for the group of non-musicians. Panel **(B)** shows the contrast between Imagining and Listening to the same melody. Panel **(C)** shows brain areas active when the non-musicians listen to a piano melody while looking at the musical score sheet showing the notes. Panel **(D)** show the same conditions as in Panel **(C)** but for PP only.

complexity of the structure and execution of the imagined music. Third, musical imagery is similar to music listening in both non-musicians and musicians. Fourth, there is a degree of functional similarity between playing, imagining, and listening, when comparing both the activity in the NE system, as indexed by the pupil, and by the overlap in activity between sensory and motor neural networks in the whole brain. Finally, we revealed "audiation" in an expert musician with pupillometry, given the similarity in which the pupil changed during musical listening and imagery.

## Musical Effort as Revealed in the Eye Pupils

Experiment 1 provided clear evidence for a relationship between NE activity and mental workload or attentional intensity within the domain of music cognition. That is, the conditions requiring action were more effortful, regardless of the presence of auditory feedback than the conditions where no performance took place. Experiment 2 with fMRI confirmed in part this aspect since effort-related activations appeared most clearly in the playing condition with the professional pianist (PP) within the superior part of the LC. The fact that differences in the effort were not significant in the listening and imagery conditions of Experiment 2 can be explained by the fact that the latter two conditions engaged less the LC-NE system than the motor condition, as already shown by the small responses of the pupil in Experiment 1. The act of playing is also likely to add effortful motor planning and some degree of physical effort.

As expected, we found positive correlations between pupil diameters of the professional pianist during different "executions" of the same piano piece (i.e., normal, silenced, listened, and imagined), which also indicate similar task demands and differential degrees of load on resources. This might have also indicated the engagement of similar cognitive and execution mechanisms and, possibly, similar affective processes as the music unfolded. Our finding of a close affinity in cognitive workload between standard playing and silent playing confirms an intimate link between the motor imagery of sound-producing body motions and the resultant sounds (Godøy and Jørgensen, 2001).

Comparing a group of (non-professional) pianists and non-musicians while listening to PP's performance of the three differently effortful pieces revealed several interesting aspects. Pianists attended more intensively to the most difficult piece than non-musicians since they showed larger pupils than non-musicians only for the most difficult piece. Non-musicians seemed to be the most engaged group by listening since their pupil size was larger overall than for the pianists as well as PP. This suggests that the amount of attention allocated for the same task may follow a hierarchy of expertise demanding less attentional effort in expert or performers than in novices.

Interestingly, there was only weak evidence for a commonality between subjective effort ratings and the objective effort gauged with pupil diameter during listening. The lack of a strong relationship suggests that psychophysiological methods like pupillometry index mental effort in a manner that is not "observable" in awareness or *via* introspection (Laeng et al., 2012; Laeng and Sulutvedt, 2014). Future studies should clarify to what extent subjective reports and objective measures of effort dissociate (e.g., Bruya, 2010). However, Kahneman (1973) had already pointed out that attentional effort is not identical to either "felt" effort or the observed likelihood of error. This is because effortless and overlearned tasks (e.g., telling someone one's phone number) can visibly increase the pupil, revealing that the ease of retrieving information from long-term memory, instead of the load on working memory, lies behind the feeling of "effortlessness."

## Musical Effort as Revealed by LC Activity

To our knowledge, the association between mental effort and LC activity in piano playing, albeit less in imagery, is the first reported in the literature. The brainstem region corresponding, according to current methods, to the anatomical coordinates of the LC were active when playing the "complex" piece by Grieg than the simpler piece, suggesting a role for NE activity varying with mental load (see **Figure 10B**). Caution must be taken in interpreting this result of course since this is a single subject study and there are several challenges involved both in identifying the LC areas and in interpreting fMRI results associated with LC. We, therefore, regard our findings as suggestive and promising for future studies to explore the connection between LC and mental effort.

Moreover, the fMRI results indicated that when PP imagined playing, this activated her brain's premotor, motor, and perceptual areas, while when imagining the sounds of a melody (without playing) the auditive areas become predominantly active. This is consistent with previous studies of imagery: Bangert et al. (2006) found that a distributed network involving SMA and superior temporal gyrus activated during a muted keyboard task with pianists. Also, Gerardin et al. (2000) found evidence for overlapping neural networks responsible for imagined and real movements. Moreover, Bastepe-Graya et al. (2020) showed that imagery in an "oud" musician activated sensory and motor areas similarly when playing. Thus, the present study's findings are also consistent with several previous neuroimaging studies (e.g., Langheim et al., 2002; Meister et al., 2004) by showing a substantial cortical overlap between playing and imagining music in a pianist and the presence of activity in the primary motor cortex during imagery. Importantly, these differential conjunctions in brain activity appear to mirror those exposed by pupillometry, where imagery appeared to be most similar to listening to the same piece in the professional pianist; a finding that we would like to interpret as capturing "audiation" on the fly.

## LIMITATIONS

The present findings should be considered cautiously since their generalizability may be limited given that we presented a single case, though of very high-level, pianistic, expertise. We also note that the control samples were small in both experiments, especially in the fMRI study. Hence, several of the present results are indicative rather than conclusive. Moreover, the study did not include a control experiment in which participants listened and imagined also non-musical contents, which could have thrown light on brain activity specifically related to music. The present study can be considered exploratory since it was motivated by a gap in music psychology research concerning mental effort and its role in musical performance and imagery.

## CONCLUSIONS

There is "musical effort" and it is measurable by the use of pupillometry. The use of pupillometry as a gauge of mental work in music is novel and the present findings suggest its potential. This objective method in measuring effort offers insights that may not be easy to expose by verbal reports or observing behavior. A combined and complementary approach of psychophysiology and neuroimaging seems very promising and it can provide converging evidence that considerably strengthens interpretations. In particular, activity in the brainstem's LC modulated by task complexity is consistent with changes in the level of mental effort and NE neuromodulation as indexed by pupil size. Musical imagery has a strong commonality with music listening in both experts and naïve individuals.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The study was reviewed and approved by the IRB of the Department of Psychology, University of Oslo (Ref. number: 3568281). The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

BL, RG, and TE contributed to conception and design of the study. TE, MS, TH, AB, and BL performed the experiments and statistical analysis. TH and BL pre-processed the pupillometry data and TE pre-processed the fMRI data. BL wrote the first draft of the manuscript. TE wrote the fMRI section of the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Research Council of Norway through its Centers of Excellence scheme to RITMO, project number 262762.

## REFERENCES

Alnæs, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., and Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *J. Vis.* 14:1. doi: 10.1167/14.4.1

Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *NeuroImage* 26, 839–851. doi: 10.1016/j.neuroimage.2005.02.018

Bailes, F. (2006). The use of experience-sampling methods to monitor musical imagery in everyday life. *Musicae Scientiae* 10, 173–187. doi: 10.1177/102986490601000202

Bailes, F. (2015). Music in mind? An experience sampling study of what and when, towards an understanding of why. *Psychomusicol. Music Mind Brain* 25, 58–68. doi: 10.1037/pmu0000078

Bailes, F. (2019). "Empirical musical imagery beyond the "mind's ear"" in *The Oxford Handbook of Sound and Imagination (Vol. 2),* eds

M. Grimshaw-Aagaard, M. Walther-Hansen, and M. Knakkergaard (Oxford Handbooks Online). doi: 10.1093/oxfordhb/9780190460242.013.1

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., et al. (2006). Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *NeuroImage* 30, 917–926. doi: 10.1016/j.neuroimage.2005.10.044

Bastepe-Graya, S. E., Acer, N., Gumus, K. Z., Gray, J., and Degirmencioglue, L. (2020). Not all imagery is created equal: a functional Magnetic resonance imaging study of internally driven and symbol driven musical performance imagery. *J. Chem. Neuroanat.* 104:101748. doi: 10.1016/j.jchemneu.2020.101748

Battisti, F. (2007). *On Becoming A Conductor: Lessons and Meditations on the Art of Conducting.* Galesville: Meredith Music Publications.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.* 91, 276–292. doi: 10.1037/0033-2909.91.2.276

Betts, G. H. (1909). *The Distribution and Functions of Mental Images.* New York, NY: Teachers College, Columbia University.

Bishop, L., Bailes, F., and Dean, R. T. (2013). Musical expertise and the ability to imagine loudness. *PLoS One* 8:e56052. doi: 10.1371/journal.pone.0056052

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., and Bendayan, R. (2017). Non-normal data: is ANOVA still a valid option? *Psicothema* 29, 552–557. doi: 10.7334/psicothema2016.383

Borst, G., and Kosslyn, S. M. (2008). Visual mental imagery and visual perception: structural equivalence revealed by scanning processes. *Mem. Cogn.* 36, 849–862. doi: 10.3758/mc.36.4.849

Bowling, D. L., Graf Ancochea, P., Hove, M. J., and Fitch, W. T. (2019). Pupillometry of groove: evidence for noradrenergic arousal in the link between music and movement. *Front. Neurosci.* 12:1039. doi: 10.3389/fnins.2018.01039

Brodsky, W., Henik, A., Rubinstein, B., and Zorman, M. (2003). Auditory imagery from musical notation in expert musicians. *Percept. Psychophys.* 65, 602–612. doi: 10.3758/bf03194586

Brodksy, W., Kessler, Y., Rubenstein, B. S., Ginsborg, J., and Henik, A. (2008). The mental representation of music notation: notational audiation. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 427–445. doi: 10.1037/0096-1523.34.2.427

Bruya, B. (2010). *Effortless Attention: A New Perspective in the Cognitive Science of Attention and Action.* Cambridge, MA: MIT Press.

Cahn, D. (2008). The effects of varying ratios of physical and mental practice and task difficulty on performance of a tonal pattern. *Psychol. Music* 36, 179–191. doi: 10.1177/0305735607085011

Cebrian, A. N., and Janata, P. (2010). Electrophysiological correlates of accurate mental image formation in auditory perception and imagery tasks. *Brain Res.* 1342, 39–54. doi: 10.1016/j.brainres.2010.04.026

Chaffin, C. R. (2009). *An Examination of the Cognitive Workload Associated With Conducting in an Instrumental Music Context.* Ann Arbor, MI: UMI Dissertation Publishing.

Christensen, J. (2019). "Improvisation: an ideal display of imagination in embodied perception," in *The Oxford Handbook of Sound and Imagination (Vol. 2)*, eds M. Grimshaw-Aagaard, M. Walther-Hansen, and M. Knakkergaard (Oxford Handbooks Online). doi: 10.1093/oxfordhb/9780190460242.013.7

Clark, T., and Williamon, A. (2011). Imagining the music: methods for assessing musical imagery ability. *Psychol. Music* 40, 471–493. oxford, UK: doi: 10.1177/0305735611401126

Clark, T., Williamon, A., and Aksentijevic, A. (2012). "Musical imagery and imagination: the function, measurement and application of imagery skills for performance," in *Musical Imaginations: Multidisciplinary Perspectives on Creativity, Performance and Perception*, eds D. Hargreaves, D. Miell, and R. MacDonald (New York, NY: Oxford University Press), 351–365.

Coffman, D. (1990). Effects of mental practice, physical practice and knowledge of results on piano performance. *J. Res. Music Educ.* 38, 187–196. doi: 10.2307/3345182

Combarieu, J. (1907). *La Musique: Ses Lois, Son évolution.* Paris: Flammarion.

Cotter, K. N. (2019). Mental control in musical imagery. A dual component model. *Front. Psychol.* 10:1904. doi: 10.3389/fpsyg.2019.01904

Cox, A. (2016). *Music and Embodied Cognition.* Bloomington, IN: Indiana University Press.

Davidson-Kelly, K., Schaeffer, R. S., Moran, N., and Overy, K. (2015). "Total Inner Memory": deliberate uses of multimodal musical imagery during performance preparation. *Psychomusicol. Music Mind Brain* 25, 83–92. doi: 10.1037/pmu0000091

Diedrichsen, J. (2007). A spatially unbiased atlas template of the human cerebellum. *NeuroImage* 33, 127–138. doi: 10.1016/j.neuroimage.2006.05.056

Driskell, J. E., Copper, C., and Moran, A. (1994). Does mental practice enhance music performance? *J. Appl. Psychol.* 79, 481–492. doi: 10.1037/0021-9010.79.4.481

Farrugia, N., Jakubowski, K., Cusack, R., and Stewart, L. (2015). Tunes stuck in your brain: the frequency and affective evaluation of involuntary musical imagery correlate with cortical structure. *Conscious. Cogn.* 35, 66–77. doi: 10.1016/j.concog.2015.04.020

Fidler, F., and Loftus, G. R. (2009). Why figures with error bars should replace *p* values. *J. Psychol.* 217, 27–37. doi: 10.1027/0044-3409.217.1.27

Fine, P. A., Wise, K. J., Goldemberg, R., and Bravo, A. (2015). Performing musicians' understanding of the terms "mental practice" and "score analysis". *Psychomusicol. Music Mind Brain* 25, 69–82. doi: 10.1037/pmu0000068

Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* 13, 361–377. doi: 10.3758/bf03193857

Gallagher, S. (2005). *How the Body Shapes the Mind.* Oxford: Clarendon.

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479. doi: 10.1080/02643290442000310

Gallese, V., and Metzinger, T. (2003). Motor ontology: the representational reality of goals, actions and selves. *Philos. Psychol.* 16, 365–388. doi: 10.1080/0951508032000121760

Gerardin, E., Sirigu, A., and Lehéricy, S. (2000). Partially overlapping neural networks for real and imagined hand movements. *Cereb. Cortex* 10, 1093–1104. doi: 10.1093/cercor/10.11.1093

Gingras, B., Marin, M. M., Puig-Waldmüller, E., and Fitch, W. T. (2015). The eye is listening: music-induced arousal and individual differences predict pupillary responses. *Front. Hum. Neurosci.* 9:619. doi: 10.3389/fnhum.2015.00619

Glenberg, A. M., and Gallese, V. (2012). Action-based language: a theory of language acquisition, comprehension, and production. *Cortex* 48, 905–922. doi: 10.1016/j.cortex.2011.04.010

Glover, S., Bibby, E., and Tuom, E. (2020). Executive functions in motor imagery: support for the motor-cognitive model over the functional equivalence model. *Exp. Brain Res.* 238, 931–944. doi: 10.1007/s00221-020-05756-4

Godøy, R. I. (2001). "Imagined action, excitation and resonance," in *Musical Imagery*, eds R. I. Godøy and H. Jørgensen (New York, NY: Taylor and Francis), 237–250.

Godøy, R. I. (2010a). "Gestural affordances of musical sound," in *Musical Gestures: Sound, Movement and Meaning*, eds R. I. Godøy and M. Leman (New York, NY: Routledge), 105–127.

Godøy, R. I. (2010b). Images of sonic objects. *Organised Sound* 15, 54–62. doi: 10.1017/s1355771809990264

Godøy, R. I., Haga, E., and Jensenius, A. (2006). "Playing 'air instruments': mimicry of sound-producing Gestures by Novices and Experts," in *Gesture in Human-Computer Interaction and Simulation: 6th International Gesture Workshop. LNAI 3881*, eds S. Gibet, N. Courty, and J.-F. Kamp (Berlin: Springer-Verlag), 256–267.

Godøy, R. I., and Jørgensen, H. (2001). *Musical Imagery.* Lisse: Swets and Zeitlinger.

Godøy, R. I., and Leman, M. (2010). *Musical Gestures: Sound, Movement and Meaning.* New York, NY: Routledge.

Gracyk, T. (2019). "Imaginative listening to music," in *The Oxford Handbook of Sound and Imagination (Vol. 2)*, eds M. Grimshaw-Aagaard, M. Walther-Hansen, and M. Knakkergaard (Oxford Handbooks Online). doi: 10.1093/oxfordhb/9780190460242.013.11

Gregg, M. J., Clark, T. W., and Hall, C. R. (2008). Seeing the sound: an exploration of the use of mental imagery by classical musicians. *Musicae Scientiae* 12, 231–247. Oxford, UK: doi: 10.1177/102986490801200203

Halpern, A. A., and Overy, K. (2019). "Voluntary auditory imagery and music pedagogy," in *The Oxford Handbook of Sound and Imagination (Vol. 2)*, eds M. Grimshaw-Aagaard, M. Walther-Hansen, and M. Knakkergaard (Oxford Handbooks Online). doi: 10.1093/oxfordhb/9780190460242.013.49

Halpern, A. R., and Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cereb. Cortex* 9, 697–704. doi: 10.1093/cercor/9.7.697

Halpern, A. R., Zatorre, R. J., Bouffard, M., and Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia* 42, 1281–1292. doi: 10.1016/j.neuropsychologia.2003.12.017

Harris, R., and de Jong, B. M. (2015). Differential parietal and temporal contributions to music perception in improvising and score-dependent musicians, an fMRI study. *Brain Res.* 1624, 253–264. doi: 10.1016/j.brainres.2015.06.050

Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., and Ceballos-Baumann, A. O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *J. Cogn. Neurosci.* 17, 282–293. doi: 10.1162/0898929053124893

Herholz, S. C., Lappe, C., Knief, A., and Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *Eur. J. Neurosci.* 28, 2352–2360. doi: 10.1111/j.1460-9568.2008.06515.x

Hess, E. H., and Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science* 140, 1190–1192. doi: 10.1126/science.143.3611.1190

Highben, Z., and Palmer, C. (2004). Effects of auditory and motor mental practice in memorized piano performance. *Bull. Council Res. Music Educ.* 159, 58–65.

Hodges, D. A. (2009). "Bodily responses to music," in *The Oxford Handbook of Music Psychology*, 2nd Edn, eds S. Hallam, I. Cross, and M. Thaut (New York, NY: Oxford University Press), 183–196.

Høffding, S. (2019). *A Phenomenology of Musical Absorption.* Cham, Switzerland: Palgrave MacMillan.

Holmes, P. (2005). Imagination in practice: a study of the integrated roles of interpretation, imagery and technique in the learning and memorisation processes of two experienced solo performers. *Br. J. Music Educ.* 22, 217–235. doi: 10.1017/s0265051705006613

Hubbard, T. L. (2010). Auditory imagery: empirical findings. *Psychol. Bull.* 136, 302–329. doi: 10.1037/a0018436

Hubbard, T. L. (2013). "Auditory imagery contains more than audition," in *Multisensory Imagery*, eds S. Lacey and R. Lawson (New York, NY: Springer), 221–247.

Hubbard, T. L. (2017). Momentum in music: musical succession as physical motion. *Psychomusicol. Music Mind Brain* 27, 14–30. doi: 10.1037/pmu0000171.

Hubbard, T. L. (2018). "Neural mechanisms of musical imagery," in *The Oxford Handbook of Music and Brain*, eds M. H. Thaut and D. A. Hodges (New York, NY: Oxford University Press), 521–549.

Hubbard, T. L., and Stoeckig, K. (1988). Musical imagery: generation of tones and chords. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 656–667. doi: 10.1037//0278-7393.14.4.656

Janata, P. (2001). Brain electrical activity evoked by mental formation of auditory expectations and images. *Brain Topogr.* 13, 169–193. doi: 10.1023/a:1007803102254

Janata, P., and Paroo, K. (2006). Acuity of auditory images in pitch and time. *Percept. Psychophys.* 68, 829–844. doi: 10.3758/bf03193705

Jäncke, L., Baumann, S., Koeneke, S., Meyer, M., Laeng, B., Peters, M., et al. (2006). Neural control of playing a reversed piano: empirical evidence for an unusual cortical organisation of musical functions. *Neuroreport* 17, 447–451. doi: 10.1097/01.wnr.0000204978.91253.33

Jäncke, L., Shah, N. J., and Peters, M. (2000). Cortical activations in primary and secondary motor areas for complex bimanual movements in professional pianists. *Cognitive Brain Res.* 10, 177–183. doi: 10.1016/s0926-6410(00)00028-8

Jensen, M. S., Heggli, O. A., Alves, D. M., and Vuust, P. (2017). "A low-cost MRI compatible keyboard," in *Proceedings of the New Interfaces in Musical Experiments*, (Copenhagen: Aalborg University), 257–260.

Joshi, S., Li, Y., Kalwani, R. M., and Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron* 89, 221–234. doi: 10.1016/j.neuron.2015.11.028

Jourdain, R. (1997). *Music, the Brain and Ecstasy: How Music Captures Our Imagination.* New York, NY: HarperCollins.

Just, M. A., and Carpenter, P. A. (1993). The intensity dimension of thought: pupillometric indices of sentence processing. *Can. J. Exp. Psychol.* 47, 310–339. doi: 10.1037/h0078820

Just, M. A., Carpenter, P. A., and Miyake, A. (2003). Neuroindices of cognitive workload: neuroimaging, pupillometric and event-related potential studies of brain work. *Theor. Issues Ergon. Sci.* 4, 56–88. doi: 10.1080/14639220210159735

Kahneman, D. (1973). *Attention and Effort.* Englewood Cliffs, NJ: Prentice Hall.

Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science* 154, 1583–1585. doi: 10.1126/science.154.3756.1583

Kahneman, D., Tursky, B., Shapiro, D., and Crider, A. (1969). Pupillary, heart rate, and skin resistance changes during a mental task. *J. Exp. Psychol.* 79, 164–167. doi: 10.1037/h0026952

Kang, O., and Wheatley, T. (2015). Pupil dilation patterns reflect the contents of consciousness. *Conscious. Cogn.* 35, 128–135. doi: 10.1016/j.concog.2015.05.001

Keller, P. E. (2001). Attentional resource allocation in musical ensemble performance. *Psychol. Music* 29, 20–38. doi: 10.1177/0305735601291003

Keller, P. E. (2012). Mental imagery in musical performance: underlying mechanisms and potential benefits. *Ann. N Y Acad. Sci.* 1252, 206–213. doi: 10.1111/j.1749-6632.2011.06439.x

Keren, N. I., Lozar, C. T., Harris, K. C., Morgan, P. S., and Eckert, M. A. (2009). *In vivo* mapping of the human locus coeruleus. *NeuroImage* 47, 1261–1267. doi: 10.1016/j.neuroimage.2009.06.012

Koeneke, S., Lutz, K., Wüstenberg, T., and Jäncke, L. (2004). Long-term training affects cerebellar processing in skilled keyboard players. *Neuroreport* 15, 1279–1282. doi: 10.1097/01.wnr.0000127463.10147.e7

Korsakova-Kreyn, M. (2018). Two-level model of embodied cognition in music. *Psychomusicology: Music, Mind, and Brain* 28, 240–259. doi: 10.1037/pmu0000228

Kosslyn, S. M. (1980). *Image and Mind.* Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate.* Cambridge, MA: MIT Press.

Kosslyn, S. M., Ball, T. M., and Reiser, B. J. (1978). Visual images preserve metric spatial information: evidence from studies of image scanning. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 47–60. doi: 10.1037//0096-1523.4.1.47

Kraemer, D. J. M., Macrae, C. N., Green, A. E., and Kelly, W. M. (2005). Musical imagery: sound of silence activates auditory cortex. *Nature* 434:158. doi: 10.1038/434158a

Krings, T., Töpper, R., Foltys, H., Erberich, S., Sparing, R., Willmes, K., et al. (2000). Cortical activation patterns during complex motor tasks in piano players and controls subjects. A functional magnetic resonance imaging study. *Neurosci. Lett.* 278, 189–193. doi: 10.1016/s0304-3940(99)00930-1

Laeng, B., and Alnæs, D. (2019). "Pupillometry," in *Eye Movement Research*, eds C. Klein and U. Ettinger (Switzerland: Springer Nature), 449–502.

Laeng, B., Eidet, L. M., Sulutvedt, U., and Panksepp, J. (2016). Music chills: the eye pupil as a mirror to music's soul. *Conscious. Cogn.* 44, 161–178. doi: 10.1016/j.concog.2016.07.009

Laeng, B., Sirois, S., and Gredebäck, G. (2012). Pupillometry: a window to the preconscious? *Perspect. Psychol. Sci.* 7, 18–27. doi: 10.1177/1745691611427305

Laeng, B., and Sulutvedt, U. (2014). The eye pupil adjusts to imaginary light. *Psychol. Sci.* 25, 188–197. doi: 10.1177/0956797613503556

Langheim, F. J., Callicott, J. H., Mattay, V. S., Duyn, J. H., and Weinberger, D. R. (2002). Cortical systems associated with covert music rehearsal. *NeuroImage* 16, 901–908. doi: 10.1006/nimg.2002.1144

Leaver, A. M., van Lare, J., Zielinski, B., Halpern, A. R., and Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *J. Neurosci.* 29, 2477–2485. doi: 10.1523/JNEUROSCI.4921-08.2009

Leman, M. (2008). *Embodied Music Cognition and Mediation Technology.* Cambridge, MA: The MIT Press.

Levitin, D., and Grafton, S. T. (2016). Measuring the representational space of music with fMRI: a case study with sting. *Neurocase* 22, 548–557. doi: 10.1080/13554794.2016.1216572

Liao, H.-I., Yoneya, M., Kashino, M., and Furukawa, S. (2018). Pupillary dilation response reflects surprising moments in music. *J. Eye Mov. Res.* 11:13. doi: 10.16910/jemr.11.2.13

Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36. doi: 10.1016/0010-0277(85)90021-6

Libby, W. L., Lacey, B. C., and Lacey, J. I. (1973). Pupillary and cardiac activity during visual attention. *Psychophysiology* 10, 270–294. doi: 10.1111/j.1469-8986.1973.tb00526.x

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461. doi: 10.1037/h0020279

Limb, C. J., and Braun, A. R. (2008). Neural substrates of spontaneous musical performance: an fMRI study of jazz improvisation. *PLoS One* 3:e1679. doi: 10.1371/journal.pone.0001679

Loewenfeld, I. (1999). *The Pupil: Anatomy, Physiology, and Clinical Applications.* Oxford: Butterworth-Heinemann, 2278.

Loftus, G. R., and Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychon. Bull. Rev.* 1, 476–490. doi: 10.3758/BF03210951

Lotze, M., and Halsband, U. (2006). Motor imagery. *J. Physiol.* 99, 386–395. doi: 10.1016/j.jphysparis.2006.03.012

Lotze, M., Scheler, G., Tan, H.-R. M., Braun, C., and Birbaumer, N. (2003). The musician's brain: functional imaging of amateurs and professionals during performance and imagery. *NeuroImage* 20, 1817–1829. doi: 10.1016/j.neuroimage.2003.07.018

Lu, J., Yang, H., He, H., Jeon, S., Hou, C., Evans, A. C., et al. (2017). The multiple-demand system in the novelty of musical improvisation: evidence from an MRI study on composers. *Front. Neurosci.* 11:695. doi: 10.3389/fnins.2017.00695

Mäki-Marttunen, V., Hagen, T., Bruno Laeng, B., and Espeseth, T. (2019). Distinct neural mechanisms meet challenges in dynamic visual attention due to either load or object spacing. *J. Cogn. Neurosci.* 32, 65–84. doi: 10.1162/jocn_a_01469

Mathôt, S., Fabius, J., Van Heusden, E., and Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behav. Res. Methods* 50, 94–106. doi: 10.3758/s13428-017-1007-2

Meister, I. G., Krings, T., Foltys, H., Boroojerdi, B., Müller, M., Töpper, R., et al. (2004). Playing piano in the mind—an fMRI study on music imagery and performance in pianists. *Cogn. Brain Res.* 19, 219–228. doi: 10.1016/j.cogbrainres.2003.12.005

Mesaros, H. (2008). *Bravo Fortissimo Glenn Gould: The Mind of a Canadian Virtuoso.* Baltimore, MD: American Literary Press, Inc.

Mikumo, M. (1994). Motor encoding strategy for pitches of melodies. *Music Percept.* 12, 175–197. doi: 10.2307/40285650

Milton, J., Solodkin, A., Hluštík, P., and Small, S. L. (2007). The mind of expert motor performance is cool and focused. *NeuroImage* 35, 804–813. doi: 10.1016/j.neuroimage.2007.01.003

Montero, B. G. (2016). *Thought in Action: Expertise and the Conscious Mind.* Oxford, UK: Oxford University Press.

Moseley, P., Alderson-Day, B., Kumar, S., and Fernyhough, C. (2018). Musical hallucinations, musical imagery, and earworms: a new phenomenological survey. *Conscious. Cogn.* 65, 83–94. doi: 10.1016/j.concog.2018.07.009

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., and Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Hum. Brain Mapp.* 35, 4140–4154. doi: 10.1002/hbm.22466

Naito, E., and Hirose, S. (2014). Efficient foot motor control by Neymar's brain. *Front. Hum. Neurosci.* 8:594. doi: 10.3389/fnhum.2014.00594

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., and Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* 15, 1040–1046. doi: 10.1038/nn.3130

Nunnally, J. C., Knott, P. D., Duchnowski, A., and Parker, R. (1967). Pupillary response as a general measure of activation. *Percept. Psychophys.* 2, 149–155. doi: 10.3758/bf03210310

O'Shea, H., and Moran, A. (2016). Chronometric and pupil-size measurements between motor execution and motor imagery in expert pianists. *Psychol. Music* 44, 1289–1303. doi: 10.1177/0305735615616286

Pearson, J., Naselaris, T., Holmes, E. A., and Kosslyn, S. M. (2015). Mental imagery: functional mechanisms and clinical applications. *Trends Cogn. Sci.* 19, 590–602. doi: 10.1016/j.tics.2015.08.003

Peñalba, A. A. (2011). "Towards a theory of proprioception as a bodily basis for consciousness in music," in *Music and Consciousness: Philosophical, Psychological and Cultural Perspectives*, eds D. Clarke and E. Clarke (Oxford: Oxford University Press), 215–230.

Platz, F., and Kopiez, R. (2012). When the eye listens: a meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Percept.* 30, 71–83. doi: 10.1525/mp.2012.30.1.71

Reed, C. L. (2002). Chronometric comparisons of imagery to action: visualising versus physically performing springboard dives. *Mem. Cogn.* 30, 1169–1178. doi: 10.3758/bf03213400

Reisberg, D. (1992). *Auditory Imagery.* Hillsdale, NJ: Erlbaum.

Repp, B. H. (1999). Control of expressive and metronomic timing in pianists. *J. Mot. Behav.* 31, 145–164. doi: 10.1080/00222899909600985

Reybrouck, M. (2001). "Musical imagery between sensory processing and ideomotor simulation," in *Musical Imagery*, eds R. I. Godøy and H. Jørgensen (New York, NY: Taylor and Francis), 117–135.

Rorden, C., Karnath, H. O., and Bonilha, L. (2007). Improving lesion-symptom mapping. *J. Cogn. Neurosci.* 19, 1081–1088. doi:10.1162/jocn.2007.19.7.1081

Sacks, O. (2007). *Musicophilia: Tales of Music and the Brain.* New York, NY: First Vintage Books.

Sara, S. J., and Bouret, S. (2012). Orienting and reorienting: the locus coeruleus mediates cognition through arousal. *Neuron* 76, 130–141. doi: 10.1016/j.neuron.2012.09.011

Satoh, M., Takeda, K., Nagata, N., Hatazawa, J., and Kuzuhara, S. (2001). Activated brain regions in musicians during an ensemble: a PET study. *Cogn. Brain Res.* 12, 101–108. doi: 10.1016/s0926-6410(01)00044-1

Schiavio, A., Menin, D., and Matyja, J. (2014). Music in the flesh: embodied simulation in musical understanding. *Psychomusicol. Music Mind Brain* 24, 340–343. doi: 10.1037/pmu0000052

Schön, D., Anton, J. L., Roth, M., and Besson, M. (2002). An fMRI study of music sight-reading. *Neuroreport* 13, 2285–2289. doi: 10.1097/00001756-200212030-00023

Schürmann, M., Raija, T., Fujikia, N., and Hari, R. (2002). Mind's ear in a musician: where and when in the brain. *NeuroImage* 16, 434–440. doi: 10.1006/nimg.2002.1098

Shenhav, A., Musslick, S., Falk, F., Kool, W., Griffiths, T. L., Cohen, J. D., et al. (2017). Toward a rational and mechanistic account of mental effort. *Ann. Rev. Neurosci.* 40, 99–124. doi: 10.1146/annurev-neuro-072116-031526

Simoens, V. L., and Tervaniemi, M. (2013). Auditory short-term memory activation during score reading. *PLoS One* 8:e53691. doi: 10.1371/journal.pone.0053691

Simpson, H. M., and Climan, M. H. (1969). Pupillary and electromyographic changes during an imagery task. *Psychophysiology* 8, 483–490. doi: 10.1111/j.1469-8986.1971.tb00480.x

Simpson, H. M. (1969). Effects of a task-relevant response on pupil size. *Psychophysiology* 6, 115–121. doi: 10.1111/j.1469-8986.1969.tb02890.x

Smith, J. D., Wilson, M., and Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia* 33, 1433–1454. doi: 10.1016/0028-3932(95)00074-d

Swann, C., Moran, A., and Piggott, D. (2015). Defining elite athletes: issues in the study of expert performance in sport psychology. *Psychol. Sport Exerc.* 16, 3–14. doi: 10.1016/j.psychsport.2014.07.004

Tona, K.-D., Keuken, M. C., de Rover, M., Lakke, E., Forstmann, B. U., Nieuwenhuis, S., et al. (2017). *In vivo* visualization of the locus coeruleus in humans: quantifying the test-retest reliability. *Brain Struct. Funct.* 222, 4203–4217. doi: 10.1007/s00429-017-1464-5

Tsay, C.-J. (2013). Sight over sound in the judgment of music performance. *Proc. Natl. Acad. Sci. U S A* 110, 14580–14585. doi: 10.1073/pnas.1221454110

Turker, H. B., Riley, E., Luh, W. M., Colcombe, S. J., and Swallow, K. M. (2019). Estimates of locus coeruleus function with functional magnetic resonance imaging are influenced by localization approaches and the use of multi-echo data. *BioRxiv* [Preprint]. doi: 10.1101/731620

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience.* Cambridge, MA: MIT Press.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., and Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition* 101, 80–113. doi: 10.1016/j.cognition.2005.09.003

Vlek, R. J., Schaefer, R. S., Gielen, C. C. A. M., Farquhar, J. D. R., and Desain, P. (2011). Shared mechanism in perception and imagery of auditory accents. *Clin. Neurophysiol.* 122, 1526–1532. doi: 10.1016/j.clinph.2011.01.042

Vuvan, D. T., and Schmuckler, M. A. (2011). Tonal hierarchy representations in auditory imagery. *Mem. Cogn.* 39, 477–490. doi: 10.3758/s13421-010-0032-5

Weiss, M. W., Trehub, S. E., Schellenberg, G., and Habashi, P. (2016). Pupils dilate for vocal or familiar music. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 1061–1065. doi: 10.1037/xhp0000226

Williams, T. I. (2015). The classification of involuntary musical imagery: the case for earworms. *Psychomusicology* 25, 5–13. doi: 10.1037/pmu0000082

Wöllner, C., and Williamon, A. (2007). An exploratory study of the role of performance feedback and musical imagery in piano playing. *Res. Stud. Music Educ.* 29, 39–54. doi: 10.1177/1321103x07087567

Wolpert, D. M., and Flanagan, J. R. (2001). Motor prediction. *Curr. Biol.* 11, 729–732.

Zatorre, R. J. (1999). Brain imaging studies of musical perception and musical imagery. *J. New Music Res.* 28, 229–236. doi: 10.1076/jnmr.28.3.229.3112

Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* 8, 547–558. doi: 10.1038/nrn2152

Zatorre, R. J., and Halpern, A. R. (2005). Mental concerts: musical imagery and the auditory cortex. *Neuron* 47, 9–12. doi: 10.1016/j.neuron.2005.06.013

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., and Evans, A. C. (1996). Hearing in the mind's ear: a PET investigation of musical imagery and perception. *J. Cogn. Neurosci.* 8, 29–46. doi: 10.1162/jocn.1996.8.1.29

Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., and Kramer, S. E. (2014). The eye as a window to the listening brain: neural correlates of pupil size as a measure of cognitive listening load. *NeuroImage* 101, 76–86. doi: 10.1016/j.neuroimage.2014.06.069

Zhang, Y., Chen, G., Wen, H., Lu, K. H., and Liu, Z. (2017). Musical imagery involves Wernicke's area in bilateral and anti-correlated network interactions in musicians. *Sci. Rep.* 7:17068. doi: 10.1038/s41598-017-17178-4

Zimmerman, E., and Lahav, A. (2012). The multisensory brain and its ability to learn music. *Ann. N Y Acad. Sci.* 1252, 179–184. doi: 10.1111/j.1749-6632.2012.06455.x

**Conflict of Interest**: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 October 2020
doi: 10.3389/fpsyg.2020.557225

# Effect of Background Music on Attentional Control in Older and Young Adults

*Amélie Cloutier[1]*, Natalia B. Fernandez[2,3], Catherine Houde-Archambault[1] and Nathalie Gosselin[1]**

[1]International Laboratory for Brain, Music and Sound Research (BRAMS), Center for Research on Brain, Language and Music (CRBLM) and Laboratory for Music, Emotions and Cognition Research (MUSEC), Department of Psychology, University of Montreal, Montreal, QC, Canada, [2]Laboratory of Behavioral Neurology and Imaging of Cognition (LabNIC) and Swiss Center for Affective Sciences (CISA), Department of Neuroscience, University of Geneva, Geneva, Switzerland, [3]Laboratory of Cognitive and Affective Neuroscience (CANEURO), Department of Psychology, University of Zurich, Zurich, Switzerland

OPEN ACCESS

**Edited by:**
Cunmei Jiang,
Shanghai Normal University, China

**Reviewed by:**
Lutz Jäncke,
University of Zurich, Switzerland
Clara Eline James,
University of Applied Sciences and Arts of Western Switzerland, Switzerland

***Correspondence:**
Nathalie Gosselin
nathalie.gosselin@umontreal.ca
Amélie Cloutier
amelie.cloutier.1@umontreal.ca

**Specialty section:**
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology

**Received:** 29 April 2020
**Accepted:** 15 September 2020
**Published:** 20 October 2020

**Citation:**
Cloutier A, Fernandez NB, Houde-Archambault C and Gosselin N (2020) Effect of Background Music on Attentional Control in Older and Young Adults. Front. Psychol. 11:557225. 10.3389/fpsyg.2020.557225

Healthy aging may be accompanied by cognitive decline that includes diminished attentional control, an executive function that allows us to focus our attention while inhibiting distractors. Previous studies have demonstrated that background music can enhance some executive functions in both young and older adults. According to the *Arousal-Mood Theory*, the beneficial influence of background music on cognitive performance would be related to its ability to increase the arousal level of the listeners and to improve their mood. Consequently, stimulating and pleasant music might enhance attentional control. Therefore, the aims of this study were (1) to determine if the influence of background music, and more specifically its arousal level, might improve attentional control in older adults and (2) whether this effect is similar across older and young adults. Older and young adults performed a visuo-spatial flanker task during three auditory conditions: stimulating music, relaxing music, and silence. Participants had to indicate as fast and as accurately as possible the direction of a central arrow, which was flanked by congruent or incongruent arrows. As expected, reaction times were slower for the incongruent compared to congruent trials. Interestingly, this difference was significantly greater under the relaxing music condition compared to other auditory conditions. This effect was the same across both age groups. In conclusion, relaxing music seems to interfere with visuo-spatial attentional control compared to stimulating music and silence, regardless of age.

**Keywords: healthy aging, executive functions, attentional control, flanker task, background music, musical emotions, arousal, neuropsychology**

## INTRODUCTION

Music listening induces strong and consistent emotions in the listener (Blood and Zatorre, 2001; Hunter and Schellenberg, 2010; Chanda and Levitin, 2013). As recommended by Eerola and Vuoskoski (2011), these musical emotions are often studied into valence (i.e., positive and negative emotions) and arousal (i.e., stimulating and relaxing) dimensions (Vieillard et al., 2008), which are taken from Russell's model (Russell, 1980). Both music-resulting emotions and dimensions have been convincingly demonstrated to be associated with different musical parameters (for a review, see Juslin and Laukka, 2004). For example, most of the time, a fast tempo is associated with a high level of arousal, whereas a slow tempo is associated with a low level of arousal. Similarly, music composed in a major mode is typically associated with a high level of valence and with positive emotions like joy and peace. Also, it has been demonstrated that loudness is related to the perceived level of arousal and valence in music (Schubert, 2004; Dean et al., 2011; Olsen et al., 2015). Further studies also showed a positive correlation between the tempo of background music and reading speed (Kallinen, 2002), perceptual motor abilities (Nittono et al., 2000), and visual attention tasks (Bolger et al., 2013; Trost et al., 2014), giving an evident support to the impact of musical parameters on cognition. In Bolger et al. (2013) study, the targets of the visual attention task were presented across four selected metrical positions of the auditory stimulus in order to observe the entrainment effect of the rhythm of the music. In Trost et al. (2014) study, targets appeared time-locked to either strong or weak beats of the background music. The tempo of a musical stimulus presented before the cognitive task was also correlated to spatial ability (Husain et al., 2002).

Valence and arousal dimensions also seem to interact in inducing musical emotions (Salimpoor et al., 2009; van den Bosch et al., 2013). Previous work has demonstrated that music inducing a higher level of arousal generates more pleasure in the listener (Salimpoor et al., 2009). More precisely, if the listener likes the musical excerpt, subjective felt arousal ratings and the listener's arousal state (as measured by electrodermal activity) increase with pleasure ratings (Salimpoor et al., 2009). These authors also specify that this link is unidirectional, since an increase in arousal does not always lead to pleasure (Salimpoor et al., 2009). Similarly, the familiarity of the music (i.e., how well someone knows the musical piece) has been positively correlated with the level of arousal and the pleasantness rated by the listener (van den Bosch et al., 2013). Moreover, it has been demonstrated that perception of positive emotional valence in music increases with age, with older adults tending to find music more pleasant on average than young adults (Cohrdes et al., 2020). It is also important to note that, in general, older adults have more positive emotional well-being than young adults (Carstensen et al., 2011).

Background music has been shown to have both beneficial and detrimental effects on a variety of cognitive functions in healthy young adults (see Kämpfe et al., 2010 for a meta-analysis). According to the most well-known theory regarding the link between music and cognitive performance, the *Arousal-Mood Theory*, a musical stimulus presented before the task and characterized by a high level of arousal (i.e., stimulating music) and a high level of valence (i.e., pleasant music) would increase the arousal level of the listener and improve his or her mood, thereby enhancing subsequent cognitive performance (Thompson et al., 2001). Other studies have demonstrated this effect when music was presented simultaneously with a variety of executive tasks, such as cognitive flexibility, working memory and attentional control (Thompson et al., 2005; Mammarella et al., 2007; Jefferies et al., 2008; Jiang et al., 2011; Bottiroli et al., 2014; Shih et al., 2016; Fernandez et al., 2020). However, not all of the available research findings fit well within this theoretical relationship between music and cognitive performance. For example, some research suggests that highly pleasant music requires more attentional resources and thus may impair cognitive performance in the context of attentional tasks (Nemati et al., 2019).

In particular, the impact of background music on attentional control, an executive function that allows one to focus attention on a specific stimulus, while inhibiting distractors from the environment (Theeuwes, 2010; Diamond, 2013), is somewhat ambiguous, with previous research generating heterogeneous results and not always controlling for levels of arousal and/or valence. One study demonstrated that, compared to silence, personally chosen background music enhanced young adults' attentional control performance (Darrow et al., 2006), while another showed that sad music enhanced selective attention performance compared to calm, happy, and scary music (Jefferies et al., 2008). Furthermore, a study that used music prior to a flanker task, which measures attentional control (Eriksen and Eriksen, 1974), demonstrated that positive affect (i.e., higher valence level) induces a larger flanker effect [difference in reaction time (RT) between incongruent and congruent trials], suggesting that background music characterized by positive valence impairs attentional control performance (Rowe et al., 2007). A recent study reported improved perceptual judgment in a flanker task in young adults when they were listening to joyful and arousing background music, compared to sad and tender music as well as to silence (Fernandez et al., 2020); however, they did not find any effect of background music on attentional control performance *per se*. Also, Burkhard et al. (2018) showed that, compared to a silent condition, relaxing and exciting music did not have any effect on young adults' inhibitory performance on the go/no-go task (Nosek and Banaji, 2001), nor on the event-related components underlying inhibitory processing. In sum, studies in young adults still show heterogeneous results concerning the effect of background music on attentional control, and the role of arousal in this effect is still not completely understood.

The role of music in attentional control in elderly populations has only been investigated very recently (Fernandez et al., 2020). In this work, they demonstrated that, compared to silence as well as sad and tender music, joyful and highly arousing background music enhanced perceptual judgments in a flanker task in both older and young adults (Fernandez et al., 2020). No background music effect was found on older adults' attentional control performance (Fernandez et al., 2020). However, this study used a modified version of the flanker task, taken from the Attention Network Test, which measures several components of attention and includes cues before the trials. Thus, a more challenging task measuring attentional control specifically might produce different results regarding the effect of background music in older adults. In sum, it is apparent that the existing findings about the effect of background music on attentional control in both older and young adults are not always in accordance with the *Arousal-Mood Theory* (Thompson et al., 2001).

It is important to study the effect of background music in older adults since attentional control can be impaired in normal cognitive aging (Buckner, 2004). Indeed, older adults have been reported to be generally more sensitive to distractors (Hasher and Zacks, 1988; Hasher et al., 1991; Gazzaley et al., 2005; Darowski et al., 2008), showing slower RT in the flanker task compared to young adults (Zeef et al., 1996; Salthouse, 2010). Also, a speed-accuracy trade-off can be observed in older adults' performance at the flanker task (Wild-Wall et al., 2008; Hsieh and Fang, 2012; Hsieh and Lin, 2014). In other words, compared to younger adults, older adults make fewer errors but present slower RT. It is important to note that this slowing in older adults' performance could also be caused by vision and/or hearing loss, since a relationship has been demonstrated between vision and hearing and cognitive performance in this population (Li and Lindenberger, 2002; Wahl and Heyl, 2003; Lin et al., 2013). A recent study investigating the cerebral substrates associated with this age-related slowdown in the flanker task, demonstrated additional brain activations in posterior parieto-occipital areas which were linked to greater efforts to process the central target in incongruent trials (Fernandez et al., 2019). In sum, the literature indicates that older adults struggle more in inhibiting distractors and probably recruit different brain areas to compensate for their difficulties.

In addition to attentional declines in the elderly, it is possible that the presentation of a visual or auditory stimulus during the completion of a cognitive task might be more distracting for older adults than young adults, even if they are told to ignore the distraction (Guerreiro et al., 2010). Studies conducted by Alain and Woods (1999) and by Andrés et al. (2006) demonstrated that adding irrelevant sounds to a visual discrimination task impairs older adults more than young adults in their RT, as well as in the amplitude of the event-related potential linked to the processing of distraction (N1 and MMN). In the same manner, adding background music to a visual task could potentially be more distracting for older adults than for young adults.

However, it is also possible that background music added to a visual task could impair the performance of young adults. Indeed, for attentional tasks that are time-critical, as well as for spatial attentional tasks, shared attentional resources are involved when processing stimuli from different modalities (i.e., auditory and visual), and this is the case for adults of all ages (for a review, see Wahn and König, 2017). This can lead to impairment in the processing of one or both modalities. For example, when auditory and visual stimuli are presented simultaneously in an attentional task, both auditory and visual processing are slowed down (Dunifon et al., 2016). However, the effect of background music on attentional control in young and older adults is still not fully understood and needs further investigation.

In sum, normal aging is accompanied by cognitive decline that affects attentional control. Thus, it is important to find easy and pleasant ways for older adults to maximize their attentional control in everyday situations, for example with background music. However, the beneficial effect of background music on different executive functions is not fully understood, possibly due to the fact that the arousal level of music is not always controlled in previous studies. More particularly, the comparison between young and older adults in the effect of background music on attentional control specifically needs more investigation.

This study aimed to determine if the influence of background music, and more specifically its arousal level, might improve visuo-spatial attentional control in older adults and whether this effect is similar across older and young adults. To do this, we compared the effect of stimulating and relaxing music on performance on the flanker task, with a silence condition representing the base level performance.

Regarding the effect of background music, we expected faster answers and fewer errors for older adults under the stimulating music condition compared to both the relaxing music and silence conditions. As for young adults, knowing that results in the literature about the effect of background music on attentional control are still heterogeneous, there were no hypotheses concerning the effect of background music on their performance on the flanker task.

## MATERIALS AND METHODS

### Participants

Nineteen older adults and 21 younger adults participated in this experiment. They all provided informed consent and received financial compensation for their participation. All participants were francophone Quebecers and reported to have normal audition, as well as normal or corrected-to-normal visual acuity. They also reported information about their music listening habits. None reported neurological, neurodevelopmental, or diagnosed psychiatric disorders. Depression and anxiety questionnaires were used to ensure that participants did not have clinically significant levels of anxio-depressive symptoms.

Young adults completed both the Beck Anxiety Inventory (BAI; Beck et al., 1988) and Beck Depression Inventory II (BDI-II; Beck et al., 1996), for which scores over critical thresholds (26/63 and 29/63, respectively) were considered exclusion criteria. All young adults presented scores of 12 or lower ($M$ = 4.36; $SD$ = 3.83) for the BAI and scores of 20 or lower ($M$ = 7.24; $SD$ = 5.32) for the BDI-II.

Older adults completed the Geriatric Anxiety Inventory (GAI; Pachana et al., 2007), as well as the short form of the Geriatric Depression Scale (GDS-SF; Burke et al., 1991). Scores over critical thresholds (9/20 and 5/15, respectively) were considered exclusion criteria. Older participants had scores of 8 or lower ($M$ = 2.05, $SD$ = 2.59) on the GAI, as well as of 3 or lower for the GDS-SF ($M$ = 0.79; $SD$ = 1.08). In addition, general cognitive state was evaluated using the Mini Mental State Examination (MMSE; Folstein et al., 1975; Commenges et al., 1992) to ensure no deficits (e.g., mild cognitive impairment). Based on previous studies (Folstein et al., 1975; Hudon et al., 2009), older participants whose scores were over the threshold of 27/30 were retained in the study. All older participants had scores of 28 or more ($M$ = 29.26; $SD$ = 0.73).

In addition, basic executive functioning was assessed using the color-word interference test, from the Delis-Kaplan Executive Function System battery (D-KEFS; Delis et al., 2001). The color-word interference condition consists of naming the color of the ink with which each word is printed, thus permitting an evaluation of inhibition processes. No inhibition deficits were observed in either group, with standard scores in the average range when compared to age-based norms ($M = 10.29$, $SD = 1.82$ for young adults; $M = 10.42$, $SD = 1.98$ for older adults).

The two groups were significantly different in age (see **Table 1**). They were matched in terms of sex, years of schooling, and years of musical training, with a similar proportion of men and women and equivalent years of schooling and years of musical training (see **Table 1**). However, our participants were mainly women (i.e., 18 older women for one man and 19 young women for two men). For musical expertise, none of the participants were professional musicians.

The music listening habits of our sample did not appear to be different between young and older adults, neither as principal activity (reported by 13/21 young adults with a mean of 2.69 h/week and 13/19 older adults with a mean of 2.76 h/week) nor as background music (reported by all young adults with a mean time of 8.1 h/week and 13/19 older adults with a mean time of 8.79 h/week).

## Flanker Task

All participants performed an arrow version of Eriksen's flanker task (Eriksen and Eriksen, 1974) from a viewing distance of 100 cm from the screen. We followed previous recommendations for size and spacing parameters (Zeef et al., 1996; Maylor and Lavie, 1998; Hsieh and Lin, 2014). Participants were asked to focus their attention on the central arrow (0.4° of visual angle vertically and 0.6° horizontally) of a series of five and to indicate the direction in which it pointed, as quickly and accurately as possible. The target was flanked by two arrows on the left and two arrows on the right and could either point the same direction (congruent condition: right > > > > > or left < < < < <) or the opposite direction (incongruent condition: right < < > < < or left > > < > >) as the central arrow (see **Figure 1**).

Each trial contained five steps (see **Figure 1**). First, a fixation cross was displayed in the center of the screen for 500 ms, followed by an array of arrows in the middle of the screen for a duration of 250 ms. Next, a black screen was presented, and participants had a maximum of 2,000 ms to provide their answer regarding the direction of the central arrow. After the answer was given or the time limit was over, the screen remained black for 500 ms. Finally, the symbol "--" was presented in the middle of the screen during the inter-trial interval (duration between 850 and 950 ms). Depending on the participant's RT and the duration of the inter-trial interval, the total duration of one trial varied between 2,500 and 4,200 ms.

The experiment comprised 21 blocks containing 32 trials each (with an equal number of congruent and incongruent trials) for a total of 672 trials. Both blocks and trials were presented in a randomized order for each participant. Of the 21 blocks, seven were allocated to each of the three auditory conditions (stimulating music, relaxing music, and silence). For the two musical conditions, each block was associated with a different musical excerpt. Participants could take breaks between each block to rest. The total duration of one block varied between 80 and 90 s, depending on the RT of the participant. Without the breaks between each block, the total duration of the entire task was approximately 30 min. To familiarize participants with the task, it was preceded by a practice block that included feedback to inform the participants about their performance. The practice block was presented with background music characterized by an intermediate tempo (i.e., 110 beats per minute, BPM). The flanker task and the music were presented using MATLAB (MATLAB Release 2018a, The MathWorks, Inc., Natick, Massachussetts, United States) with the "Psychophysics Toolbox Version 3" extension (Brainard, 1997; Kleiner et al., 2007).

**TABLE 1** | Comparison between older and younger adults on demographic variables.

| | Age groups | | | | | |
|---|---|---|---|---|---|---|
| | Older adults | Young adults | *df* | $t/\chi^2$ | *p* | Effect size (*r*) |
| N (M, F) | 19 (1, 18) | 21 (2, 19) | 1 | 0.26 | =0.61 | =0.08 |
| Age (years) | 67.26 (3.16) | 23.95 (3.51) | 38 | −40.82 | <0.001 | =0.99 |
| Years of education | 16.16 (2.69) | 16.48 (1.86) | 38 | 0.44 | =0.66 | =0.07 |
| Years of musical training | 1.37 (2.17) | 2.81 (4.69) | 38 | 1.23 | =0.23 | =0.2 |

*Except for sex, this table presents means (and standard deviations). M = male, F = female. Group composition was compared for sex, using a chi square test, and for age, years of education, and years of musical training using independant t-tests.*

## Musical Stimuli

All participants performed the flanker task under three auditory conditions: stimulating music, relaxing music, and silence. The music was pleasant sounding instrumental works composed in a major mode, chosen from the classical repertoire. Inter-rater agreement between three researchers was used to select the seven most stimulating (e.g., William Tell Overture: Final, composed by Giochino Rossini), as well as the seven most relaxing (e.g., Suite Bergamasque, Clair de Lune composed by Claude Debussy), musical excerpts from a larger pool of musical material in use in our laboratory. Excerpts of 100 s were chosen from the original pieces, so that the arousal and valence levels, as well as the tempi, were stable throughout each excerpt. The stimulating musical excerpts had a mean tempo of 153.14 BPM ($SD = 23.35$), while the relaxing musical excerpts had a mean tempo of 59.29 BPM ($SD = 11.34$). All excerpts were normalized at peak value (90% of maximum amplitude) and logarithmic fade-ins and fade-outs of 500 ms were added at the beginning and end of each excerpt, using Adobe Audition 3.0 software (Adobe Systems, Inc. San Jose, CA, United States).

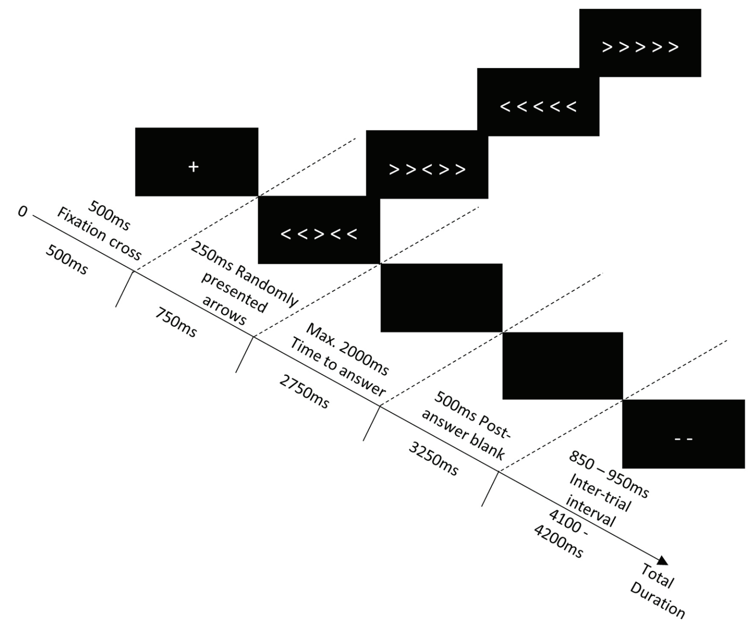



**FIGURE 1 |** The course of a flanker task trial. Symbols are not to scale; they have been enlarged to be visible in this diagram.

The music was presented *via* Beyer Dynamic Headphones (Model DT 770 Professional, 250 OHM).

## Musical Evaluation

After completing the flanker task, participants were asked to listen carefully to each musical excerpt without time restriction and to evaluate how much the piece was considered to be (a) arousing, i.e., relaxing or stimulating, (b) unpleasant or pleasant, and (c) unfamiliar or familiar, using a continuous visual analogue scale from 0 (extreme left) to 100 (extreme right). Thus, a low score on the arousal dimension would mean that the musical excerpt was judged as relaxing.

## Data Analysis

Group composition was compared for sex, using a chi square test, and for age, years of education, and years of musical training using independent *t*-tests.

Performance on the flanker task was analyzed using RT and error rate (ER). Average RT values for successful trials were calculated in milliseconds for each flanker congruency type of trial (i.e., congruent and incongruent), each auditory condition (stimulating music, relaxing music, and silence), and each participant separately. The averages and standard deviations of ER as percentages (excluding missed trials) were also calculated for each participant, as well as for each flanker congruency and auditory condition. RT and ER scores were entered into separate mixed-design analyses of variance (ANOVAs) with Age Group (older and young adults) as a between-subject factor, Auditory Condition (stimulating music, relaxing music, and silence), and Flanker Congruency trial type (congruent and incongruent) as within-subject factors. When interactions between repeated measure factors were significant, a standard contrasts analysis was used to determine if the difference between congruent and incongruent trials (i.e., flanker effect) was the same between auditory conditions.

To confirm that the musical conditions differed in perceived arousal level and to explore whether there was a difference between older and younger adults' judgments, a mixed-design ANOVA with the between subject factor Age Group (older and younger adults) and the within subject factor Music Condition (stimulating music and relaxing music) was conducted. Two other exploratory mixed design ANOVAs were conducted with the judgments of valence and familiarity.

All statistical analyses were performed using IBM SPSS Statistics 24 (IBM Corp., 2016). Behavioral mean results (music evaluation and flanker performance) as well as statistical results are presented in **Tables 1–4**.

# RESULTS

## Musical Stimuli Evaluation

As expected, stimulating music was judged to be significantly more arousing than relaxing music by both older and young adult groups, the size of this effect being large (see **Figure 2** and **Table 2**). There was no difference between older and young adults in their evaluation of the arousal level of musical excerpts and no significant interaction between Music Condition and Age Group.

Relaxing music was considered significantly more pleasant than stimulating music by both older and young adults, with this effect being large (see **Figure 2** and **Table 2**). Older adults generally judged the musical excerpts to be more pleasant

**TABLE 2 |** Results of the analyses of variance (ANOVA) for the evaluation of arousal, valence, and familiarity.

| Predictor | *df* | *F* | *p* | $\eta^2$ |
|---|---|---|---|---|
| **Arousal** | | | | |
| Music Condition | 1, 38 | 1453.3 | <0.001 | 0.98 |
| Age Group | 1, 38 | 0.016 | =0.9 | 0.00 |
| Music Condition × Age Group | 1, 38 | 2 | =0.165 | 0.05 |
| **Valence** | | | | |
| Music Condition | 1, 38 | 32.28 | <0.001 | 0.46 |
| Age Group | 1, 38 | 7.53 | =0.009 | 0.17 |
| Music Condition × Age Group | 1, 38 | 0.009 | =0.926 | 0.00 |
| **Familiarity** | | | | |
| Music Condition | 1, 38 | 1.32 | =0.258 | 0.033 |
| Age Group | 1, 38 | 21.48 | <0.001 | 0.36 |
| Music Condition × Age Group | 1, 38 | 0.08 | 0.778 | 0.002 |

**TABLE 3 |** Results of the ANOVA for the flanker task RT.

| Predictor | *df* | *F* | *p* | $\eta^2$ |
|---|---|---|---|---|
| **Omnibus analysis** | | | | |
| Age Group | 1, 38 | 55.02 | <0.001 | 0.59 |
| Flanker Congruency | 1, 38 | 418.75 | <0.001 | 0.92 |
| Auditory Condition × Flanker Congruency | 2, 67 | 3.995 | 0.027 | 0.095 |
| **Contrasts analysis** | | | | |
| Relaxing vs. Stimulating | 1, 38 | 10.61 | =0.002 | 0.22 |
| Relaxing vs. Silence | 1, 38 | 4.29 | =0.045 | 0.1 |
| Stimulating vs. Silence | 1, 38 | 0.116 | 0.735 | 0.003 |

than young adults, with this effect also being large. There was no significant interaction between Music Condition and the Age Group.

Older adults were significantly more familiar ($M$ = 84.41, $SD$ = 16.73) with the musical excerpts than young adults ($M$ = 62.32, $SD$ = 17.41), with the size of this effect being large (see **Table 2**). There was no difference between stimulating and relaxing music in their level of familiarity. Finally, there was no significant interaction between Music Condition and the Age Group.

## Flanker Task

Reaction time performance on the flanker task revealed a significant and general slowing in older adults compared to young adults (large effect, see **Table 3** and **Figure 3**). For both older and young adults, RT was significantly slower in the incongruent trials than in the congruent ones, i.e., a flanker effect was clearly observed. Moreover, an interaction between Auditory Condition and Flanker Congruency showed that the difference in RT between incongruent and congruent trials varied between the three conditions. More specifically, the influence of background music revealed a greater flanker effect for relaxing music than for stimulating music or silence, these

**TABLE 4 |** Results of the ANOVA for the flanker task ER.

| Predictor | *df* | *F* | *p* | $\eta^2$ |
|---|---|---|---|---|
| **Omnibus analysis** | | | | |
| Age Group | 1, 38 | 9.86 | =0.003 | 0.21 |
| Flanker Congruency | 1, 38 | 66.28 | <0.001 | 0.64 |
| Age Group × Flanker Congruency | 1, 38 | 14.05 | <0.001 | 0.27 |
| Auditory Condition | 2, 38 | 2.46 | =0.097 | 0.056 |
| **Post-hoc analysis (ANOVAs)** | | | | |
| Difference between older and young adults for congruent trials | 1, 38 | 1.13 | =0.294 | 0.03 |
| Difference between older and young adults for incongruent trials | 1, 38 | 11.45 | =0.002 | 0.23 |

effects being, respectively, large and average (see **Figure 4** and **Table 3**). These two latter conditions did not differ in terms of flanker effect (see **Figure 4** and **Table 3**).

Older adults made fewer errors overall compared to young adults, and this was a large effect (see **Figure 3** and **Table 4**). Also, for both older and young adults, ER was significantly higher for the incongruent than the congruent trials, this effect also being large. Moreover, an interaction between Age Group and Flanker Congruency showed that there was no age-related difference for the congruent trials, while older adults made significantly fewer errors in the incongruent trials compared to young adults (see **Figure 5** and **Table 4**). There were no significant differences between the three experimental conditions in ER.

# DISCUSSION

This study aimed to explore the effect of the arousal level of background music on visuo-spatial attentional control in young and older adults. To do this, both groups performed an arrow version of the flanker task under three auditory conditions: stimulating music, relaxing music, and silence.

## Error Rates and Reaction Times in the Flanker Task

The effects on ER seemed quite limited, probably due to the ceiling effect observed in both older and young adults (success rate > 95% for all participants). All participants presented expected slower RT and increased errors in the incongruent trials compared to the congruent ones, suggesting greater difficulty in inhibiting the distracting and incongruent arrows (Eriksen and Eriksen, 1974; Eriksen, 1995). Importantly, older adults had overall slower RT and lower ER than young adults. They seemed to favor accuracy over speed in their performance on the flanker task, while young adults favored speed over accuracy, which is consistent with previous studies using the same task (Wild-Wall et al., 2008; Hsieh and Fang, 2012;

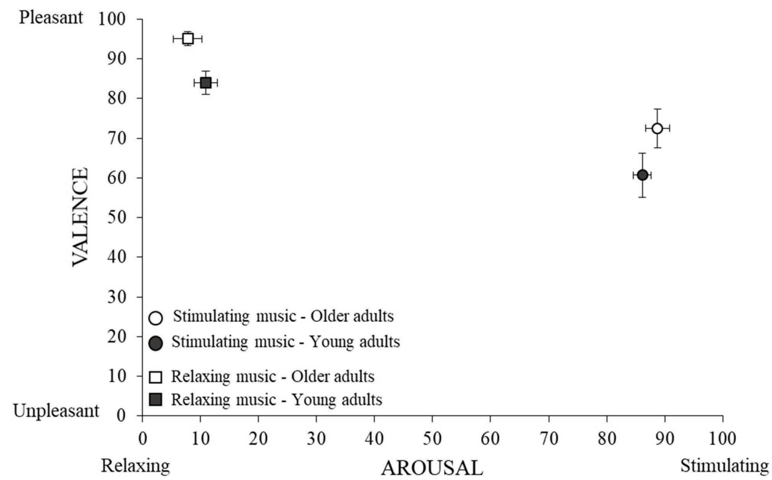



**FIGURE 2 |** Judgments of arousal and valence. Mean rating (and standard errors) are presented as a function of music conditions and age groups on both valence and arousal dimensions.

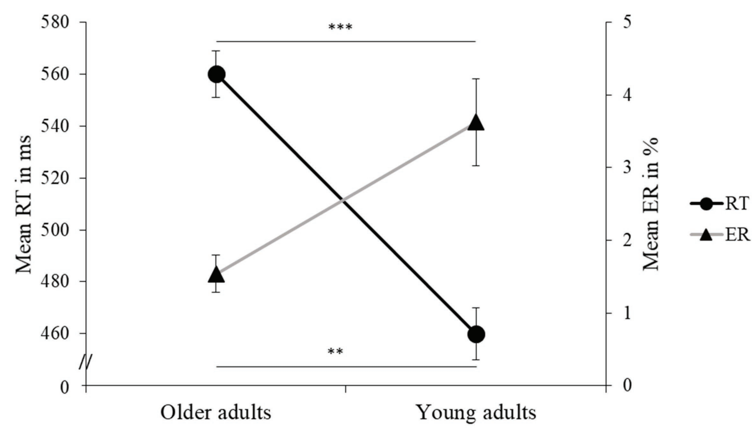



**FIGURE 3 |** Mean reaction time (RT) in ms and error rate (ER) in % (and standard errors) are presented for older and young adults. Values of *p* (asterisk): $^{**}p < 0.01$ and $^{***}p < 0.001$.

Hsieh and Lin, 2014). Although normal aging has been linked to impairments in attentional control (Hasher and Zacks, 1988; Hasher et al., 1991; Buckner, 2004; Gazzaley et al., 2005; Darowski et al., 2008), the slower results observed in our older adults during incongruent trials might be explained by compensatory mechanisms adopted to adequately complete the task (Wild-Wall et al., 2008; Hsieh and Fang, 2012; Hsieh and Lin, 2014), or decreased eyesight and hearing, which has been linked to cognitive performance deficits (Li and Lindenberger, 2002; Wahl and Heyl, 2003; Lin et al., 2013).

In addition, results obtained in the flanker task might have been influenced by the female dominance of our sample. A previous study demonstrated that visual selective attention performance of women is more affected by invalid cues, while men benefit from those invalid cues (Merritt et al., 2007). Also, women are more influenced by irrelevant spatial cues compared to men (Bayliss et al., 2005). Finally, a study demonstrated that incongruent flankers impair women's performance more than men's, showing a gender difference in visuo-spatial selective attention (Stoet, 2010).

## The Effect of Background Music on Attentional Control

In our study, the influence of background music during a visuo-spatial attention task revealed impaired attentional control performance during relaxing music exposure compared to silence and stimulating music. These results are not consistent with the *Arousal-Mood Theory* (i.e., stimuli rated as pleasant and stimulating can increase the arousal level and improve the mood in listeners) and other previous studies demonstrating that stimulating and pleasant music enhances cognitive

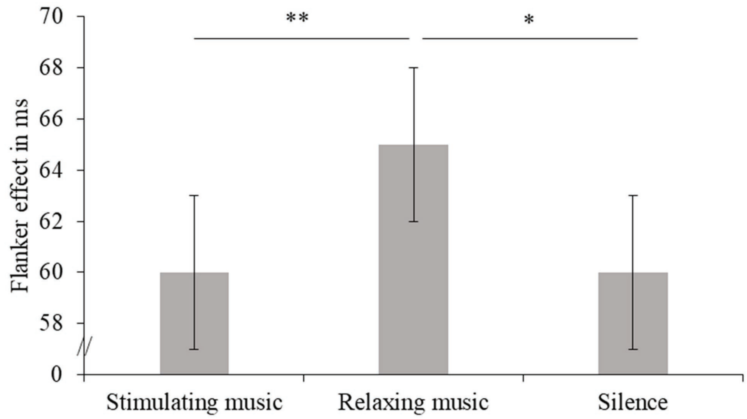



**FIGURE 4 |** Flanker effects in ms (and standard errors) are presented for all participants (combined across age groups) as a function of Auditory Condition. Values of *p* (asterisk): **p* < 0.05 and ***p* < 0.01.

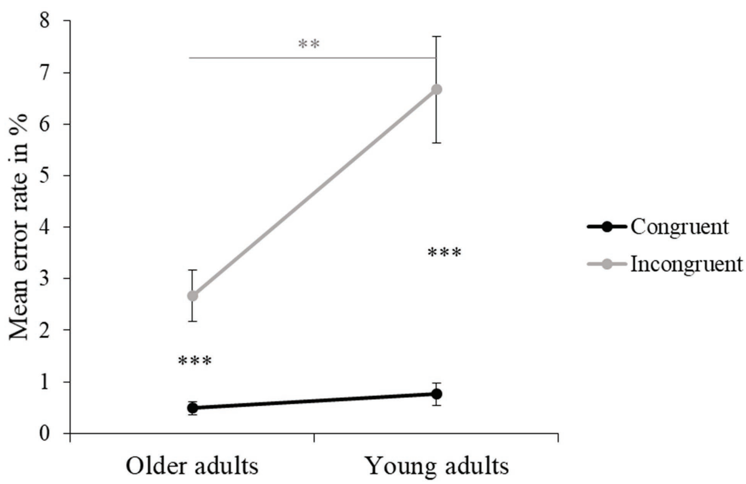



**FIGURE 5 |** Interaction between Age Group and Flanker Congruency in ER. Mean ER in % (and standard errors) were obtained for congruent and incongruent trials and separately for older and young adults. Values of *p* (asterisk): ***p* <0.01 and ****p* < 0.001.

performance (Thompson et al., 2001; Mammarella et al., 2007; Jiang et al., 2011; Bottiroli et al., 2014; Shih et al., 2016; Fernandez et al., 2020). They are also inconsistent with recent research that suggests no influence of relaxing or stimulating background music on inhibitory processing, albeit using a different type of inhibition task (go/no-go; Burkhard et al., 2018).

It is difficult to reconcile these results demonstrating a difference between relaxing and stimulating music with the existing literature. However, based on previous studies demonstrating positive correlations between the tempo of music and cognitive performance across a number of domains, including reading speed (Kallinen, 2002), perceptual motor speed (Nittono et al., 2000), spatial ability (Husain et al., 2002), and visual attention tasks (Bolger et al., 2013; Trost et al., 2014), we hypothesize that this influence of relaxing music on participants' reaction times might be associated with the tempi of our musical excerpts. Greater flanker effects are observed for the relaxing music condition, which is characterized by slower tempi, compared to the stimulating music condition, associated with faster tempi.

The contradiction between our results and the *Arousal-Mood Theory* may also be explained by the fact that, in this study, stimulating music was judged to be less pleasant than relaxing music, while usually it is judged to be more pleasant (Salimpoor et al., 2009). In line with this, a previous study that used music to induced different moods prior the flanker task, showed that pleasant music induced a general slowdown in RT, as well as a greater flanker effect, compared to both neutral condition and unpleasant music (Rowe et al., 2007). Although all of the musical excerpts in our study were judged to be pleasant, the fact that relaxing music was seen as significantly more pleasant than stimulating music might explain why the former induced a greater flanker effect than the latter. It has also been demonstrated that listening to highly pleasurable music involves more attentional resources, and leads to a decline in cognitive performance (Nemati et al., 2019). Thus, it is possible that our relaxing music impaired the flanker task performance compared to stimulating music and silence by inducing a higher positive valence in our participants and by the same token, involving more attentional resources, leaving less for the execution of the task (Dunifon et al., 2016; Wahn and König, 2017; Nemati et al., 2019). It is possible that the sharing of attentional resources when processing both auditory and visual stimuli at the same time is even more difficult when the auditory stimuli is relaxing music (Dunifon et al., 2016; Wahn and König, 2017; Nemati et al., 2019).

## The Effect of Background Music Across Age

Regarding age, the greater flanker effects observed for the relaxing music condition was similar between both groups. This is in line with very recent work demonstrating that classical music with different arousal and valence levels has the same impact on attention processing in both young and older adults, even if the latter experience a decline in this particular executive function (Fernandez et al., 2020). This also mirrors a study by Alain and Woods (1999), who reported no differences between older and young adults in their performance on a visual attention task, while listening to irrelevant auditory stimuli. However, in their study, the irrelevant sounds presented simultaneously with the visual attention task provoked greater event-related potential amplitude in older adults compared to young adults. These results demonstrate distinct background sounds processing across age for similar performances in a visual attention task. The behavioral similarities between older and young adults in the current study cannot rule out the possibility of age-related differences in the underlying neural networks for attentional control. Moreover, given that older and young adults differed in rating the musical excerpts for valence and familiarity, it is possible that, with similar ratings in these two dimensions, we would observe a difference between the two age groups in the effect of background music on attentional control.

## Musical Stimuli Evaluation

We analyzed participants' arousal, valence, and familiarity evaluations of the musical excerpts. Our results indicated that the arousal level of the musical excerpts was judged as expected by both age groups. Also, all musical excerpts were evaluated as pleasant, but unexpectedly, relaxing music pieces were felt to be more pleasant than stimulating music pieces. This finding is inconsistent with previous studies demonstrating that stimulating music generated higher ratings of pleasantness by listeners (Salimpoor et al., 2009; van den Bosch et al., 2013). However, those studies used musical material chosen by the participants or that resembled participants' favorite music, from various musical genres (classic, jazz, rock, etc.), whereas in our study, only classical music selected by the researchers was used. Indeed, it has been demonstrated that, when listening to familiar music, there is a strong positive correlation between the pleasure felt by the listener and his level of arousal, but when the music is not familiar, there is no longer a clear relation between pleasure and arousal (van den Bosch et al., 2013). This might explain why we obtained different results, since our participants listened to music that they did not choose and were thus not as familiar as they would have been with personally chosen music.

Relaxing and stimulating music did not differ in terms of familiarity level here, suggesting that the observed effect of background music on attentional control is likely due to the variations in arousal and valence levels only. However, we did find that music excerpts used in the current experiment were rated as more familiar and more pleasant for older adults than young adults. These results might be explained by the fact that older adults listen more to classical music, while young adults listen mostly to popular music (Savage, 2006). Thus, our musical excerpts might have matched older adults' tastes and habits better than young adults'. Another putative explanation is related to the fact that perception of positive emotional valence in music increases with age; in other words, older adults tend to find music more pleasant on average than young adults (Cohrdes et al., 2020). The authors of this study interpret this finding in relation to other research that suggests that older adults' emotional well-being is more positive than young adults' (Carstensen et al., 2011). Hence, it is possible that our older participants were more inclined to find the music pleasant than our younger participants due to their age.

Moreover, our participants were mostly women, and gender is known to have a moderate influence on the emotions induced by music (Aljanaki et al., 2016). Indeed, women tend to feel more amazed by classical music than men (Aljanaki et al., 2016). It has been showed that the brain activity linked to music-induced pleasantness and self-reported feelings of happiness were significantly greater in women than in men (Diaz et al., 2011). Another study demonstrated that women show elevated electrophysiological activity for arousing and unpleasant music, compared to men (Nater et al., 2006). Thus, the female dominance in our study might have influenced the results concerning musical stimuli evaluation.

Although the evaluation of the valence dimension showed unexpected results [i.e., (1) relaxing music evaluated as more pleasant than stimulating music and (2) overall higher valence scores in the older compared to younger adults], the evaluation of the arousal dimension of our stimulating/relaxing musical excerpts were judged as expected. This allowed us to evaluate the effect of the arousal dimension of background music on attentional control, as measured by the flanker task.

## Conclusion

In conclusion, we observed the expected performance of older and young adults in the flanker task, with slower RT and greater ER for incongruent trials compared to congruent trials. Our results, not supported by the *Arousal-Mood Theory*, suggest that relaxing pleasant background music can impair visuo-spatial attentional control performance, inducing a greater flanker effect or RT than that observed for stimulating pleasant music and silence. This effect was the same for older and young adults despite the typical decrement in attentional control associated with healthy aging.

### Limitations

This study presents some limitations. First, only classical music was used and older adults found it more pleasant and familiar than young adults, potentially inducing a differential impact of these stimuli on our participants as a function of age. Future studies should control for the impact of age on emotional judgments of musical stimuli when comparing older and young adults. Second, although we screened older adults for cognitive impairments, given that hearing and vision loss can impact the cognitive performance of older adults, ideally these perceptual functions should be measured as well and, if necessary, entered as covariates in the analysis. Third, the music conditions differed not only in arousal but also in valence. Even if those two dimensions often interact together (Jefferies et al., 2008), it would be interesting to observe their separate effects on attentional control performance through manipulation of each of these factors independently. Fourth, given that a single visuo-spatial task was used to assess attentional control, the conclusions of this study are limited to visuo-spatial attentional control.

### Future Perspectives

In order to improve the evaluation of arousal, future studies should use real-time objective measurements of arousal through the recording of electrodermal activity, while participants listen to the music and execute the task. Also, it would be interesting to control for inter-individual variability, in general, arousal level by measuring it before the beginning of the experiment. It would also be important to take into account the gender of participants in studying music-induced emotions. Moreover, to improve the ecological validity of the results, future work could also investigate the influence of longer periods of background music listening (in contrast to our 100 s excerpts) on visuo-spatial attentional control performance. An important extension to the current research would be the inclusion of other modalities of attentional control during background music listening, in order to draw more general conclusions about attentional control, and not limited to visuo-spatial attentional control as in the current paper. As mentioned previously, since some studies found an effect of background music on cortical activity in absence of a behavioral effect (Alain and Woods, 1999; Jäncke and Sandmann, 2010), it would also be interesting to investigate the impact of background music on EEG measures. Finally, if future studies support the present findings demonstrating a detrimental effect of relaxing background music on visuo-spatial attentional control performance and reproduce this effect with other executive functions, this study should be used as guideline in recommending or not the use of background music, while performing a cognitive task.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comité d'éthique de la recherche en arts et en sciences, Université de Montréal. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

AC elaborated the theoretical frame and formulated the research question, as well as the objectives and hypotheses. AC contributed to the creation of the research protocol and methodology. AC contributed to the data collection and analysis and wrote the article. CH-A contributed to the creation of the research protocol and methodology, as well as the data collection. CH-A contributed to the revision and correction of the article. NF contributed to the establishment of the flanker task parameters, as well as to the revision and correction of the article. NG contributed to the elaboration of the theoretical frame and the formulation of the research question, objectives, and hypotheses. NG supervised the creation of the research protocol and methodology, as well as the data collection and analysis and the article redaction. All authors contributed to the article and approved the submitted version.

## FUNDING

The study is funded by the Fonds de recherche du Québec société et culture (FRQSC), 2019-B1Z-256735, and the Center for Research on Brain, Language and Music (CRBLM). The CRBLM is funded by the Fonds de recherche nature et technologies du Québec (FRQNT) and by the FRQSC.

## ACKNOWLEDGMENTS

We would like to thank Falco Enzler for his help in programming the MATLAB script for the flanker task. We also want to thank Simone Dalla Bella, Ph.D., and Simona Brambati, Ph.D. who evaluated Amélie's master's thesis. And most importantly, we would like to thank our participants, without whom the research would be impossible.

## REFERENCES

Alain, C., and Woods, D. L. (1999). Age-related changes in processing auditory stimuli during visual attention: evidence for deficits in inhibitory control and sensory memory. *Psychol. Aging* 14, 507–519. doi: 10.1037/0882-7974.14.3.507

Aljanaki, A., Wiering, F., and Veltkamp, R. C. (2016). Studying emotion induced by music through a crowdsourcing game. *Inf. Process. Manag.* 52, 115–128. doi: 10.1016/j.ipm.2015.03.004

Andrés, P., Parmentier, F. B., and Escera, C. (2006). The effect of age on involuntary capture of attention by irrelevant sounds: a test of the frontal hypothesis of aging. *Neuropsychologia* 44, 2564–2568. doi: 10.1016/j.neuropsychologia.2006.05.005

Bayliss, A. P., di Pellegrino, G., and Tipper, S. P. (2005). Sex differences in eye gaze and symbolic cueing of attention. *Q. J. Exp. Psychol. A* 58, 631–650. doi: 10.1080/02724980443000124

Beck, A. T., Epstein, N., Brown, G., and Steer, R. A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *J. Consult. Clin. Psychol.* 56, 893–897. doi: 10.1037//0022-006x.56.6.893

Beck, A. T., Steer, R. A., and Brown, G. K. (1996). Manual for the beck depression inventory-II. San Antonio, TX: Psychological Corporation.

Blood, A. J., and Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. U. S. A.* 98, 11818–11823. doi: 10.1073/pnas.191355898

Bolger, D., Trost, W., and Schön, D. (2013). Rhythm implicitly affects temporal orienting of attention across modalities. *Acta Psychol.* 142, 238–244. doi: 10.1016/j.actpsy.2012.11.012

Bottiroli, S., Rosi, A., Russo, R., Vecchi, T., and Cavallini, E. (2014). The cognitive effects of listening to background music on older adults: processing speed improves with upbeat music, while memory seems to benefit from both upbeat and downbeat music. *Front. Aging Neurosci.* 6:284. doi: 10.3389/fnagi.2014.00284

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Buckner, R. L. (2004). Memory and executive function in aging and AD: multiple factors that cause decline and reserve factors that compensate. *Neuron* 44, 195–208. doi: 10.1016/j.neuron.2004.09.006

Burke, W. J., Roccaforte, W. H., and Wengel, S. P. (1991). The short form of the Geriatric Depression Scale: a comparison with the 30-item form. *J. Geriatr. Psychiatry Neurol.* 4, 173–178. doi: 10.1177/089198879100400310

Burkhard, A., Elmer, S., Kara, D., Brauchli, C., and Jäncke, L. (2018). The effect of background music on inhibitory functions: an ERP study. *Front. Hum. Neurosci.* 12:293. doi: 10.3389/fnhum.2018.00293

Carstensen, L. L., Turan, B., Scheibe, S., Ram, N., Ersner-Hershfield, H., Samanez-Larkin, G. R., et al. (2011). Emotional experience improves with age: evidence based on over 10 years of experience sampling. *Psychol. Aging* 26, 21–33. doi: 10.1037/a0021285

Chanda, M. L., and Levitin, D. J. (2013). The neurochemistry of music. *Trends Cogn. Sci.* 17, 179–193. doi: 10.1016/j.tics.2013.02.007

Cohrdes, C., Wrzus, C., Wald-Fuhrmann, M., and Riediger, M. (2020). "The sound of affect": age differences in perceiving valence and arousal in music and their relation to music characteristics and momentary mood. *Music. Sci.* 24, 21–43. doi: 10.1177/1029864918765613

Commenges, D., Gagnon, M., Letenneur, L., Dartigues, J. -F., Barberger-Gateau, P., and Salamon, R. (1992). Statistical description of the Mini-Mental State Examination for French elderly community residents. *J. Nerv. Ment. Dis.* 180, 28–32. doi: 10.1097/00005053-199201000-00007

Darowski, E. S., Helder, E., Zacks, R. T., Hasher, L., and Hambrick, D. Z. (2008). Age-related differences in cognition: the role of distraction control. *Neuropsychology* 22, 638–644. doi: 10.1037/0894-4105.22.5.638

Darrow, A.-A., Johnson, C., Agnew, S., Fuller, E. R., and Uchisaka, M. (2006). Effect of preferred music as a distraction on music majors' and nonmusic majors' selective attention. *Bull. Counc. Res. Music. Educ.* 170, 21–31.

Dean, R. T., Bailes, F., and Schubert, E. (2011). Acoustic intensity causes perceived changes in arousal levels in music: an experimental investigation. *PLoS One* 6:e18591. doi: 10.1371/journal.pone.0018591

Delis, D. C., Kaplan, E., and Kramer, J. H. (2001). Delis-Kaplan executive function system.

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

Diaz, J.-L., Flores-Gutiérrez, E. O., Rio-Portilla, Y., and Cabrera, M. C. (2011). "Musical emotion assessment, brain correlates, and gender differences" in *Music: Composition, interpretation and effects*. ed. T. A. Ivanova (New York, USA: Nova Science Pub Inc.), 31–56.

Dunifon, C. M., Rivera, S., and Robinson, C. W. (2016). Auditory stimuli automatically grab attention: evidence from eye tracking and attentional manipulations. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 1947–1958. doi: 10.1037/xhp0000276

Eerola, T., and Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychol. Music* 39, 18–49. doi: 10.1177/0305735610362821

Eriksen, C. W. (1995). The flankers task and response competition: a useful tool for investigating a variety of cognitive problems. *Vis. Cogn.* 2, 101–118. doi: 10.1080/13506289508401726

Eriksen, B. A., and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16, 143–149.

Fernandez, N. B., Hars, M., Trombetti, A., and Vuilleumier, P. (2019). Age-related changes in attentional control and their relationship with gait performance in older adults with high risk of falls. *Neuroimage* 189, 551–559. doi: 10.1016/j.neuroimage.2019.01.030

Fernandez, N. B., Trost, W. J., and Vuilleumier, P. (2020). Brain networks mediating the influence of background music on selective attention. *Soc. Cogn. Affect. Neurosci.* 14, 1441–1452. doi: 10.1093/scan/nsaa004

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Gazzaley, A., Cooney, J. W., Rissman, J., and D'esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nat. Neurosci.* 8, 1298–1300. doi: 10.1038/nn1543

Guerreiro, M. J., Murphy, D. R., and Van Gerven, P. W. (2010). The role of sensory modality in age-related distraction: a critical review and a renewed view. *Psychol. Bull.* 136, 975–1022. doi: 10.1037/a0020731

Hasher, L., Stoltzfus, E. R., Zacks, R. T., and Rypma, B. (1991). Age and inhibition. *J. Exp. Psychol. Learn. Mem. Cogn.* 17, 163–169. doi: 10.1037//0278-7393.17.1.163

Hasher, L., and Zacks, R. T. (1988). "Working memory, comprehension, and aging: a review and a new view" in *Psychology of learning and motivation*. *Vol. 22*. ed. G. H. Bower (New York, USA: Elsevier), 193–225.

Hsieh, S., and Fang, W. (2012). Elderly adults through compensatory responses can be just as capable as young adults in inhibiting the flanker influence. *Biol. Psychol.* 90, 113–126. doi: 10.1016/j.biopsycho.2012.03.006

Hsieh, S., and Lin, Y. -C. (2014). The boundary condition for observing compensatory responses by the elderly in a flanker-task paradigm. *Biol. Psychol.* 103, 69–82. doi: 10.1016/j.biopsycho.2014.08.008

Hudon, C., Potvin, O., Turcotte, M. -C., D'Anjou, C., Dubé, M., Préville, M., et al. (2009). Normalisation du Mini-Mental State Examination (MMSE) chez les Québécois francophones âgés de 65 ans et plus et résidant dans la communauté. *Can. J. Aging* 28, 347–357. doi: 10.1017/S0714980809990171

Hunter, P. G., and Schellenberg, E. G. (2010). "Music and emotion" in *Music perception*. eds. M. R. Jones, R. R. Fay and A. N. Popper (New York, NY: Springer), 129–164.

Husain, G., Thompson, W. F., and Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music. Percept.* 20, 151–171. doi: 10.1525/mp.2002.20.2.151

Jäncke, L., and Sandmann, P. (2010). Music listening while you learn: no influence of background music on verbal learning. *Behav. Brain Funct.* 6, 1–14. doi: 10.1186/1744-9081-6-3

Jefferies, L. N., Smilek, D., Eich, E., and Enns, J. T. (2008). Emotional valence and arousal interact in attentional control. *Psychol. Sci.* 19, 290–295. doi: 10.1111/j.1467-9280.2008.02082.x

Jiang, J., Scolaro, A. J., Bailey, K., and Chen, A. (2011). The effect of music-induced mood on attentional networks. *Int. J. Psychol.* 46, 214–222. doi: 10.1080/00207594.2010.541255

Juslin, P. N., and Laukka, P. (2004). Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. *J. New Music Res.* 33, 217–238. doi: 10.1080/0929821042000317813

Kallinen, K. (2002). Reading news from a pocket computer in a distracting environment: effects of the tempo of background music. *Comput. Hum. Behav.* 18, 537–551. doi: 10.1016/S0747-5632(02)00005-5

Kämpfe, J., Sedlmeier, P., and Renkewitz, F. (2010). The impact of background music on adult listeners: a meta-analysis. *Psychol. Music* 39, 424–448. doi: 10.1177/0305735610376261

Kleiner, M., Brainard, D., and Pelli, D. (2007). What's new in Psychtoolbox-3. *Perception* 36, 1–16.

Li, K. Z., and Lindenberger, U. (2002). Relations between aging sensory/sensorimotor and cognitive functions. *Neurosci. Biobehav. Rev.* 26, 777–783. doi: 10.1016/s0149-7634(02)00073-8

Lin, F. R., Yaffe, K., Xia, J., Xue, Q. L., Harris, T. B., Purchase-Helzner, E., et al. (2013). Hearing loss and cognitive decline in older adults. *JAMA Intern. Med.* 173, 293–299. doi: 10.1001/jamainternmed.2013.1868

Mammarella, N., Fairfield, B., and Cornoldi, C. (2007). Does music enhance cognitive performance in healthy older adults? The Vivaldi effect. *Aging Clin. Exp. Res.* 19, 394–399. doi: 10.1007/BF03324720

Maylor, E. A., and Lavie, N. (1998). The influence of perceptual load on age differences in selective attention. *Psychol. Aging* 13, 563–573. doi: 10.1037//0882-7974.13.4.563

Merritt, P., Hirshman, E., Wharton, W., Stangl, B., Devlin, J., and Lenz, A. (2007). Evidence for gender differences in visual selective attention. *Pers. Individ. Differ.* 43, 597–609. doi: 10.1016/j.paid.2007.01.016

Nater, U. M., Abbruzzese, E., Krebs, M., and Ehlert, U. (2006). Sex differences in emotional and psychophysiological responses to musical stimuli. *Int. J. Psychophysiol.* 62, 300–308. doi: 10.1016/j.ijpsycho.2006.05.011

Nemati, S., Akrami, H., Salehi, S., Esteky, H., and Moghimi, S. (2019). Lost in music: neural signature of pleasure and its role in modulating attentional resources. *Brain Res.* 1711, 7–15. doi: 10.1016/j.brainres.2019.01.011

Nittono, H., Tsuda, A., Akai, S., and Nakajima, Y. (2000). Tempo of background sound and performance speed. *Percept. Mot. Skills* 90:1122. doi: 10.2466/PMS.90.3.1122-1122

Nosek, B. A., and Banaji, M. R. (2001). The go/no-go association task. *Soc. Cogn.* 19, 625–666. doi: 10.1521/soco.19.6.625.20886

Olsen, K. N., Dean, R. T., Stevens, C. J., and Bailes, F. (2015). Both acoustic intensity and loudness contribute to time-series models of perceived affect in response to music. *Psychomusicology* 25, 124–137. doi: 10.1037/pmu0000087

Pachana, N. A., Byrne, G. J., Siddle, H., Koloski, N., Harley, E., and Arnold, E. (2007). Development and validation of the Geriatric Anxiety Inventory. *Int. Psychogeriatr.* 19, 103–114. doi: 10.1017/S1041610206003504

Rowe, G., Hirsh, J. B., and Anderson, A. K. (2007). Positive affect increases the breadth of attentional selection. *Proc. Natl. Acad. Sci. U. S. A.* 104, 383–388. doi: 10.1073/pnas.0605198104

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714

Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., and Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PLoS One* 4:e7487. doi: 10.1371/journal.pone.0007487

Salthouse, T. A. (2010). Is flanker-based inhibition related to age? Identifying specific influences of individual differences on neurocognitive variables. *Brain Cogn.* 73, 51–61. doi: 10.1016/j.bandc.2010.02.003

Savage, M. (2006). The musical field. *Cult. Trends* 15, 159–174. doi: 10.1080/09548960600712975

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Percept.* 21, 561–585. doi: 10.1525/mp.2004.21.4.561

Shih, Y. -N., Chien, W. -H., and Chiang, H. -S. (2016). Elucidating the relationship between work attention performance and emotions arising from listening to music. *Work* 55, 489–494. doi: 10.3233/WOR-162408

Stoet, G. (2010). Sex differences in the processing of flankers. *Q. J. Exp. Psychol.* 63, 633–638. doi: 10.1080/17470210903464253

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta Psychol.* 135, 77–99. doi: 10.1016/j.actpsy.2010.02.006

Thompson, R. G., Moulin, C., Hayre, S., and Jones, R. (2005). Music enhances category fluency in healthy older adults and Alzheimer's disease patients. *Exp. Aging Res.* 31, 91–99. doi: 10.1080/03610730590882819

Thompson, W. F., Schellenberg, E. G., and Husain, G. (2001). Arousal, mood, and the Mozart effect. *Psychol. Sci.* 12, 248–251. doi: 10.1111/1467-9280.00345

Trost, W., Frühholz, S., Schön, D., Labbé, C., Pichon, S., Grandjean, D., et al. (2014). Getting the beat: entrainment of brain activity by musical rhythm and pleasantness. *Neuroimage* 103, 55–64. doi: 10.1016/j.neuroimage.2014.09.009

van den Bosch, I., Salimpoor, V., and Zatorre, R. J. (2013). Familiarity mediates the relationship between emotional arousal and pleasure during music listening. *Front. Hum. Neurosci.* 7:534. doi: 10.3389/fnhum.2013.00534

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., and Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cogn. Emot.* 22, 720–752. doi: 10.1080/02699930701503567

Wahl, H. -W., and Heyl, V. (2003). Connections between vision, hearing, and cognitive function in old age. *Generations* 27, 39–45.

Wahn, B., and König, P. (2017). Is attentional resource allocation across sensory modalities task-dependent? *Adv. Cogn. Psychol.* 13, 83–96. doi: 10.5709/acp-0209-2

Wild-Wall, N., Falkenstein, M., and Hohnsbein, J. (2008). Flanker interference in young and older participants as reflected in event-related potentials. *Brain Res.* 1211, 72–84. doi: 10.1016/j.brainres.2008.03.025

Zeef, E. J., Sonke, C. J., Kok, A., Buiten, M. M., and Kenemans, J. (1996). Perceptual factors affecting age-related differences in focused attention: performance and psychophysiological analyses. *Psychophysiology* 33, 555–565. doi: 10.1111/j.1469-8986.1996.tb02432.x003


**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 03 November 2020
doi: 10.3389/fpsyg.2020.538194

# The Mozart Effect on the Episodic Memory of Healthy Adults Is Null, but Low-Functioning Older Adults May Be an Exception

***Susana Silva**, Filipa Belim and São Luís Castro***

*Center for Psychology at University of Porto, Faculty of Psychology and Education Sciences, University of Porto, Porto, Portugal*

Literature on the effects of passive music listening on cognitive performance is mixed, showing negative, null or positive results depending on cognitive domain, age group, temporal relation between music and task (background music vs. music before task, the latter known as Mozart effect), or listener-dependent variables such as musical preference. Positive effects of background music on the two components of episodic memory – item and source memory - for verbal materials seem robust and age-independent, and thus deserve further attention. In the current study, we investigated two potential enhancers of music effects on episodic memory: stopping music before task performance (Mozart effect) to eliminate music-related distraction and using preferred music to maximize reward. We ran a main study on a sample of 51 healthy younger adults, along with a pilot study with 12 older adults, divided into low- vs. high functioning according to cognitive performance in a screening test. Against our expectations, Bayesian analyses showed strong evidence that music had no advantage over silence or environmental sounds in younger adults. Preferred music had no advantage either, consistent with the possibility that music-related reward had no impact on episodic memory. Among older adults, low- but not high-functioning participants' item memory was improved by music – especially by non-preferred music - compared to silence. Our findings suggest that, in healthy adults, prior-to-task music may be less effective than background music in episodic memory enhancement despite decreased distraction, possibly because reward becomes irrelevant when music is stopped before the task begins. Our pilot findings on older adults raise the hypothesis that low-functioning older participants relate to prior-to-task auditory stimulation in deviant ways when it comes to episodic memory enhancement.

**Keywords: music listening, Mozart effect, episodic memory, preference, aging**

OPEN ACCESS

***Edited by:***
*Franco Delogu, Lawrence Technological University, United States*

***Reviewed by:***
*Mireille Besson, Institut de Neurosciences Cognitives de la Méditerranée (INCM), France*
*Thomas Schäfer, Medical School Berlin, Germany*

****Correspondence:***
*Susana Silva*
*susanamsilva@fpce.up.pt; zanasilva@gmail.com*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology*

***Received:*** *13 March 2020*
***Accepted:*** *08 October 2020*
***Published:*** *03 November 2020*

***Citation:***
*Silva S, Belim F and Castro SL (2020) The Mozart Effect on the Episodic Memory of Healthy Adults Is Null, but Low-Functioning Older Adults May Be an Exception. Front. Psychol. 11:538194. doi: 10.3389/fpsyg.2020.538194*

## INTRODUCTION

Music is incorporated in many aspects of everyone's life. The act of listening to music is often motivated by the search for aesthetic experiences or affective regulation (Groarke and Hogan, 2018), but there is also an increasing awareness that passive music listening – i.e., listening without analytical intentions – may enhance cognitive performance (for cognitive effects of music training, see Kraus and Chandrasekaran, 2010; Besson et al., 2011). Along with research on younger adults, attention has been paid to the benefits that passive music listening may have for older adults, including those with cognitive impairments (Thaut, 2010; Peck et al., 2016; Fang et al., 2017. Determining when and how passive music listening enhances cognitive performance in younger adults is an important practical goal, in that it may provide directions for the optimization of their study/work conditions. Doing the same for older age groups is perhaps an even more critical goal: as the world tries to deal with the consequences of increased longevity, cognitive improvement in older ages becomes a priority.

Despite the potential that lies in promoting cognitive enhancement through passive music listening, research findings remain mixed. These findings fall into three main research topics: sung vs. spoken words, background music effects, and neuropsychological priming via music (related to the Mozart effect tradition, see below). In all these topics, we find evidence of negative, null and positive impact of music across different cognitive domains and populations.

Research on *sung vs. spoken words* evaluates the impact of presenting sung versions of verbal material to be encoded and later recalled or recognized. Studies with younger adults have pointed to disadvantages (Racette and Peretz, 2007) or limited advantages (Wallace, 1994) of sung words over regular speech in subsequent word recall. In healthy older adults, the impact of singing words during encoding on subsequent verbal recognition seems null (Simmons-Stern et al., 2010). In contrast, Alzheimer's disease patients seem to benefit from listening to sung words for later recognition (Simmons-Stern et al., 2010) and recall of verbal material (Moussard et al., 2012).

A second research topic relates to *background music*, i.e., the presence of music during task performance. For younger participants, negative effects of music on reading, memory (Cassidy and MacDonald, 2007; Kämpfe et al., 2011; Thompson et al., 2012; Yang et al., 2015) and complex tasks (Gonzalez and Aiello, 2019) have been widely reported. These have been associated to distraction, dual-tasking and cognitive load. An extreme scenario of detrimental music-effects on younger adults may be found in studies using the Irrelevant Sound Effect paradigm (Perham and Vizard, 2011), in which the task to be accomplished requires serial ordering and thus may conflict with the ordinal information (sequence of events) of background music itself. Null effects of background music on working memory have been reported for younger adults (Chew et al., 2016), and positive effects on various domains may be found for both age groups: effects on foreign language learning (Kang and Williamson, 2014, younger adults), category fluency (Thompson et al., 2005, older and Alzheimer's disease patients), working memory (Mammarella et al., 2007; Mammarella, 2017, older adults), autobiographical memory (Irish et al., 2006, Alzheimer's disease; Foster and Valentine, 2001, mild and moderate dementia), semantic memory, processing speed (Bottiroli et al., 2014, older adults), verbal encoding (Ferreri et al., 2013, younger adults) and episodic memory (Bottiroli et al., 2014; Ferreri et al., 2014, 2015, older and younger adults).

Finally, a third approach tests the effects of listening to music prior to task performance, in the tradition of the so-called *Mozart effect* (Rauscher et al., 1993) and in line with the idea of neuropsychological priming (Cassidy and MacDonald, 2007). Again, we find mixed results: negative effects on working memory (Giannouli et al., 2019, younger and older adults), null effects on working memory training (Borella et al., 2019, older adults) and attention (Lake and Goldstein, 2011, mild cognitive impairment patients), as well as positive effects on word fluency (Giannouli et al., 2019, younger adults) and image encoding (Carr and Rickard, 2016, younger adults). Positive effects of music – whether as embedding sung words, as background, or as prime – have been related to the idea that music may be a source of mood improvement, arousal, or reward (Blood and Zatorre, 2001; Schellenberg, 2005; Salimpoor et al., 2013; Ferreri and Verga, 2016). While mood improvement tends to enhance cognitive performance due to increased dopamine levels in the brain (Ashby et al., 1999), arousal is known to act according to an inverted-U shape, where both extremely low and extremely high arousal damage performance while moderate levels benefit it (see Thompson et al., 2001, for a review). As for reward, the key to cognitive improvement seems to lie in the role of mesolimbic reward system in memory formation during reward-motivated learning (see Ferreri and Verga, 2016).

When facing this heavily mixed picture, identifying patterns of effectiveness is crucial, namely those concerning the cognitive domains that seem to respond most sensitively to music stimulation. Among reports of positive music-effects, studies on episodic memory for verbal materials (Ferreri et al., 2013, 2014, 2015; Bottiroli et al., 2014) stand out for their consistency: positive effects were replicated, and they were found both in healthy younger (Ferreri et al., 2015) and older adults (Bottiroli et al., 2014; Ferreri et al., 2014). Moreover, these studies showed that background music facilitates the encoding of printed verbal materials not only when music is compared to a silent context, but also when compared to non-musical auditory contexts such as environmental sounds or noise. This suggests that the impact of music on episodic memory is specific, rather than reflecting a general advantage of sound over silence. Episodic memory for verbal materials seems, thus, like a promising domain to test when it comes to considering the potential benefits of passive music listening. However, positive music-effects on episodic memory for verbal materials have been investigated only from the perspective of background music. Therefore, an important question is whether such positive effects extend to priming paradigms, i.e., when music is presented before the memory task and participants perform in silence. Both background and priming paradigms are thought to enhance arousal, reward and mood (Blood and Zatorre, 2001; Schellenberg, 2005; Salimpoor et al., 2013; Ferreri and Verga, 2016), but priming (non-task-concurrent music) is potentially less detrimental in terms of dual-task, cognitive load and distraction (Borella et al., 2019). From this viewpoint, it is possible that positive effects from music on episodic memory become more visible under priming than under concurrent stimulation: priming could keep the benefits of music, while minimizing its costs.

Another pattern that can be found in the literature concerns the moderating role of musical preference – the extent to which listeners enjoy the music style that is used as background or prime – in boosting the positive effects of music. Preference does not seem to change the Irrelevant Sound Effect (Perham and Vizard, 2011), but – at least in younger adults – it boosts the positive effects of background music on reading comprehension (Mcdonald, 2013), as well as those of musical primes on image encoding (Carr and Rickard, 2016): in both cases, preferred music outperforms silence, while non-preferred music does not. One explanation for the advantage of preferred music is that it is rewarding (Blood and Zatorre, 2001; Ferreri and Verga, 2016), and reward may be one mechanism subtending the positive effects of music on cognition (Ferreri and Verga, 2016). Despite the possibility that preference is a boosting factor because it maximizes reward, its moderating role has not been considered while approaching the impact of music on episodic memory.

In the present study, we investigated the impact of prior-to-task music on episodic memory in a sample of 51 healthy younger adults (Experiment 1) and we examined how preference modulates this impact. Following Ferreri et al. (2015)'s paradigm, we manipulated the auditory context associated to encoding printed words: silence, environmental sounds, musical excerpts. Unlike Ferreri et al. (2015), who presented the auditory context 20 s before the task began and kept it going as background while the task unfolded, we turned off the sound as the task began (prior-to-task stimulation). For comparison with Ferreri et al. (2015), we used the same pre-task 20-sec stimulation period, even though experiments on the Mozart effect tend to use longer stimulation times (a few minutes, see, e.g., Rauscher et al., 1993). We were aware that this could limit the impact of music but, in the absence of evidence regarding minimum auditory stimulation times to grant the Mozart effect, we moved on with our choice, which would grant maximum comparability with Ferreri et al.'s (2015) settings. We then tested auditory context effects on later recognition considering the two components of episodic memory: item memory and source memory (Easterbrook, 1959; Glisky et al., 1995). While item memory stands for the content of the encoding context, source memory represents the context in which the content is being encoded. In Ferreri et al.'s (2015) study, music had an advantage over environmental sounds and silence in both item (recognizing a printed word as old or new) and source memory (indicating the auditory context where the word appeared), suggesting that music is beneficial for episodic memory in its two dimensions. To grant the presence of a source component – an encoding context – in our study, we presented the auditory context not only before, but also after task performance: by doing this, we prevented the association of words to the following auditory context. To investigate the moderating role of preference, we then asked participants to rate their preference for the three musical excerpts, and we compared silence, environmental sounds, preferred and non-preferred music. Based on the idea that non-task-concurrent music (prime) could keep the benefits of background music (arousal, reward or mood improvement) while minimizing its distraction-related costs, we predicted that music would have strong advantage over silence and/or environmental sounds. Concerning preference, we predicted an advantage of preferred over non-preferred music, based on previous findings (Mcdonald, 2013; Carr and Rickard, 2016) as well as on the principle that preference could maximize the reward component of music.

The literature on episodic memory suggests that background music may have similar positive effects on younger and healthy older adults (Bottiroli et al., 2014; Ferreri et al., 2014, 2015). However, there is also evidence that, among older adults, those with cognitive impairments may be more responsive to music effects on cognition due to impairment-related mechanisms that add to the general mechanisms of arousal, reward or mood improvement: on the one hand, music may recruit brain areas spared after degeneration in cognitively impaired individuals and elicit compensatory mechanisms in the brain (Ferreri and Verga, 2016) which are not activated in healthy participants under the same music stimulation; on the other hand, music may act by reducing task-related anxiety, which is expected to be higher in cognitively impaired participants (Irish et al., 2006). For instance, Simmons-Stern et al. (2010) found benefits in sung (vs. spoken) words in Alzheimer's patients, but not in younger healthy participants. In order to make a preliminary approach to the possibility that low-functioning older adults could show increased sensitivity to the Mozart effect on episodic memory, we ran a second, pilot, experiment (Experiment 2), with a small group of older adults, split into low- vs. high-functioning concerning cognitive status.

Comparisons between the three groups – younger, high-functioning older and low-functioning older adults - were complemented with the analysis of sensitivity to possible semantic associations between music and words and to possible serial order effects (primacy/recency) on item memory.

# EXPERIMENT 1

## Methods

### Participants

Fifty-one undergraduate Psychology students (45 women, mean age $\pm$ *SD* = 19.9 $\pm$ 1.9 years) participated in the experiment. All participants had normal hearing and normal or corrected-to-normal vision, and did not report any psychiatric, neurological or cognitive problems. All signed informed consent according to the Declaration of Helsinki.

### Stimulus Materials

Verbal stimuli consisted of 45 + 45 = 90 words selected from the PORLEX database (Gomes and Castro, 2003, see **Supplementary Appendix A**). One set of 45 words was presented at the encoding phase (old words, to be remembered), and both sets (45 old + 45 new) were presented at the test phase. Old and new words were matched for length, frequency and morphological status (verb, noun or adjective).

Audio stimuli (auditory contexts before and after encoding) consisted of five 20-s audio files containing silence, environmental sounds (water running and birds, simultaneously) and three instrumental (non-vocal) music excerpts. We chose instrumental music because it may be less detrimental to performance in verbal tasks than vocal music (Crawford and Strapp, 1994). The audio file containing environmental sounds was extracted from a recording that was available online[1].

The three music excerpts were selected from an initial pool of 12 instrumental pieces, following an online pre-test with 20 university students (see **Supplementary Appendix B**). The pre-test was run with the goal of selecting maximally contrasting music stimuli in terms of preference. The initial set contained excerpts from four different and potentially contrasting music genres (3 examples per genre, 3 × 4 = 12): metal, hip hop, electronic and jazz. For each music excerpt, participants were asked to rate their level of preference (scale with 10 levels, where 1 means "I don't like it" and 10 "I like it a lot"). We analyzed the pre-test data per subject, with the initial aim of determining preference contrasts: we listed the most contrasting pairs of excerpts per subject and then counted the frequency of occurrence of such pairs across subjects. The final selection included "John and the Creatures – Here's to the Crazy Ones" (metal genre), "Robert Miles – Children" (electronic genre) and ("Thelonious Monk - Blue Monk") from jazz genre (**Supplementary Appendix B**).

All audio files except the silent one were normalized to 70 dB. The start and end points of music excerpts coincided with structural breaks in the musical piece, thus avoiding abrupt transitions.

## Procedure

For the encoding phase, the 45 old words were randomly divided into five lists of nine words (5 × 9 = 45), and each of these five word-lists was presented in a different auditory context (silence, environmental sounds, music 1/metal, music 2/electronic, music 3/jazz).

We created three versions of the encoding phase, each with a different pairing between word lists and auditory context. We did this by keeping the order of words constant while switching the order of auditory contexts: silence-environmental sounds-music vs. environmental sounds-music-silence vs. music-silence-environmental sounds. The three versions of the experiment were balanced across subjects. The goal of having multiple versions of the experiment - each with its own pairing between words and auditory contexts - was to control for two types of confounds that could act upon the auditory context effects we were interested in. One confound related to possible pre-existing semantic associations between a specific word and a given auditory context: for instance, the word "alumínio" [aluminum] could be associated to the metal music genre more easily than to silence or nature sounds (see Ferreri and Verga, 2016) due to the word's relation with the name of the genre (metal). If all participants saw the word coupled with metal music, they could have better encoding of both the word (item memory) and the auditory context (source memory) due to the semantic association, thus falsely increasing the advantage of music contexts. Different pairings between words and auditory contexts, distributed across participants, would counteract this type of confound. The other confound we wanted to control for concerned primacy/recency effects on item memory, i.e., the possibility of words at the top/bottom of the list being more memorable and lending an advantage to the first/last context presented. Again, while a single pairing between words and auditory contexts could lend a spurious advantage to words coupled with the first (primacy) or last context (recency), different versions of the experiment would dilute this influence.

The 20-s auditory context (music, environmental sounds or silence) was presented before and after each word list. Single words were presented on the screen for 5 s, always preceded by a 200-ms fixation cross (46 800 ms for each nine-word list). Thus, participants started each of the five blocks of the experiment by listening to the auditory context for 20 s, then they saw the words in silence (46.8 s) and, finally, they listened again to the auditory context. At the end of each block (auditory context + word list + auditory context), they saw an instruction on the monitor to use the space bar of the computer keyboard to move on to the next block. Participants were instructed to read the words silently and try to memorize as many as they could, since they would be later tested on this. They were not instructed to pay attention to the association between words and auditory context.

At the end of the encoding phase, participants were given a visual discrimination task (XO letter comparison task, Salthouse et al., 1997), lasting 5 min and working as an interference task. This interference task was critical for testing long-term episodic memory performance and not just working memory (Ferreri et al., 2015). The task consisted of examining a pair of letters (including only X and/or O) and deciding as quickly as possible whether the letters were the same (e.g., XX) or different (e.g., XO). Participants responded by pressing one of two keys in the computer keyboard.

After the interference task, participants were tested for discrimination between old and new words (item memory), as well as for their memory of the auditory context in which the word was presented (source memory). Thus, in this test phase, participants saw each of the 45 old + 45 new words, presented in pseudorandomized order: first, we did a full randomization using algorithmic procedures, and then we adjusted the list in order to avoid patterns concerning the alternation between old and new. They were then asked two different questions: first, "did you see this word before? Yes or No?", and second, "In which circumstances did you see it?". Here, there were four response options – Silence, Environmental sounds, Music or Did not see it before. Participants responded Yes or No by pressing either the left or the right Control (Ctrl) key in the computer keyboard. Half the participants saw "Yes No" on the monitor and were instructed to use the left Ctrl key for Yes. The other half saw "No Yes" and used the right one for Yes. As for the second question, the options were numbered as 1, 2, 3, 4, respectively, and participants pressed the corresponding number-key.

At the end of the experimental session, participants listened again to the three music excerpts and were asked to rate each of these for preference (1 to 10). These ratings allowed us to determine the contexts of preferred-music and non-preferred-music for each participant. The experiment was performed in a quiet room, and stimuli were delivered using Presentation

[1]https://www.youtube.com/watch?v=8myYyMg1fFE

software[2]. We used a 15-inch monitor for visual display and high-quality headphones for audio reproduction. Each session lasted approximately 40 min.

### Analysis

Preference ratings were first tested for significant contrasts across musical excerpts, to validate the idea that we were dealing with different levels of preference.

To determine the effects of auditory context, subject-level d-prime values (Stanislaw and Todorov, 1999) were computed for each of the five contexts (*silence, environmental sounds, metal, electronic, and jazz music*). D-prime is a measure of discrimination, and it is based on differences between hit rates (correct classification of the target) and false alarms (classification of non-targets as targets). A value of zero indicates no discrimination, and negative values point to reversed classifications (non-targets dominantly classified as targets and vice-versa). Values were calculated for item memory (discrimination between old and new verbal items) and source memory (discrimination between the target auditory context and the other ones). By averaging d-prime values for the three musical excerpts, we obtained d-prime for *music*. Based on each participant's preference ratings, another set of d-prime values were calculated: for the music context(s) with highest preference (*d-prime preferred music*), for the one(s) with lowest preference (*d-prime non-preferred music*). In some cases, participants rated a single excerpt as the most liked (e.g., score of 9 against 7 and 5) and a single excerpt as the least liked (the one with a score of 5, in the previous example). In other cases, two of the excerpts had the same score (e.g., 9, 9, 5): in these cases, we averaged the d-prime values of those two excerpts in order to define the d-prime (for preferred music, in this example).

Effects from auditory context were tested three times: first, comparing silence, environmental sounds and music (*general auditory context*); second, comparing silence, environmental sounds, preferred and non-preferred music (*preference-related auditory context*); finally, considering the five *specific auditory contexts* (silence, environmental sounds, metal music, electronic music, jazz music). For each analysis, we considered the effects on both item memory (discrimination between old and new words as a function of auditory context) and source memory (discriminant identification of contexts). All within-subject comparisons were made with repeated measures ANOVAs, using auditory context as factor and d-prime as dependent variable.

We made a complementary mixed ANOVA using *experiment version* (silence-environmental sounds-music vs. environmental sounds-music-silence vs. music-silence-environmental sounds) as between-subjects factor and general auditory context (silence, environmental sounds, music) as within-subjects. As we mentioned in the Procedure section (see above), experiment version manipulations were made to dilute potential confounds from *semantic association* or *serial order effects*, and thus they were not a critical part of our research question. However, examining effects from experiment version could allow us to identify such confounds, contributing to future improvements of the current paradigm and expanding the comparative analysis of younger vs. older adults' susceptibility to these. The use of *extended semantic associations* between words and auditory contexts by participants would be indicated by main effects of experiment version on both item and source memory – meaning that a particular experiment version carried semantic associations between the globality of word lists and their respective auditory contexts (e.g., version where the word 'petrify' was paired with silence, 'aluminum' with metal music and 'seagull' with environmental sounds), favoring both item and source recognition in that particular version. The use of *local semantic associations* would be indexed by interactions between experiment version and auditory context on both item and source memory, meaning that one and only one particular word list could be perceived as semantically related to its encoding context (e.g., a word list containing 'petrify' paired with silence in a given version, without pairings between 'aluminum' with metal music or 'seagull' with environmental sounds), favoring silence as an encoding context in that version. As for *serial order effects,* they would be manifested by interactions between experiment version and auditory context on item memory: if an auditory context elicited higher levels of performance in item memory when it appeared early in the experiment this could index *primacy* effects; in the reverse case (auditory context favoring item memory only when appearing late in the experiment), *recency* effects could be considered.

Unless otherwise specified, the adopted critical level of significance was 0.05. Violations of sphericity were compensated with Greenhouse-Geisser corrections. When significant effects from auditory context were observed, we carried out pairwise comparisons across the different levels of the factor using Bonferroni corrections for multiple comparisons. For the mixed ANOVA, significant interactions were broken down into auditory context effects per experiment version. The null-hypothesis-significance-testing procedure was complemented with the estimation of Bayes factors for null results, as delivered by JASP (JASP Team, 2020). Bayes factors for null results [null/alternative]) indicate the likelihood of observing the null data under the null hypothesis compared to the alternative hypothesis, thus quantifying the strength of evidence in favor of absent effects (Hoekstra et al., 2018; Wagenmakers et al., 2018). A rule of thumb is to consider a Bayes factor between 1 and 3 as weak evidence, though favoring the null hypothesis, between 3 and 10 as substantial, between 10 and 30 strong, and between 30 and 100 as decisive evidence (Jarosz and Wiley, 2014). Bayes factors can also be used to estimate the likelihood of the data under the alternative hypothesis compared to the null one (Bayes factor [alternative/null].

## Results

### Preference Ratings

Average preference ratings were 5.12 ($SD$ = 2.48) for the metal excerpt, 3.86 for electronic music ($SD$ = 1.97) and 5.76 for jazz ($SD$ = 2.13). Average rating contrasts were 1.25 for metal vs. electronic ($SD$ = 2.62), 1.90 for jazz vs. electronic ($SD$ = 2.66) and 0.65 for jazz vs. metal ($SD$ = 3.03). Ratings of electronic

[2]https://www.neurobs.com/

music were significantly lower than ratings of metal ($t(50) = 3.40$, $p = 0.001$, $d = 0.57$) and jazz ($t(50) = -5.10$, $p < 0.001$, $d = 0.89$). Metal and jazz excerpts did not differ from each other ($p = 0.13$).

### Effects of General Auditory Context

General auditory context (silence vs. environmental sounds vs. music) had no effects on item memory ($p = 0.63$, $\eta^2p = 0.009$) and source memory ($p = 0.51$, $\eta^2p = 0.013$). Bayes factors (null/alternative) indicated that data for item memory and source memory were 10.39 and 8.61 times more likely to occur under the null hypothesis than under the alternative one, showing strong and substantial evidence for the null hypothesis, respectively (Jarosz and Wiley, 2014).

### Effects of Preference-Related Auditory Context

Comparisons between silence, environmental sounds, preferred music and non-preferred music did not reveal significant effects, neither on item memory ($p = 0.67$, $\eta^2p = 0.010$) nor on source memory ($p = 0.88$, $\eta^2p = 0.004$). Bayes factors (null/alternative) of 21.86 for item memory and 31.29 for source memory indicated strong and very strong support to the null hypothesis, respectively.

### Effects of Specific Auditory Context

When comparing silence, environmental sounds and each of the three music excerpts, effects on item memory were null ($p = 0.68$, $\eta^2p = 0.011$: Bayes factor, BF (null/alternative) = 39.10), and those on source memory were significant ($F(4,200) = 3.72$, $p = 0.006$, $\eta^2p = 0.069$, BF [alternative/null] = 4.06, **Figure 1**). Metal music was significantly better identified than jazz as the auditory context of old words ($p = 0.003$, $d = 0.67$).

### Semantic Associations and Serial Order Effects as Revealed by Experiment Version

There was no evidence that semantic associations between words and auditory contexts were being exploited by participants: Main effects of experiment version (silence, environmental sounds vs. music; environmental sounds, music, silence vs. music, silence, environmental sounds) – which could indicate *extended* semantic associations if present for both memory types – were significant for source memory ($F(2,48) = 3.39$, $p = 0.042$, $\eta^2p = 0.124$, **Figure 2**) but null for item memory ($p = 0.41$, $\eta^2p = 0.036$; BF[null/alternative] = 1.04). In source memory, participants who completed the silence – environmental sounds (Es) – music version outperformed those who listened first to music, then silence, and then Es ($p = 0.045$, $d = 0.74$). Please note that, although the Bayes factor for the null data in item memory was weak (1.04), the inverted Bayes factors (alternative/null) differed substantially for item (BF = 0.957) vs. source memory (BF = 66.72).

As for the interaction between experiment version and auditory context, it was significant for both item ($F(4,96) = 5.14$, $p = 0.001$, $\eta^2p = 0.177$, BF [alternative/null] = 3.74) and source memory ($F(4,96) = 7.06$, $p < 0.001$, $\eta^2p = 0.227$, BF [alternative/null] = 178.66, **Figure 2**). At first sight, this could point to *local* semantic associations – i.e., the possibility that one specific word list fits semantically with a specific auditory context, favoring subsequent recognition of words and contexts in the experiment version where this particular combination occurred. However, the interactions showed different patterns for the two types of memory (see **Figure 2**): in the music-silence-Es version, words presented under silence and music were better recognized than those presented in the Es context (item memory, Es vs. silence: $p = 0.016$, $d = 1.01$; Es vs. music: $p = 0.016$, $d = 0.80$), while auditory context had no effect on source memory. In the Es-music-silence version, music outperformed silence ($p = 0.012$, $d = 1.08$) in item memory, while environmental sounds outperformed both music ($p = 0.005$, $d = 1.08$) and silence ($p < 0.001$, $d = 1.46$) in source memory. In the silence-Es-music version, auditory context had no effects on item memory, while music outperformed Es in source memory ($p = 0.001$, $d = 0.60$).

Although local semantic associations do not seem to account for the interactions between auditory context and experiment version on item vs. source memory, *primacy-related effects* (expected for item memory only) are consistent with the observed pattern of interactions: whenever there were context effects on item memory, the last context (last word list) always elicited the worst performance (see **Figure 2**). Therefore, there seems to be evidence that recency was detrimental to performance in younger adults.

## Discussion

Unlike our prediction, music contexts did not enhance episodic memory compared to silence or environmental sounds, suggesting that the advantage of background music that had been seen in previous studies (Bottiroli et al., 2014; Ferreri et al., 2014, 2015) may vanish if music is stopped before task performance. Also unlike our prediction – but consistent with the possibility that music-related reward did not improve episodic memory - music preference had no moderating role.

Additional comparisons, considering specific musical pieces, indicated an advantage of heavy metal over jazz music in source memory. One possibility is that this was due to increased physiological arousal (Rickard, 2004), since the jazz excerpt was lighter in terms of texture and event density compared to heavy metal. From this viewpoint, it is possible that physiological arousal overrided reward in the enhancement of younger adults' source memory. Given that we did not measure arousal in participants, this is, however, just a hypothesis.

From the analysis of experiment version effects, we found no evidence that younger participants were responding based on semantic associations between words and auditory contexts. However, the interactive effects of experiment version and auditory context on item memory suggested that the most recently presented items may have been encoded more poorly, possibly revealing a primacy effect on item memory, or a decrease in attention levels due to fatigue.

Experiment version had a main effect on source memory: participants had better recognition of the auditory context associated with a given word when auditory contexts were ordered as silence, environmental sounds and music during encoding. As we already pointed out in the Results section, this pattern is not consistent with effects from semantic associations (otherwise it should be present for item memory

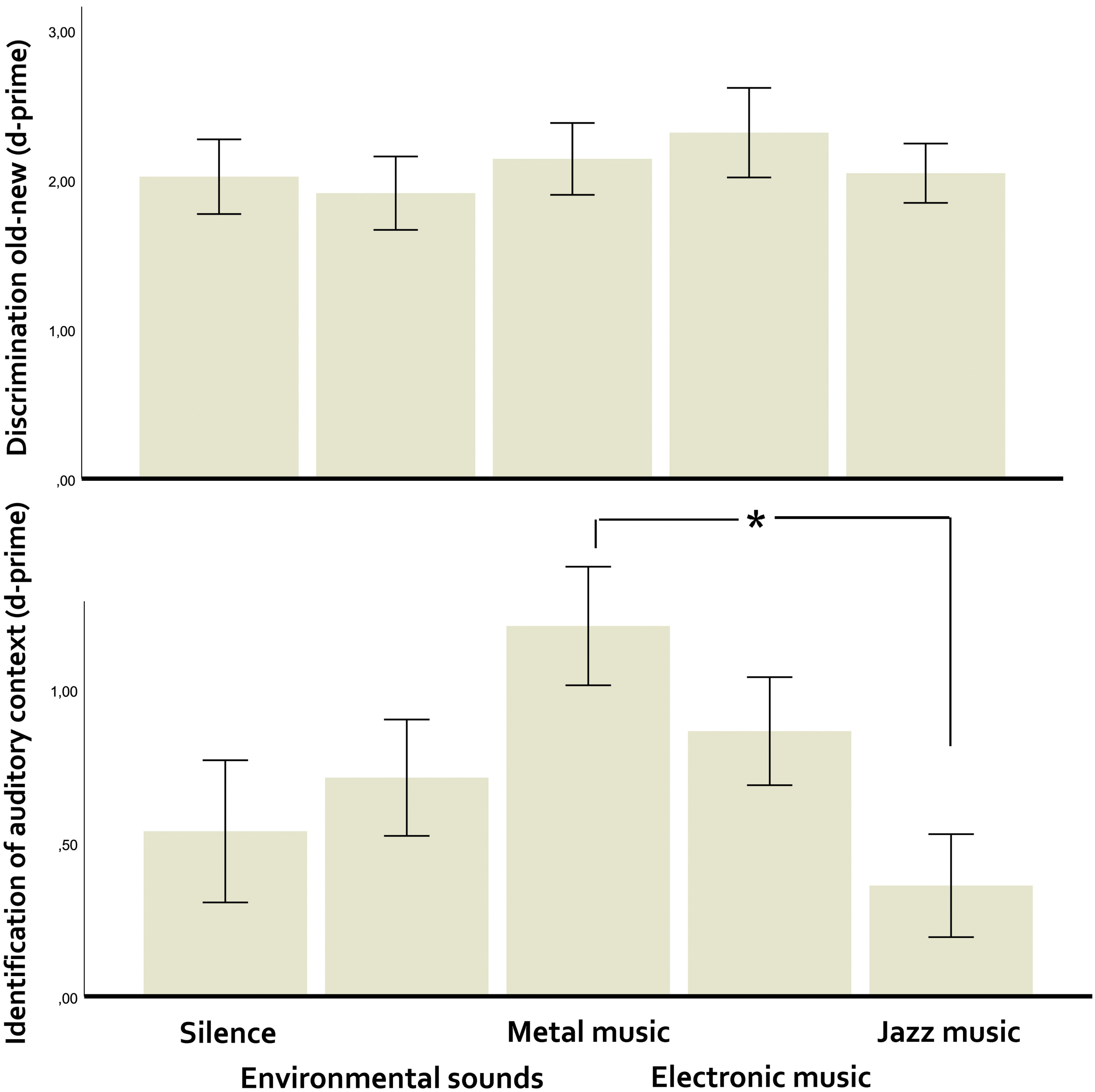



**FIGURE 1 |** Discrimination old-new (item memory, above) and discriminant identification of all auditory contexts (source memory, below) as a function of specific auditory context (including contrasts across music genres). Vertical bars represent the standard error of the mean. The asterisk indicates significant differences.

too). So, what else may account for the positive impact of ordering auditory contexts as silence, environmental sounds and music on source memory? One possibility is that participants were accurate in recency memory – one measure of source memory referring to awareness of the early vs. late appearance of words in the encoding phase (Czernochowski et al., 2008) – but they were not aware of the order of sources (e.g., music appearing early, followed by silence and then by environmental sounds). Under uncertainty, participants might have assumed that contexts were ordered as silence, environmental sounds and music, since this corresponds to a logical order that starts with no information (silence), follows into low-structured information (environmental sounds) and ends with high-structured information (music). Therefore, when confronted with a word they knew had been presented early in the experiment, they might have assumed the word appeared in a silent context; if they thought the word appeared at the end, they would respond 'music'. When auditory contexts were ordered as silence, environmental sounds and music, this strategy would grant high performance in source memory. In the other two versions, it would not work. In Experiment 2, we looked again into these

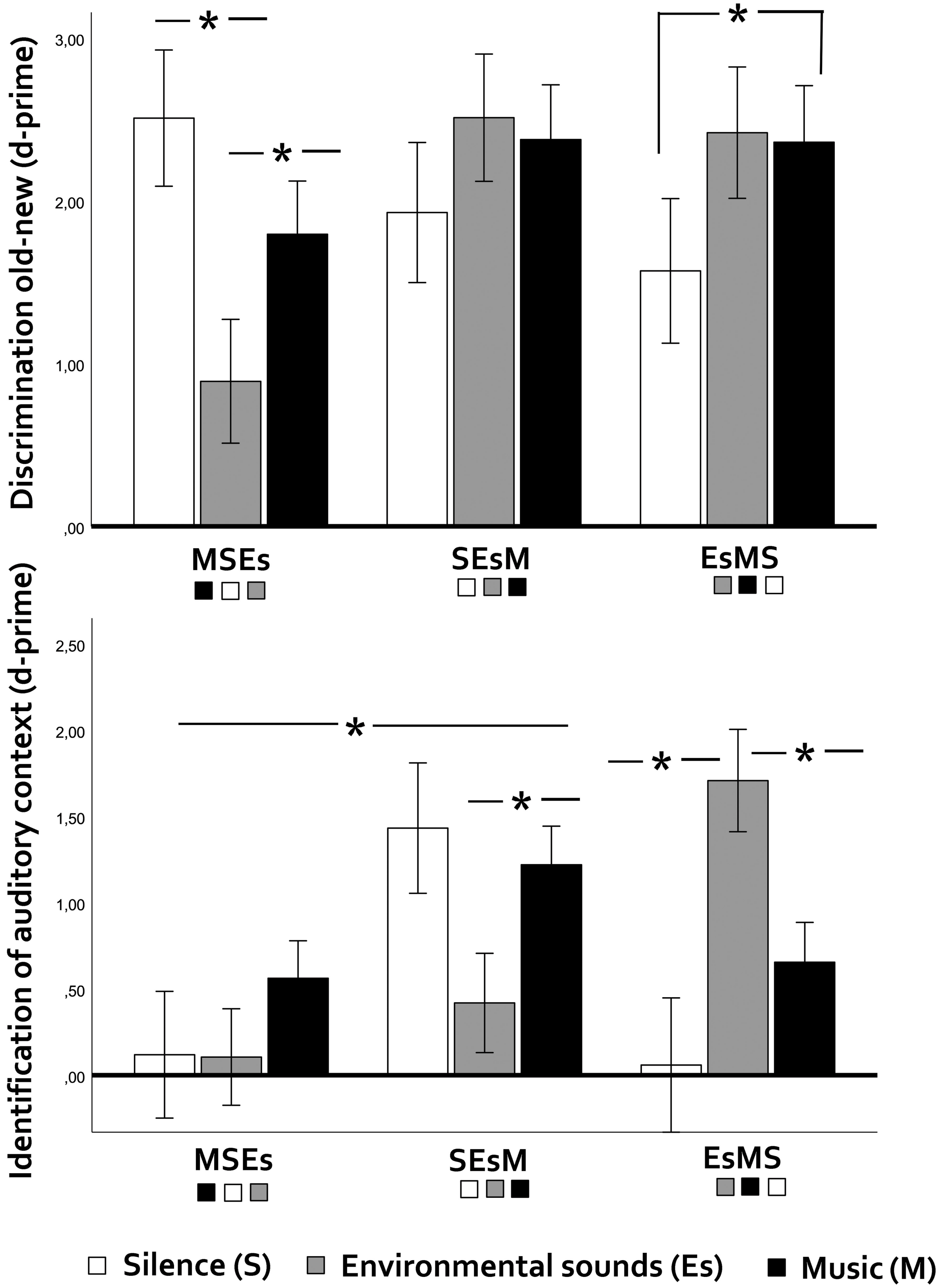



**FIGURE 2 |** Discrimination old-new (item memory, above) and discriminant identification of preference-related auditory contexts (source memory, below) in younger adults as a function of general auditory context (silence-environmental sounds, music) and experiment version (three different orderings of silence, environmental sounds and music blocks). Vertical bars indicate the standard error of the mean. Asterisks indicate significant differences. Note that – when contexts differ significantly, the last context presented in each version elicits the worst performance in item memory, indicating primacy effects. Asterisks indicate significant differences.

experiment version effects to see whether and how they act in older adults.

# EXPERIMENT 2

This experiment was designed as a pilot study to test the hypothesis that low-functioning older adults may be more sensitive than healthy younger and older adults to music effects. Since we were interested in a homogenous sample in terms of socio-economical and cultural background, we recruited participants from a nursing-home located at a small community. The available sample was already small, and eligibility criteria (see below) made it even smaller (n = 12). Considering the effects of normal aging on cognitive functioning and the low educational levels of our participants (see below), we simplified the stimulus set with the goal of avoiding experimental stress and/or floor effects. In order to understand how the effects of musical context may depend on lower vs. higher cognitive status, we administered the Mini Mental-State Examination (MMSE) adapted to the Portuguese population (Guerreiro et al., 1994).

## Methods

### Participants

Twelve healthy older adults (7 women, mean age $\pm$ *SD* = 75.25 $\pm$ 8.3 years; mean schooling $\pm$ *SD* = 4.92 $\pm$ 2.39 years) from a nursing home participated in the experiment. These participants were selected by the local health technician, based on the absence of incapacitating deficits (able to read, speak, communicate, corrected vision and hearing). Our examination of cognitive functioning using MMSE indicated normal levels of performance in seven participants (high-functioning), and five cases below the cut-off score (low-functioning): 17, 19, 21, 22, 22, cut-off of 24 (Morgado et al., 2009).

Prior to any contact with the participants, the local ethics committee approved the experiment. All participants signed informed consent according to the Declaration of Helsinki.

### Stimulus Materials

Given that participants had very low educational levels, we reduced substantially the size of the stimulus set. Verbal stimuli consisted of 10 + 10 words (20) words, taken from the list used in Experiment 1 (see **Supplementary Appendix C**). We selected words with lower frequency and length to facilitate encoding. One set of 10 words was presented at the encoding phase (old words, to be remembered), and both sets (old and new, 10 old + 10 new) were presented at the test phase.

Silence and environmental sounds stimuli were the same as in Experiment 1. We selected music stimuli based on our previous knowledge about the socio-economic, generational and cultural background of participants: we considered two genres likely to be familiar and/or preferred (fado and traditional local music), contrasting with hip-hop – highly likely to be non-preferred. Again, the idea of contrasting musical styles served to prevent the possibility of the same person having the same preference for all pieces in the list. Auditory stimuli were processed in the same way as in Experiment 1.

The Mini Mental State Examination was administered to assess cognitive functioning of each participant. MMSE is one of the most commonly used screening tools and assesses global cognitive functions in clinical or research contexts, and it is suited to individuals with low educational levels (Guerreiro et al., 1994).

### Procedure

The procedure was similar to Experiment 1, except that there were only 10 old words pseudo-randomly divided into five lists of 2 words each. Each list was presented in-between silence, environmental sounds, music 1/hip hop, music 2/fado music and 3/traditional local music.

Mini Mental State Examination (MMSE) was administered before the experimental task. To minimize difficulties associated to the interaction with the computer, the experimenter pressed the keyboard keys after participants provided their responses vocally.

### Analysis

We ran the same analyses as in Experiment 1, adding cognitive functioning (low-, below MMSE cut-off score, $n$ = 5, vs. high-functioning participants, equal or above cut-off, $n$ = 7) as between-subjects factor to all. Given that the two groups were differently distributed across the three versions of the experiment (low-functioning: one participant in version 1, one in version 2, three in version 3 vs. high-functioning: three in version 1, three in version 2, one in version 3), we paid critical attention to experiment version effects when interpreting the effects of cognitive functioning.

Due to the limited size of our sample, and also because cognitive status as dichotomous variable generated imbalanced groups (5 vs. 7), we did a cross-check analysis with MMSE scores as covariate.

## Results

### Preference Ratings

Average preference ratings for fado, hip-hop and traditional music were 7.5 (*SD* = 1.93), 4.17 (*SD* = 2.55), and 8 (*SD* = 1.95), respectively. Average rating contrasts were 3.33 (*SD* = 2.93) for fado vs. hip-hop, 0.50 (*SD* = 2.84) for fado vs. traditional music and 3.83 (*SD* = 3.61) for hip-hop vs. traditional music. Hip-hop ratings were significantly lower than fado ($t(11)$ = 3.93, $p$ = 0.002, $d$ = 1.49) and traditional music ($t(11)$ = 3.67, $p$ = 0.004, $d$ = 1.70).

### Effects of General Auditory Context and Cognitive Functioning

Cognitive functioning had a marginal main effect on item memory ($F(1,10)$ = 3.41, $p$ = 0.094, $\eta^2p$ = 0.255, BF [alternative/null] = 2.47, **Figure 3**), with low-functioning participants showing poorer performance than high-functioning ones. Effects of cognitive functioning on source memory did not reach significance ($p$ = 0.105, $\eta^2p$ = 0.241, BF [null/alternative] = 1.99]).

Main effects of auditory context were non-significant for item ($p$ = 0.732, $\eta^2p$ = 0.041, BF [null/alternative] = 0.869) or source memory ($p$ = 0.741, $\eta^2p$ = 0.030, BF [null/alternative] = 4.18). For mean hit rates and d-prime values, please see **Supplementary**

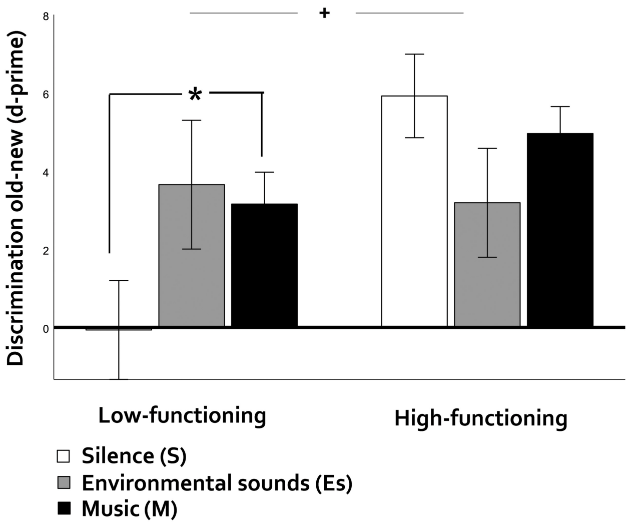



**FIGURE 3 |** Discrimination old-new (item memory) in older adults as a function of general auditory context (silence-environmental sounds, music) and cognitive functioning (low vs. high). Vertical bars represent the standard error of the mean. The asterisk indicates significant differences and the symbol '+' marginal ones. In the low-functioning group, music, but not environmental sounds outperformed silence.

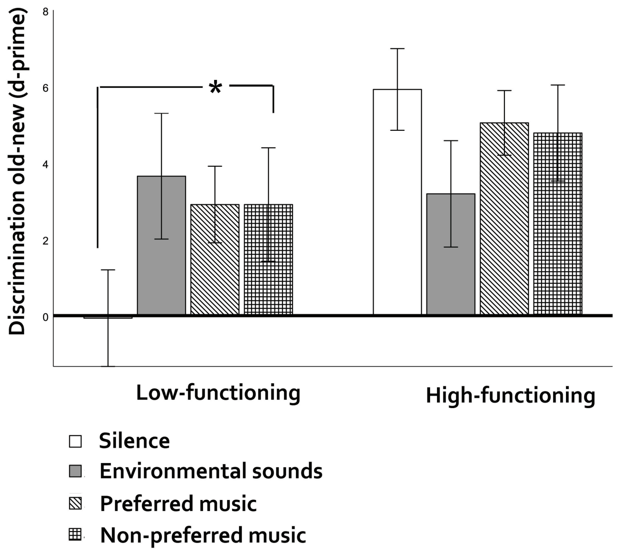



**FIGURE 4 |** Discrimination old-new (item memory) in older adults as a function of preference-related auditory context (preferred vs. non-preferred music) and cognitive functioning (low vs. high). Vertical bars represent the standard error of the mean. Asterisks indicate significant differences.

**Appendix D**). However, the interaction between cognitive functioning and auditory context was significant for item memory ($F(29,3) = 6.71$, $p = 0.006$, $\eta^2p = 0.402$). Among low-functioning participants, music, but not environmental sounds, outperformed silence (music: $p = 0.009$, $d = 1.43$, BF [alternative/null] = 17.34; environmental sounds: $p = 0.42$, $d = 0.96$, BF [alternative/null] = 1.07) while high-functioning participants did not show cross-context differences ($ps > 0.24$, BFs [alternative/null] < 1.39). For source memory, the interaction was non-significant ($p = 0.387$, $\eta^2p = 0.091$, BF [null/alternative] = 4.05). Please note that experiment version did not interact with auditory context (please see below), indicating that the effects of auditory context on low- but not high-functioning participants were not due to group differences in the allocation to different experiment versions.

Cross-check analyses with MMSE scores as covariate showed a significant interaction between auditory context and MMSE on item memory ($F(2,20) = 5.78$, $p = 0.010$, $\eta^2p = 0.336$). Follow-up correlations indicated a strong negative correlation between subject-level advantage of music over silence and MMSE scores ($r(10) = -0.802$, $p = 0.002$), thus reinforcing the idea that music increases its benefits as cognitive status becomes lower. For source memory, the interaction between auditory context and MMSE scores was not significant ($p = 0.262$, $\eta^2p = 0.125$).

## Effects of Preference-Related Auditory Context and Cognitive Functioning

Comparisons across silence, environmental sounds, preferred and non-preferred music (**Figure 4**) showed non-significant results for item ($p = 0.732$, $\eta^2p = 0.041$, BF [null/alternative] = 3.82) and source memory ($p = 0.701$, $\eta^2p = 0.030$, BF [null/alternative] = 7.48).

Cognitive functioning interacted significantly with auditory context for item memory ($F(1,10) = 3.43$, $p = 0.029$, $\eta^2p = 0.255$). Pairwise comparisons for low-functioning participants showed a marginal advantage of non-preferred music over silence ($p = 0.097$, $d = 1.08$, BF [alternative/null] = 4.99 vs. preferred music over silence, BF [alternative/null] = 1.36), while high-functioning participants showed no context effects ($ps > 0.53$, BFs [null/alternative) > 1.05). The cognitive functioning x context interaction was non-significant for source memory ($p = 0.665$, $\eta^2p = 0.050$, BF [null/alternative] = 10.99).

Cross-check analyses with MMSE as covariate confirmed these results, showing a significant interaction between auditory context and MMSE ($F(1,10) = 3.86$, $p = 0.019$, $\eta^2p = 0.279$). Follow-up comparisons showed a marginal negative correlation between the advantage of non-preferred music over silence and MMSE scores ($r(12) = -0.544$, $p = 0.067$). For source memory, there was no significant interaction ($p = 0.295$, $\eta^2p = 0.144$).

## Effects of Specific Auditory Context and Cognitive Functioning

Comparisons across silence, environmental sounds, fado, hip-hop and traditional music (**Figure 5**) showed non-significant effects on item ($p = 0.161$, $\eta^2p = 0.148$, BF [null/alternative] = 1.80) or source memory ($p = 0.489$, $\eta^2p = 0.080$, BF [null/alternative] = 5.35), but the interaction between cognitive functioning and context on item memory was significant ($F(4,40) = 2.84$, $p = 0.036$, $\eta^2p = 0.221$, BF [alternative/null] = 1.32). In the low-functioning group, silence showed a significant disadvantage regarding fado ($p = 0.033$,


**FIGURE 5 |** Discrimination old-new (item memory) in older adults as a function of specific auditory context and cognitive functioning (low vs. high). Vertical bars represent the standard error of the mean. Asterisks indicate significant differences.

*d* = 1.75) and hip-hop (*p* = 0.016, *d* = 1.18). In the high-functioning group, pairwise comparisons showed no significant differences (*p*s > 0.089).

The interaction between auditory context and cognitive functioning was non-significant for source memory (*p* = 0.762, $\eta^2$p = 0.044, BF [null/alternative] = 11.12).

### Semantic Associations and Serial Order Effects as Revealed by Experiment Version

As in younger adults, there was no evidence of extended (main effect of experiment version) or local (interaction between experiment version and auditory context) semantic associations between words and auditory contexts, which should be reflected into experiment version effects on both item and source memory: Main effects of experiment version were null (*p* = 0.746, $\eta^2$p = 0.093, BF [null/alternative] = 3.06) for item memory with no interactions with cognitive functioning (*p* = 0.593, $\eta^2$p = 0.194, BF [null/alternative] = 2.15), but they were significant for source memory ($F(2,6)$ = 8.15, *p* = 0.019, $\eta^2$p = 0.731, BF [alternative/null] = 1.30) with further interactions with cognitive functioning ($F(2,6)$ = 10.36, *p* = 0.011, $\eta^2$p = 0.755, BF [alternative/null] = 1.50, **Figure 6**). Interactions between experiment version and general auditory context were non-significant for item (*p* = 0.755, $\eta^2$p = 0.136, BF [null/alternative] = 3.99) and source memory (*p* = 0.844, $\eta^2$p = 0.102, BF [null/alternative] = 2.22). For both memory types, there were no significant further interactions with cognitive functioning (item: *p* = 0.234, $\eta^2$p = 0.350, BF [null/alternative] = 2.39; source: *p* = 0.218, $\eta^2$p = 0.359, BF [null/alternative] = 0.782).

The lack of interaction between auditory context and experiment version on item memory (BF [null/alternative] = 3.99) indicates that the serial order (primacy) effects which had been observed for younger participants did not hold for older adults.

As for the main effect of experiment version on source memory only (see above, [$F(2,6)$ = 8.15, *p* = 0.019, $\eta^2$p = 0.731]), it seems to replicate what we saw in experiment 1 – a possible confound between the real order of contexts and a logical (but unreal) order, but now we must take into consideration the interaction with cognitive function (see above, $F(2,6)$ = 10.36, *p* = 0.011, $\eta^2$p = 0.755). Similar to younger adults, both high and low functioning older participants showed improved source memory when listening to sound-Es-music compared to music-silence-sound (**Figure 6**; high: *p* = 0.009; low: *p* = 0.007). However, in the low functioning group, the most logical order was outperformed by Es-music-silence (*p* = 0.004). So, once again, low-functioning adults deviated from the general pattern.

## Discussion

Unlike younger and older high-functioning participants, music contexts had a positive impact on item memory of low-functioning older participants, showing an advantage over silence. Interestingly, the advantage of music was not driven by preferred music. Instead, when preference-related context effects were tested, it was non-preferred music that showed an advantage over silence. Consistent with this, the least preferred music genres (fado and hip-hop) showed significantly increased benefits to item memory compared to silence.

The order of auditory sources (experiment version) had again an effect on source memory. In older high-functioning adults, the effect was the same as in younger adults (source memory was maximal when auditory contexts were ordered as silence, environmental sounds and music). In the low-functioning group, the order silence-environmental sounds-music was only the second most effective.

Unlike younger adults, item recency was not detrimental to item memory performance of older adults, both high-and low-functioning. This could indicate that either primacy effects were not active in older adults, or that older adults – unlike younger ones - did not get tired or decrease their levels of attention across task blocks. Since the latter possibility is unlikely, we will favor the first interpretation (primacy effects in younger but nor older participants) in the general discussion.

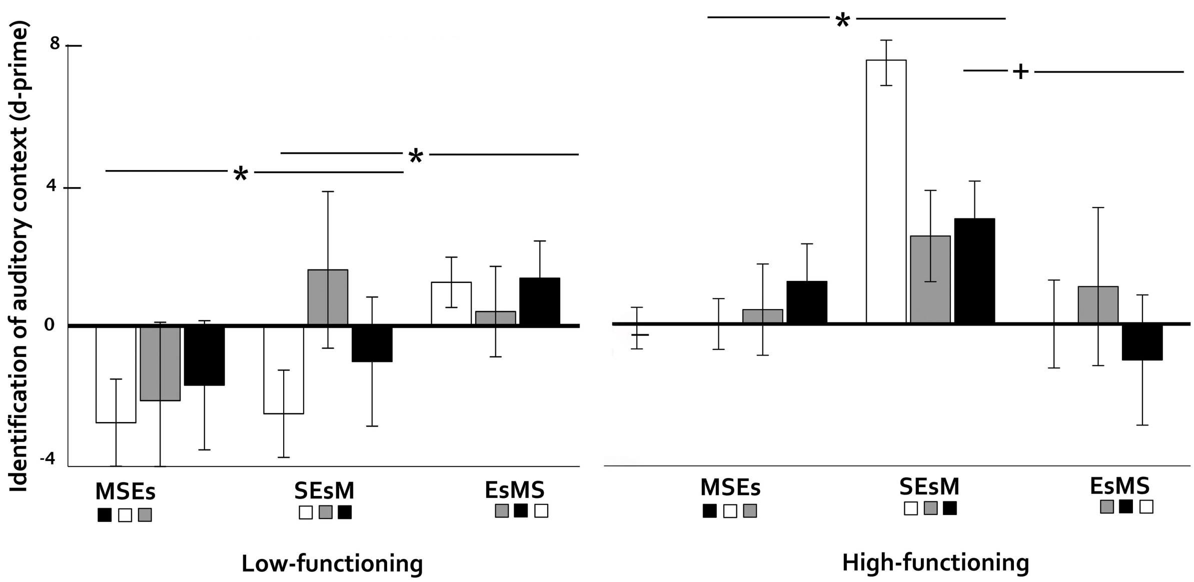



**FIGURE 6 |** Discriminant identification of general auditory context (silence-environmental sounds, music) in older adults as a function of cognitive status (low – vs. high functioning) and experiment version (three different orderings of silence, environmental sounds and music blocks). Vertical bars represent the standard error of the mean. Asterisks indicate significant differences and the symbol '+' marginal ones.

Paralleling younger adults, there was no evidence that semantic associations between words and auditory contexts were being used in older adults.

## GENERAL DISCUSSION

We aimed to determine whether the positive effects of background music on episodic memory that have been reported for both younger and older adults (Ferreri et al., 2013, 2014, 2015; Bottiroli et al., 2014) are observed when music is stopped before task performance (music as neuropsychological priming, or Mozart effect), and whether preferred music has an advantage over non-preferred. We predicted that music would show strong positive effects due to its non-distractive presence (not working as an interfering background), and that preferred music would have an advantage due to its role in maximizing reward. We ran our main study on a sample of 51 younger adults, and we complemented our approach with a pilot study on 12 older participants, divided into low- and high-functioning according to their performance on the Mini Mental-State Examination test. We predicted that low-functioning older participants would show increased sensitivity to positive music effects compared to healthy younger and older adults.

Unlike our predictions, the main study on younger adults (Experiment 1) showed null advantages of music over silence or environmental sounds, and no advantages of preferred music over non-preferred. This applied to both item and source memory. Null results were supported by the outcomes of Bayesian analyses. In addition, a sensitivity power analysis carried out with G*Power (Faul et al., 2007) showed that the size of our younger adults sample ($n$ = 51) would reliably detect small effect sizes for general auditory context ($\eta^2p = 0.031$) and preference-related auditory context ($\eta^2p = 0.026$). Therefore, evidence for lack of effects on item or source memory seems solid in this group. Contrasting these null effects of prior-to-task music with the positive effects of background music that have been seen before (Ferreri et al., 2013, 2014, 2015; Bottiroli et al., 2014) suggests that eliminating the distraction caused by background music (Cassidy and MacDonald, 2007; Kämpfe et al., 2011; Thompson et al., 2012; Yang et al., 2015) may yield other losses. Specifically, it is possible that, for episodic memory, the arousal, mood or reward effects afforded by music backgrounds (Blood and Zatorre, 2001; Schellenberg, 2005; Salimpoor et al., 2013; Ferreri and Verga, 2016) are lost or attenuated when music is stopped before the task begins. Given that preference also had null effects, and preference is strongly linked to reward, it is possible that reward may be the key loss: music-related reward may no longer favor episodic memory if music is stopped before the task begins. In contrast, arousal may not have been totally irrelevant to performance: although heavy metal music (potentially arousing at the physiological level) did not outperform non-musical auditory contexts, it differed significantly from jazz (a less arousing genre) concerning effects on source memory. Determining the roles of music-related reward vs. physiological arousal as prior-to-task enhancers of episodic memory remains, thus, a challenge for future research.

Alternative explanations for our null results based on sample characteristics do not seem likely: it is known that musical expertise attenuates or eliminates the Mozart effect (Twomey and Esgate, 2002), but our participants were not musicians; it is also known that men are less sensitive to this effect than women (Gilleta et al., 2003), but the majority of our participants was female. One alternative explanation that may deserve future approaches relates to the short exposure time to auditory contexts we used in our study. Typically, the Mozart effect has been implemented with a few minutes of musical stimulation, while we used only 20 s. Our intention was to maximize the similarity with Ferreri et al.'s (2015) study on background music to optimize the comparison between background and prior-to-task music, but it is possible that our choice weakened the effects of priming auditory contexts. Future studies could work on this exposure time variable to shed light on this matter.

Our pilot study with low- vs. high-functioning older adults was less powered than we could wish and its results should be taken with caution: for effects of general auditory context, G*power sensitivity analyses showed that our study could reliably detect medium main effects of auditory context (our effects were small – $\eta^2p = 0.04$ and.03 for item and source memory, respectively), large effects of cognitive functioning (our effect on item memory was large, $\eta^2p = 0.26$ but medium on source memory, $\eta^2p = 0.24$), and large interactive effects between auditory context and cognitive functioning (our effect on item memory was large, $\eta^2p = 0.40$ and medium for source memory, $\eta^2p = 0.09$). For preference-related auditory context, we could reliably detect only large main effects of auditory context and the observed effects were small - $\eta^2p = 0.04$ and.03 for item and source memory, respectively. The interaction between preference-related auditory context and cognitive functioning could be reliably detected also only under large effect sizes. We found a large effect size for item memory ($\eta^2p = 0.25$), but a small one ($\eta^2p = 0.05$) for source memory. Admitting that low power did not make us miss small, yet real effects, our findings point to a deviant pattern in low-functioning older adults: Unlike younger and older high-functioning adults, music had an advantage over silence in item memory, and non-preferred, rather than preferred music, carried this advantage. In terms of specific music genres, fado and hip-hip (the least preferred) outperformed the silence condition, while local traditional music did not. A striking aspect in low-functioning older participants was their very low performance under silence (see **Figures 3–5**). Finally, the advantage that ordering auditory contexts as silence, environmental sounds and music had on the source memory of healthy younger and older adults was not obvious in the low-functioning group.

The advantage of music over silence only in low-functioning older adults may have been due to recruitment of music-induced compensatory mechanisms (Ferreri and Verga, 2016) that were not activated in younger and healthy older adults. It may also have been related to the presence of extremely high - hence detrimental - levels of task-related arousal (Thompson et al., 2001) in low-functioning older adults compared to the other groups: anxiety may have been mitigated by music (Irish et al., 2006), highlighting the beneficial effects of music in low-functioning, but not in other adults. The possibility that mitigating anxiety was a key mechanism in enhancing performance would be consistent with low-functioning participants' very low performance under silence (potentially anxiogenic), and their facilitated performance under the two least physiologically arousing music pieces (fado and hip-hop, with softer timbres compared to local traditional music) – which coincided with the least preferred genres. From this viewpoint, relaxation rather than lack of preference would be the key to the positive impact of music. Future studies adding anxiety measures before and after music listening may shed light on this matter. Concerning the findings that low-functioning participants did not commit the same error as the other groups in source memory - responding to memory for temporal order (recency memory, which item was seen most recently?) based on a logical order of sources, instead of responding to the specific question we asked them (in which auditory context?, see section "Discussion" on Experiment 1), this may point to emerging problems with recency memory in this population.

Low-functioning older participants shared some behaviors with the other two groups. First, primacy effects on item memory were observed in younger, but not in older adults – low- or high-functioning. From a methodological viewpoint, this means that the use of multiple experiment versions may be critical when studying younger adults, and the difference between low- and high-functioning older participants may be irrelevant in this respect. In addition, we saw no evidence that semantic associations between words and auditory contexts were being used in any of the groups. This may indicate that prior-to-task music is not submitted to semantic processing, or, more likely, that our stimulus materials provided little or no opportunities for semantic associations. Future experiments manipulating the amount of semantic associations between music and word lists could clarify this matter.

The main contribution of our study was to highlight the null impact of prior-to-task music on the episodic memory of healthy younger adults. Unlike extant research using background, task-concurrent music, we saw no advantage from music over silence or environmental sounds in a priming paradigm where music was stopped before the task began, possibly because music-related reward loses strength under these circumstances. Although our findings require future validation with direct comparisons between background vs. prior-to-task music effects on episodic memory, they contribute to raise the awareness that background and prior-to-task music effects may engage different mechanisms (see also Lake and Goldstein, 2011) and they should not be approached interchangeably.

As for our pilot study contrasting low- with high-functioning older adults, it helped raise a number of hypothesis – namely the possible role of music in low-functioning participants' anxiety reduction in a prior-to-task music stimulation scenario, but future research using large samples is mandatory before conclusions can be drawn.

## DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

## ETHICS STATEMENT

Ethical review and approval was not required for Experiment 1 in accordance with the local legislation and institutional requirements. Experiment 2 was reviewed and approved by Instituto da Segurança Social da Madeira. The patients/participants provided their written informed consent to participate in these studies.

## AUTHOR CONTRIBUTIONS

SS, FB, and SC conceptualized the study. SS and FB created the stimulus set, analyzed the data, and wrote the first draft of the manuscript. FB collected the data. SC revised the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was funded by the Portuguese Foundation for Science and Technology (CPUP UID/PSI/00050/2013; FEDER/COMPETE2020 POCI-01-0145-FEDER-007294).

## ACKNOWLEDGMENTS

We are grateful to Bela Vista nursing home and to all our participants. We leave a special thank you note to Dr. Nuno Gaspar for fruitful discussions on this topic.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.538194/full#supplementary-material

**Supplementary Table 1 |** Appendices **A–D**.

**Supplementary Table 2 |** Database older adults.

**Supplementary Table 3 |** Database younger adults.

## REFERENCES

Ashby, F. G., Isen, A. M., and Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychol. Rev.* 106, 529–550. doi: 10.1037/0033-295X.106.3.529

Besson, M., Chobert, J., and Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *Front. Psychol.* 2:94. doi: 10.3389/fpsyg.2011.00094

Blood, A. J., and Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci.* 98, 11818–11823. doi: 10.1073/pnas.191355898

Borella, E., Carretti, B., Meneghetti, C., Carbone, E., Vincenzi, M., Madonna, J. C., et al. (2019). Is working memory training in older adults sensitive to music? *Psychol. Res.* 83, 1107–1123. doi: 10.1007/s00426-017-0961-8

Bottiroli, S., Rosi, A., Russo, R., Vecchi, T., and Cavallini, E. (2014). The cognitive effects of listening to background music on older adults: processing speed improves with upbeat music, while memory seems to benefit from both upbeat and downbeat music. *Front. Aging Neurosci.* 6:284. doi: 10.3389/fnagi.2014.00284

Carr, S. M., and Rickard, N. S. (2016). The use of emotionally arousing music to enhance memory for subsequently presented images. *Psychol. Music* 44, 1145–1157. doi: 10.1177/0305735615613846

Cassidy, G., and MacDonald, R. A. R. (2007). The effect of background music and background noise on the task performance of introverts and extraverts. *Psychol. Music* 35, 517–537. doi: 10.1177/0305735607076444

Chew, A. S.-Q., Yu, Y.-T., Chua, S.-W., and Gan, S. K.-E. (2016). The effects of familiarity and language of background music on working memory and language tasks in Singapore. *Psychol. Music* 44, 1431–1438. doi: 10.1177/0305735616636209

Crawford, H. J., and Strapp, C. M. (1994). Effects of vocal and instrumental music on visuospatial and verbal performance as moderated by studying preference and personality. *Pers. Individ. Diff.* 16, 237–245. doi: 10.1016/0191-8869(94)90162-7

Czernochowski, D., Fabiani, M., and Friedman, D. (2008). Use it or lose it? SES mitigates age-related decline in a recency/recognition task. *Neurobiol. Aging* 29, 945–958. doi: 10.1016/j.neurobiolaging.2006.12.017

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.* 66, 183–201. doi: 10.1037/h0047707

Fang, R., Ye, S., Huangfu, J., and Calimag, D. P. (2017). Music therapy is a potential intervention for cognition of Alzheimer's disease: a mini-review. *Transl. Neurodegen.* 6:2. doi: 10.1186/s40035-017-0073-9

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Ferreri, L., Aucouturier, J., Muthalib, M., Bigand, E., and Bugaiska, A. (2013). Music improves verbal memory encoding while decreasing prefrontal cortex activity: an fNIRS study. *Front. Hum. Neurosci.* 7:779. doi: 10.3389/fnhum.2013.00779

Ferreri, L., Bigand, E., and Bugaiska, A. (2015). The positive effect of music on source memory. *Musicae Sci.* 19, 402–411. doi: 10.1177/1029864915604684

Ferreri, L., Bigand, E., Perrey, S., Muthalib, M., Bard, P., and Bugaiska, A. (2014). Less effort, better results: how does music act on prefrontal cortex in older adults during verbal encoding? An fNIRS study. *Front. Hum. Neurosci.* 8:301. doi: 10.3389/fnhum.2014.00301

Ferreri, L., and Verga, L. (2016). Benefits of music on verbal learning and memory: how and when does it work? *Music Percept. Interdiscip. J.* 34, 167–182. doi: 10.1525/mp.2016.34.2.167

Foster, N. A., and Valentine, E. R. (2001). The effect of auditory stimulation on autobiographical recall in dementia. *Exp. Aging Res.* 27, 215–228. doi: 10.1080/036107301300208664

Giannouli, V., Kolev, V., and Yordanova, J. (2019). Is there a specific Vivaldi effect on verbal memory functions? Evidence from listening to music in younger and older adults. *Psychol. Music* 47, 325–341. doi: 10.1177/0305735618757901

Gilleta, K. S., Vrbancic, M. I., Elias, L. J., and Saucier, D. M. (2003). A Mozart effect for women on a mental rotations task. *Percept. Mot. Skills* 96(3 Pt 2), 1086–1092. doi: 10.2466/pms.2003.96.3c.1086

Glisky, E. L., Polster, M. R., and Routhieaux, B. C. (1995). Double dissociation between item and source memory. *Neuropsychology* 9, 229–235. doi: 10.1037/0894-4105.9.2.229

Gomes, I., and Castro, S. L. (2003). Porlex, a lexical database in European Portuguese. *Psychologica* 32, 91–108.

Gonzalez, M. F., and Aiello, J. R. (2019). More than meets the ear: investigating how music affects cognitive task performance. *J. Exp. Psychol. Appl.* 25, 431–444. doi: 10.1037/xap0000202

Groarke, J. M., and Hogan, M. J. (2018). Development and psychometric evaluation of the adaptive functions of music listening scale. *Front. Psychol.* 9:516. doi: 10.3389/fpsyg.2018.00516

Guerreiro, M. P. S. A., Silva, A. P., Botelho, M. A., Leitão, O., Castro-Caldas, A., and Garcia, C. (1994). Adaptação à população portuguesa da tradução do mini mental state examination (MMSE). *Rev. Portug. Neurol.* 1, 9–10.

Hoekstra, R., Monden, R., van Ravenzwaaij, D., and Wagenmakers, E.-J. (2018). Bayesian reanalysis of null results reported in medicine: strong yet variable evidence for the absence of treatment effects. *PLoS One* 13:e0195474. doi: 10.1371/journal.pone.0195474

Irish, M., Cunningham, C. J., Walsh, J. B., Coakley, D., Lawlor, B. A., Robertson, I. H., et al. (2006). Investigating the enhancing effect of music on autobiographical memory in mild Alzheimer's disease. *Dement. Geriatr. Cogn. Disord.* 22, 108–120. doi: 10.1159/000093487

Jarosz, A. F., and Wiley, J. (2014). What Are the Odds? A practical guide to computing and reporting Bayes factors. *J. Problem Solv.* 7:2. doi: 10.7771/1932-6246.1167

JASP Team (2020). *JASP (Version 0.13.1)[Computer software]*.

Kämpfe, J., Sedlmeier, P., and Renkewitz, F. (2011). The impact of background music on adult listeners: a meta-analysis. *Psychol. Music* 39, 424–448. doi: 10.1177/0305735610376261

Kang, H. J., and Williamson, V. J. (2014). Background music can aid second language learning. *Psychol. Music* 42, 728–747. doi: 10.1177/0305735613485152

Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605. doi: 10.1038/nrn2882

Lake, J. I., and Goldstein, F. C. (2011). An examination of an enhancing effect of music on attentional abilities in older persons with mild cognitive impairment. *Percept. Mot. Skills* 112, 267–278. doi: 10.2466/04.10.15.PMS.112.1.267-278

Mammarella, N. (2017). Commentary: emotion effects on implicit and explicit musical memory in normal aging. *Front. Psychol.* 8:2208. doi: 10.3389/fpsyg.2017.02208

Mammarella, N., Fairfield, B., and Cornoldi, C. (2007). Does music enhance cognitive performance in healthy older adults? The Vivaldi effect. *Aging Clin. Exp. Res.* 19, 394–399. doi: 10.1007/BF03324720

Mcdonald, J. (2013). The effect of music preference on complex task performance. *Glob. Tides* 7:10.

Morgado, J., Rocha, C. S., Maruta, C., Guerreiro, M., and Martins, I. P. (2009). Novos valores normativos do mini-mental state examination. *Sinapse* 9, 10–16.

Moussard, A., Bigand, E., Belleville, S., and Peretz, I. (2012). Music as an aid to learn new verbal information in Alzheimer's disease. *Music Percept.* 29, 521–531. doi: 10.1525/mp.2012.29.5.521

Peck, K. J., Girard, T. A., Russo, F. A., and Fiocco, A. J. (2016). Music and memory in Alzheimer's disease and the potential underlying mechanisms. *J. Alzheimers Dis.* 51, 949–959. doi: 10.3233/JAD-150998

Perham, N., and Vizard, J. (2011). Can preference for background music mediate the irrelevant sound effect? *Appl. Cogn. Psychol.* 25, 625–631. doi: 10.1002/acp.1731

Racette, A., and Peretz, I. (2007). Learning lyrics: to sing or not to sing? *Mem. Cogn.* 35, 242–253. doi: 10.3758/BF03193445

Rauscher, F. H., Shaw, G. L., and Ky, C. N. (1993). Music and spatial task performance. *Nature* 365, 611–611. doi: 10.1038/365611a0

Rickard, N. S. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychol. Music* 32, 371–388. doi: 10.1177/0305735604046096

Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., and Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science* 340, 216–219. doi: 10.1126/science.1231059

Salthouse, T. A., Toth, J. P., Hancock, H. E., and Woodard, J. L. (1997). Controlled and automatic forms of memory and attention: process purity and the uniqueness of age-related influences. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 52, 216–228.

Schellenberg, E. G. (2005). Music and cognitive abilities. *Curr. Direct. Psychol. Sci.* 14, 317–320. doi: 10.1111/j.0963-7214.2005.00389.x

Simmons-Stern, N. R., Budson, A. E., and Ally, B. A. (2010). Music as a memory enhancer in patients with Alzheimer's disease. *Neuropsychologia* 48, 3164–3167. doi: 10.1016/j.neuropsychologia.2010.04.033

Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Instr. Comput.* 31, 137–149. doi: 10.3758/bf03207704

Thaut, M. H. (2010). Neurologic music therapy in cognitive rehabilitation. *Music Percept.* 27, 281–285. doi: 10.1525/mp.2010.27.4.281

Thompson, R. G., Moulin, C. J. A., Hayre, S., and Jones, R. W. (2005). Music enhances category fluency in healthy older adults and Alzheimer's disease patients. *Exp. Aging Res.* 31, 91–99. doi: 10.1080/03610730590882819

Thompson, W. F., Schellenberg, E. G., and Husain, G. (2001). Arousal, mood, and the mozart effect. *Psychol. Sci.* 12, 248–251. doi: 10.1111/1467-9280.00345

Thompson, W. F., Schellenberg, E. G., and Letnic, A. K. (2012). Fast and loud background music disrupts reading comprehension. *Psychol. Music* 40, 700–708. doi: 10.1177/0305735611400173

Twomey, A., and Esgate, A. (2002). The Mozart effect may only be demonstrable in nonmusicians. *Percept. Mot. Skills* 95(3 Pt 1), 1013–1026. doi: 10.2466/pms.2002.95.3.1013

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., et al. (2018). Bayesian inference for psychology. Part II: example applications with JASP. *Psychon. Bull. Rev.* 25, 58–76. doi: 10.3758/s13423-017-1323-7

Wallace, W. T. (1994). Memory for music: effect of melody on recall of text. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 1471–1485. doi: 10.1037/0278-7393.20.6.1471

Yang, J., McClelland, A., and Furnham, A. (2015). The effect of background music on the cognitive performance of musicians: a pilot study. *Psychol. Music* 44, 1202–1208. doi: 10.1177/0305735615592265

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

OPINION
published: 08 December 2020
doi: 10.3389/fpsyg.2020.557162

# Come on Baby, Light My Fire: Sparking Further Research in Socio-Affective Mechanisms of Music Using Computational Advancements

*Ilana Harris*[1]* *and Mats B. Küssner*[2]

[1] Centre for Music and Science, Faculty of Music, University of Cambridge, Cambridge, United Kingdom, [2] Department of Musicology and Media Studies, Humboldt-Universität zu Berlin, Berlin, Germany

**Keywords: socio-affective behavior, computational methods, joint music-making, prosocial behavior, emotional contagion**

OPEN ACCESS

***Edited by:***
*Marta Olivetti Belardinelli, Sapienza University of Rome, Italy*

***Reviewed by:***
*Jonna Katariina Vuoskoski, University of Oslo, Norway*

***Correspondence:***
*Ilana Harris*
*ilanaharris@alumni.harvard.edu*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology*

***Received:*** *29 April 2020*
***Accepted:*** *30 October 2020*
***Published:*** *08 December 2020*

***Citation:***
*Harris I and Küssner MB (2020) Come on Baby, Light My Fire: Sparking Further Research in Socio-Affective Mechanisms of Music Using Computational Advancements. Front. Psychol. 11:557162. doi: 10.3389/fpsyg.2020.557162*

## MUSICAL ENGAGEMENT: SOCIO-AFFECTIVE UNDERPINNINGS

Socio-affective behavior is entangled in our experience of music (Devroop, 2012; Koelsch, 2014; Aucouturier and Canonne, 2017; Saarikallio, 2019). Joint musical engagement, or making and listening to music with others, was found to result in increased prosocial tendencies (Kirschner and Tomasello, 2010; Rabinowitch et al., 2013; Cirelli et al., 2014) and is thought to occur due to overlapping mechanisms underpinning interactive musical behavior and empathically driven prosocial behaviors (Rabinowitch et al., 2012; Clarke et al., 2015; Saarikallio, 2019). In this paper, we present opportunities for experimental investigation of emotional contagion, a specific subprocess hypothesized to lie at this overlap, and highlight ways to improve understanding of *how* joint musical engagement may promote prosocial behaviors.

Socio-affective components of joint musical engagement have been postulated following empirical investigation of joint music-making and group music-listening (Egermann et al., 2011; Rabinowitch et al., 2013) and hinge on subprocesses including affective alignment, where joint expression of emotion among interlocutors allows for facilitated transfer of semantic and affective content (Cross, 2005; Bharucha et al., 2012; Rabinowitch et al., 2012; Vesper et al., 2017). In this sense, affective alignment may contribute to higher-level processes of musical interaction such as shared intentionality by ensuring that members are working toward a common musical goal in real time and have "coordinated action roles for pursuing that shared goal" (Tomasello et al., 2005, p. 680) through upregulating constituent socio-affective behaviors (e.g., other-directed behaviors) that help individuals ascertain their interlocutor's internal state and align their behavior accordingly (Cross et al., 2012). Joint music-making's positive influence on socio-affective behaviors in non-musical contexts suggests that psychosocial processes underpinning musical interaction may overlap with those involved in non-musical interaction, and that co-activation of these overlapping structures may result in prosocial transfer effects (Kirschner and Tomasello, 2010; Cross et al., 2012; Saarikallio, 2019).

Scientific inquiry probing the effects of musical engagement on prosociality has risen in prevalence in recent years; particularly, musical engagement's influence on prosocial behaviors underscored by empathy has gained considerable traction in music psychology and related fields (King and Waddington, 2017; Davis, 2018; Riess et al., 2018). Empathy may be defined as "the ability to produce emotional and experiential responses to the situations of others that approximate their responses and experiences" (Rabinowitch et al., 2013, p. 485) and is a core component of social cognition comprising both slow (controlled) and fast (automatic) psychological subprocesses (i.e., dual process theory; Lieberman, 2007; Batson, 2009) that ultimately "constitute a causal force in motivating prosociality towards other conspecifics" (Decety et al., 2016, p. 371). Slow processes are "evaluative," requiring top-down cognitive assessment, while fast processes are immediate, "automatic detection" of social signals; separate neural representations for fast and slow processing of social information have been proposed accordingly (i.e., "mirror neuron system" and "mentalizing system"; Vogeley, 2017). Emotional contagion, a subprocess of empathy, is defined as automatic mimicry of another's behavioral cues associated with a particular affective state; it is thought to foster survival through increasing recognition of and successful communication between conspecifics, and underpins the capacity to build and maintain human attachment bonds (de Vignemont and Singer, 2006; Feldman, 2017; Prochazkova and Kret, 2017). Though theoretical study of emotional contagion in music has begun, there is a lack of experimental study that causally explains how automatic detection of socio-affective signals influences our experience of music (Miu and Vuoskoski, 2017). Investigation of emotional contagion during musical interaction is critical to understanding how relationships between co-performers may be similar to other types of social relationships (e.g., through attachment bonds) and, consequently, how joint musical engagement may lead to upregulated other-directed behaviors such as those that arise within a particular social relationship.

Experimental investigation of emotional contagion is practically difficult because it requires simultaneous measurement of interlocutors' complex emotion states in interactionist paradigms; this matter is further complicated in the context of music, where substantial ecological validity is needed to elicit behaviors of interest (e.g., empathy-promoting musical components; Rabinowitch et al., 2013). In the following two sections, we introduce research from related fields incorporating computational techniques for measuring behavioral and physiological correlates of emotional contagion; we situate such techniques in the context of music psychology and suggest avenues by which they may be incorporated into existing experimental paradigms to triangulate investigation of socio-affective processes using behavioral, physiological, and social signal processing, as has been done across numerous subfields of psychology (Pantic and Vinciarelli, 2015; Azevedo et al., 2016; Sutherland et al., 2017; de Barbaro, 2019; Oswald et al., 2020).

## THE FRONTIER: BEHAVIORAL CUES

The following section outlines possibilities for investigating socio-affective components in music using computational methods for behavior recognition drawn from research in computer science and the behavioral sciences. First, facial cues are an important behavioral cue for emotional expression in music (Thompson et al., 2008, 2010; Livingstone et al., 2009; Waddell and Williamon, 2017). Approximately 95% of automatic emotion recognition literature relies on facial cues, which has led to applicability of these techniques to an expanding number of datasets (Noroozi et al., 2018). Computational emotion recognition using facial cues has been incorporated into the study of empathic behavior in group settings. For instance, Kumano et al. (2011, 2014, 2015) conducted a series of experiments to see if empathic interactions could be predicted based on facial data from video recordings of four-person meetings. Their Naive Bayes Network Model was able to predict empathy state given facial expression information across time and improved when parameters such as reaction time in mirrored expression between interlocutors and head gesture annotations were added. Scientific study of music has not yet incorporated computational determination of joint emotional expression epochs from facial cues; this is likely to be a fruitful area of inquiry, involving relay of complex affective information at the intersection of individually, socially, and musically driven systems.

Though literature examining communication of emotion through body as opposed to facial cues in non-musical contexts is lacking, several studies have found that determination of emotion state is modulated by body posture/movement (Aviezer et al., 2008a,b; Martinez et al., 2016). In musical settings, visual content, often in the form of body movement, plays a critical role in conveying affective information (Vines et al., 2011; Vuoskoski et al., 2014, 2016). Computational analysis of body position/gesture using motion capture experiments has incurred important findings with respect to joint emotional expression and audience-perceived emotion (Burger et al., 2013; Chang et al., 2019). Still, analyses of body postures/gestures across various cultural and developmental contexts and further determination of indices that convey socio-affective information are necessary to better understand their role in both musical interaction and potential transference to non-musical interaction.

Following research on prosocial behaviors as a consequence of joint music-making, similar outcomes of music listening have begun to be studied (Ruth and Schramm, 2020). Continuous self-report of emotion by audience members during live concert settings is a promising experimental tool (Egermann, 2019). These measures collect rating data simultaneously with and continuously throughout the stimuli's presentation and may be able to achieve the temporal specificity needed in order to determine instances of affective alignment between participants. Moreover, continuous measurement of self-reported affect supports various forms of rating interfaces, including linear potentiometers (Vines et al., 2011; Baytaş et al., 2016), binary trigger buttons (Baytaş et al., 2019), and four-quadrant valence-arousal joysticks (Sharma et al., 2020; or its digital analog in Egermann, 2019). Furthermore, such interfaces may be attached to the participant (i.e., wearables) such that implementation in a paradigm involving movement is possible. In addition, top-down cameras can provide useful visual displays of crowd behavior, as evinced in analyses of pedestrian movement (Xu et al., 2020); concerning non-coordinated movement of audiences, this line of research within computer science could nicely complement existing methods in motion tracking (e.g., analysis of head movement in Swarbrick et al., 2019).

## THE FUTURE: EMERGING DATA SOURCES AND ANALYSES

Music serves a number of social functions in everyday listening (Sloboda and O'Neill, 2001). Recently, social surrogacy was added as a potential reason for musical listening, extrapolating online listener behavior to internal processes (Schäfer and Eerola, 2020). Greenberg and Rentfrow (2017) list a number of avenues by which social media and streaming data can be used within music psychology; implementing several of these analyses in tandem could be well-suited to studying socio-affective behavior. For example, combining analyses of song-specific emotion data from Spotify APIs, listeners' comments on social media and self-report data gathered from online surveys could help determine individual differences in socio-emotional components of music listening. In addition, experience sampling methods (ESMs) have become increasingly popular for administering repeated surveys of everyday musical experience (e.g., Juslin et al., 2008; de Barbaro, 2019), with more recent ESM interfaces allowing for user mobility and nuanced user input (e.g., Randall and Rickard, 2017). Housing state-of-the-art digital self-assessment scales for emotion within existing ESMs could help bolster outcome evaluations and uncover relationships between socio-affective components in everyday musical behaviors (Betella and Verschure, 2016; Juslin, 2016).

Detecting emotion from acoustic properties of music has been extensively researched; for instance, tempo and mode tend to be good indicators of perceived emotion (Eerola et al., 2013). However, computational methods for music emotion recognition have tended to favor certain features over others (e.g., timbre accounting for over 60%; Yang et al., 2019). Recently, researchers have begun to develop software packages for emotion recognition in music, which include fine-grained features such as specific textural shifts and articulations (Panda et al., 2018). Such software could provide important contextual affective information in existing joint music-making paradigms. In addition, natural language processing (NLP) of song lyrics is a burgeoning area of research in music psychology (Anglada-Tort et al., 2019). NLP could be useful for identifying empathic tendencies in group songwriting, a prevalent music therapy intervention (e.g., using language style synchrony as a proxy for empathy as in Lord et al., 2015).

Lastly, behavioral tasks that can robustly quantify socio-emotional components in various populations after engagement in social music activities are needed. Several studies to date have developed novel tasks or adapted tasks from other disciplines to achieve this end (Rabinowitch et al., 2013; Reddish et al., 2014; Brown, 2017). In social neuroscience, researchers have investigated mechanisms underlying evolutionarily advantageous socio-affective behavior through experimental paradigms targeting social modulation of threat response (DeVries et al., 2003; Coan et al., 2006). Automated stress recognition *via* analyses of multimodal physiological and motion data has begun to show potential for validated use in social science research (Hovsepian et al., 2015). Further, higher-order pattern detection of heart rate variability (HRV) has been used to predict interpersonal affective alignment at levels above chance (McCraty, 2017). In the near future, such methods could be incorporated into behavioral cooperation tasks following joint music-making paradigms in order to assess transfer effects on socio-affective processing in non-musical contexts.

## CONCLUDING THOUGHTS

Several precautions should be taken when incorporating emotion detection techniques into scientific study of music. A general framework for ethical (e.g., intrinsic biases due to demographically limited training) and practical (e.g., overfitted algorithms) considerations for using computational techniques in social science research is covered in a review paper by Martinez (2019). Concerns specific to music psychology include the following. First, non-verbal displays of emotion in musical settings may present differently than in idealized non-musical settings (e.g., differing behavioral cues for real vs. acted emotions as in Wilting et al., 2006; overlapping basic emotions as in Juslin et al., 2011; Juslin, 2013; Akkermans et al., 2019); an affective taxonomy appropriate for the given research question should be carefully determined and cross-checked with each session of algorithmic fine-tuning. Furthermore, it is likely advantageous to limit stimuli to a particular genre or song in order to conserve behavioral cue utilization among both performers and listeners (Juslin, 2000) and de-escalate computational complexity (Lange and Frieler, 2018).

This article has summarized recent developments in music psychology and related fields that may be applied to detecting emotional contagion in music. We have discussed this research in terms of how it may be incorporated into existing experimental paradigms in scientific studies of music. We hope to encourage further findings regarding the means by which various forms of musical engagement can result in positive prosocial consequences for a broader population.

## AUTHOR CONTRIBUTIONS

IH conducted literature reviews in order to gather necessary evidence and to generate initial drafts. MBK assisted argument development and expanded source material. Both authors approved the final version of the manuscript.

## FUNDING

This research was supported by the Konrad Adenauer Stiftung for Postgraduate Research at Humboldt-Universität zu Berlin, and the Herchel Smith Fellowship for Postgraduate Research at Cambridge.

## ACKNOWLEDGMENTS

The authors would like to acknowledge Konrad Adenauer Stiftung for their financial and administrative support provided throughout the process that this article was conceived. In addition, the first author would like to thank the Institut für Musikwissenschaft und Medienwissenschaft, in particular Professor Sebastian Klotz and the members of the Transkulturelle

Musikwissenschaft Kolloquium, as well as members of the Cognitive Neurobiology Current Topics Seminar, particularly Dr. Vladislav Nachev, Dr. Katharina Stumpenhorst, and Professor York Winter, for providing crucial feedback during the development of this research. Lastly, the first author would like to thank her current mentor, Professor Ian Cross, for immensely promoting consolidation of this research through critical feedback and discussion.

## REFERENCES


Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., et al. (2019). Decoding emotions in expressive music performances: a multi-lab replication and extension study. *Cogn. Emot.* 33, 1099–1118. doi: 10.1080/02699931.2018.1541312

Anglada-Tort, M., Krause, A. E., and North, A. C. (2019). Popular music lyrics and musicians' gender over time: a computational approach. *Psychol. Music*. doi: 10.1177/0305735619871602

Aucouturier, J.-J., and Canonne, C. (2017). Musical friends and foes: the social cognition of affiliation and control in improvised interactions. *Cognition* 161, 94–108. doi: 10.1016/j.cognition.2017.01.019

Aviezer, H., Hassin, R. R., Bentin, S., and Trope, Y. (2008a). "Putting facial expressions back in context," in *First Impressions,* eds N. Ambady, and J. J. Skowronski (New York, NY: Guilford Publications), 255–286.

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., et al. (2008b). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychol. Sci.* 19, 724–732. doi: 10.1111/j.1467-9280.2008.02148.x

Azevedo, R., Taub, M., Mudrick, N., Farnsworth, J., and Martin, S. A. (2016). "Interdisciplinary research methods used to investigate emotions with advanced learning technologies," in *Methodological Advances in Research on Emotion and Education*, eds M. Zembylas, and P. A. Schutz (Cham: Springer International Publishing), 231–243. doi: 10.1007/978-3-319-29049-2_18

Batson, C. D. (2009). *These Things Called Empathy: Eight Related but Distinct Phenomena*. The MIT Press. Available online at: https://mitpress.universitypressscholarship.com/view/10.7551/mitpress/9780262012973.001.0001/upso-9780262012973-chapter-2 (accessed April 5, 2020).

Baytaş, M. A., Göksun, T., and Özcan, O. (2016). The perception of live-sequenced electronic music via hearing and sight. *Proc. Int. Conf. New Interfaces Music. Expr.* 16, 194–199. doi: 10.5281/zenodo.1175978

Baytaş, M. A., Obaid, M., La Delfa, J., Yantaç, A. E., and Fjeld, M. (2019). "Integrated apparatus for empirical studies with embodied autonomous social drones," in *1st International Workshop on Human-Drone Interaction* (Glasgow: Ecole Nationale de l'Aviation Civile [ENAC]). Available online at: https://hal.archives-ouvertes.fr/hal-02128387 (accessed February 19, 2020).

Betella, A., and Verschure, P. F. M. J. (2016). The affective slider: a digital self-assessment scale for the measurement of human emotions. *PLoS ONE* 11:e0148037. doi: 10.1371/journal.pone.0148037

Bharucha, J., Curtis, M., and Paroo, K. (2012). "Musical communication as alignment of brain states," in *Language and Music as Cognitive Systems* (New York, NY: Oxford University Press), 139–155. doi: 10.1093/acprof:oso/9780199553426.003.0016

Brown, L. S. (2017). The influence of music on facial emotion recognition in children with autism spectrum disorder and neurotypical children. *J. Music Ther.* 54, 55–79. doi: 10.1093/jmt/thw017

Burger, B., Saarikallio, S., Luck, G., Thompson, M. R., and Toiviainen, P. (2013). Relationships between perceived emotions in music and music-induced movement. *Music Percept. Interdiscip. J.* 30, 517–533. doi: 10.1525/mp.2013.30.5.517

Chang, A., Kragness, H. E., Livingstone, S. R., Bosnyak, D. J., and Trainor, L. J. (2019). Body sway reflects joint emotional expression in music ensemble performance. *Sci. Rep.* 9:205. doi: 10.1038/s41598-018-36358-4

Cirelli, L. K., Einarson, K. M., and Trainor, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants. *Dev. Sci.* 17, 1003–1011. doi: 10.1111/desc.12193

Clarke, E., DeNora, T., and Vuoskoski, J. (2015). Music, empathy and cultural understanding. *Phys. Life Rev.* 15, 61–88. doi: 10.1016/j.plrev.2015.09.001

Coan, J. A., Schaefer, H. S., and Davidson, R. J. (2006). Lending a hand: social regulation of the neural response to threat. *Psychol. Sci.* 17, 1032–1039. doi: 10.1111/j.1467-9280.2006.01832.x

Cross, I. (2005). "Music and meaning, ambiguity and evolution," in *Musical Communication*, eds D. Miell, R. MacDonald, and D. Hargreaves (Oxford, UK: Oxford University Press), 27–43.

Cross, I., Laurence, F., and Rabinowitch, T.-C. (2012). "Empathy and creativity in group musical practices: towards a concept of empathic creativity," in *The Oxford Handbook of Music Education, Vol. 2*, eds G. E. McPherson, and G. F. Welch (New York, NY: Oxford University Press), 336–353. doi: 10.1093/oxfordhb/9780199928019.013.0023

Davis, M. H. (2018). "Chapter 1: history and definitions," in *Empathy: A Social Psychological Approach,* eds G. E. McPherson, and G. F. Welch (New York, NY; Oxon: Routledge). doi: 10.4324/9780429493898-1

de Barbaro, K. (2019). Automated sensing of daily activity: a new lens into development. *Dev. Psychobiol.* 61, 444–464. doi: 10.1002/dev.21831

de Vignemont, F., and Singer, T. (2006). The empathic brain: how, when and why? *Trends Cogn. Sci.* 10, 435–441. doi: 10.1016/j.tics.2006.08.008

Decety, J., Bartal, I. B.-A., Uzefovsky, F., and Knafo-Noam, A. (2016). Empathy as a driver of prosocial behaviour: highly conserved neurobehavioural mechanisms across species. *Philos. Trans. R. Soc. B Biol. Sci.* 371. doi: 10.1098/rstb.2015.0077

DeVries, A. C., Glasper, E. R., and Detillion, C. E. (2003). Social modulation of stress responses. *Physiol. Behav.* 79, 399–407. doi: 10.1016/S0031-9384(03)00152-5

Devroop, K. (2012). The social-emotional impact of instrumental music performance on economically disadvantaged South African students. *Music Educ. Res.* 14, 407–416. doi: 10.1080/14613808.2012.685456

Eerola, T., Friberg, A., and Bresin, R. (2013). Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Front. Psychol.* 4:487. doi: 10.3389/fpsyg.2013.00487

Egermann, H. (2019). *Aesthetic Judgement and Emotional Processing in Contemporary Music*, (Cambridge).

Egermann, H., Sutherland, M. E., Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2011). Does music listening in a social context alter experience? *A* physiological and psychological perspective on emotion. *Music. Sci.* 15, 307–323. doi: 10.1177/1029864911399497

Feldman, R. (2017). The neurobiology of human attachments. *Trends Cogn. Sci.* 21, 80–99. doi: 10.1016/j.tics.2016.11.007

Greenberg, D. M., and Rentfrow, P. J. (2017). Music and big data: a new frontier. *Curr. Opin. Behav. Sci.* 18, 50–56. doi: 10.1016/j.cobeha.2017.07.007

Hovsepian, K., al'Absi, M., Ertin, E., Kamarck, T., Nakajima, M., and Kumar, S. (2015). cStress: towards a gold standard for continuous stress assessment in the mobile environment. *Proc. ACM Int. Conf. Ubiquitous Comput.* 2015, 493–504. doi: 10.1145/2750858.2807526

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: relating performance to perception. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1797–1812. doi: 10.1037/0096-1523.26.6.1797

Juslin, P. N. (2013). What does music express? Basic emotions and beyond. *Front. Psychol.* 4:596. doi: 10.3389/fpsyg.2013.00596

Juslin, P. N. (2016). "Emotional reactions to music," in *The Oxford Handbook of Music Psychology*, eds I. Cross, S. Hallam, and M. Thaut (Oxford: Oxford University Press), 197–214. doi: 10.1093/oxfordhb/9780198722946.013.17

Juslin, P. N., Liljeström, S., Laukka, P., Västfjäll, D., and Lundqvist, L.-O. (2011). Emotional reactions to music in a nationally representative sample of Swedish adults: prevalence and causal influences. *Music Sci.* 15, 174–207. doi: 10.1177/1029864911401169

Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., and Silva, A. (2008). An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion* 8, 668–683. doi: 10.1037/a0013505

King, E., and Waddington, C. (Eds.). (2017). *Music and Empathy*. Abingdon, Oxon; New York, NY: Routledge. doi: 10.4324/9781315596587

Kirschner, S., and Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evol. Hum. Behav.* 31, 354–364. doi: 10.1016/j.evolhumbehav.2010.04.004

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666

Kumano, S., Otsuka, K., Matsuda, M., and Yamato, J. (2014). Analyzing perceived empathy based on reaction time in behavioral mimicry. *IEICE Trans. Inf. Syst.* E97.D, 2008–2020. doi: 10.1587/transinf.E97.D.2008

Kumano, S., Otsuka, K., Mikami, D., Matsuda, M., and Yamato, J. (2015). Analyzing interpersonal empathy via collective impressions. *IEEE Trans. Affect. Comput.* 6, 324–336. doi: 10.1109/TAFFC.2015.2417561

Kumano, S., Otsuka, K., Mikami, D., and Yamato, J. (2011). Analyzing empathetic interactions based on the probabilistic modeling of the co-occurrence patterns of facial expressions in group meetings. *Face Gesture* 2011, 43–50. doi: 10.1109/FG.2011.5771440

Lange, E. B., and Frieler, K. (2018). Challenges and opportunities of predicting musical emotions with perceptual and automatized features. *Music Percept. Interdiscip. J.* 36, 217–242. doi: 10.1525/mp.2018.36.2.217

Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol.* 58, 259–289. doi: 10.1146/annurev.psych.58.110405.085654

Livingstone, S. R., Thompson, W. F., and Russo, F. A. (2009). Facial expressions and emotional singing: a study of perception and production with motion capture and electromyography. *Music Percept. Interdiscip. J.* 26, 475–488. doi: 10.1525/mp.2009.26.5.475

Lord, S. P., Sheng, E., Imel, Z. E., Baer, J., and Atkins, D. C. (2015). More than reflections: empathy in motivational interviewing includes language style synchrony between therapist and client. *Behav. Ther.* 46, 296–303. doi: 10.1016/j.beth.2014.11.002

Martinez, A. M. (2019). The promises and perils of automated facial action coding in studying children's emotions. *Developmental Psychology,* 55, 1965–1981. doi: 10.1037/dev0000728

Martinez, L., Falvello, V. B., Aviezer, H., and Todorov, A. (2016). Contributions of facial expressions and body language to the rapid perception of dynamic emotions. *Cogn. Emot.* 30, 939–952. doi: 10.1080/02699931.2015.1035229

McCraty, R. (2017). New frontiers in heart rate variability and social coherence research: techniques, technologies, and implications for improving group dynamics and outcomes. *Front. Public Health* 5:267. doi: 10.3389/fpubh.2017.00267

Miu, A., and Vuoskoski, J. (2017). "The social side of music listening: empathy and contagion in music-induced emotions," in *Music and Empathy*, eds E. King and C. Waddington (New York, NY: Routledge), 124–138. doi: 10.4324/9781315596587-6

Noroozi, F., Corneanu, C., Kamińska, D., Sapiński, T., Escalera, S., and Anbarjafari, G. (2018). Survey on emotional body gesture recognition. *IEEE Trans. Affect. Comput.* 9:1. doi: 10.1109/TAFFC.2018.2874986

Oswald, F. L., Behrend, T. S., Putka, D. J., and Sinar, E. (2020). Big data in industrial-organizational psychology and human resource management: forward progress for organizational research and practice. *Annu. Rev. Organ. Psychol. Organ. Behav.* 7, 505–533. doi: 10.1146/annurev-orgpsych-032117-104553

Panda, R., Malheiro, R. M., and Paiva, R. P. (2018). Novel audio features for music emotion recognition. *IEEE Trans. Affect. Comput.* 9:1. doi: 10.1109/TAFFC.2018.2820691

Pantic, M., and Vinciarelli, A. (2015). "Social signal processing," in *The Oxford Handbook of Affective Computing,* eds R. Calvo, S. D'Mello, J. Gratch, and A. Kappas (Oxford: Oxford University Press).

Prochazkova, E., and Kret, M. E. (2017). Connecting minds and sharing emotions through mimicry: a neurocognitive model of emotional contagion. *Neurosci. Biobehav. Rev.* 80, 99–114. doi: 10.1016/j.neubiorev.2017.05.013

Rabinowitch, T.-C., Cross, I., and Burnard, P. (2012). "Musical group interaction, intersubjectivity and merged subjectivity," in *Kinesthetic Empathy in Creative and Cultural Practices*, eds D. Reynolds and M. Reason (Bristol; Chicago, IL: Intellect Books), 109–120.

Rabinowitch, T.-C., Cross, I., and Burnard, P. (2013). Long-term musical group interaction has a positive influence on empathy in children. *Psychol. Music* 41, 484–498. doi: 10.1177/0305735612440609

Randall, W. M., and Rickard, N. S. (2017). Personal music listening: a model of emotional outcomes developed through mobile experience sampling. *Music Percept. Interdiscip. J.* 34, 501–514. doi: 10.1525/mp.2017.34.5.501

Reddish, P., Bulbulia, J., and Fischer, R. (2014). Does synchrony promote generalized prosociality? *Relig. Brain Behav.* 4, 3–19. doi: 10.1080/2153599X.2013.764545

Riess, H., Neporent, L., and Alda, A. (2018). "Shared mind intelligence," in *The Empathy Effect: Seven Neuroscience-Based Keys for Transforming the Way We Live, Love, Work, and Connect Across Differences* (Boulder, CO: Sounds True).

Ruth, N., and Schramm, H. (2020). Effects of prosocial lyrics and musical production elements on emotions, thoughts and behavior. *Psychology of Music.* doi: 10.1177/0305735620902534

Saarikallio, S. (2019). Access-awareness-agency (AAA) model of music-based social-emotional competence (MuSEC). *Music Sci.* 2:2059204318815421. doi: 10.1177/2059204318815421

Schäfer, K., and Eerola, T. (2020). How listening to music and engagement with other media provide a sense of belonging: an exploratory study of social surrogacy. *Psychol. Music* 48, 232–251. doi: 10.1177/03057356187 95036

Sharma, K., Castellini, C., Stulp, F., and van den Broek, E. L. (2020). Continuous, real-time emotion annotation: a novel joystick-based analysis framework. *IEEE Trans. Affect. Comput.* 11, 78–84. doi: 10.1109/TAFFC.2017.27 72882

Sloboda, J. A., and O'Neill, S. A. (2001). "Emotions in everyday listening to music," in *Music and Emotion: THEORY and Research. Series in Affective Science,* eds P. N. Juslin, and J. A. Sloboda (New York, NY: Oxford University Press), 415–429.

Sutherland, C. A. M., Rhodes, G., and Young, A. W. (2017). Facial image manipulation: a tool for investigating social perception. *Soc. Psychol. Personal. Sci.* 8, 538–551. doi: 10.1177/1948550617697176

Swarbrick, D., Bosnyak, D., Livingstone, S., Bansal, J., Marsh-Rollo, S., Woolhouse, M., et al. (2019). How live music moves us: head movement differences in audiences to live versus recorded music. *Front. Psychol.* 9:2682. doi: 10.3389/fpsyg.2018.02682

Thompson, W. F., Russo, F. A., and Livingstone, S. R. (2010). Facial expressions of singers influence perceived pitch relations. *Psychon. Bull. Rev.* 17, 317–322. doi: 10.3758/PBR.17.3.317

Thompson, W. F., Russo, F. A., and Quinto, L. (2008). Audio-visual integration of emotional cues in song. *Cogn. Emot.* 22, 1457–1470. doi: 10.1080/02699930701813974

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691. doi: 10.1017/S0140525X05000129

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., et al. (2017). Joint action: mental representations, shared information and general mechanisms for coordinating with others. *Front. Psychol.* 7:2039. doi: 10.3389/fpsyg.2016.02039

Vines, B., Krumhansl, C., Wanderley, M., Dalca, I., and Levitin, D. (2011). Music to my eyes: cross-modal interactions in the perception of emotions in musical performance. *Cognition* 118, 157–170. doi: 10.1016/j.cognition.2010. 11.010

Vogeley, K. (2017). Two social brains: neural mechanisms of intersubjectivity. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160245. doi: 10.1098/rstb.2016.0245

Vuoskoski, J. K., Gatti, E., Spence, C., and Clarke, E. F. (2016). Do visual cues intensify the emotional responses evoked by musical performance? A psychophysiological investigation. *Psychomusicol. Music Mind Brain* 26, 179–188. doi: 10.1037/pmu0000142

Vuoskoski, J. K., Thompson, M. R., Clarke, E. F., and Spence, C. (2014). Crossmodal interactions in the perception of expressivity in musical performance. *Atten. Percept. Psychophys.* 76, 591–604. doi: 10.3758/s13414-013-0582-2

Waddell, G., and Williamon, A. (2017). Eye of the beholder: stage entrance behavior and facial expression affect continuous quality ratings in music performance. *Front. Psychol.* 8:513. doi: 10.3389/fpsyg.2017.00513

Wilting, J., Krahmer, E., and Swerts, M. (2006). Real vs. acted emotional speech. *Int. J. Web Eng. Technol* 3, 805–808.

Xu, T., Shi, D., Chen, J., Li, T., Lin, P., and Ma, J. (2020). Dynamics of emotional contagion in dense pedestrian crowds. *Phys. Lett. A* 384:126080. doi: 10.1016/j.physleta.2019.126080

Yang, F., Zhao, X., Jiang, W., Gao, P., and Liu, G. (2019). Multi-method fusion of cross-subject emotion recognition based on high-dimensional EEG features. *Front. Comput. Neurosci.* 13:53. doi: 10.3389/fncom.2019.00053

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 09 December 2020
doi: 10.3389/fpsyg.2020.557398

# A Comparison of Human and Computational Melody Prediction Through Familiarity and Expertise

*Matevž Pesek[1]*, Špela Medvešek[1], Anja Podlesek[2], Marko Tkalčič[3] and Matija Marolt[1]*

[1] *Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia,* [2] *Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia,* [3] *Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Koper, Slovenia*

Melody prediction is an important aspect of music listening. The success of prediction, i.e., whether the next note played in a song is the same as the one predicted by the listener, depends on various factors. In the paper, we present two studies, where we assess how music familiarity and music expertise influence melody prediction in human listeners, and, expressed in appropriate data/algorithmic ways, computational models. To gather data on human listeners, we designed a melody prediction user study, where familiarity was controlled by two different music collections, while expertise was assessed by adapting the Music Sophistication Index instrument to Slovenian language. In the second study, we evaluated the melody prediction accuracy of computational melody prediction models. We evaluated two models, the SymCHM and the Implication-Realization model, which differ substantially in how they approach melody prediction. Our results show that both music familiarity and expertise affect the prediction accuracy of human listeners, as well as of computational models.

**Keywords: music similarity, music perception, music information retrieval, implication-realization model, compositional hierarchical model, melody prediction**

OPEN ACCESS

***Edited by:***
*Cunmei Jiang,*
*Shanghai Normal University, China*

***Reviewed by:***
*Chang Liu,*
*University of Texas at Austin, United States*
*Peter Van Kranenburg,*
*Utrecht University, Netherlands*

****Correspondence:***
*Matevž Pesek*
*matevz.pesek@fri.uni-lj.si*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology*

***Received:*** *30 April 2020*
***Accepted:*** *13 November 2020*
***Published:*** *09 December 2020*

***Citation:***
*Pesek M, Medvešek Š, Podlesek A, Tkalčič M and Marolt M (2020) A Comparison of Human and Computational Melody Prediction Through Familiarity and Expertise. Front. Psychol. 11:557398. doi: 10.3389/fpsyg.2020.557398*

## 1. INTRODUCTION

One of the main aspects of listening to music is the tendency of the brain to constantly predict the upcoming melodic events. How human listeners perform the ongoing prediction of music is influenced by (i) their general music expertise and by (ii) their familiarity with the type of music they are listening to. These two concepts are two facets of the knowledge that listeners possess. Music knowledge, in the widest sense, is acquired in various ways, either through formal music training or just by listening to music regularly. The research topic of melody prediction has been extensively studied from the perspective of psychology (e.g., Rohrmeier and Koelsch, 2012; Egermann et al., 2013) and neuroscience (e.g., Zatorre et al., 1994; Thiessen and Saffran, 2009; Lappe et al., 2013). In recent years, research on understanding human melody prediction has crossed over to the development of computational models that perform melody prediction. The knowledge that such models use, typically stems from a dataset that the researchers develop or train their models on. The question is how do models, trained on a dataset, and humans, trained through years of listening to music, compare in terms of melody prediction. We conjecture that the algorithms, just like humans, perform better at melody prediction on familiar music, i.e., on music that resembles the music they have been trained on. We also conjecture that algorithms that are agnostic of music culture perform equally well on music from different cultures. We compare melody prediction performance of the algorithms to that of humans and shed light on how much expertise and familiarity is required to train an algorithm to perform comparably well to humans.

### 1.1. Problem Formulation

The problem of melody prediction can be posed and evaluated in two ways. In a strict way, the predictor (human or algorithm) needs to predict the exact note that follows an initial set of notes of a song. In this case, there is only one possible correct prediction: i.e., the next note in that particular song. However, we can relax the requirements for prediction and treat as correct each prediction that makes sense from a musical perspective. For example, the songs *If You Don't Know Me By Now* by Harold Melvin and the Bluenotes and *Take it to the Limit* by The Eagles start with an identical sequence of notes but at some point begin to diverge[1]. If a predictor was asked to predict the first note that diverges, the correctness of the prediction would depend on whether the ground truth was the first or the second song. In this work we use both strict and relaxed evaluation—for the latter, we devise a **methodology for evaluating melody prediction based on music theory**.

Before proceeding to the next research question, we need to define the terms familiarity and expertise. They are both related to knowing music, but in ways that make them two separate concepts. The term expertise is related either to a more formal way of obtaining broad musical knowledge or intense engagement with music, whereas familiarity is related to the informal acquisition of musical knowledge. One could argue that being an expert in music makes one also familiar with (a specific type of) music. However, with different music cultures, we may encounter different expertise-familiarity scenarios. If we consider European and Chinese listeners and Chinese music, four scenarios are possible: (i) low general music expertise and low familiarity with Chinese music (e.g., a European listener with low music education), (ii) low general music expertise and high familiarity with Chinese music (e.g., a Chinese listener with low music education), (iii) high general music expertise and low familiarity with Chinese music (e.g., a European listener with high musical knowledge), and (iv) high general music expertise and high familiarity with Chinese music (e.g., a Chinese listener with high music education). While the cases (i) and (iv) are intuitive, (ii) and (iii) require more attention. In the case of (ii), the listener has not received formal music education but has been exposed to Chinese music throughout her life. The assumption is that this exposure has given the listener implicit music knowledge. This knowledge could allow her to continue correctly a melody, if it is similar to the melodic patterns (i.e., sequences of notes with similar characteristics) that she has been exposed to although she might not be able to describe why. In the case of (iii), the listener has high general musical knowledge but has not been exposed to Chinese music. In the melody continuation scenario, the listener could likely choose a tone that is correct from the musical theory perspective, but happens rarely in Chinese music. Hence, expertise and familiarity, although correlated, should be treated as two separate variables.

The examples in the previous paragraph are illustrative only and do not reflect our experimental design. In section 3, we describe in detail which familiarity categories (with European or Chinese music) and which expertise categories the subjects fall into. Furthermore, we must distinguish between the concepts of expertise and familiarity on one hand and how we operationalize them for the purpose of our study on the other hand. As described in section 3, we operationalize the expertise using the Musical Sophistication Index instrument (Müllensiefen et al., 2014). In order to operationalize the familiarity we used the cultural background, e.g., we assume Europeans are not familiar with Chinese music but are familiar with European music.

In addition to the question about the melody prediction, we also address the following question: **Do expertise and familiarity influence the melody prediction of humans?** In order to answer this question we devised an experiment where we asked subjects from a homogeneous cultural background (Slovenians with different levels of musical expertise) to predict the melody of Western (i.e., familiar) and non-Western (Chinese, i.e., non-familiar) songs. In order to measure their expertise, we **adapted the Music Sophistication Index (MSI) instrument (Müllensiefen et al., 2014) to the Slovenian language**, which is an additional contribution of this paper.

Finally, we evaluate **how algorithms perform in melody prediction**. We evaluated the performance of two algorithms, (i) the Implication-Realization (I-R) model developed by Narmour (1990), which is agnostic of musical culture, and (ii) the Compositional Hierarchical Model for symbolic music representations (SymCHM) developed by Pesek et al. (2017b), which is trained on a dataset of songs and hence biased toward familiar songs.

## 2. RELATED WORK

In this section, we first survey work on human music prediction from psychology and neuroscience, which demonstrates that familiarity and expertise influence melody prediction, supporting our rationale for the work presented. We then proceed on surveying algorithms for melody prediction, in particular how they use existing knowledge and what methodologies have been proposed for melody prediction. Lastly, we survey work on tonal hierarchies that will be used as the theoretical background for constructing the relaxed evaluation criterion.

### 2.1. Culturally-Dependent Human Melody Prediction

Music from different cultures activates similar brain areas, as shown by Morrison and others in their functional magnetic resonance imaging experiment (Morrison et al., 2003), but there are differences in brain activity between the situations in which one listens to the music of their own cultural background vs. that of a foreign culture. The cultural influence has also been demonstrated for genre preference (Soley and Hannon, 2010),

[1]The reader may find a lot of similar sets of songs, for example The Kinks: All Day And All Of The Night vs. The Doors: Hello, I Love You or the triple Mike Oldfield: Tubular Bells vs. Death Angel: The Ultra-Violence vs. Possessed: The Exorcist

perceived mood estimation (Balkwill and Thompson, 1999), and musical memory (Demorest et al., 2008), among other aspects. Children already show preferences for music similar to their cultural and musical tradition (Demorest et al., 2008; Soley and Hannon, 2010). The preference also affects musical memory, which is better for culturally similar music in comparison to unknown music (Demorest et al., 2008). Also the recognition of rhythm is influenced by culture, which is why foreign rhythms seem more complex and are more difficult to recognize (Cross, 2001).

The cognitive processes which occur in music perception are also employed in language perception, which is supported by the research performed by Nan et al. (2006) and Maess et al. (2001). In the latter, magnetoencephalography was used to analyze the Broca's area in the human brain, which plays an important role in syntactic analysis during auditory language comprehension. The authors showed it also analyzes the incoming harmonic sequences. Nan et al. (2006) further explored the shared functionality of this area and analyzed the structure of a musical phrase. They discovered that the brain response produced by musical phrase inconsistencies are similar to those produced when observing syntactic inconsistencies. Specifically, the *closure positive shift* (CPS), which is a universal phrase perception mechanism and occurs in both music and language domains, was observed. In music, the CPS occurred between 100 and 450 ms after the phrase end. The CPS occurred earlier when the participants listened to music of their culture and later when listening to music of a foreign culture (Nan et al., 2006).

Mood and emotion recognition in music depend on culture-specific and universal structural characteristics of the music played. Balkwill and Thompson (1999) explored whether Western listeners recognize the *intended* emotions in music in the Hindi tonal system, which was unknown to the listeners. The results showed that the listeners are able to recognize basic emotions, e.g., joy, sadness, and anger, however, they are unable to identify more complex emotions which may be different from culture to culture, e.g., "peace" (Balkwill and Thompson, 1999). Fritz et al. (2009) also carried out a survey in the reverse direction and reached a similar conclusion: in Western music, the expression of basic emotions is universally recognizable, as well.

Several studies analyze the cultural influence on rhythmic and melodic complexity (Eerola et al., 2006; de Fleurian et al., 2017). Groups of participants with different cultural backgrounds (African, Western) perceived music of their own culture as less complex. Interestingly, African participants perceived Western music as less complex, compared to the perception of African music by European participants. The perception of Western music as less complex is most likely the consequence of the prevalence of Western culture. People are acquainted with Western music, regardless of their culture. This influence of Western music also results in the disappearance of less robust cultures or, in less radical cases, the infiltration of Western musical foundations into other musical cultures (Huron, 2008).

The topic of musical expectation is also important in both perceptual and cultural contexts. One of the biggest factors accrediting to the differences in music expectations is the culture (Castellano et al., 1984; Krumhansl, 1999; Krumhansl et al., 2000). The studies performed by Castellano et al. (1984) and Krumhansl (1999), Krumhansl et al. (2000) took into consideration two groups of participants—one from a Western culture (European or American) and one from a non-Western culture (African or Asian). The task in these experiments was to predict the continuation of musical excerpts (i.e., a subsequence of notes from a longer sequence). Most commonly, the participants from different cultures have decided very similarly with regard to the continuation of their own and foreign cultures. In one of the studies by Krumhansl (1999), the non-European participants achieved higher accuracy for European music, for the previously known European classical works. Consequently, their responses were more in line with the music theory, compared to the European group, who supposedly possessed more general knowledge of Western music.

Listeners also provide diverse responses to music structure of unknown music styles. They can quickly adapt to a different style and are able to adjust their expectations after a short exposure to an unknown style. There are currently two different prevailing explanations for such behavior: (1) when listening to music, we continuously learn the characteristics of the style as we pay attention to the stylistic tendencies in music and (2) there are basic psychological principles or universal qualities of music that can be applied to various music styles (Krumhansl et al., 2000; Huron, 2008).

The aforementioned research shows the interconnection between music perception and culture, focusing on comparing the perception of different music features: from "low-level" rhythmic and melodic complexity, to "high-level" features, such as mood and emotion. Through exploration of perception between different cultures, users' familiarity with culture-influenced music style was evaluated and analyzed. However, the users' expertise, which could affect perception, was seldom explored in these studies. In this paper, we explore both the effect of expertise and familiarity, and their influence on human perception of music by observing user responses in a melody prediction task.

## 2.2. Computational Melody Prediction

In music theory and information retrieval, music patterns and their repetitions have been studied by a number of groups, both in theory (e.g., Lerdahl and Jackendoff, 1983; Margulis, 2013) and computational approaches (e.g., Marsden, 2010; Ren, 2016; Pesek et al., 2017b). Several tasks emerged throughout the years in the Music Information Retrieval Evaluation eXchange (MIREX), which is a community-based framework for formal evaluation of algorithms and techniques related to music information retrieval (MIR) (Downie, 2008). Among a variety of available MIREX tasks, the *Discovery of repeated themes & sections* (Collins et al., 2014) and *Patterns for prediction* became popular in the last decade. The aim of the *Discovery of repeated themes & sections* task is to find repetitions, which represent one of the more significant aspects of a music piece (Meredith et al., 2002). The MIREX task definition states "the algorithms take a piece of music as input, and output a list of patterns repeated within that piece" (Collins, 2016). Based on this discovery task, the *Patterns for prediction* task was created as an offshoot of the pattern discovery task (Janssen et al., 2019). The goal of this task is to predict a continuation of a given excerpt. There are two subtasks defined within the *Patterns for prediction* task: in the *explicit task*, the algorithm is given an excerpt and should generate a continuation; in the *implicit task*, the algorithm is given an excerpt and a candidate continuation, and should return the probability that this is the true continuation of the provided prime.

Several authors tacked the *Pattern for prediction* subtasks in recent years. de Reuse (2019) proposed a CopyForward algorithm for the explicit task, inspired by the Cosiatec approach by Meredith (2013). The CopyForward algorithm selects a section of the excerpt, and translates and copies it as a continuation. Colombo (2018) also proposed an approach named BachProp for the explicit task, which uses a recurrent neural network to determine note probabilities for the continuation. Ycart and Benetos (2019) proposed an LSTM model for the implicit task, which discriminates the real and fake continuations, while Ens and Pasquier (2019) proposed the GenDetect algorithm, which generates a collection of categorical distributions for each music excerpt and the uses a Gradient Boosting Classifier, to predict whether a continuation is real or fake.

We proceed by describing two computational models for modeling musical expectations that we used in our study: the Implication-Realization (I-R) model by Narmour (1990) and the Compositional Hierarchical Model for symbolic music representations (SymCHM) by Pesek et al. (2017b). The I-R model represents one of the most well known models of melodic expectation. It is based on musicological rules, so we deem it has high expertise and is culturally agnostic. As an alternative, we chose the SymCHM model, which is based on compositional modeling through unsupervised learning. It thus deduces the underlying rules from the music itself and does not incorporate any direct musicological knowledge in its structure. The two models represent two opposite poles and well-fit our goal of exploring the differences between expertise and familiarity in prediction.

#### 2.2.1. Implication-Realization (I-R) Model

Eugene Narmour studied Leonard Meyer's theory of musical expectation, based on the understanding of musical structure and the perception of musical emotions and meaning. He then developed a complex theory of melodic perception, which he called the Implication-Realization Model (Narmour, 1990).

The model considers *implicative* intervals, by which one forms an expectation about the continuation of the melody, and the *realized* intervals, which (presumably) fulfill these expectations (Toiviainen and Eerola, 2016). The model therefore observes the perceptive systems which process information in a *top-down* manner, as well as those which process information in a *bottom-up* manner. While the latter approach represents the expectations, the former mimics the realization of the representations, which are learned and depend on the musical knowledge and culture of an individual (Pearce, 2003).

The I-R model contains five criteria on the basis of which the suitability of the realized interval is estimated:

- registral direction: implicit intervals larger than 8 semitones imply a change in the direction of the melody, and those smaller imply preservation of the direction,
- registral return: it prefers returns to the first tone of the implicit interval or a deviation from the latter by a maximum of 2 semitones in any direction,
- intervallic difference: Implicit intervals of 5 or more semitones imply similarly large realized intervals (with a deviation of up to 2 semitones in any direction when changing the direction of the melody, or three semitones in the direction of conservation); while the implicit intervals which are $>5$ semitones imply smaller realized intervals,
- proximity: this criterion prefers the realized intervals of five semitones or smaller,
- closure: implies a change in the direction of the melody or a smaller realized interval than the implicit interval, if the latter was large (at least 3 semitones larger than the former).

The model prefers small realized intervals and preserves the direction of the melody or stays on the same tone, and in the case of larger realized intervals, a change in the direction of the melody.

#### 2.2.2. Updated I-R Model

The five basic criteria of Narmour's model were later updated by additional criteria by different researchers, most notably Schellenberg (1997), who reduced the number of factors. These criteria are not all based on the realization of implicit intervals. The added criteria are:

- consonance: the preferred realized intervals are the consonant intervals: unison, perfect forth, perfect fifth, and octave (Krumhansl, 1995),
- tonality: tonally more stable tones are preferred (Krumhansl, 1999),
- melodic attraction: the tonal ratio of tonality of both tones in the realized interval (Lerdahl, 1996),
- tessitura: predictions of tones that are close to the median pitch of the melody (Hippel, 2000),
- mobility: on the basis of the auto-correlation between consecutive tones, this criterion estimates how predictive the individual tone is in terms of previous tones and the median position (Hippel, 2000).

In the experiment, we used the bottom-up I-R model's implementation provided by Toiviainen and Eerola (2016) to compute the values for the selected criteria.

#### 2.2.3. Compositional Hierarchical Model for Symbolic Music Representations (SymCHM)

In recent years, deep architectures based on neural networks have become prominent in the field of machine learning and pattern recognition. Such architectures have the ability to learn and model characteristics of the underlying data on multiple levels of abstraction, with simple structures being modeled on low levels and more complex concepts on higher layers.

The current implementations of neural network-based deep architectures need large amounts of data for training, and are thus less usable in cases where the amount of data is limited. Moreover, these approaches operate as black boxes, as insight into the learned structures is difficult and the ability to use these models for observation and analysis is therefore limited.

The Compositional Hierarchical Model for music information retrieval was first introduced by Pesek et al. (2014) as a deep architecture that can learn to model the input data on multiple layers with increasing levels of complexity, but also has the ability to learn from small datasets, and enables insight into the learned structures, including automated chord estimation and multiple fundamental frequency estimation (Pesek et al., 2017a). The model was built on a time-frequency-magnitude input and produced frequently co-occurring compositions of harmonic structures in the input signal. The symbolic version of the model SymCHM (Pesek et al., 2017b) has been applied to the task of finding repeated melodic patterns and sections by learning for compositions of symbolic events in the time-pitch-onset domain. The model has also recently been applied to the extraction of rhythmic patterns (Pesek et al., 2020). The SymCHM learns a hierarchical representation of patterns occurring in the input, where patterns encoded by the parts on higher layers are compositions of the patterns on lower layers. "Part activations" expose the learned patterns (and their variations) in the input data. Shorter and more trivial patterns naturally occur more frequently, longer patterns less frequently. On the other hand, longer patterns may entirely subsume shorter patterns.

The motivation for such model originates in the idea of decomposition of complex signals into simpler parts of the signal. These parts possess various levels of granularity and can be distributed across several layers, depending on their complexity. Starting from the first layer, which contains parts representing individual events (such as a frequency presence, or a note event in symbolic representation), new consecutive layers of frequently co-occurring compositions are created. The consecutive layers thus contain compositions of parts on previous layers. Since statistics is employed as the driving force behind the training procedure of such model, the structure can be learned in an unsupervised manner.

The learned hierarchy is also transparent. By observing the learned concepts encoded in individual compositions, the structure can be transparently observed without a specialized process, which is needed in black-box architectures. The learned concepts in the compositional hierarchy are encoded relatively. Again, originating from the idea of signal decomposition, parts with the same structure occur in different signals, as well as at multiple time- and pitch/frequency-shifted locations in a single signal. By relatively encoding the structures, a single composition is activated at multiple locations of the learned concept occurring in the input signal. The activations therefore represent the instantiations of the relatively encoded concepts.

#### *2.2.3.1. Training and using the SymCHM*

Starting at its input, the model observes the event co-occurrence frequencies and relatively models the relations between them. In terms of melodic sequences, if two or more events co-occur on a specific interval in several locations, both events can be joined into a composition. The latter represents a newly composed part on a consecutive layer. The composition is relatively encoded, meaning that should two events co-occur at one pitch location and again at a different one, the same composition is formed. This procedure is repeated layer-by-layer until the desired complexity of the learned parts is achieved. In contrast to the first layer where the model observes individual events in the input, co-occurrences of compositions are observed on higher layers. These then form new relatively encoded compositions on a consecutive layer, based on the previous layer.

Each part may occur at several locations in the input. Since the part is relatively encoded, the occurrences are defined by temporal placement and pitch attributes. The occurrence of a part in the input is denoted "activation," which contains the information about the time and pitch of the occurrence. The parts learned by the model can be observed as melodic patterns and their activations as pattern occurrences.

Once the model is built, it can be inferred over another (or over the original input). The inference may be exact or approximate, where in the latter case biologically-inspired hallucination and inhibition mechanisms enable the model to find variants of part occurrences with deletions, changes, or insertions, thus increasing its predictive power and robustness. The hallucination mechanism provides means to activate a part even when the input is incomplete or changed. In symbolic music representations, such changes often occur in melodic variations and ornamentation. The hallucination enables the model to robustly identify patterns with variations. The inhibition mechanism is also essential in SymCHM for the removal of redundant co-occurrences. As the model does not rely on any musicological rules, parts may produce a large number of competing patterns. Inhibition may be used to reduce the number of activations and find the patterns that best correspond to the learned hierarchy.

SymCHM therefore learns a hierarchical representation of patterns occurring in the input, where patterns encoded by the parts on higher layers are compositions of the patterns on lower layers. The inference produces "part activations" which expose the learned patterns (and their variations) in the input data. Shorter and more trivial patterns naturally occur more frequently, longer patterns less frequently. On the other hand, longer patterns may entirely subsume shorter patterns.

## 2.3. Tonal Hierarchies

The theory of tonal hierarchies by Krumhansl and Cuddy (2010) is based on the assumption that statistically frequent musical patterns (in most cases) provide reliable guidelines for the listener's abstraction of the tone hierarchy. The listeners should therefore successfully orient themselves to the actual tonal hierarchy. Their perception should also coincide with the frequencies of the occurrence of tones and their combinations.

Musical context establishes the tonal hierarchy. Certain tones are more specific, more stable and more important to the structure than others. In classical Western tonal-harmonic music of the eighteenth and nineteenth centuries, tonic is the main tone in the tonal hierarchy, followed by the dominant, dominant parallels, the remaining tones of the scale, and lastly the tones that are not part of the scale. This hierarchy reflects the influence of a

**TABLE 1 |** The major and minor tonal hierarchies, obtained by the probe tone method.

| Note | C | C# | D | D# | E | F | F# | G | G# | A | A# | B |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C-major | 6.35 | 2.23 | 3.48 | 2.33 | 4.38 | 4.09 | 2.52 | 5.19 | 2.39 | 3.66 | 2.29 | 2.88 |
| C-minor | 6.33 | 2.68 | 3.52 | 5.38 | 2.60 | 3.53 | 2.54 | 4.78 | 3.98 | 2.69 | 3.34 | 3.17 |

triadic (acordic) structure in which consonant chords dominate. Krumhansl and Cuddy (2010) used the *probe tone* method to quantify tone hierarchies. The participants were asked to listen to incomplete scales and evaluated how well the individual tones completed the scale. The results are shown in **Table 1**.

Tones higher in the tonal hierarchy appear more often and last longer and in the stressed metric positions (Krumhansl, 1999). Moreover, the higher their position in the hierarchy, the more quickly these tones are recognized as part of the scale (Janata and Reisberg, 1988).

In addition to the musical reference points that lead the musical perception, musical memory and understanding, the listeners are also sensitive to frequently occurring sound sequences (Saffran et al., 1999; Saffran and Griepentrog, 2001). By repeated implicit listening, they develop mental representations that reflect musical consistency, through which they then encrypt and memorize musical patterns, and generate expectations. Sensitivity to these consistencies allows for a relatively quick adaptation to new musical styles. The concept that one central tone is a reference point for a multitude of hierarchically connected tones is not limited to the Western tonic-harmonic style, but also to other styles and cultures. Unique hierarchies can even be found within individual songs.

Western listeners quickly adapted to the tonal hierarchies of an unknown (Indian) style in an experiment performed by Castellano et al. (1984). It turned out that the more important tones were played many times, which allows the listeners who are not familiar with the style to find the appropriate tonal hierarchy. Even the inexperienced listeners are flexible and adapt quickly to tone sequences in unknown musical contexts, while for the musically educated participants, the statistical processing of music becomes even more evident (Oram and Cuddy, 1995).

## 3. METHODOLOGY

In this paper, we present two studies on how melody prediction is affected by music familiarity and expertise in (1) human listeners and (2) computational approaches. In this section, we first describe two datasets of music excerpts we collected and used in both studies. We also describe the evaluation metrics used to assess prediction accuracy, and the translation and validation of the Music Sophistication Index, used as an instrument to assess human music expertise. Finally, we describe how we used the SymCHM by adjusting it for the melody prediction task.

### 3.1. Datasets

To control for music familiarity, we collected two datasets of folk songs to be used in our studies: Chinese and European. The Chinese music excerpts (used with permission of the authors) are part of the music database used in the study of the perception of musical phrases by Nan et al. (2006). The European sections were taken from the freely available online collection Robokopp[2], which contains folk and war songs and anthems from German and English speaking environments. Since we conducted the first study on Slovenian participants, we limited the selection to German songs, due the German influence on Slovenian folk songs and the Slovenian musical heritage (Vodušek, 1959; Vidakovič and Delo, 2003).

#### 3.1.1. Generating the Excerpts

We initially randomly selected 30 non-polyphonic musical fragments in each of the datasets. We converted the MIDI song representations into audio using the Midi Sheet Music and MuseScore 2 programs. We re-synthesized all of the MIDI excerpts in order to avoid variations in sound and timbre quality between both collections.

As the average length of the songs was 18.7 s (about 8 bars), we created shorter song excerpts, representing individual phrases within the songs, to make the music prediction task more user-friendly. An example of a full song is shown in **Figure 1A**, while its shortened excerpt is depicted in **Figure 1B**.

We shortened the songs in two different ways: some of the songs were cut after the penultimate tone in the phrase (complete-phrase excerpts—the participants had to predict the last tone in the phrase), while others were cut at random (incomplete-phrase excerpts). The dataset contained 75% of the complete-phrase excerpts and 25% of the incomplete-phrase excerpts.

From the initial 30, we chose 20 excerpts per dataset based on the following criteria: (1) the number of tones in the excerpt, (2) the maximum interval occurring in the excerpt, and (3) the tonal range of the excerpt.

In order to ensure a homogeneous structure of musical events, we selected fragments that were within two standard deviations of each criterion across the dataset. The values for the four excerpts described in **Figure 1** are given in **Table 2**.

#### 3.1.2. Analysis of the Generated Datasets

We used one-way ANOVA to compare the chosen criteria of the musical excerpts between the two datasets. The differences between the datasets were not statistically significant (**Table 3**).

The musical excerpts contained 15 events on average (Chinese: $\mu = 15.6$, $\sigma = 3.3$, German: $\mu = 13.9$, $\sigma = 4.3$). The lowest number of events in the Chinese dataset was 10, while the highest was 22. In the German dataset, the lowest number of the events was 7, while the highest was 24.

[2]http://www.musicanet.org/robokopp/Volksong.html (accessed February 15, 2018)


**FIGURE 1 |** **(A)** A full Chinese folk song; **(B)** A Chinese song excerpt; complete (excerpt 1); **(C)** A Chinese song excerpt, incomplete (excerpt 2); **(D)** A European song excerpt, complete (excerpt 3); **(E)** A European song excerpt, incomplete (excerpt 4).

**TABLE 2 |** Values of the three criteria for the music excerpts b–e, depicted in **Figure 1**.

| Excerpt | Number of events | Largest interval | Range in semitones |
|---|---|---|---|
| b | 13 | VIII (octave) | 12 |
| c | 17 | VI (sixth) | 17 |
| d | 15 | VI (sixth) | 12 |
| e | 12 | IV (fourth) | 11 |

**TABLE 3 |** ANOVA comparison of the German and Chinese datasets.

| Criterion | *MS* | $F_{(1,38)}$ | *p* | $\omega^2$ |
|---|---|---|---|---|
| Number of events | 28.9 | 1.88 | 0.179 | 0.021 |
| Largest interval | 0.1 | 0.04 | 0.845 | −0.025 |
| Tonal range | 6.4 | 0.54 | 0.469 | −0.012 |
| Duration in seconds | 4.2 | 1.04 | 0.315 | 0.001 |

$1 - \beta = 0.11$ *at* $\alpha = 0.05$.

The largest interval in the majority of songs was the major sixth. Across the Chinese dataset, the largest intervals ranged from a minor third to the octave. On the European dataset, the songs' largest interval ranged from fourth to the tenth (decima). The datasets differed only slightly in the tonal range. The average range of the songs was 12 semitones (Chinese: $\mu = 12.4, \sigma = 3.8$; German: $\mu = 11.6, \sigma = 2.8$). In the Chinese dataset, the smallest range was five semitones and the largest range was 20 semitones. In the German dataset, and the smallest range was 7 semitones and the largest 17.

## 3.2. Evaluation Metrics

As mentioned earlier in the problem formulation, a sequence of melodic events can have different continuations that make sense from a music theory perspective. This is demonstrated in different songs that share a part of the melody, which at some point diverge. Hence, a predicted tone, which makes musical sense, but is not the exactly the same as in the original melody, should be considered as correctly predicted.

In our experiment, we devised two evaluation metrics: strict evaluation, which considers only the correct note, and relaxed evaluation, where a prediction, which is part of the tonal hierarchy (scale) of the music excerpt, is taken as correct. Thus, for excerpts in the European dataset in a major scale, seven tones are correct (see **Figure 2**), while for those in a minor scale, nine tones are correct (including the augmented sixth and seventh scale degrees appearing in the harmonic and melodic scales). For the excerpts in the Chinese dataset, the five suitable continuation notes are based on the pentatonic scale (**Figure 3**).

## 3.3. Slovenian Translation of the Music Sophistication Index

In order to measure the music expertise of the subjects, we used the Goldsmiths Music Sophistication Index (Gold-MSI) instrument (Müllensiefen et al., 2014). The Gold-MSI is a questionnaire with 38 items that measure various aspects of music sophistication. It was developed for English speaking subjects. Because the subjects in our study were Slovenian-speaking and the instrument was not available in Slovenian, we needed to adapt it.

When a questionnaire is adapted to another cultural environment it has to be validated (Sousa and Rojjanasrirat,


**FIGURE 2 |** A list of suitable octave-invariant responses for the selected excerpt from the European dataset in C-major.


**FIGURE 3 |** A list of suitable octave-invariant responses for the selected excerpt from the Chinese dataset in the pentatonic scale starting in B♭.

2010). The intercultural differences can affect the validity of the participant's responses in a questionnaire in a situation where non-native speakers are asked to respond, even if they possessed good knowledge of the language (Blažica and Lewis, 2015). Moreover, the differences in the participants' responses can occur even when using the same questionnaire in two cultures with a single language, but with completely different cultures (e.g., the USA and New Zealand) (Brown et al., 2017). Therefore, in addition to translation, adaptation and the assessment of validity and reliability (test–retest reliability and internal consistency) are obligatory for the adaptation of an instrument for a new cultural environment (Arafat et al., 2016).

### 3.3.1. Translation and Implementation

The Gold-MSI questionnaire contains 38 self-report items about musical engagement and education. The first 31 items contain statements and the individuals assess on a 7-point Likert-type scale how much they agree with each of the statements, while the last seven items ask about their music education, the number of instruments played by the individual and similar.

The questionnaire was translated into Slovenian with multiple quality checks, as suggested by Sousa and Rojjanasrirat (2010). Two translators independently translated the questionnaire and the third translator reviewed the translation. We combined both translations into the first Slovenian version. This version was back-translated by the fourth (independent) translator into English. Finally, we compared back-translation with the original English text and implemented minor changes to the first version. The final Slovenian translation was implemented as an online questionnaire, using the PHP framework CodeIgniter. In addition to the Gold-MSI questionnaire, four short demographic questions (gender, age, education, status) were asked. The data were processed using the statistical analysis tool R.

### 3.3.2. Participants

The questionnaire was completed by 231 people (79 men, 152 women). The participants were mostly students (136) and employees (75), aged between 16 and 58 ($\mu = 26.7$, $\sigma = 7.3$). Almost all participants (96.5%) had education higher than secondary school: 83 had undergraduate degree (3-year programme equivalent to the first Bologna cycle), 86 completed the graduate programme (second Bologna cycle), and 54 the postgraduate programme (third bologna cycle).

The vast majority of the participants had at least a few years of music education. Only 66 participants never attended a music school (28.6%). The majority ($f = 139$, 60.2%) were enrolled in some formal type of education for at least three years, of which 44 participants had 10 or more years of music education. One third of the participants ($f = 75$) never learned about music theory, whereas 128 participants (55.4%) were trained in this field for three years or more (of which 41 participants had more than six years of experience). In Slovenian music-school system, the elementary music education most commonly takes 6 years and involves both learning a music instrument and music theory courses. Only 38 participants (16%) had no music experience at all—they never attended a music school, nor did they learn to play any instruments by themselves.

Most of the participants answered that they a attended music school for singing ($f = 47$), followed by piano ($f = 43$), guitar ($f = 33$), flute ($f = 15$), and violin ($f = 11$). The other instruments had a frequency of 3 or less, and 65 participants (28.1%) did not play any instrument.

### 3.3.3. Confirmatory Factor Analysis

First, we compared the characteristics of the translated and original version of Gold-MSI. We checked the gathered responses in terms of seven parameters—average values, dispersion (minimal and maximal values, standard deviations), Cronbach's $\alpha$, McDonald's $\omega$, and Guttman's $\lambda 6$ for five different categories reported by Müllensiefen et al. (2014):

- active engagement—A,
- perceptual abilities—P,
- musical training—M,
- singing abilities—S,
- emotions—E.

Each of the 31 questions in the questionnaire belongs to one of these five categories: A and P contain 9 questions each, M and S contain 7, and E contains 6. We also considered the general sophistication factor (GEN), which includes 18 of the 31 questions.

**TABLE 4 |** Comparison of average values, dispersion, and reliability measures between the research performed by Müllensiefen et al. (2014) (EN; $n = 147.633$) and our research (SL; $n = 231$) for five specific factors and the general factor of musical sophistication.

| | Active engagement (A) | | Perceptual abilities (P) | | Musical training (M) | | Singing abilities (S) | | Emotions (E) | | General sophistication (GEN) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | EN | SL | EN | SL | EN | SL | EN | SL | EN | SL | EN | SL |
| *M* | 41.52 | 37.67 | 50.20 | 50.85 | 26.52 | 28.65 | 31.67 | 32.22 | 34.66 | 34.77 | 81.58 | 82.16 |
| *SD* | 10.36 | 11.69 | 7.86 | 9.67 | 11.44 | 12.20 | 8.72 | 10.18 | 5.04 | 5.83 | 20.62 | 23.93 |
| Max | 63 | 62 | 63 | 63 | 49 | 48 | 49 | 49 | 42 | 42 | 126 | 124 |
| Min | 9 | 9 | 9 | 22 | 7 | 7 | 7 | 8 | 6 | 14 | 18 | 22 |
| $\alpha$ | 0.87 | 0.87 | 0.87 | 0.88 | 0.90 | 0.92 | 0.87 | 0.88 | 0.79 | 0.76 | 0.93 | 0.94 |
| $\omega$ | 0.87 | 0.88 | 0.87 | 0.89 | 0.90 | 0.92 | 0.87 | 0.88 | 0.79 | 0.77 | 0.93 | 0.94 |
| G6 | 0.86 | 0.87 | 0.87 | 0.88 | 0.91 | 0.93 | 0.87 | 0.88 | 0.77 | 0.74 | 0.94 | 0.96 |

**TABLE 5 |** Results of one-way *t*-test for average values of individual factors.

| Factor | $t_{(230)}$ | *p* | *d* |
|---|---|---|---|
| Active engagement | −5.01 | < 0.001 | −0.661 |
| Perceptual abilities | 1.02 | 0.309 | 0.135 |
| Musical training | 2.65 | 0.009 | 0.349 |
| Singing abilities | 0.82 | 0.415 | 0.108 |
| Emotions | 0.29 | 0.774 | 0.038 |
| General sophistication | 0.37 | 0.713 | 0.049 |

We cross-examined the values reported in the initial research and compared them to the results we had obtained. The cross-examination is shown in **Table 4**. The reliability measures indicate good internal consistency in all factors, both for the English and the Slovenian versions.

A one-way *t*-test was used to compare the results of the Slovenian sample on all five specific factors of musical sophistication and the general factor with the average for each factor obtained in the original survey. The values were statistically significant ($p < 0.05$) for the active engagement factor and the musical training factor (**Table 5**), indicating the presence of the differences between the Slovenian and original sample in these types of musical sophistication, measurement non-invariance (e.g., differences in interpretation of values on the Likert scale) and similar.

We also performed a confirmatory analysis of the Slovenian questionnaire, as the authors of the initial research report a poor fit to the one-factor model (Müllensiefen et al., 2014). Confirmatory factor analysis of the one-factor model of the Slovenian questionnaire has shown that the data did not fit well, $\chi^2(665)$=2901, $p < 0.001$; CFI = 1.00; TLI = 1.00; RMSEA = 0.150, 90% CI = [0.146, 0.154]; SRMR = 0.128. The fit was even worse with the 5-factor model (the same factors were used as in the original survey), $\chi^2(655)$=9690, $p < 0.001$; CFI = 0.328; TLI = 0.278; RMSEA = 0.244, 90% CI = [0.240, 0.249]; SRMR = 0.085, which was one of the reasons why we had to re-construct the questionnaire.

### 3.3.4. Exploratory Factor Analysis

The aim of the exploratory analysis was to reduce the number of items that would still give sufficient information about the music sophistication of the participants. Since the experimental task in our research was quite long, we intended to shorten the Gold-MSI questionnaire significantly.

We extracted one general factor with the eigenvalue of 13.1. We then selected the items with absolute factor loading greater than 0.70; the questionnaire was thus reduced to eight items (its length was reduced by 79%).

Based on these eight items, we created a new index of musical sophistication (**Table 6**), which coincides well with the general sophistication index of Gold-MSI ($r = 0.95$). The correlation was calculated on the basis of the score obtained by adding the weighted values of individual items (the number of points for the individual index of musical sophistication was the sum of the items no. 5, 7, 10, 12, 19, 22, 27, and 32, weighted by their loading on the extracted factor; Gold-MSI was also calculated using the same procedure). The final version of the short Slovenian Gold-MSI contained items on the ability to make judgments about good singing, abilities to play music/sing by heart, sing proper notes, compare two versions of the same song, recognize specifics of the music piece and detect wrong notes, as well as items on identifying with being a musician and the amount of practicing an instrument.

## 3.4. Adapting the SymCHM for Melody Prediction

The SymCHM model was initially designed for pattern discovery in symbolic music data, so we adapted it for the melody prediction task. In the following subsection we describe the adaptation of the model used in the experiment.

The implementation of SymCHM works with a comma-separated-values (.csv) input, commonly used in the MIREX pattern discovery task. We therefore transformed the MIDI files from the dataset to the desired CSV files as follows. A song is represented by a single CSV file. Each line in the CSV file contains the following three elements $\{T_o, P_1, D\}$, where individual variables represent the following features:

**TABLE 6 |** Selected items with the highest loading.

| No. | Item | Factor loading |
|---|---|---|
| 5 | Dobro znam presoditi, ali je nekdo dober ali slab pevec. (*I am able to judge whether someone is a good singer or not*) | 0.734 |
| 7 | Na pamet lahko pojem ali igram skladbe. (*I can sing or play music from memory*) | 0.753 |
| 10 | Ob spremljavi glasbenega posnetka sem sposoben zapeti prave note. (*I am able to hit the right notes when I sing along with a recording*) | 0.794 |
| 12 | Zmožen sem primerjati in razpravljati o razlikah med dvema izvedbama ali različicama iste pesmi. (*I can compare and discuss differences between two performances or versions of the same piece of music*) | 0.795 |
| 19 | Zmožen sem prepoznati posebnosti poslušane skladbe. (*I am able to identify what is special about a given musical piece*) | 0.802 |
| 22 | Opazim, kadar nekdo poje ali igra napačne tone. (*I can tell when people sing or play out of tune*) | 0.724 |
| 27 | Ne bi rekel, da sem glasbenik. (*I would not consider myself a musician*) | 0.746 |
| 32 | Koliko časa ste redno, dnevno vadili glasbeni inštrument? (*I engaged in regular, daily practice of a musical instrument (including voice) for X years*) | |

*The first column corresponds to the item number in the original Gold-MSI questionnaire. The second column describes the individual items in Slovenian.*

**TABLE 7 |** The procedure of pattern matching and weight calculation.

| Position | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MIDI Pitch | 64 | 62 | 64 | 62 | 64 | 62 | 71 | 71 | 71 | 72 | 71 | 69 | 74 | 74 | 72 | ? |
| Segment | 0 | b | 0 | b | 0 | b | G | G | G | H | G | E | J | J | H | * |
| Pattern 1 | 0 | | 0 | | 0 | | | | | | | | | | | **D** |
| Segment | | | | | | | 0 | 0 | 0 | A | 0 | b | C | C | A | * |
| Pattern 2 | | | | | | | 0 | 0 | 0 | | 0 | b | C | C | | **A** |
| Segment | | | | | | | | | | | | 0 | E | E | C | * |
| Pattern 3 | | | | | | | | | | | | 0 | E | E | | **L** |

*Letters represent relatively encoded pitch values.*

- $T_o$: onset time
- $P_1$: pitch
- $D$: duration time.

As SymCHM performs pattern matching, we needed to convert the discovered patterns, represented by the part structures in SymCHM, into melody predictions. We thus perform prediction by searching for occurrences of the patterns, which the model learned during training. We search for individual learned patterns in a given excerpt, and find the continuation of the melody from the best fitting patterns.

To perform this search, we converted the found patterns into regular expressions of relatively encoded patterns with gaps. These gaps represent differences between the learned pattern and the identified pattern. For example, a relatively encoded pattern $\{0, 5, -3, 0\}$ was transformed into the following regular expression: $\rightarrow$ $\{0[0a - zA - Z] * E[0a - zA - Z] * c[0a - zA - Z] * 0\}$. The pattern represented a melodic structure in which the second event occurs five semitones above the first, third event occurs three semitones below the first event, and the fourth event occurs at the same position as the first event. The positive semitone offsets were encoded into upper-case letters (e.g., $5 \rightarrow E$), and the negative offsets into lower-case letters (e.g., $-3 \rightarrow c$). The $[0a - zA - Z]*$ segments (i.e., sequences of notes) represented gaps of indefinite length. These gaps allowed the discovered patterns to match with potential variations in the patterns.

For each excerpt, we searched for the learned patterns of the SymCHM model, encoded as regular expressions. By excluding the last event and matching all the remaining events in the pattern, we were able to obtain the "predicted" pitch. The procedure is shown in **Table 7**, where the midi pitch represents the MIDI sequence of melodic events. Since there were a number of gaps in the regular expressions, we assigned weights to the predictions using the following criteria, increasing their prediction probability:

- the length of a pattern,
- the proximity of the beginning of the pattern toward the end of the excerpt,
- the total length of matched gaps.

Using the criteria, the following formula was established to calculate the weight (relevance) of the individual predictions

$$w = \frac{N_{eventsinpattern}}{N_{segmentlength} - N_{beginningofthematch} + 1} \quad (1)$$

For a pattern with 32 events which would completely match the last 32 events in the segment, the calculated weight is $w = \frac{32}{33} = 0.97$. For a 4-event pattern, matching the last 4 events in a segment, the weight would be smaller—$w = \frac{4}{5} = 0.60$). A 4-event pattern which would match the first event in the 20-event segment, the weight is considerably lower—$w = \frac{4}{21} = 0.19$.


**FIGURE 4 |** The user interface developed for the prediction part of the questionnaire. When the play button was depressed, the excerpt was played. By dragging the vertical scroll-bar, a pitch in between ±12 semitones was played. By clicking on the note button on the right side, the chosen pitch was added to the end of the excerpt. When added, the excerpt with the added pitch could be played back by clicking on the play button on the left.

Pattern 1 in **Table 7** contains an example with a low weight: $w_1 = \frac{4}{17} = 0.24$. None of the shown patterns matched completely, with patterns 2 and 3 matching better than pattern 1—$w_2 = \frac{8}{11} = 0.73$ and $w_3 = \frac{4}{6} = 0.67$.

## 4. STUDIES

In this section, we present the two studies, where we assess how music familiarity and music expertise influence melody prediction in human listeners, and, expressed in appropriate data/algorithmic ways, computational models. The datasets of music excerpts presented in the Methodology section were used in both studies. The Slovenian translation of the Music Sophistication index was used in Study 1. In both studies we used both the strict and relaxed evaluation measures.

### 4.1. Study 1: Human Melody Prediction

In this study, we researched the influence of the participants' expertise on their predictions. We first collected their responses and, based on their music sophistication, split them into two groups of musicians and non-musicians. We explore the differences in their prediction through several types of prediction evaluation. Based on their self-report on familiarity with Chinese music, we also compare their responses to assess the impact of familiarity.

#### 4.1.1. Data Acquisition

To acquire the data, we created a web interface that consisted of two parts: in the first part we gathered demographic information, musical expertise (using the shortened Gold-MSI questionnaire), music preferences and the frequency of listening to Chinese music. In the second part we asked the participants to perform the melody prediction task.

During the first part the participants were asked about their age, gender and level of education, followed by eight questions from the short Gold-MSI questionnaire, as described in **Table 6**. The participants were also asked to choose up to three preferred genres from a list of 20 genre labels. Additionally, they reported on the amount of time (daily, a few times a month, a few times a year, never) they listened to Chinese music.

In the second part, the participants' task was to predict the continuation of melodic sequences from the database using an interface shown in **Figure 4**. The task of the participants was to listen to a short music excerpt, and then select the pitch they believed best continued the excerpt. The participants could choose a pitch between ± 12 semitones of the last pitch in the excerpt. If necessary, the participants could listen to the excerpt and their selected continuation combined. There were no time limit or number of replays imposed on the participants. In the end, they had to indicate whether or not they had recognized the song. By asking this question, we wanted to avoid the noise induced by responses of participants who knew the songs in the dataset. In case of known songs, the participants would most likely correctly predict the continuation of the excerpt. In this experiment, we evaluated how shared melodic patterns of European folk songs contribute to the familiarity, and avoid known prediction of known songs.

The participants first learned how to use the interface with three trial music excerpts. We added this step to give the participants the time to familiarize themselves with the interface, in order to minimize the influence of the interface on the responses. The chosen songs were a well-known children's song (*Kuža pazi*), the national anthem (*Zdravljica*) and the international version of "Happy Birthday." Because these songs are well known in their environment, the participants were able to focus on the interface during the trial usage.

After the initial interaction, the participants proceeded to the experimental part, during which they listened to the excerpts from the Chinese and the German song datasets. The order of the datasets and the excerpts within was randomized to exclude the bias of the dataset order. The questionnaire was distributed via email and social networks.

#### 4.1.2. Participants

Fifty-seven participants, 26 male and 31 female, completed the questionnaire. The participants were between 16 and 54 years old ($\mu = 26.7, \sigma = 7.5$). The majority of the participants (59.6%) had the education level higher than a high-school diploma, of which 14 obtained a bachelor's degree, 16 had a master's degree and 10 had a PhD degree. All participants were Slovenian.

A large majority of the participants frequently practiced a musical instrument in the past. Only 17.5% never practiced an instrument. The average practice span was 6.8 years ($\sigma = 6.5$). A third of the participants (33.3%) practiced for more than 10 years (the maximum was 25 years), and about three quarters (73.7%) practiced at least for three years. The most popular music genres chosen by the participants were classical music—13 participants chose this genre as their favorite, 5 participants chose it as second favorite and 15 as their third favorite genre. Other most popular genres were rock (in order of top three favorite genres: 16, 6, and 4), and pop (6, 9, 5).

The participants mostly did not listen to Chinese music: 39 participants (68%) never and 13 rarely listened to it. Only two participants listened to Chinese music regularly (a few times a month) and three participants listened to it almost daily.

#### 4.1.3. Participant Groups

To distinguish between familiarity and expertise, we split the participants into two groups. The first group contained participants with high music sophistication (we refer to them as *musicians*), based on the music sophistication index and their performance in the trial attempts of the melody prediction questionnaire. The participants in the second group (the *non-musicians*) scored lower on the MSI questionnaire (below 39 points, the average MSI score was 42.2 out of 56) or missed the prediction for either of the three commonly-known songs in the trial attempts of the second part of the questionnaire. There were 36 participants assigned into the first group (17 male, 19 female, MSI: $\mu = 46.8$, $\sigma = 3.3$; years of music education: $\mu = 9.6$, $\sigma = 6.3$), while the second group contained 21 participants (9 male, 12 female, MSI: $\mu = 34.4$, $\sigma = 6.8$; years of music education: $\mu = 1.8, \sigma = 2.2$).

#### 4.1.4. Influence of Expertise on Predictions

On average, the participants correctly predicted 58% of European and 34% of Chinese continuations. For the European excerpts, they were statistically significantly better ($V = 135, p < 0.01$) at predicting the complete sections (**Table 8**, *European complete excerpts*) in comparison to the incomplete sections, while for the Chinese dataset they were better at solving the incomplete excerpts, but the difference was not significant ($V = 571, p = 0.523$).

The musicians performed much better than the non-musicians, for both the European (68 vs. 41%) and the Chinese dataset (39 vs. 25%). The differences between both groups were statistically significant for music of both datasets and both excerpt types. The greatest difference between musicians and non-musicians occurred between the prediction of the European complete excerpts (Wilcoxon test—European dataset: complete excerpts, $W = 613.5, p < 0.01$; incomplete $W = 536.5, p = 0.01$) and the Chinese incomplete excerpts (Wilcoxon test—Chinese dataset: complete excerpts, $W = 537.5$, $p = 0.01$; incomplete excerpts, $W = 609.5, p < 0.01$).

#### 4.1.5. Relaxed Evaluation

We also analyzed the responses through relaxed evaluation. The participants were better at correctly predicting both the European and the Chinese excerpts, with 94.5% success for European dataset, and 89.7% for Chinese dataset. For the subgroups on the European dataset, the results were in favor of musicians (97.9%) vs. non-musicians (88.6%); both groups were slightly more successful when predicting the continuations of the complete musical excerpts (98.3 and 89.5% for the musicians and non-musicians, respectively) than the incomplete ones (96.7 and 85.7%). On the Chinese dataset the musicians achieved 95.6% (with 95.7% of the complete excerpts and 95.0% of the incomplete excerpts continued correctly according to the relaxed evaluation), while the non-musicians achieved 79.7% (80.9% for complete and 76.2% success rate for incomplete excerpts).

#### 4.1.6. Assessing Familiarity With Music of Foreign Culture

One of the questions on the demographic part of the questionnaire was how often the participants listened to Chinese music. Three responded that they listen to it practically every day, two responded a few times a month, and 13 occasionally (a few times a year). Therefore, 18 participants reported that they listen (at least occasionally) to Chinese music, and 39 reported that they never listen to it.

To analyze the differences between familiarity and expertise, we split the participants into two subgroups, depending on their self-report about the familiarity with Chinese music. Considering their sophistication, there was a similar ratio (2:1) of musicians vs. non-musicians in both groups (**Table 9**).

Differences between the musicians and the non-musicians of Chinese music were not statistically significant. This indifference can imply that we chose a prevalent culture with a very specific

**TABLE 8 |** The performance of the participants and the computational models in melody prediction task, using strict evaluation.

| | European excerpts | | | Chinese excerpts | | |
|---|---|---|---|---|---|---|
| | All (%) | Complete (%) | Incomplete (%) | All (%) | Complete (%) | Incomplete (%) |
| All participants | 58 | 63 | 42 | 34 | 34 | 35 |
| Musicians | 68 | 74 | 47 | 39 | 37 | 44 |
| Non-musicians | 41 | 43 | 33 | 25 | 27 | 21 |
| baseline *(all notes)* | 4 | 4 | 4 | 4 | 4 | 4 |
| baseline *(scale notes)* | 6.7 | 6.7 | 6.7 | 9.1 | 9.1 | 9.1 |
| SymCHM–eu | 60 | 73 | 20 | 30 | 33 | 20 |
| SymCHM–cn | 45 | 53 | 20 | 30 | 40 | 0 |
| Adjusted I-R model | 50 | 60 | 20 | 35 | 25 | 40 |

*The table also includes a baseline, taking into account all possible 25 semitons, which could be picked from the interface, and a scale notes baseline, which includes 15 semitones for major/minor scales (European excerpts), and 11 semitones for pentatonic scale (Chinese excerpts), both within ± one octave.*

musical style, which the participants have heard before at least a few times (for example, in Chinese restaurants, popular movies, etc.), and that, with minimum exposure, the participants were able to memorize the characteristics of the style well-enough to perform this task.

## 4.2. Study 2: Computational Melody Prediction

In this subsection, we analyze the performance of the two selected computational models: the SymCHM and the I-R model. We also compare their melody prediction performance with the participants' results from study 1.

The SymCHM model needs to be trained before it can be used for prediction, and we decided to train two different models, each on a separate folk song dataset: one only on European folk songs (SymCHM–eu) and one only on Chinese folk songs (SymCHM–cn). In this way, we can estimate how (un)familiarity with a certain culture influences the model. The training sets contained approximately 14,000 events (tones) from the Essen folk song collection[3] and did not contain the songs from the generated excerpts used for the melody prediction task. In the study, we additionally evaluate the influence of the training set's size on prediction, thus controlling for the different amounts of "familiarity" with a music culture.

The SymCHM model extracts all knowledge from its training set. In contrast, the I-R model needs no training, as it contains universal music-theoretical rules, derived from human expertise. The I-R model therefore represents an expert system without any culture-specific familiarity.

### 4.2.1. Comparison of Expertise and Familiarity in Computational Models

#### *4.2.1.1. SymCHM*

The SymCHM, trained on the set of European songs (SymCHM–eu), correctly predicted 60% of the European excerpts and 30% of the Chinese excerpts. It also correctly predicted 73% of complete excerpts in the European dataset and 20% of the incomplete excerpts. For the Chinese dataset, the distribution was more uniform: the SymCHM correctly predicted 33% of the complete and 20% of the incomplete excerpts.

The SymCHM trained on the set of Chinese songs (SymCHM–cn), was less successful than the SymCHM–eu. It is interesting that SymCHM–cn performed better on the European dataset than the Chinese dataset, correctly predicting 45% of the European excerpts and, the same as SymCHM–eu, only 30% of the Chinese excerpts.

Considering the complete and incomplete excerpts, the SymCHM–eu performed better on the complete excerpts: 73% complete and only 20% incomplete excerpts were correctly predicted on the European dataset; and 33 % complete and 20% incomplete excerpts were correctly predicted on the Chinese dataset.

#### *4.2.1.2. Influence of the Learning Dataset Size on SymCHM's Results*

All the reported results of the SymCHM–eu and SymCHM–cn models were obtained using models trained on approximately 15,000 events (300 short European songs for SymCHM–eu, and 189 longer Chinese songs for SymCHM–cn). Considering the learned familiarity, we evaluated the SymCHM model's performance using different smaller dataset sizes. Initially, we trained the SymCHM–eu model with only 50 songs. To assess the model's performance with respect to the dataset size, we trained the model using larger datasets, thus increasing its musical knowledge. The results are shown in **Table 10**.

By increasing the training dataset from 50 to 100 songs, SymCHM's performance was not significantly improved. Moreover, the incomplete sequences were less likely predicted correctly. With the further enlargement of the training set, the model's performance increases faster in the "all" and "complete excerpts," while it the "incomplete excerpts" increased with the 300-song training set. The performance of the SymCHM is therefore impacted by the training set. The model's performance does gradually increase for all subtypes of exerpts. However, due to the small size of the incomplete excerpts subset, the results vary between training set.

[3] http://kern.ccarh.org/browse (accessed June 22, 2018)

**TABLE 9 |** Comparing the performance of the participants who listen to Chinese music to those who do not listen to Chinese music.

| Listens to Chinese | Musicians | | | Non-musicians | | |
|---|---|---|---|---|---|---|
| | All (%) | Complete (%) | Incomplete (%) | All (%) | Complete (%) | Incomplete (%) |
| Yes (12 musicians, 6 non-musicians) | 38 | 37 | 40 | 24 | 26 | 20 |
| No (24 musicians, 15 non-musicians) | 40 | 37 | 49 | 25 | 26 | 21 |

**TABLE 10 |** The impact of the SymCHM–eu's training dataset size, repeated 5 times (for training sizes 50–200).

| No. of songs | 50 | 100 | 200 | 300* |
|---|---|---|---|---|
| Avg. no. of events | 2k | 4.5k | 9.5k | 15k |
| Avg. – all excerpts | 37% $\sigma = 6.8$ | 39% $\sigma = 11.0$ | 42% $\sigma = 9.7$ | 53% $\sigma = 9.4$ |
| Avg. – complete excerpts | 45% $\sigma = 7.8$ | 45% $\sigma = 4.9$ | 55% $\sigma = 10.9$ | 64% $\sigma = 12.3$ |
| Avg. – incomplete excerpts | 12% $\sigma = 9.8$ | 4% $\sigma = 10.0$ | 8% $\sigma = 9.8$ | 20% $\sigma = 0$ |

*The training was repeated only 3 times for the training set of 300 songs (marked with *).*

**TABLE 11 |** Average values for the mean reciprocal ranks (MRR) for all three models on each of the two datasets.

| Model | Dataset | |
|---|---|---|
| | European | Chinese |
| SymCHM–eu | 0.720 | 0.491 |
| SymCHM–cn | 0.635 | 0.520 |
| Adjusted I-R model | 0.648 | 0.490 |

### 4.2.2. Narmour's Implication-Realization Model

We first evaluated the I-R model. The model's predictions showed, the rules of the initial (non-extended) I-R model are too restrictive and, consequently resulted in poor performance on the melody prediction task. The model it did not correctly predict any continuation on the European dataset, and only three in the Chinese dataset.

We analyzed the ground-truth continuations of the excerpts in the European dataset. These can be summarized three rules regarding the predicted events:

- they are high on the tonal hierarchy,
- if they are on the top or the bottom limit of the excerpt's range, they change direction,
- the distance between the predicted events and the starting point is 7 or less semitones.

These rules are also included in the extended Implication-Realization Model. The initial model listed the 0 (repeating the last event) as the best answer, which is in most cases an acceptable answer from the point of view of conformity with the tonic hierarchy in the song, but it is mostly incorrect. In addition, the I-R model puts too much weight to the answers in the immediate vicinity of the starting point (± a few semitones).

For the purposes of the task, we fine-tuned the model, using a subset of the criteria from the extended Implication-Realization Model, since the initial model emphasized very small intervals and preferred predicting the same tone too often. We retained the *registral return*, *proximity*, *tonality*, *melodic attraction*, and *tessitura* criteria. We obtained the score of each possible outcome by averaging the normalized values of the five criteria.

Using this adjustment of the extended Implication-Realization model, the evaluation yielded significantly better results. The adjusted Implication-Realization Model correctly predicted 50% of European and 35% of Chinese excerpts (**Table 8**).

### 4.2.3. Relaxed Evaluation

Taking into account the correct responses in relaxed evaluation, both SymCHM–eu and SymCHM–cn models correctly classified all excerpts in the European dataset. The models therefore achieved better results than the participants. For the Chinese dataset, SymCHM–cn also achieved 100%, while SymCHM–eu achieved only 80%.

### 4.2.4. Comparison Between the SymCHM and the Adjusted I-R Model

The adjusted I-R model, which was implemented for this experiment using aforementioned libraries, returns a probability score for each of the possible continuations across five different criteria: registral return, proximity, tonality, melodic attraction, and tessitura. All criteria are equally represented and a combined probability is provided for each continuation, given the normalized sum of probabilities across all five criteria. The SymCHM's responses can also be ranked using the weights of the individual responses. We therefore compared both models by observing the ranks that the true continuations receive, with regard to all possibilities. We used the mean reciprocal rank to compare both models.

The mean reciprocal rank (MRR) is a statistical measure for evaluating the accuracy of the order of elements in a list sorted by a criterion (in our case, this is the weight for each possible answer for each song), using the following formula:

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}, \tag{2}$$

where $Q$ represents the list of all possible responses, and $rank_i$ represents the position of the first relevant response. Given the rank of the correct response is 1 (i.e., the first response is the relevant response), its reciprocal rank (RR) equals 1. If the rank of the relevant response is $n$, the RR equals $\frac{1}{n}$.

We calculated the average values for the RR across all three models SymCHM–eu, SymCHM–cn and the I-R model, for the European and the Chinese datasets individually. The results are shown in **Table 11**.

The SymCHM–eu obtained the highest MRR, meaning that among the selected models, it was the most successful in providing the responses. For the European dataset, it received an average rank of 0.72, while for the Chinese dataset it received the rank of 0.491. The adjusted I-R model received the RR of 0.648 on the European dataset and 0.49 on the Chinese dataset.

The performance of the adjusted I-R model was more similar across both datasets, compared to the SymCHM's results. Since it contains rules, which are culture agnostic, this behavior was somewhat expected. Nevertheless, the results on the European dataset were still higher. However, the best-performing SymCHM–cn's results were also lower on the Chinese dataset. The performance of the adjusted I-R model and the SymCHM–cn was therefore more consistent across both datasets, compared to the SymCHM–eu model.

## 5. DISCUSSION

During this research, two datasets were collected, each containing 20 music excerpts. The European dataset contained excerpts from the European folk song collection, while the Chinese dataset contained Chinese folk songs. Both the Chinese and the European datasets contained 75% of complete excerpts, in which the final note in the melody or phrase was predicted, and 25% of incomplete excerpts, where the phrase ended at a random location and the following note was to be predicted.

**In study 1**, we first analyzed the collected participants' melody predictions. We further split them in two groups: musicians and non-musicians, based on their MSI score, to analyze the differences between the participants. The MSI questionnaire was translated to Slovenian language and validated (section 3.3).

On average, the European participants correctly predicted 58% of European and 34% of Chinese continuations. For the European excerpts, they were statistically significantly better ($V = 135, p < 0.01$) at predicting the complete (**Table 8**) than the incomplete sections, while for the Chinese dataset they were better at solving the incomplete excerpts but the difference was not significant ($V = 571, p = 0.523$).

The participants in the musician group performed much better than the non-musicians, for both the European and the Chinese dataset. The differences between the groups of musicians and non-musicians were statistically significant in the music of both datasets and both complete and incomplete excerpt types. The greatest difference between the musicians and non-musicians occurred between the prediction of the European complete excerpts and the Chinese incomplete excerpts (**Table 8**). The results confirm the MSI as a credible instrument for music

**TABLE 12 |** Results of the Wilcoxon signed-rank test (comparing European and Chinese datasets; W) and Wilcoxon rank sum test (comparing musicians and non-musicians; V).

| Participants | Dataset | *W* | *V* | *p* |
|---|---|---|---|---|
| Musicians | European : Chinese | | 0 | $< 0.01$ |
| Non-musicians | European : Chinese | | 32 | $< 0.01$ |
| Musicians : non-musicians | European | 370 | | 0.90 |
| Musicians : non-musicians | Chinese | 152 | | $< 0.01$ |

sophistication validation. Since the end of a phrase within an excerpt is more predictable due to the musicological rules of the melodic form, we hypothesized the participants, as well as the models would be more successful in correctly predicting the complete than the incomplete excerpts. The results concur with our hypothesis.

In comparing the responses between the European and Chinese datasets, there were statistically important differences in both groups of participants (musicians and non-musicians). There were also statistical differences in the participants' performance between the musicians and non-musicians within the Chinese, but not European, dataset. The results of significance tests between different groups are shown in **Table 12**. The better performance of musicians in both datasets was expected in this experiment. We attribute the significance of the difference on the Chinese dataset to the musicians' expertise, whereas the familiarity of the European dataset influenced the relatively better performance of the non-musicians on the European dataset. In this aspect, the comparison of the non-musicians performance on both datasets also unveiled the underlying difference between the listeners' expertise vs. familiarity.

Regardless of the poorer performance of the participants in both groups on the Chinese dataset (compared to the European dataset), their performance was quite high—the participants predicted a suitable response in almost 90% of cases, while the musicians achieved an almost perfect score.

**In study 2**, we performed an evaluation of the prediction of the two computational models, and compared them to the participants' prediction performance. In general, the participants performed better in the prediction task than the compared models on both datasets. It is evident that SymCHM, especially SymCHM–eu, came very close to the performance of the participants. We can conclude that the SymCHM model's performance lays between non-musicians' and musicians' performance in this prediction task. The extended I-R model first needed fine-tuning to perform in this task. After the adjustment, the model performed significantly better. However, it seems the expertise implemented in this model does not outperform the participants, nor the SymCHM, which was trained and thus familiar with the background.

In the relaxed evaluation of the predictions, both the SymCHM–eu and SymCHM–cn models correctly classified all excerpts in the European dataset, and therefore achieved better results than the participants. This can be attributed to the learning process in which the SymCHM extracts the common patterns from the training dataset. If the dataset only contains a major/minor or pentatonic scales, the model will output only predictions matching the scales. In this aspect, these results should not be generalized. In a similar manner, the relative simplicity of the pentatonic scale (compared to the major/minor scales) influenced the SymCHM–cn model's results. The SymCHM–cn achieved a 100% success in the relaxed evaluation on the European dataset and the SymCHM–eu achieved lower results on the Chinese dataset. We attribute these results to in the structure of the scales—the tones of the pentatonic scale are a subset of the diatonic scale. In the relaxed evaluation, both the adjusted I-R model and the SymCHM model performed worse on the Chinese dataset than on the European dataset. We did not expect this difference in performance, since the SymCHM model learns the patterns from the training set, while the adjusted I-R model employs rules, which are universal to music. In this sense, we could attribute this difference to potential Western-music bias, although further research is needed to confirm this assumption.

We also compared the predictions using the mean reciprocal rank. On the European dataset, the SymCHM–eu performed significantly better than the adjusted I-R model (second best) and the SymCHM–cn. On the contrary, the SymCHM–cn received the best score on the Chinese dataset, whereas the SymCHM–eu and the adjusted I-R model performed similarly worse. The results for the difference between the SymCHM–eu and the SymCHM–cn models were expected: the models were trained, and therefore familiarized, with each music type, and were therefore expected to perform better when applied to the music of the same cultural background. On the other hand, the adjusted I-R model performed similarly on both datasets, proving the I-R model is agnostic of music culture. Therefore, the difference between the experience and the familiarity in computational approaches is clearly visible in their performance across different datasets.

Additionally, the SymCHM models were also evaluated by training on different dataset sizes. It is evident, that the model performs better, when trained on a larger dataset, thus increasing its "familiarity" with the underlying patterns, which are shared among the songs with similar cultural background.

The role of familiarity due to cultural background has been discussed in related work. Although some works support the assumption that cultural background plays a role in experiencing music (Balkwill and Thompson, 1999; Cross, 2001; Morrison et al., 2003; Demorest et al., 2008; Soley and Hannon, 2010) recent work has shown that within-culture variance of music is higher than between-culture variance (Mehr et al., 2019). This would indicate that the specific choice of music from a culture may influence how much variance the familiarity variable accounts for.

## 6. CONCLUSION

In this paper, we explored the influence of the listeners' cultural background and their music sophistication on melody prediction. This was done on two datasets consisting of musical excerpts of European and Chinese folk songs. The melody prediction data was gathered on 57 participants. The participants were asked to predict the possible melody continuation of each music excerpt from the two datasets. The responses were split into two groups: (i) musicians (high music sophistication), and (ii) non-musicians (low music sophistication). The music sophistication was acquired using the MSI instrument, which was adapted to Slovenian-speaking participants. We compared the participants' responses of the two groups. Musicians performed better than non-musicians on both the familiar (European) dataset, and the less familiar (Chinese) dataset.

In addition, we compared the melody prediction performance of two computational models: (i) the adjusted I-R model and (ii) the symbolic compositional hierarchical model (SymCHM). The SymCHM was trained twice, once for each melody prediction task (on a set of Chinese songs and European songs, separately). The SymCHM outperformed the adjusted I-R model in the strict melody prediction task. We also compared the predictions of the SymCHM and the adjusted I-R models with the melody prediction performances of human listeners. Musicians outperformed both the SymCHM and the adjusted I-R model.

In both studies, the experiment results showed that the music excerpts which ended at the end of a phrase (complete excerpts) were more predictable than those which ended in the middle of a phrase (incomplete excerpts). Both the participants and the computational models correctly predicted less than half of the incomplete excerpts; they were both more successful in predicting complete excerpts. As the product of this research, we also developed the Slovenian version of the Musical sophistication index questionnaire, evaluated on 230 participants. Additionally, we collected the responses of 57 participants in the prediction task. Both the dataset and the translated MSI questionnaire are made publicly available.

Based on the described work, we are planning on extending this experiment with participants from different cultural backgrounds and with datasets of less-known folk music. We also plan on performing an inverted experiment to further assess the computational models, by using the models' predicted responses and having the participants evaluating their subjective correctness of the responses. Another planned extension of the current research will be exploring the participants' latent factors which influence their implicit expertise in predicting. Furthermore, we plan on further exploring the error patterns of human listeners and evaluate the underlying decision process in comparison to their music sophistication.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author/s.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

MP and MM wrote a plan for research. ŠM executed experiments. AP examined the results in Study 1. MP, MT, and MM evaluated the results in Study 1. MP, MT, and ŠM wrote the article. All authors contributed to the article and approved the submitted version.

## REFERENCES


Arafat, S., Chowdhury, H., Qusar, M., and Hafez, M. (2016). Cross cultural adaptation and psychometric validation of research instruments: a methodological review. *J. Behav. Health* 5:129. doi: 10.5455/jbh.20160615121755

Balkwill, L.-L., and Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: psychophysical and cultural cues. *Music Percept*. 17, 43–64. doi: 10.2307/40285811

Blažica, B., and Lewis, J. R. (2015). A slovene translation of the system usability scale: the SUS-SI. *Int. J. Hum. Comput. Interact*. 31, 112–117. doi: 10.1080/10447318.2014.986634

Brown, G. T., Harris, L. R., O'Quin, C., and Lane, K. E. (2017). Using multi-group confirmatory factor analysis to evaluate cross-cultural research: identifying and understanding non-invariance. *Int. J. Res. Method Educ*. 40, 66–90. doi: 10.1080/1743727X.2015.1070823

Castellano, M. A., Bharucha, J. J., and Krumhansl, C. L. (1984). Tonal hierarchies in the music of north India. *J. Exp. Psychol. Gen*. 113, 394–412. doi: 10.1037/0096-3445.113.3.394

Collins, T. (2016). *2015:Discovery of Repeated Themes & Sections - MIREX Wiki*.

Collins, T., Tillmann, B., Barrett, F. S., Delbé, C., and Janata, P. (2014). A combined model of sensory and cognitive representations underlying tonal expectations in music: from audio signals to behavior. *Psychol. Rev*. 121, 33–65. doi: 10.1037/a0034695

Colombo, F. (2018). "Generating and discriminating symbolic music continuations with bachprop," in *19th International Society for Music Information Retrieval Conference* (Paris).

Cross, I. (2001). Music, cognition, culture, and evolution. *Ann. N. Y. Acad. Sci*. 930, 28–42. doi: 10.1111/j.1749-6632.2001.tb05723.x

de Fleurian, R., Blackwell, T., Ben-tal, O., and Daniel, M. (2017). Information-theoretic measures predict the human judgment of rhythm complexity. *Cogn. Sci*. 41, 800–813. doi: 10.1111/cogs.12347

de Reuse, T. (2019). "Copyforward: point-set matching for predicting patterns," in *Submission to the MIREX 2019 Patterns for Predicton Task* (Delft), 1–2.

Demorest, S. M., Morrison, S. J., Beken, M. N., and Jungbluth, D. (2008). Lost in translation: an enculturation effect in music memory performance. *Music Percept*. 25, 213–223. doi: 10.1525/mp.2008.25.3.213

Downie, J. S. (2008). The music information retrieval evaluation exchange (2005–2007): a window into music information retrieval research. *Acoust. Sci. Technol*. 29, 247–255. doi: 10.1250/ast.29.247

Eerola, T., Himberg, T., Toiviainen, P., and Louhivuori, J. (2006). Perceived complexity of western and African folk melodies by western and African listeners. *Psychol. Mus*. 34, 337–371. doi: 10.1177/0305735606064842

Egermann, H., Pearce, M. T., Wiggins, G. A., and McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cogn. Affect. Behav. Neurosci*. 13, 533–553. doi: 10.3758/s13415-013-0161-y

Ens, J., and Pasquier, P. (2019). "Discriminating symbolic continuations with gendetect," in *Submission to the MIREX 2019 Patterns for Predicton Task2* (Delft), 1–2.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., et al. (2009). Universal recognition of three basic emotions in music. *Curr. Biol*. 19, 573–576. doi: 10.1016/j.cub.2009.02.058

Hippel, P. V. (2000). Redefining pitch proximity: tessitura and mobility as constraints on melodic interval size. *Music Percept*. 17, 315–327. doi: 10.2307/40285820

Huron, D. (2008). Science & music: lost in music. *Nature* 453, 456–457. doi: 10.1038/453456a

Janata, P., and Reisberg, D. (1988). Response-time measures as a means of exploring tonal hierarchies. *Music Percept*. 6, 161–172. doi: 10.2307/40285423

Janssen, B., Ren, I. Y., Owers, J., and Collins, T. (2019). *Patterns for Prediction - MIREX Wiki*.

Krumhansl, C. L. (1995). Music psychology and music theory: problems and prospects. *Music Theor. Spectr*. 17, 53–80. doi: 10.2307/745764

Krumhansl, C. L., and Cuddy, L. L. (2010). "A theory of tonal hierarchies in music," in *Music Perception, Vol. 36*, eds M. Riess Jones, R. Fay, and A. Popper (New York, NY: Springer). doi: 10.1007/978-1-4419-6114-3_3

Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., and Eerola, T. (1999). Melodic expectation in finnish spiritual folk hymns: convergence of statistical, behavioral, and computational approaches. *Music Percep*. 17, 151–195. doi: 10.2307/40285890

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., and Louhivuori, J. (2000). Cross-cultural music cognition: cognitive methodology applied to North Sami yoiks. *Cognition* 76, 13–58. doi: 10.1016/S0010-0277(00)00068-8

Lappe, C., Steinsträter, O., and Pantev, C. (2013). Rhythmic and melodic deviations in musical sequences recruit different cortical areas for mismatch detection. *Front. Hum. Neurosci*. 7:260. doi: 10.3389/fnhum.2013.00260

Lerdahl, F. (1996). Calculating tonal tension. *Music Percept*. 13, 319–363. doi: 10.2307/40286174

Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge: MIT Press.

Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nat. Neurosci*. 4, 540–545. doi: 10.1038/87502

Margulis, E. H. (2013). *On Repeat*. Oxford, UK: Oxford University Press. doi: 10.1093/acprof:oso/9780199990825.001.0001

Marsden, A. (2010). Schenkerian analysis by computer: a proof of concept. *J. New Music Res*. 39, 269–289. doi: 10.1080/09298215.2010.503898

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., et al. (2019). Universality and diversity in human song. *Science* 366:eaax0868. doi: 10.1126/science.aax0868

Meredith, D. (2013). "Cosiatec and siateccompress: pattern discovery by geometric compression," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)* (Curitiba), 1–6.

Meredith, D., Lemstrom, K., and Wiggins, G. A. (2002). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *J. New Music Res*. 31, 321–345. doi: 10.1076/jnmr.31.4.321.14162

Morrison, S. J., Demorest, S. M., Aylward, E. H., Cramer, S. C., and Maravilla, K. R. (2003). fMRI investigation of cross-cultural music comprehension. *NeuroImage* 20, 378–384. doi: 10.1016/S1053-8119(03)00300-8

Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PLoS ONE* 9:e89642. doi: 10.1371/journal.pone.0089642

Nan, Y., Knösche, T. R., and Friederici, A. D. (2006). The perception of musical phrase structure: a cross-cultural ERP study. *Brain Res*. 1094, 179–191. doi: 10.1016/j.brainres.2006.03.115

Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago, IL: University of Chicago Press.

Oram, N., and Cuddy, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychol. Res*. 57, 103–118.

Pearce, M. T. (2003). *Melodic Expectation in Music: A Review of Theoretical Accounts and Empirical Research*. Neobjavljeno delo. Available online at: http://webprojects.eecs.qmul.ac.uk/marcusp/notes/expectation.pdf

Pesek, M., Leonardis, A., and Marolt, M. (2014). "A compositional hierarchical model for music information retrieval," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)* (Taipei), 131–136.

Pesek, M., Leonardis, A., and Marolt, M. (2017a). Robust real-time music transcription with a compositional hierarchical model. *PLoS ONE* 12:e0169411. doi: 10.1371/journal.pone.0169411

Pesek, M., Leonardis, A., and Marolt, M. (2017b). SymCHM-an unsupervised approach for pattern discovery in symbolic music with a compositional hierarchical model. *Appl. Sci.* 7:1135. doi: 10.3390/app7111135

Pesek, M., Leonardis, A., and Marolt, M. (2020). An analysis of rhythmic patterns with unsupervised learning. *Appl. Sci.* 10:178. doi: 10.3390/app10010178

Ren, I. Y. (2016). "Closed patterns in folk music and other genres," in *6th International Workshop on Folk Music Analysis* (Dublin), 56–58.

Rohrmeier, M. A., and Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *Int. J. Psychophysiol.* 83, 164–175. doi: 10.1016/j.ijpsycho.2011.12.010

Saffran, J. R., and Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: evidence for developmental reorganization. *Dev. Psychol.* 37, 74–85. doi: 10.1037/0012-1649.37.1.74

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52. doi: 10.1016/S0010-0277(98)00075-4

Schellenberg, E. (1997). Simplifying the implication-realization model of melodic expectancy. *Music Percept.* 14, 295–318. doi: 10.2307/40285723

Soley, G., and Hannon, E. E. (2010). Infants prefer the musical meter of their own culture: a cross-cultural comparison. *Dev. Psychol.* 46, 286–292. doi: 10.1037/a0017555

Sousa, V. D., and Rojjanasrirat, W. (2010). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *J. Eval. Clin. Pract.* 17, 268–274. doi: 10.1111/j.1365-2753.2010.01434.x

Thiessen, E. D., and Saffran, J. R. (2009). How the melody facilitates the message and vice versa in infant learning and memory. *Ann. N.Y. Acad. Sci.* 1169, 225–233. doi: 10.1111/j.1749-6632.2009.04547.x

Toiviainen, P., and Eerola, T. (2016). *MIDI Toolbox 1.1*. Available online at: https://github.com/miditoolbox/

Vidakovič, J., and Delo (2003). *Sto Slovenskih Ljudskih Pesmi*. Prešernova družba.

Vodušek, V. (1959). Arhaični slovanski peterec-deseterec v slovenski ljudski pesmi. *Slovenski Etnograf* 12, 181–202.

Ycart, A., and Benetos, E. (2019). "Polyphonic music sequence classification with LSTM networks," in *Submission to the MIREX 2019 Patterns for Predicton Task* (Urbana-Champaign, IL), 1–4.

Zatorre, R. J., Evans, A. C., and Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* 14, 1908–1919. doi: 10.1523/JNEUROSCI.14-04-01908.1994

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 11 December 2020
doi: 10.3389/fpsyg.2020.566373

# What Makes Musical Prodigies?

*Chanel Marion-St-Onge, Michael W. Weiss, Megha Sharda and Isabelle Peretz**

*Department of Psychology, International Laboratory for Brain, Music, and Sound Research, University of Montreal, Montreal, QC, Canada*

Musical prodigies reach exceptionally high levels of achievement before adolescence. Despite longstanding interest and fascination in musical prodigies, little is known about their psychological profile. Here we assess to what extent practice, intelligence, and personality make musical prodigies a distinct category of musician. Nineteen former or current musical prodigies (aged 12–34) were compared to 35 musicians (aged 14–37) with either an early (mean age 6) or late (mean age 10) start but similar amount of musical training, and 16 non-musicians (aged 14–34). All completed a Wechsler IQ test, the Big Five Inventory, the Autism Spectrum Quotient, the Barcelona Music Reward Questionnaire, the Dispositional Flow Scale, and a detailed history of their lifetime music practice. None of the psychological traits distinguished musical prodigies from control musicians or non-musicians except their propensity to report flow during practice. The other aspects that differentiated musical prodigies from their peers were the intensity of their practice before adolescence, and the source of their motivation when they began to play. Thus practice, by itself, does not make a prodigy. The results are compatible with multifactorial models of expertise, with prodigies lying at the high end of the continuum. In summary, prodigies are expected to present brain predispositions facilitating their success in learning an instrument, which could be amplified by their early and intense practice happening at a moment when brain plasticity is heightened.

**Keywords: musical prodigies, musical talent, expertise, achievement, practice, intelligence, personality**

OPEN ACCESS

***Edited by:***
*Elvira Brattico,*
*Aarhus University, Denmark*

***Reviewed by:***
*Joyce L. Chen,*
*University of Toronto, Canada*
*Patrick K. A. Neff,*
*University of Regensburg, Germany*

****Correspondence:***
*Isabelle Peretz*
*isabelle.peretz@umontreal.ca*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology*

***Received:*** *27 May 2020*
***Accepted:*** *28 October 2020*
***Published:*** *11 December 2020*

***Citation:***
*Marion-St-Onge C, Weiss MW, Sharda M and Peretz I (2020) What Makes Musical Prodigies? Front. Psychol. 11:566373. doi: 10.3389/fpsyg.2020.566373*

## INTRODUCTION

CH plays the violin exceptionally well. He's a 26-year-old acclaimed professional musician who studied at Juilliard, has won numerous national and international competitions, and currently plays on a Stradivarius violin. He made his orchestral debut at 7 years old. A musician like CH, who showed "superior performance within a specific domain" before adolescence, is considered to be a musical prodigy in the present study (see **Supplementary Table 1** for definitions). Here, in the largest sample of exceptional musicians considered so far, we examine non-musical traits, such as practice, autistic traits, and intelligence, that have been associated with musical prodigiousness.

In doing so, we endorse the Multifactorial Gene–Environment Interaction Model proposed by Ullén et al. (2016) (**Figure 1**), which assumes complex interactions between genes, environment, practice behavior, and psychological traits (Mosing et al., 2014).

Practice is obviously central to the development of any skill, and musical skill in particular. From the influential deliberate practice perspective, practice is the only important factor in acquiring

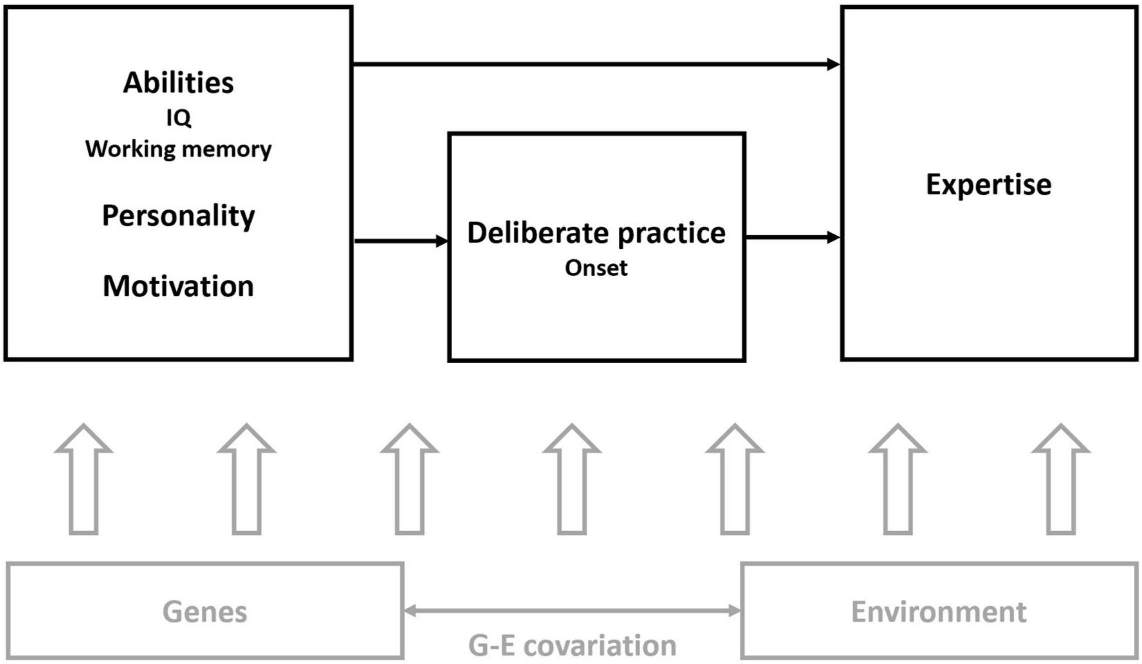



**FIGURE 1 |** Adaptation from the Multifactorial Gene–Environment Interaction Model proposed by Ullén et al. (2016) in which the factors assessed here are highlighted in black. The arrows represent the influence between psychological traits, practice behavior, and expertise (or achievement). Below are the complex influences of genes, environment, and their interaction on all the variables depicted above.

expertise (Ericsson et al., 1993). Other perspectives hold that practice alone is not sufficient. In a meta-analysis on the relationship between practice and performance, Macnamara et al. (2014) found that the variance in music performance explained by deliberate practice is 21%, which leaves the majority of variance unexplained. Complicating matters further, individuals vary considerably in the amount of practice needed to reach expert-level performance (Ackerman, 2014). For example, in chess, the minimum amount of deliberate practice required to achieve master level is around 3,000 h, but some players accumulate as many as 20,000 h without reaching that status (Campitelli and Gobet, 2011). Thus, the relation between practice and performance is not straightforward.

Practice is not a purely environmental factor. Genetic predispositions also come into play. There is no difference, for example, in music perception abilities of monozygotic twins with differing amounts of musical practice (Mosing et al., 2014). The age of onset of musical training can also interact with genetic differences in brain structure and function (Herholz and Zatorre, 2012). A confluence of neurogenetic factors might influence practice, as well as musical abilities like the precision of motor timing in sequential tapping, complicating the relationship between practice and musical achievement (Ullén et al., 2015).

The Multifactorial Gene–Environment Interaction Model (MGIM; Ullén et al., 2016; **Figure 1** for an adapted version) of music proficiency and expertise is arguably the most comprehensive model of musical talent proposed so far. The model is evidence-driven in the sense that it emerges from recent findings in the field of expertise. It incorporates the roles of multiple factors in expertise development, such as practice required to reach a certain level of performance, personality traits, IQ, and working memory.

Motivation to practice is another psychological dimension considered in the model suggested by Ullén et al. (2016), but often ignored in neurogenetic studies of musicality. This trait seems especially relevant to prodigies, who have been described as possessing a "rage to master," or a drive fueling their interest and capacity to practice for extended periods of time (Winner, 2000). As Gagné and McPherson (2016) note, the terminology used by Winner encompasses various concepts such as flow, obsessive passion, and intrinsic motivation. Indeed, the tendency to experience flow may contribute to prodigies' motivation. Flow is a psychological state characterized by intense concentration and a heightened sense of control, and it constitutes an experience that is inherently rewarding (Nakamura and Csikszentmihalyi, 2009). The experience of flow when playing music correlates with amount of music practice (Butkovic et al., 2015; Marin and Bhattacharya, 2013), but flow itself is not a predictor of achievement (i.e., which musician will win a competition; Marin and Bhattacharya, 2013). Moreover, personality traits like openness to experience and musical flow share genetic influence (Butkovic et al., 2015). Accordingly, intrinsic motivation, frequency of practice, propensity to experience flow during practice, and reward experienced with music, will be examined here.

Besides practice and motivation, the presence of autistic traits could distinguish prodigies from their peers. Autistic traits are measured by metrics such as the Autism Spectrum Quotient (AQ; Baron-Cohen et al., 2001). A defining autistic trait is attention to detail, which refers to the propensity to focus attention on detailed aspects of sensory information, and which may be more prevalent among musical prodigies (Ruthsatz and Urbach, 2012). Because autistic traits are independent from the personality components of the Big Five inventory (Wakabayashi et al., 2006; Austin, 2005), all participants in the present study will complete the Autism Spectrum Quotient questionnaire in addition to the Big Five Inventory.

Enhanced intelligence is another trait often associated with musical training (for reviews, Schellenberg and Weiss, 2013; Miendlarzewska and Trost, 2014; Swaminathan and Schellenberg, 2019), but most research has focused on typical musicians. Whether musical prodigies, who represent the extreme of musical achievement, would obtain correspondingly high IQ scores is unclear. Support for this idea comes from the study of a relatively large sample of prodigies ($n$ = 18), of which eight were musical prodigies. The musical prodigies obtained a high IQ ($M$ = 129) compared to the general population, with especially high scores for working memory (Ruthsatz et al., 2014). A more recent case study conducted with a musical prodigy also showed superior working memory (Comeau et al., 2018).

Empirical research on musical prodigies is scarce. Case studies have investigated aspects of musical and cognitive abilities in individual musical prodigies (Comeau et al., 2018; Ruthsatz and Detterman, 2003; Dalla Bella et al., 2016). The typical method compares a prodigy to a control group matched on age or musical training, or uses normalized tests rather than a control group. For example, there are many reports of prodigies who possess absolute pitch – the ability to automatically identify a note without prior reference (Gagné and McPherson, 2016). However, its prevalence in prodigies relative to non-prodigy musicians has not been empirically assessed (Comeau et al., 2018). To our knowledge only one research group has recruited multiple prodigies for study, and these samples were recruited across different domains of expertise (e.g., music, visual arts, and maths), and were not compared to a control group (Ruthsatz and Urbach, 2012; Ruthsatz et al., 2014). No study to date has compared a group of musical prodigies to control groups matched on musical experience.

In the present study, we assess the extent to which prodigious talent exists on a continuum with the trajectory of typical musicians, or alternatively, constitutes a distinct category. We may assume that predispositions play an outsized role in the achievements of prodigies because they achieve so much so early in life, but the nature of those predispositions and their link with behavior and eventual achievement is unknown. In keeping with the MGIM framework (Ullén et al., 2016), we ask whether the prodigies' expertise (or achievement) is influenced by psychological traits like cognitive abilities, personality, motivation, and deliberate practice behavior, and whether there is a link between practice and psychological traits. We also consider, as an alternative view, whether the prodigy phenomenon can be explained by a simpler framework such as deliberate practice (Ericsson et al., 1993).

The study of prodigies may help to identify which ingredients are critical to reach exceptional performance in typical musicians. To answer these questions, we compared four groups of adolescent or adult participants, former or current prodigies, musicians who started training early in childhood, musicians who started training later in childhood, and non-musicians.

## MATERIALS AND METHODS

### Participants

We recruited 19 current or former prodigies. Six of them were aged 12 to 14 at the moment of testing and 13 were adult participants who were prodigies in their youth (hereafter, *prodigies*). They were recruited through online searches, references from professional musicians and music teachers, and public announcements. Detailed demographic and musical experience information are listed in **Tables 1**, **2**, respectively. Classification as prodigy was established by meeting at least one of the following criteria before age 14: (1) high achievement in performance, like winning a first prize in a national or international competition, or winning multiple regional competitions, or (2) special recognition of talent through television or documentary appearances, or orchestral debut (as used in Ruthsatz and Urbach, 2012). Their achievements, listed in **Supplementary Table 2**, were confirmed in a semi-structured interview. Two prodigies (siblings) were diagnosed with autism spectrum disorder early in life.

There were three control groups, with each group differing in their musical experience. Early-trained musicians ($N$ = 16; hereafter, *early-trained*) were similar to prodigies in age of onset of musical training and years of musical experience but did not show exceptional talent before the age of 14. Late-trained musicians ($N$ = 19; hereafter, *late-trained*) began to play their instrument later than the prodigies and early-trained musicians, on average, while accumulating a similar number of years of musical training at the time of testing. Early-trained musicians were matched individually to prodigies on age of onset of musical experience (±2 years). Late-trained musicians had a delayed onset of training after age 7 and were also matched on years of musical experience. Before 18 years old, the majority of control musicians (30 out of 35 control musicians) did not report any achievements such as those considered for the prodigy criteria.

During the interview conducted with each musician, we collected practice data on the daily or weekly estimated number of hours of deliberate practice. For participants under age 16, parents were present during the interview. Yearly estimated number of hours of practice were calculated by summing the number of hours of daily or weekly practice reported by each participant for each year of musical experience, as in other research (Ericsson et al., 1993). For example, if a participant reported practicing 20 min per day and 6 days per week, this amounts to 2 h per week for 52 weeks, and 104 h for that particular year. For each musician, we also calculated accumulated deliberate practice by summing the yearly amount of practice from the onset of musical experience. Detailed information is listed in **Table 2**.

**TABLE 1 |** Demographics.

| Group | Prodigies | Early-trained | Late-trained | Non-musicians | Statistics |
|---|---|---|---|---|---|
| *N* | 19 | 16 | 19 | 16 | |
| Sex (*F* = female; *M* = male) | 7 F, 12 M | 7 F, 9 M | 7 F, 12 M | 9 F, 7 M | $X^2(3, N = 70) = 1.74, p = 0.628$ |
| Age (years) | 21.3 ± 7.4 (12–34) | 23.3 ± 6.2 (14–33) | 25.2 ± 7.0 (14–37) | 24.4 ± 6.9 (14–36) | $F(3,66) = 1.10, p = 0.356$ |
| Education (years) | 14.0 ± 5.0 (6–21) | 15.4 ± 4.0 (8–21) | 16.8 ± 4.3 (8–25) | 16.9 ± 3.9 (9–25) | $F(3,66) = 1.82, p = 0.153$ |

*Values are reported in mean ± standard deviation with range in parentheses.*

**TABLE 2 |** Musical experience.

| Group | Prodigies | Early-trained | Late-trained | Statistics |
|---|---|---|---|---|
| *N* | 19 | 16 | 19 | |
| Age of onset (years) | 4.9 ± 1.3 (3–8) | 5.5 ± 1.5 (4–9) | 10.3 ± 2.5 (7–15) | $F(2,51) = 46.59, p < 0.001$ |
| Musical experience (years) | 17.2 ± 7.6 (8–31) | 18.1 ± 6.4 (9–28) | 15.2 ± 6.5 (7–28) | $F(2,51) = 0.84, p = 0.438$ |
| Lifetime practice (hours) | 12,710 (836–35,788) | 11,576 (628–34,192) | 11,005 (732–50,372) | $F(2,51) = 0.13, p = 0.876$ |

*Values are reported in mean ± standard deviation with range in parentheses.*

Sixteen non-musicians who had less than three years of musical experience and were not currently active musically were also tested. All non-musicians performed within the normal range on the online test for the evaluation of amusia (Peretz and Vuvan, 2017). Because musical aptitude may vary among non-musicians, we used a test of basic musical perception skills, the Musical Ear Test (Wallentin et al., 2010). Non-musicians obtained a mean of 72.2% correct (SD = 11.9) in the melody perception subtest and a mean of 72.7% correct (SD = 8.3) in the rhythm perception subtest. Their performance is comparable to the non-musicians in the original paper, with means of 69.7% (SD = 11.1) and 70.6% (SD = 8.0), respectively (Wallentin et al., 2010).

Other factors known to affect performance on behavioral tests and questionnaires, such as age, sex, and education, were matched across all groups (see **Table 1**). Most of the sample was Caucasian (48 out of 70). Seven out of 19 prodigies reported being of Asian ethnicity (South or East).

Due to time constraints and early changes in the protocol, there is missing data for one late-trained musician (Barcelona Music Reward Questionnaire), one early-trained musician (visual working memory), and one prodigy (motivation). Moreover, one prodigy and one late-trained musician were administered an abbreviated version of the IQ measure (WASI) instead of the full-scale IQ (WAIS-IV) because of time constraints. Accordingly, IQ index values are unavailable for these two participants. There are missing data for 8 participants on the measure of flow (2 prodigies, 4 early-trained, and 2 late-trained), because the measure was administered remotely and some did not reply.

## Materials and Procedure

### Online Questionnaire

Prior to their lab visit, participants completed an online questionnaire. The first section contained consent and demographics information. The online questionnaire also contained sections on absolute pitch, reward, motivation to play their instrument, and personality traits (see descriptions below). For participants who were minors, parents completed the consent form and demographics information; the remaining sections were completed by the participants themselves.

### Reward, Motivation, and Flow Questionnaires

The Barcelona Music Reward Questionnaire (BMRQ; Mas-Herrero et al., 2013) consists of 22 questions that assess reward associated to music in five dimensions: music seeking (e.g., *I'm always looking for new music*), emotion evocation (e.g., *I get emotional, listening to certain pieces of music*), mood regulation (e.g., *Music helps me chill out*), social reward (e.g., *Music makes me bond with other people*), and sensory-motor (e.g., *Music often makes me dance*). Answers were provided on a 5-point Likert scale, with 1 meaning *Completely disagree* and 5 meaning *Completely agree*.

To assess musicians' motivation to play their instrument, we selected items from the questionnaire of Desrochers et al. (2006) which did not exhibit floor or ceiling effects (i.e., with a rate equal or lower than 40% of extreme values). These items are listed in **Table 3**.

In addition, most participants filled a questionnaire assessing flow during musical practice, the Dispositional Flow Scale 2 (Jackson and Eklund, 2004). This questionnaire consists of 36 items assessing flow, using a 5-point scale (1 = *never* to 5 = *always*). The global score was obtained by calculating the mean score of all items. Examples of items, all following the statement "*When I practice my instrument…*," include: "*My*

**TABLE 3 |** Selected items to measure motivation.

| Item | Response scale | | |
|---|---|---|---|
| I play my instrument… | | | |
| Because I would feel guilty if I did not do it | Totally disagree | 1 2 3 4 5 | Totally agree |
| Because it adds something special to my personality | Totally disagree | 1 2 3 4 5 | Totally agree |
| What was the source of motivation when you began to play your instrument? | Completely internal | 1 2 3 4 5 | Completely external |

*attention is focused entirely on what I am doing*," "*I really enjoy the experience*," "*It feels like time goes by quickly*," "*I am challenged, but I believe my skills will allow me to meet the challenge*."

### Personality Traits

The Autism Spectrum Quotient (AQ; Baron-Cohen et al., 2001) consists of 50 items meant to measure five dimensions of the autistic profile: social skill (e.g., *I would rather go to a library than a party*), attention switching (e.g., *I prefer to do things the same way over and over again*), attention to detail (e.g., *I tend to notice details that others do not*; *I am fascinated by numbers*), communication (e.g., *I frequently find that I don't know how to keep a conversation going*), and imagination (e.g., *I find it difficult to imagine what it would be like to be someone else*). Answers were provided using a scale with four options: *definitely agree*, *slightly agree*, *slightly disagree* and *definitely disagree*.

The Big Five Inventory (John et al., 1991) contains 45 questions constructed to measure five different dimensions of personality: openness to experience (e.g., *Likes artistic and creative experiences*), conscientiousness (e.g., *Does things carefully and completely*), extraversion (e.g., *Is outgoing, sociable*), agreeableness (e.g., *Is considerate and kind to almost everyone*), and neuroticism (e.g., *Worries a lot*). Answers are provided using a 5-point Likert scale, with 1 meaning *Disagree strongly* and 5 meaning *Agree strongly*.

### Intellectual Quotient and Working Memory

Standardized tests of intellectual quotient (IQ) were administered to all participants. For musicians, the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2008) was administered to participants aged 17 or older, and the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV; Wechsler, 2003) was administered to participants aged 16 or younger. These batteries provide a global IQ score as well as 4 index scores: verbal comprehension, perceptual reasoning, working memory, and processing speed. For non-musicians, an abbreviated measure of IQ was used, the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) with 2 ($N$ = 10) or 4 subtests ($N$ = 15), which provide a full-2 or full-4 IQ score, respectively. WASI versions vary because the protocol was changed for time-saving purposes. The WASI was used for non-musicians because IQ is well known in the normal population, obviating the need for more extensive evaluation. Global IQ and indices for the WAIS-IV and WISC-IV were calculated using the summation of the subtests administered, and normed using the age-appropriate tables of the WAIS-IV, WISC-IV, and WASI. The mean in the normal population is 100 points, and one standard deviation corresponds to 15 points.

Since the WAIS-IV subtests of working memory are only auditory-verbal and because visual working memory could be involved in music learning (e.g., in sight-reading; Meinz and Hambrick, 2010), all participants completed a test of spatial working memory from the Wechsler Memory Scale, Third Edition (WMS-III; Wechsler, 2001). In this task, the experimenter points to blocks on a plank in a specific sequence, and the participant must point to them in the same order. The procedure is repeated with the instruction to point in the reverse order. The number of blocks increases until the participant errs. Raw scores (i.e., number of sequences correctly recalled) were calculated and used in the analyses, with higher scores indicating better spatial working memory.

The tests were administrated individually in a quiet, closed room on the campus of the University of Montreal.

## RESULTS

Prodigy status was reached at a mean age of 10.3 years ($SD$ = 1.8; $range$ = 7–13), after a mean of 5.4 years of musical experience ($SD$ = 1.3; $range$ = 3–8) and an accumulated average amount of practice of 2,364 h, although variability was large ($range$ = 187–7,357 h). Individual data are presented in **Supplementary Figure 1**. At the time of testing, prodigies accumulated a total amount of practice that did not differ statistically from their musician peers (**Figure 2**). They also reported more frequent practice in childhood than typical musicians (**Figure 3**). By the cut-off age of 14 for the status of prodigy, prodigies accumulated twice as much practice ($M$ = 4,563, $range$ = 702–13,252 h) as early-trained musicians ($M$ = 2,027, $range$ = 378–4,004 h).

Group differences in early practice were assessed using permutation analyses. Group attribution was shuffled across participants, and $t$-tests were calculated at each age. The maximum number of consecutive years that obtained a significant group difference ($p < 0.05$) was logged, and the process was repeated 1000 times to obtain a null distribution. The observed results (i.e., 9 years of consecutive, significant differences between prodigies and early-trained musicians; 6–14 years old) were less likely than 99.8% of results in the null distribution. A similar permutation test was conducted by comparing prodigies and late-trained musicians across ages with sufficient data (7–18 years of age). The observed result (i.e., group differences from age 7–10 inclusive or four consecutive years), was less likely than 96% of the null distribution (see gray boxes in **Figure 3**). These results provide further support that prodigies differed in their practice habits in childhood and early adolescence. Visualization of practice between 6 and 14 years old by individual (**Figure 4**) shows a large variability in the prodigies group, with around half of participants practicing as much as their age-matched peers, and half practicing more.

Since musicians started practicing at different ages, we also analyzed the data by year of musical experience (i.e., years since onset of experience; **Figure 5**). Using the permutation method outlined above, prodigies were found to accumulate more hours of practice than early-trained musicians from years 3–10 inclusive, thus for eight consecutive years, which corresponds to better performance than 99.5% of the null distribution. In contrast, prodigies did not practice more than late-trained musicians during any year when measured from onset of training.

Almost half of the musicians ($n$ = 23 of 54) reported having absolute pitch, with roughly half of that group ($n$ = 11) being prodigies. However, the proportion did not differ significantly across groups, with 58% of prodigies ($n$ = 11 of 19), 44% of early-trained ($n$ = 7 of 16), and 26% of late-trained musicians ($n$ = 5 of 19), $X^2(2, N = 54) = 3.89$, $p = 0.143$.

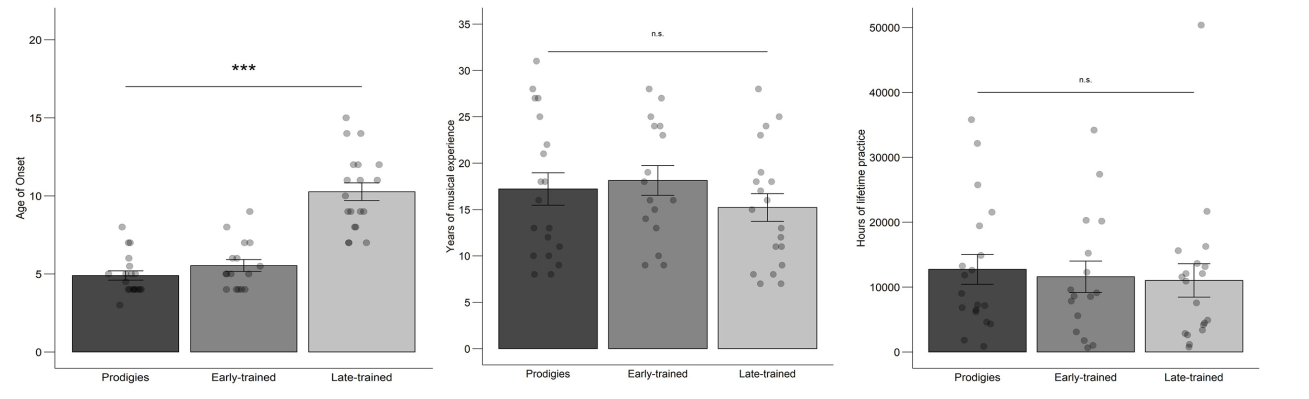



**FIGURE 2 |** Musical experience measures: mean, standard error and individual data by group. Points are jittered horizontally for visualization purposes.

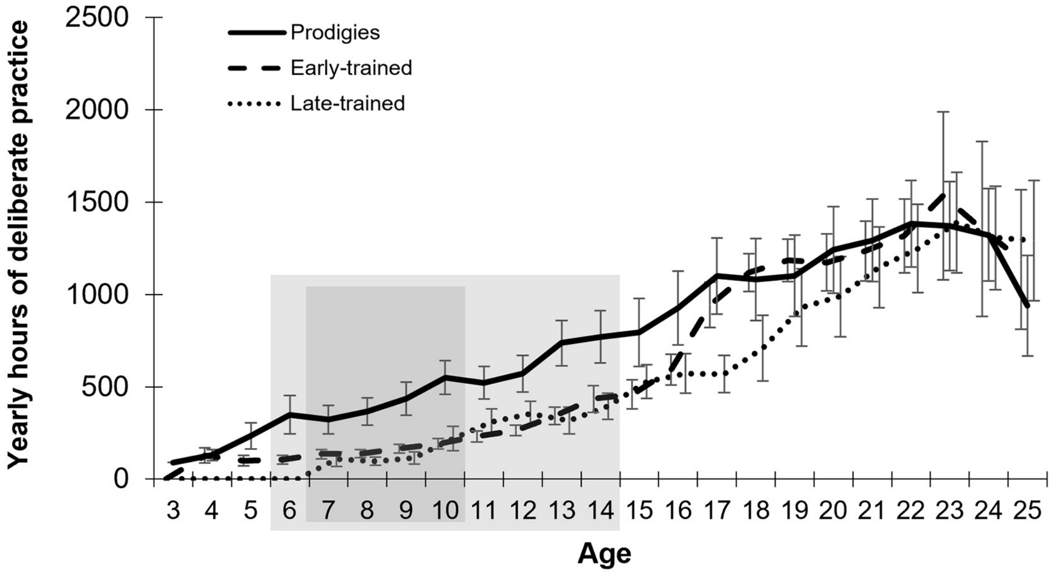



**FIGURE 3 |** Mean yearly amount and standard error of deliberate practice as a function of age, by group. Gray areas represent the longest stretches where permutation analysis showed a significant difference between prodigies and early-trained musicians (larger rectangle) and between prodigies and late-trained (smaller rectangle).

## Musical Reward and Motivation

Prodigies did not report finding music more rewarding than musicians or non-musicians. This was tested with an ANOVA computed on the BMRQ global score with group (prodigies, early-trained, late-trained, non-musicians) as a between-subjects factor, $F(3,65) = 1.14$, $p = 0.339$, $\eta^2 = 0.050$. ANOVAs were computed on the scores from each of the three motivation questions (**Table 3**), with group (prodigies, early-trained, late-trained) as a between-subjects factor. Responses to the motivation questions "*I play my instrument… Because I would feel guilty if I did not do it*" yielded no significant group effect, $F(2,50) = 1.36$, $p = 0.267$, $\eta^2 = 0.051$, and neither did responses to the question "*I play my instrument… Because it adds something special to my personality*", $F(2,50) = 0.40$, $p = 673$, $\eta^2 = 0.016$. However, responses to the question on the source of motivation when beginning to play their instrument showed a significant group effect [$F(2,50) = 4.48$, $p = 0.016$, $\eta^2 = 0.152$; **Figure 6**]. *Post hoc* pairwise comparisons using Welch's *t*-test (Bonferroni-Holm correction, three pairwise comparisons between groups) showed that prodigies ($M = 2.94$) reported a more external source of motivation when they started to play their instruments compared to late-trained musicians ($M = 1.74$), $t(26.45) = 2.90$, $p = 0.022$.

Global flow during music practice varied across groups (**Figure 7**), as shown by an ANOVA computed on the global flow score with group (prodigies, early-trained, late-trained) as a between-subjects factor, $F(2,43) = 3.62$, $p = 0.035$. *Post hoc* group comparisons showed that prodigies reported significantly more flow when they practice their instrument ($M = 3.8$, $SD = 0.5$) compared to early-trained musicians ($M = 3.3$, $SD = 0.5$,

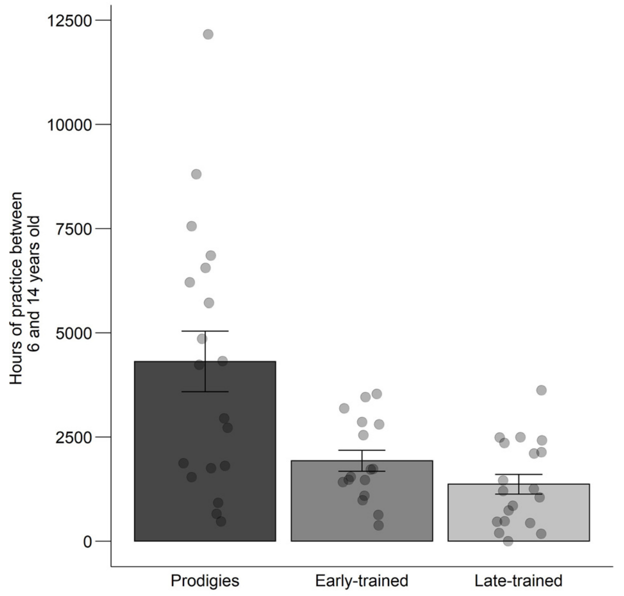



**FIGURE 4 |** Deliberate practice accumulated between 6 and 14 years old: mean, standard error and individual data, by group. Points are jittered horizontally for visualization purposes.

$p$ = 0.039, Bonferroni-Holm correction used for three pairwise comparisons between groups). Early-trained musicians did not differ significantly from late-trained musicians ($M$ = 3.7, $SD$ = 0.5, $p$ = 0.173).

## Personality Traits

There was no indication that prodigies, as a group, possessed more autistic traits than other musicians (**Figure 8**). The ANOVA computed on the AQ scores with group (prodigies, early-trained, late-trained, non-musicians) as a between-subjects factor and dimension (social, attention switching, attention to detail, communication, and imagination) as a within-subject factor did not reveal an effect of group, $F(3,66) = 1.28$, $p = 0.289$, $\eta_p^2 = 0.04$. A dimension effect was significant, $F(4,264) = 51.18$, $p < 0.001$, $\eta_p^2 = 0.44$, but there was no significant interaction with group, $F(12,264) = 1.37$, $p = 0.179$, $\eta_p^2 = 0.06$. Altogether, participants scored highest on the dimension of attention to detail (**Figure 8**, right panel). Despite the null result at the group level, there was an indication of higher prevalence of autistic traits among some individual prodigies. The three highest AQ scores (i.e., 29, 33, and 34) belonged to prodigies and one late-trained musician and may indicate clinically significant levels of autistic traits (i.e., the

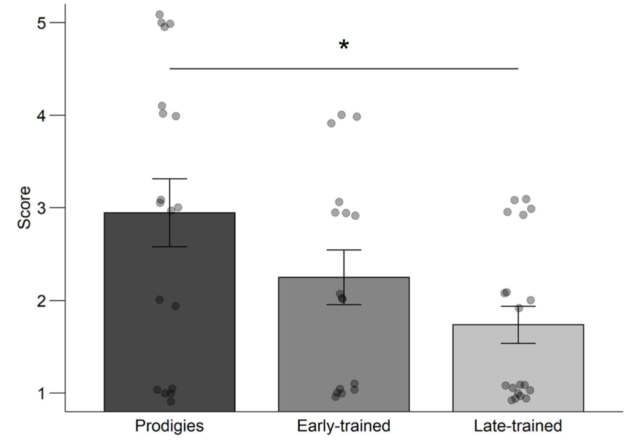



**FIGURE 6 |** Source of motivation when beginning to play: mean rating score, standard error and individual data, by group. Points are jittered for visualization purposes. See also row three of **Table 3**.

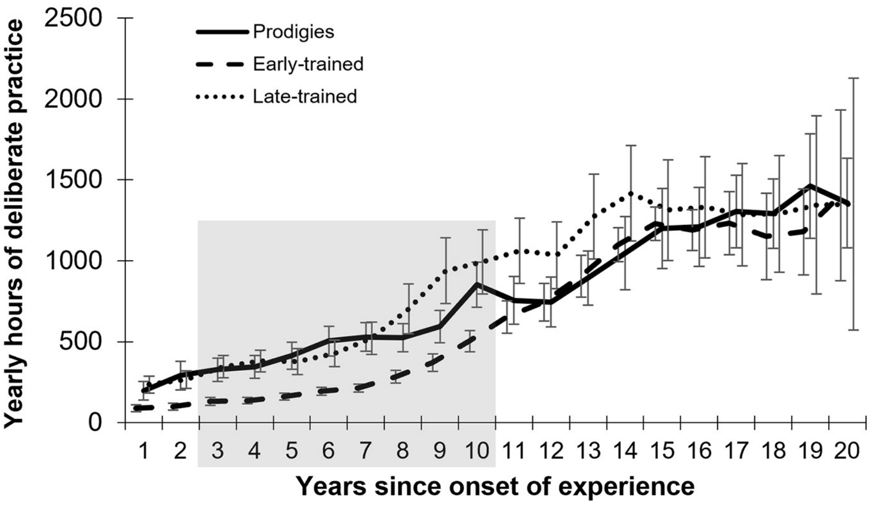



**FIGURE 5 |** Mean yearly amount of practice and standard error as a function of year since onset of musical experience, by group. The gray area represents the longest stretch where permutation analysis showed a significant difference between prodigies and early-trained musicians.

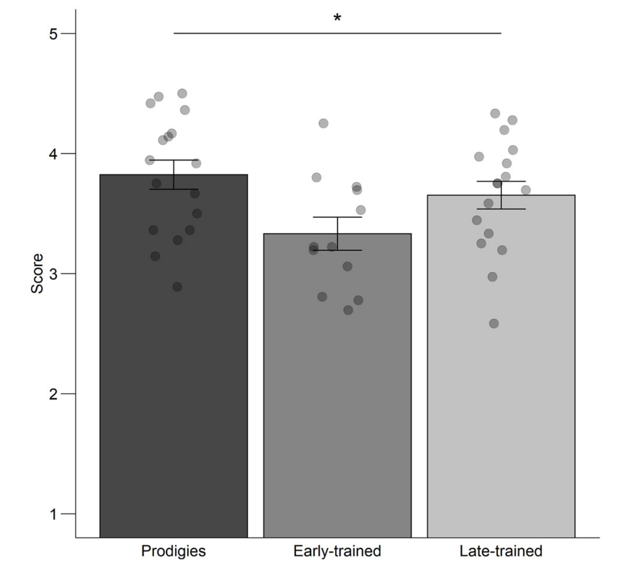



**FIGURE 7 |** Global flow: mean score, standard error and individual data, by group. Points are jittered horizontally for visualization purposes.

cut-off AQ score is 32; Baron-Cohen et al., 2001). Indeed, the disorder was formally diagnosed in two participants with AQ scores of 29 and 33 (see section "Participants").

For the Big Five Inventory, an ANOVA was computed on the mean score with group (prodigies, early-trained, late-trained, non-musicians) as a between-subjects factor and dimension or trait (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism) as a within-subject factor. The traits did not vary significantly by group, $F(3,66) = 1.92$, $p = 0.135$, $\eta_p^2 = 0.08$, and there was no interaction between group and traits, $F(12,264) = 0.74$, $p = 0.715$, $\eta_p^2 = 0.03$. However, there was a significant effect of trait, $F(4,264) = 44.66$, $p < 0.001$, $\eta_p^2 = 0.40$. Overall, participants tended to rate their openness, agreeableness, and conscientiousness high, and their extraversion and neuroticism low (**Figure 9**).

## Intellectual Quotient

Group mean IQ ranged from 113 to 120, which are above average but not exceptionally high considering that 95% of the adult participants had a university education. An ANOVA was computed on global IQ with group (prodigies, early-trained, late-trained, and non-musicians) as a between-subjects factor. There was no significant difference between groups, $F(3,66) = 1.78$, $p = 0.159$, $\eta^2 = 0.075$ (**Figure 10**), nor between musicians ($M = 116$) and non-musicians ($M = 118$), $t(36.21) = 1.08$, $p = 0.288$.

The IQ battery completed by musician participants included indices of verbal comprehension, perceptual reasoning, auditory-verbal working memory, and processing speed (**Figure 11**). An ANOVA was computed on the standardized individual index scores ($M = 100$, SD = 15, in the general population), with group (prodigies, early-trained, late-trained) as a between-subjects factor and index (verbal comprehension, perceptual reasoning, auditory-verbal working memory, and processing speed) as a within-subject factor. It revealed that verbal comprehension was better than working memory across groups, $F(3,147) = 6.00$, $p < 0.001$, $\eta_p^2 = 0.11$. The expected superiority of the prodigies was not significant in any index, as there was no group effect, $F(2,49) = 1.99$, $p = 0.147$, $\eta_p^2 = 0.08$, nor interaction between group and index, $F(6,147) = 0.75$, $p = 0.607$, $\eta_p^2 = 0.03$.

Visuo-spatial working memory, which was measured in all participants (grand mean = 19.5 of 26 trials, SD = 2.86), also did not differ according to group, $F(3,65) = 1.11$, $p = 0.350$, $\eta^2 = 0.049$, as revealed by an ANOVA with group (prodigies, early-trained, late-trained, non-musicians) as a between-subjects factor.

## Correlation With Early Musical Practice

Because early intensive practice is one of the factors that differentiated prodigies from the other musicians, we explored whether the individual amount of accumulated hours between age 6 and 14 was related to psychological traits measured here (i.e., 10 correlations; $p$-values adjusted using Bonferroni-Holm): global IQ, working memory index, processing speed index,

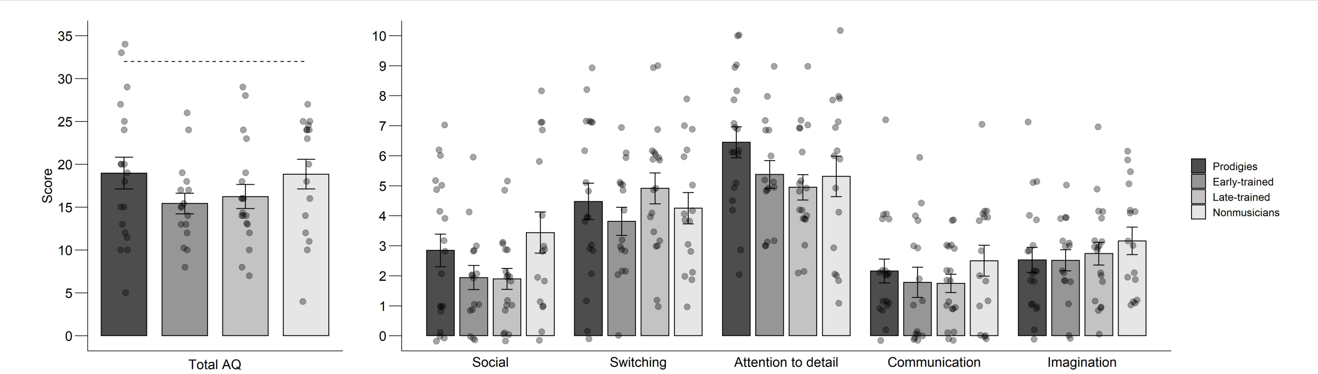



**FIGURE 8 |** Autism Spectrum Quotient (AQ) scores. In the left panel, mean scores, standard error and individual data for the total AQ score, by group. The dashed line indicates the cut-off score for clinically significant levels of autistic traits; In the right panel, mean scores, standard error and individual data by dimension and group. Points are jittered for visualization purposes.

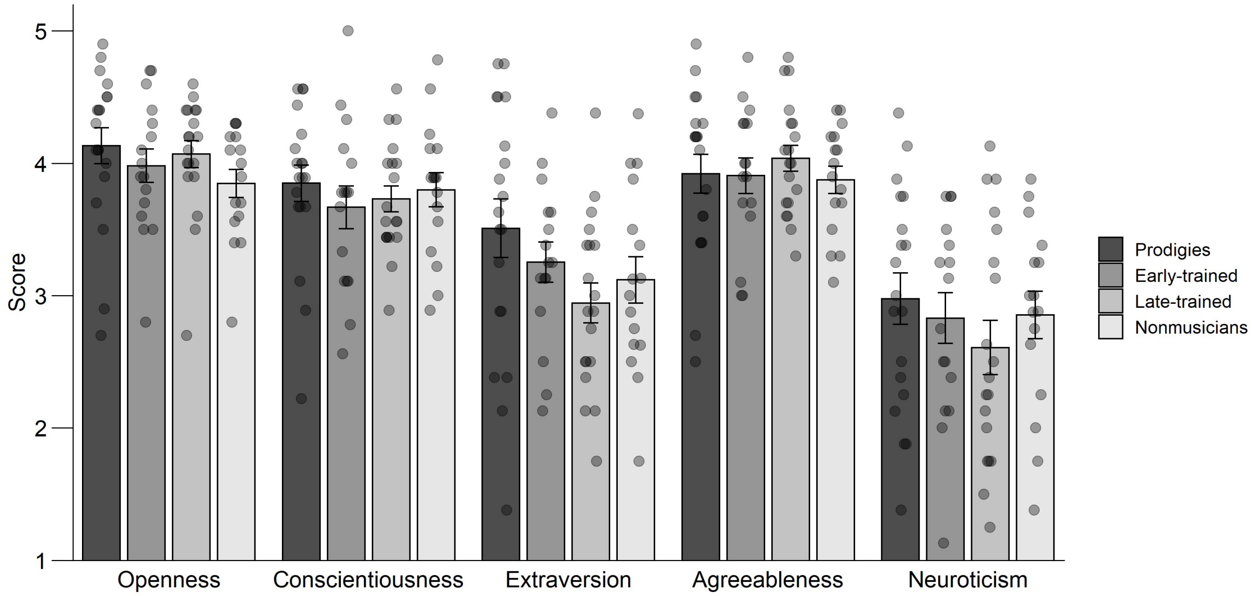



**FIGURE 9 |** Big Five Inventory: mean scores, standard error and individual data, by trait and group. Points are jittered horizontally for visualization purposes.

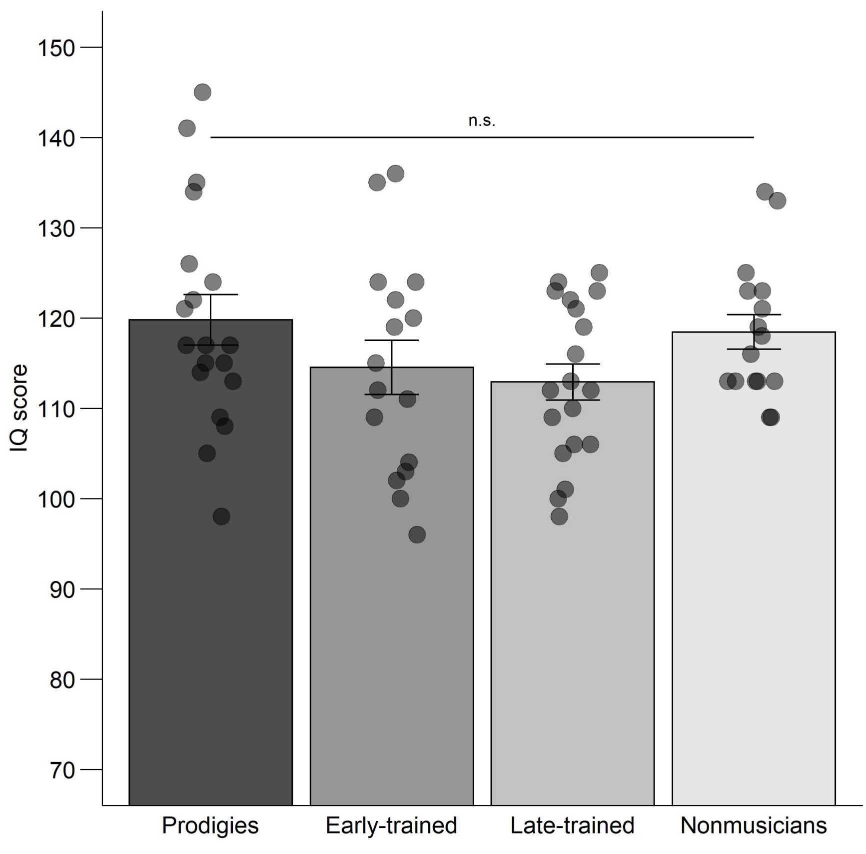



**FIGURE 10 |** Global IQ: mean scores, standard error and individual data, by group. Note that the mean score in the general population is 100 and one standard deviation is 15 points. Points are jittered horizontally for visualization purposes.

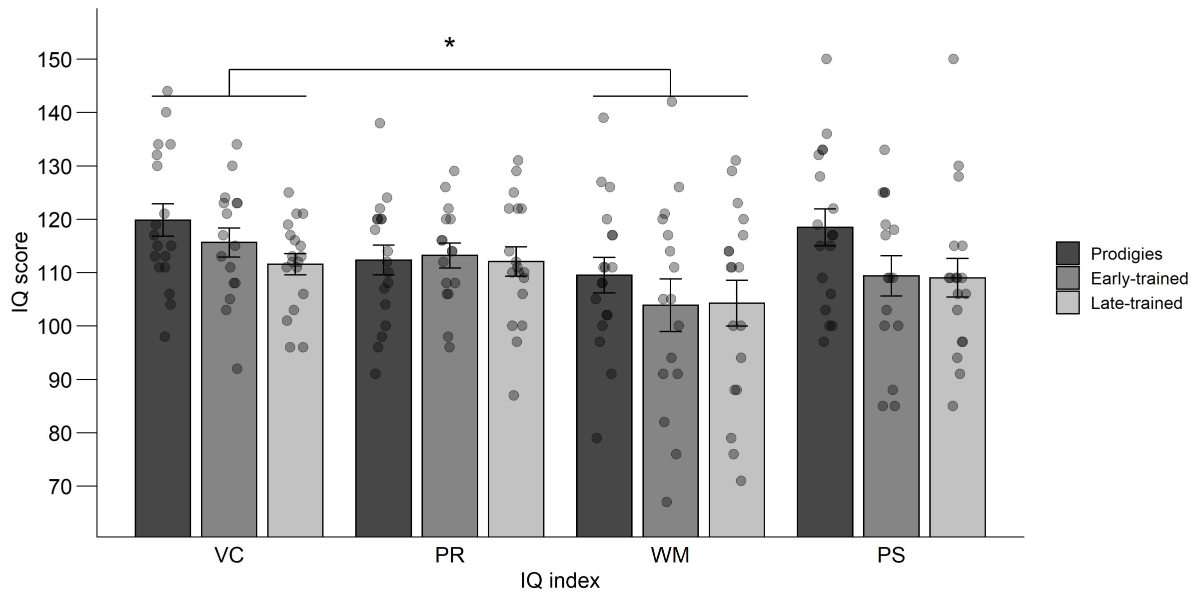



**FIGURE 11 |** Mean IQ scores, standard error and individual data for verbal comprehension (VC), perceptual reasoning (PR), working memory (WM), and processing speed (PS), by group. Points are jittered horizontally for visualization purposes.

openness to experience, conscientiousness, extraversion, music reward (BMRQ total score), autistic traits (AQ total score), attention to detail and flow. Individual amount of early practice varied considerably, especially among prodigies as mentioned previously, varying from 468–12,160 h accumulated between 6 and 14 years old. By comparison, early-trained musicians reported a range of 378–3,536 h and late-trained reported 0–3,623 h. The only trait to correlate significantly with the rate of early practice was extraversion, a dimension from the Big Five Inventory of personality, $r(52) = 0.47$, $p = 0.004$. While it appears at first glance that prodigies drive the correlation (**Figure 12**), separate correlation tests with only the prodigies ($r(17) = 0.46$, $p = 0.048$ [uncorrected]) or with only the non-prodigies (early-trained and late-trained musicians; $r(33) = 0.35$, $p = 0.040$ [uncorrected]), were significant as well. In other words, extraversion is generally correlated with amount of early practice.

## DISCUSSION

The current research examined the lifetime accumulated practice and psychological traits of musical prodigies to identify markers of their exceptionality (as described in the Multifactorial Gene-Environment Interaction Model; MGIM). Prodigies were compared with non-prodigies who began their musical training similarly early (around age 6), or later (around age 10), and non-musicians. Unlike previous studies of prodigies (e.g., Ruthsatz et al., 2014; Ruthsatz and Urbach, 2012; Comeau et al., 2018), our large sample of prodigies did not differ from other musicians in terms of intelligence, working memory, or personality, including autistic traits. Around half of the prodigies reported having absolute pitch, but the proportion of reported absolute pitch possessors did not vary significantly across groups. The characteristics that set prodigies apart were their report of more frequent practice early in life, a more external motivation to begin playing their instrument, as well as a higher tendency to experience flow during practice. Thus, models such as the MGIM are more appropriate to describe the prodigy phenomenon than the deliberate practice view. Moreover, these results suggest that prodigies are at the high extreme of the continuum of musicality rather than constituting a distinct category of musicians. Prodigies did not differ from the controls on most variables, and when they did differ, as in tendency to experience flow when practicing, their scores overlapped greatly with the ones of the other musicians.

Prodigies reported practicing twice as much as their peers from the age of 6 to 14. However, contrary to what could be expected by the deliberate practice view (Ericsson et al., 1993), there was substantial variability. Some prodigies did not practice more than their peers (**Figure 4**) and nevertheless reached higher levels of achievement. Our data also indicate that prodigies practice as much as late-trained musicians when measured from the onset of their musical experience. This means that when they begin to play their instrument, prodigies practice as much as children who are around 5 years older than themselves. Different factors could explain this phenomenon. For example, since in general, older children can better sustain their attention (Lin et al., 1999), it could be an indication that prodigies have a more advanced development of sustained attention. We speculate as well that, when they start to play, late-trained musicians must 'catch-up' to the other musicians who have already started, especially if they want to pursue a musical career. Musicians who start to play later in life might not have shown early signs of musical aptitude (i.e., predispositions) or a particular interest

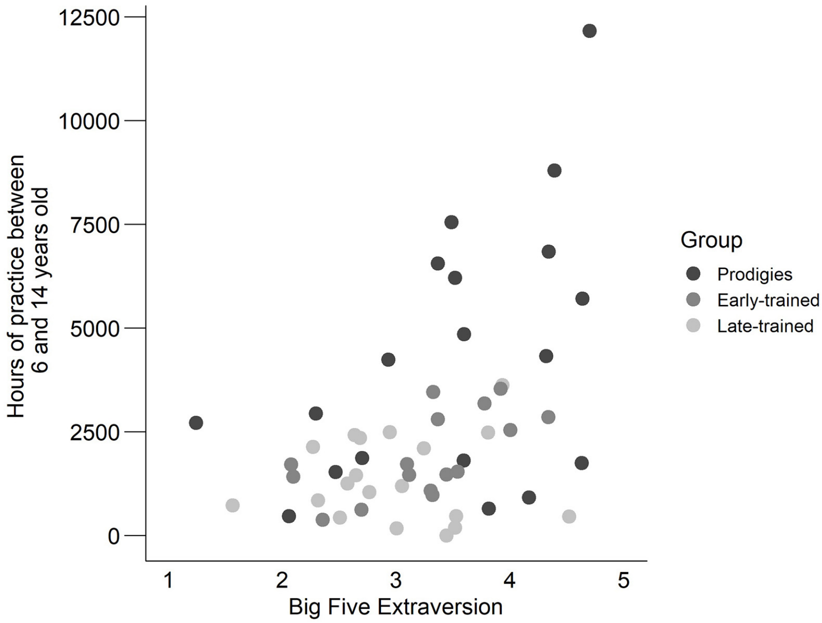



**FIGURE 12 |** Deliberate practice accumulated between 6 and 14 years old in relation to extraversion as measured by the Big Five Inventory, by group.

toward music. Those predispositions to easily learn music could also explain the fact that some of the prodigies did not practice more than their age-matched peers and still managed to reach exceptional levels of achievement early on.

Obviously, amount of practice is no guarantee of quality, and in fact there was considerable variability of early practice even in prodigies (**Figure 4**). Musicians' practice on a piece, for example, does not determine the evaluation of a newly learned piece by a jury (Williamon and Valentine, 2000). We propose that rate of progress would be a better index of the quality of practice. In our prior study of prodigies (Comeau et al., 2017), we noted that prodigies learned twice as fast as their peers, judging from the difficulty of musical pieces. For example, after 2.5 years of training, the prodigy Sarah Chang was capable of learning to play the Mendelssohn concerto on the piano whereas the typical pianist would only be capable after 10 years of training (Gagné and McPherson, 2016). Thus, prodigies not only practice more than their peers early on, but they also make more efficient use of their practice time.

Interestingly, prodigies reported that their source of motivation when beginning to play their instrument was more external compared to late-trained musicians, with early-trained musicians not significantly differing from either. Four prodigies but no early or late-trained controls reported the motivation being completely external (i.e., maximal rating). Parental investment might be one of the ingredients for fostering prodigiousness, but the relationship requires further study. For instance, parents may invest more time in response to the unusual behavior of their child. Highly invested parents have been suggested as playing a role in the development of their child's exceptional abilities (Feldman, 1993), but prodigies are also characterized as having an exceptional inner drive to master their work (Winner, 2000). The use of more comprehensive measures of motivation, for example the complete questionnaire from which we selected individual questions (Desrochers et al., 2006) or interviews with children and parents, could help clarify the sources of motivation in young prodigies and musicians.

Besides parental influences, other factors may account for their distinctive practice behavior. Prodigies were more likely to report flow during musical practice compared to early-trained musicians. Since practice requires high levels of concentration, which is hard to maintain for young children (Lin et al., 1999), any factor that influences the inherent pleasure of the activity could influence motivation to continue. Future research could measure experience of flow directly after a practice session, as well as physiological correlates (Harmat et al., 2015), rather than self-report as used here (Butkovic et al., 2015).

Autistic traits are associated with genetic factors (Miles, 2011), yet autism does not seem to characterize most musical prodigies. Only two of the 19 prodigies met the criteria for clinically significant autistic traits based on their responses to the Autism Spectrum questionnaire (Baron-Cohen et al., 2001) or formal diagnosis. Thus, we found no evidence that autism is a relevant candidate disorder in the search for common genes explaining exceptional achievements. Personality, in contrast, may play a small role. We found that the more a child practiced before adolescence, the more extraverted they reported to be. Extraversion might influence practicing indirectly because it could motivate participation in stage arts (Ullén et al., 2016). However, we note that success in competitions was a selection criterion used here and elsewhere (Ruthsatz and Urbach, 2012) for considering a child as a musical prodigy and musical prodigies were no more extraverted than other participants in our sample.

The early advantage in learning for prodigies appears to be limited to music. We found no evidence of superior intelligence or exceptional working memory in prodigies compared to other musicians, nor did we observe heightened cognitive abilities in musicians compared to non-musicians. The latter finding is in line with a recent meta-analysis obtaining no evidence for a causal effect of musical training on general cognitive abilities (Sala and Gobet, 2020). Even though most prodigies in our sample (68%) were tested as adults, age of testing does not necessarily undermine the findings. Longitudinal studies show stability of IQ scores from 6 years onward (see Yu et al., 2018; for a review).

In summary, we found that early intense practice characterizes musical prodigies during early childhood, a time when the brain is most plastic (Herholz and Zatorre, 2012). Because pre-existing differences in the recruitment of brain regions involved in auditory encoding and motor control predict success in learning to play an instrument, we may expect prodigies to be born with pre-existing differences in these brain networks. Such predispositions may be amplified by early and sustained practice. Future research should aim to identify the anatomical and functional properties of brain networks that affect exceptional learning rate and achievement. Researchers should also try to recruit prodigies while they are children, in order to better measure the traits and behaviors associated with the prodigy phenomenon as it unfolds.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, upon reasonable request.

## ETHICS STATEMENT

The studies were reviewed and approved by the Comité d'Éthique de la Recherche en Arts et en Sciences (CÉRAS), University of Montreal, Montreal, Canada. Written informed consent to participate in this study was provided by the participants or their legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

CM and IP contributed equally in the project's conception. CM, MS, and IP participated in the study design. CM performed the literature search and drafted the manuscript. CM and MW performed the statistical analysis. MW and MS provided the critical revisions. IP performed the final revisions. All authors contributed to the article and approved the submitted version.

## FUNDING

The work is funded by the program of Canada Research Chairs (IP). CM is supported by a fellowship from the Natural Sciences and Engineering Research Council of Canada. MW is supported by the Fonds de Recherche du Quebec – Nature et Technologies.

## ACKNOWLEDGMENTS

We would like to thank the research assistants who contributed to the data collection: Margot Charignon, Lucie-Maud Ménard, and Kathya Carrier. We give special thanks to our collaborator Jean-François Rivest for lively discussions and referral of talented musicians.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.566373/full#supplementary-material

## REFERENCES

Ackerman, P. L. (2014). Nonsense, common sense, and science of expert performance: talent and individual differences. *Intelligence* 45, 6–17. doi: 10.1016/j.intell.2013.04.009

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the autism spectrum quotient (AQ). *Pers. Individ. Diff.* 38, 451–460. doi: 10.1016/j.paid.2004.04.022

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism spectrum quotient: evidence from asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17. doi: 10.1023/A:1005653411471

Butkovic, A., Ullén, F., and Mosing, M. A. (2015). Personality related traits as predictors of music practice: underlying environmental and genetic influences. *Pers. Individ. Diff.* 74, 133–138. doi: 10.1016/j.paid.2014.10.006

Campitelli, G., and Gobet, F. (2011). Deliberate practice: necessary but not sufficient. *Curr. Dir. Psychol. Sci.* 20, 280–285. doi: 10.1177/0963721411421922

Comeau, G., Lu, Y., Swirp, M., and Mielke, S. (2018). Measuring the musical skills of a prodigy: a case study. *Intelligence* 66, 84–97. doi: 10.1016/j.intell.2017.11.008

Comeau, G., Vuvan, D. T., Picard-Deland, C., and Peretz, I. (2017). Can you tell a prodigy from a professional musician? *Music Percept.* 35, 200–210. doi: 10.1525/MP.2017.35.2.200

Dalla Bella, S., Sowi, J., Farrugia, N., and Berkowska, M. (2016). "Igor: a case study of a child drummer prodigy igor: a case study of a child drummer prodigy," in *Musical Prodigies: Interpretations from Psychology, Education, Musicology, and Ethnomusicology*, ed. G. E. McPherson (Oxford: Oxford University Press), 391–408.

Desrochers, A., Comeau, G., Jardaneh, N., and Green-demers, I. (2006). L'élaboration d'une échelle pour mesurer la motivation chez les jeunes élèves en piano. *Rech. Éduc. Music.* 24, 13–33.

Ericsson, K. A., Krampe, R. T., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychol. Rev.* 100, 363–406. doi: 10.1037/0033-295X.100.3.363

Feldman, D. H. (1993). Prodigies: a distinctive form of giftedness. *Gift. Child Q.* 37, 188–193.

Gagné, F., and McPherson, G. E. (2016). "Analyzing musical prodigiousness using Gagné's integrative model of talent development," in *Musical Prodigies: Interpretations from Psychology, Education, Musicology, and Ethnomusicology*, Vol. 1, ed. G. E. McPherson (Oxford: Oxford University Press), 3–114. doi: 10.1093/acprof:oso/9780199685851.003.0001

Harmat, L., de Manzano, Ö, Theorell, T., Högman, L., Fischer, H., and Ullén, F. (2015). Physiological correlates of the flow experience during computer game playing. *Int. J. Psychophysiol.* 97, 1–7. doi: 10.1016/j.ijpsycho.2015.05.001

Herholz, S. C., and Zatorre, R. J. (2012). Musical training as a framework for brain plasticity. *Neuron* 76, 486–502. doi: 10.1016/j.neuron.2012.10.011

Jackson, S. A., and Eklund, R. C. (2004). *The Flow Scales Manual*. Morgantown, WV: Fitness Information Technology Inc.

John, O. P., Donahue, E. M., and Kentle, R. L. (1991). *The Big Five Inventory—Versions 4A and 54*. Berkeley, CA: University of California.

Lin, C. C., Hsiao, C. K., and Chen, W. J. (1999). Development of sustained attention assessed using the continuous performance test among children 6–15 years of age. *J. Abnorm. Child Psychol.* 27, 403–412. doi: 10.1023/A:1021932119311

Macnamara, B. N., Hambrick, D. Z., and Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: a meta-analysis. *Psychol. Sci.* 25, 1608–1618. doi: 10.1177/0956797614535810

Marin, M. M., and Bhattacharya, J. (2013). Getting into the musical zone: trait emotional intelligence and amount of practice predict flow in pianists. *Front. Psychol.* 4:853. doi: 10.3389/fpsyg.2013.00853

Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., and Rodriguez-Fornells, A. (2013). Individual differences in music reward experiences. *Music Percept.* 31, 118–138. doi: 10.1525/mp.2013.31.2.118

Meinz, E. J., and Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: the role of working memory capacity. *Psychol. Sci.* 21, 914–919. doi: 10.1177/0956797610373933

Miendlarzewska, E. A., and Trost, W. J. (2014). How musical training affects cognitive development: rhythm, reward and other modulating variables. *Front. Neurosci.* 7:279. doi: 10.3389/fnins.2013.00279

Miles, J. H. (2011). Autism spectrum disorders–a genetics review. *Genet. Med.* 13, 278–294. doi: 10.1097/gim.0b013e3181ff67ba

Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-halkola, R., and Ullén, F. (2014). Practice does not make perfect: no causal effect of music practice on music ability. *Psychol. Sci.* 25, 1795–1803. doi: 10.1177/0956797614541990

Nakamura, J., and Csikszentmihalyi, M. (2009). "Flow theory and research," in *Oxford Handbook of Positive Psychology*, eds C. R. Snyder and S. J. Lopez (Oxford: Oxford University Press), 752.

Peretz, I., and Vuvan, D. T. (2017). Prevalence of congenital amusia. *Eur. J. Hum. Genet.* 25, 625–630. doi: 10.1038/ejhg.2017.15

Ruthsatz, J., and Detterman, D. K. (2003). An extraordinary memory: the case study of a musical prodigy. *Intelligence* 31, 509–518. doi: 10.1016/S0160-2896(03)00050-3

Ruthsatz, J., and Urbach, J. B. (2012). Child prodigy: a novel cognitive profile places elevated general intelligence, exceptional working memory and attention to detail at the root of prodigiousness. *Intelligence* 40, 419–426. doi: 10.1016/j.intell.2012.06.002

Ruthsatz, J., Ruthsatz-Stephens, K., and Ruthsatz, K. (2014). The cognitive bases of exceptional abilities in child prodigies by domain: similarities and differences. *Intelligence* 44, 11–14. doi: 10.1016/j.intell.2014.01.010

Sala, G., and Gobet, F. (2020). Cognitive and academic benefits of music training with children: a multilevel meta-analysis. *Mem. Cogn.* doi: 10.3758/s13421-020-01060-2

Schellenberg, E. G., and Weiss, M. W. (2013). "Music and cognitive abilities," in *The Psychology of Music*, ed. D. Deutsch (Cambridge, MA: Elsevier Academic Press), 499–550. doi: 10.1016/B978-0-12-381460-9.00012-2

Swaminathan, S., and Schellenberg, E. G. (2019). "Music training and cognitive abilities: associations, causes, and consequences," in *The Oxford Handbook of Music and the Brain*, eds M. H. Thaut and D. A. Hodges (Oxford: Oxford University Press). doi: 10.1093/oxfordhb/9780198804123.013.26

Ullén, F., Hambrick, D. Z., and Mosing, M. A. (2016). Rethinking expertise: a multifactorial gene–environment interaction model of expert performance. *Psychol. Bull.* 142, 427–446. doi: 10.1037/bul0000033

Ullén, F., Mosing, M. A., and Madison, G. (2015). Associations between motor timing, music practice, and intelligence studied in a large sample of twins. *Ann. N. Y. Acad. Sci.* 1337, 125–129. doi: 10.1111/nyas.12630

Wakabayashi, A., Baron-Cohen, S., and Wheelwright, S. (2006). Are autistic traits an independent personality dimension? a study of the autism-spectrum quotient (AQ) and the NEO-PI-R. *Pers. Individ. Diff.* 41, 873–883. doi: 10.1016/j.paid.2006.04.003

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., and Vuust, P. (2010). The musical ear test, a new reliable test for measuring musical competence. *Learn. Individ. Diff.* 20, 188–196. doi: 10.1016/j.lindif.2010.02.004

Wechsler, D. (1999). *Manual for the Wechsler Abbreviated Intelligence Scale (WASI)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2001). *Echelle Clinique de Memoire de Weschler MEM III (WMS-III)*. Paris: Les Editions Du Centre de Psychologie Appliquee.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-(WISC-IV) Technical and Interpretive Manual*, 4th Edn. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale (WAIS–IV)*, 4th Edn, Vol. 22. San Antonio, TX: NCS Pearson, 498.

Williamon, A., and Valentine, E. (2000). Quantity and quality of musical practice as predictors of performance quality. *Br. J. Psychol.* 91, 353–376. doi: 10.1348/000712600161871

Winner, E. (2000). The origins and ends of giftedness. *Am. Psychol.* 55, 159–169. doi: 10.1037//0003-066X.55.1.159

Yu, H., Betsy McCoach, D., Gottfried, A. W., and Gottfried, A. E. (2018). Stability of intelligence from Infancy through adolescence: an autoregressive latent variable model. *Intelligence* 69, 8–15. doi: 10.1016/j.intell.2018.03.011

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

OPINION
published: 08 January 2021
doi: 10.3389/fpsyg.2020.557807

# Music Lessons and Cognitive Abilities in Children: How Far Transfer Could Be Possible

*Franziska Degé**

*Department of Music, Max-Planck-Institute for Empirical Aesthetics, Frankfurt, Germany*

**Keywords: music lessons, music training, children, transfer, cognitive abilities**

OPEN ACCESS

***Edited by:***
*Marta Olivetti Belardinelli, Sapienza University of Rome, Italy*

***Reviewed by:***
*Aline Moussard, Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal (CRIUGM), Canada*

****Correspondence:***
*Franziska Degé*
*franziska.dege@ae.mpg.de*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology*

***Received:*** *30 April 2020*
***Accepted:*** *27 November 2020*
***Published:*** *08 January 2021*

***Citation:***
*Degé F (2021) Music Lessons and Cognitive Abilities in Children: How Far Transfer Could Be Possible. Front. Psychol. 11:557807. doi: 10.3389/fpsyg.2020.557807*

## INTRODUCTION

For many years there have been an interest in cognitive transfer effects and very controversial debates about its existence. Often, transfer is referred to as a special case of learning, but it is also absolutely possible to characterize transfer as something that is inextricably linked to learning and so happens regularly coupled with the learning process (for an overview see Klauer, 2010). Transfer is a learning or exercise effect that goes beyond the primary effect of practice. Its effects can be positive or negative. In learning and practicing processes both main and side effects can arise. Main effects can be observed in the learned domain and are summarized under the term trivial learning transfer (often also near transfer). Side effects affect domains that have been neither practiced nor learned, and these are called nontrivial learning transfer or far transfer. The main characteristic that differentiates between trivial and nontrivial transfer effects is their particular difference between the original learning situation and the transfer situation: a low difference for trivial transfer and a high difference for nontrivial transfer. The higher the distance or the range of the transfer, the greater its qualitative assessment and the less frequently it takes place. However, it is possible that nontrivial transfer occurs and Taatgen (2016) provides a very helpful framework to understand how this could happen. You can imagine the whole idea as a kind of overlap between two skills in small elements that are used to fulfill a task. People train specific cognitive skills and—as a by-product—they train a general cognitive skill as well. This so-trained general cognitive skill can be helpful for other specific skills. For example, a child learns a specific instrument (e.g., piano), and while practicing the finger movements that are important to play piano, which is a very specific skill, it also trains selective attention (a general skill that is also useful for other cognitive tasks). In Taatgen's theory, this is the case because of the small elements of information processing that can be reused. The basic parts of Taatgen's model are the so-called primitive information processing elements (PRIMs). These elements can be combined to build a PRIM rule and then PRIM rules can be compiled to include more PRIM rules and form a composite PRIM rule, and ultimately if all rules are put together, they build the rule for carrying out an operator. Since the PRIM rules refer to very small processing steps, they are independent of the overarching goal of a specific skill. In terms of transfer, this means that the operators for one task (e.g., the multicolumn addition) can be reused completely and others partially for another task (e.g., multicolumn multiplication).

With this approach it gets clearer that a complex pattern of information exchange can still be relatively independent of the specific skill or task and therefore also that specific skills that seem to be different can allow transfer. Moreover, it counts not only completely similar operators that can be shared but also partially shared ones. Therefore, also the transfer between skills that have no perfectly similar operators can be explained. Furthermore, this model gives us insight into what happens in early (only few repetitions of a task) vs. late transfer (many repetitions of a task). When you start learning a specific skill, transfer to another skill (i.e., early transfer) might be higher, because there might be a large overlap between operators (operators are not as task specific in early transfer). The longer you train a specific skill, the more you get proficient in things that concern task control and are highly specific for a particular skill. Hence, being an expert in something might limit the potential of transfer (late transfer) due to the high specificity of operators. This might be interesting in light of the paradox Schellenberg (2011) pointed out earlier in the discussion about transfer: Children taking music lessons enhance their intelligence and that musicians who make music for years are slightly better in some cognitive abilities than nonmusicians but not geniuses.

Additionally, information provided by Mähler and Stern (2006) about specific circumstances (use of mental tools, metacognitive control, analogies, and individual motivation) that might facilitate the occurrence of transfer effects supports the idea that far transfer (nontrivial transfer) might take place. Music lessons offer such specific circumstances (like training the use of mental tools or giving the opportunity to realize that deliberate practice is helpful for music learning and the next math test) that result in far-transfer effects. Above and beyond those facilitating circumstances, I argue that music lessons are special, because they are highly adaptive which might also promote transfer effects.

## MUSIC LESSONS AND COGNITIVE ABILITIES

Indeed, several studies reported an effect of music lessons on cognitive abilities in children (Schellenberg, 2004; Moreno et al., 2011; Bugos and DeMarie, 2017; Frischen et al., 2019). These beneficial (side) effects of music lessons were small with respect to effect sizes, but long-lasting (Schellenberg, 2006b) and music specific, as control groups receiving sports training (Degé and Schwarzer, 2011), painting lessons (Moreno et al., 2009), or drama lessons (Schellenberg, 2004) did not show them.

It is far beyond the scope of this article to review all literature about music lessons, music training, musical abilities, and cognitive abilities. Hence, I will highlight three important controversial issues.

First, it seems striking that numerous areas of cognition appear to improve with music lessons and/or music aptitude (Schellenberg and Weiss, 2013). For some of these cognitive abilities, a connection to music lessons seems obvious (Schellenberg, 2011), like for listening skills (Corrigall and Trainor, 2009) or sensorimotor functions (Costa-Giomi, 2005). Other specific links, e.g., to language abilities (Patel, 2008), visual–spatial abilities (Rauscher and Zupan, 2000), mathematical abilities (Bahr and Christensen, 2000), or memory (Roden et al., 2014) are less obvious. The diversity of links raises doubts about their specificity. Schellenberg (2011), taking a generalist view, suggests that an association between music lessons and a domain-general cognitive ability might explain all reported specific associations. Indeed, there is evidence that music lessons are associated with domain-general cognitive abilities like IQ (Schellenberg, 2004, 2006b) and executive functions (Degé et al., 2011). Thus, it seems likely that fewer specific links than have been suggested actually exist. It might be possible that for example the less obvious links mentioned above are the ones that are explained by a domain-general ability rather than by a specific link.

Second, the available evidence for specific as well as domain-general connections between music lessons and cognition is mostly based on correlational or quasi-experimental studies that do not provide evidence for causation. Most studies therefore cannot state unequivocally whether it is music lessons that influence cognitive abilities or whether highly skilled individuals are more inclined to take music lessons (i.e., preexisting differences). Due to their readiness of mind, these high-functioning children perform well on all tests, while still having time to take music lessons, because their day-to-day school activities (e.g., homework) are less time-consuming. Such preexisting differences seem like an appealing and parsimonious explanation. So far, preexisting differences in variables like socioeconomic status (SES), IQ, or personality have been considered. For example, children who take music lessons have a higher SES than children who do not (Sergeant and Thatcher, 1974). However, most studies controlled for SES either by randomization or by statistical means, which calls this explanation into question. Concerning IQ, it has been suggested that highly intelligent children are more likely to take music lessons than other children (Schellenberg, 2011). However, children who have decided, but have not yet started, to learn an instrument did not differ in brain structure, cognitive abilities, motor abilities, or musical abilities from children who did not intend to take music lessons (Norton et al., 2005). Regarding personality, studies with children and adults revealed that amount of music lessons was significantly positively correlated with openness to experience, conscientiousness, and sometimes agreeableness (Corrigall et al., 2013; Corrigall and Schellenberg, 2015). As with SES, it is important for (future) studies to control for potential differences in personality. Some studies already did that and still found an association between music lessons and cognitive abilities (e.g., memory) (Degé and Schwarzer, 2017). Hence, preexisting differences cannot fully explain how music lessons promote cognitive abilities. Furthermore, they fail to explain the findings of a small number of convincing and comprehensive experimental studies that have reported small but significant effects of music lessons on IQ (Schellenberg, 2004), inhibition (Moreno et al., 2011; Bugos and DeMarie, 2017; Frischen et al., 2019), or phonological awareness (Degé and Schwarzer, 2011).

Third, this small impact is supported by some meta-analyses. Regarding spatial skills, meta-analyses found an effect of music listening and music lessons (Hetland, 2000a,b). Other meta-analyses revealed either a strong correlation between music training and reading skills, but no reliable results for experimental studies (Butzlaff, 2000), or a modest but significant positive effect of music training on reading skills (Standley, 2008). However, when reading related outcome measures were investigated separately, a small impact of music training on phonological awareness (but not reading fluency) could be demonstrated (Gordon et al., 2015). Two more studies that investigated effects of music lessons on a number of cognitive domains rather than isolated aspects yielded conflicting results. While Benz et al. (2015) reported small enhancing effects on various cognitive abilities, Sala and Gobet (2017) found no influence of music lessons on cognitive abilities at all. What might be critical about the conclusion of Sala and Gobet (2017) is their viewpoint on all cognitive aspects as one. If you have a closer look at single cognitive abilities, the effect sizes are higher than for an overall measure. This might speak in favor of more solid transfer effects between music lessons and particular cognitive abilities. Moreover, small effect sizes for nontrivial transfer effects seem relatively plausible, because of their higher transfer distance. However, they still might be important. Additionally, they report that studies in this field of research could be done by more proficiently adopting a more sophisticated methodology. Although I would agree in general that for a lot of studies there is room to develop, I am wondering how far it is reasonable to use those studies to draw firm conclusions about far-transfer effects of music lessons on cognitive abilities. I think it would be rather plausible to conclude that it needs better studies to base future meta-analysis on, instead of concluding there is no effect at all.

All in all, I would summarize the field of music research by assuming that there are some small but interesting transfer effects of music lessons and cognitive abilities and that it is important to understand how they work in order to broaden our knowledge about them.

There are some frameworks or theoretical approaches that might help to understand transfer effects of music training. I will briefly mention two of them (OPERA hypothesis and multimodal integration) before discussing an explanation that seems for me very seminal to understand transfer effects of music lessons in children in a different way. The OPERA hypothesis by Patel (2011) was originally developed for the link between music and language. Patel (2011) has formulated a kind of framework that generally explains the different factors that are involved in the emergence of relationships between music and language. The five OPERA factors Overlap, Precision, Emotion, Repetition, and Attention are features of music or music making that facilitate the occurrence of transfer effects of music to language. However, this music-and-language framework might partly also hold true for transfer in general. Additionally, in a more neurological approach the idea of multimodal integration due to the multisensory activation of making music is put forward. Music making relies on multiple sensory modalities and the simultaneous integration of multisensory information (e.g., auditory, visual, and somatosensory) is needed to monitor progress and success. Hence, music practice is associated with structural and functional changes in the brain (Jäncke, 2009a). These changes can occur because neural plasticity leads to use-dependent regional growth and structural adaptation in response to intense environmental demands. Gaser and Schlaug (2003), for example, reported differences in gray matter distributions in musicians, amateur musicians, and nonmusicians in motor regions, auditory regions, and visual regions. The activation of many systems in parallel, on the one hand, results in brain changes and, on the other hand, might foster more general learning transfer (Green and Bavelier, 2008).

Both approaches consider the complex demands of music making. However, I would like to introduce an approach that emphasizes a slightly different point.

Among other explanations of potential transfer effects of music training, there is an interesting approach that might help to understand the impact of music lessons on cognitive abilities in children: the *qualitatively different experience explanation* (Schellenberg, 2006a). It is assumed that children who take music lessons make qualitatively different experiences compared to children not taking music lessons. Music lessons offer specific experiences (e.g., reading musical notation, doing auditory discrimination, memorizing musical notation and auditory passages, training of fine-motor skills, and gaining knowledge about musical structure), which train specific skills (e.g., executive functions). Regular practice of all these skills in childhood may have a positive effect on cognitive development. In short, cognitive benefits may stem from these "musically trained" abilities. What questions this postulation is the fact that some of these trained abilities are not specific to music (e.g., training of fine-motor skills); therefore other nonmusical extracurricular activities training such abilities should yield similar effects on cognition. According to Schellenberg (2004, 2006b), this was not the case. These findings point toward something special about music lessons that other nonmusical extracurricular activities cannot provide, but as yet, it is unclear what this is.

What I would like to stress a bit further are the implications of the qualitatively different experience hypothesis. It postulates that a limited set of abilities is trained by music lessons. I would argue that the set of skills subsumed under executive functions strongly overlaps with the set of abilities trained in music lessons. Music making requires executive functions (Jäncke, 2009b) and seems to have the potential to train them (Bugos et al., 2007; Bialystok and DePape, 2009; Degé et al., 2011; Moreno et al., 2011; Roden et al., 2014; Slevc et al., 2016; Jaschke et al., 2018). This idea is strikingly appealing, because it would represent a parsimonious explanation for the general and specific links between music lessons and cognition. Schellenberg and Peretz (2008) already postulated that executive functions could explain the link between music lessons and IQ, because music lessons train executive functions, and these are involved in nearly all tasks (Hannon and Trainor, 2007). This mediating hypothesis was tested and is possibly true (Degé et al., 2011), but see Schellenberg (2011). In this respect I would even go further and postulate that most of the potential associations between music lessons and cognition in children regardless of being general (i.e., IQ) or specific could be explained by executive functions (Degé et al., 2017). Furthermore, this idea is helpful in speculating what could be so special about music.

## DISCUSSION: ADAPTABILITY

Getting back to the issue of whether music is special in any regard, I would not argue that other nonmusical extracurricular activities cannot improve cognition (e.g., tennis) (Ishihara et al., 2017) but that music lessons may be more efficient. The important difference between other activities and music lessons might be their remarkable adaptability, i.e., the fit of difficulty level between a musical piece and a students' ability: Music is normally taught in one-on-one settings or small groups, and teachers attempt to adapt their teaching to the ability of the music student(s). Moreover, assignments for daily practice are typically doable and designed to become incrementally harder in adaptation to the student's progress. Importantly, research not focusing on music (Klingberg et al., 2002) has suggested that successful enhancement of an ability, in their case executive functions, depends on the adaptability of the applied training. Hence, the specific "something" about music lessons might be their extremely adaptive nature. Other leisure time activities might not be as adaptive due to larger groups (e.g., soccer), not as flexible assignments (e.g., drama lessons with a given date for a performance), and fit of difficulty level (arts lessons in which it takes time until the feedback of mastery, i.e., the picture, is finished). As already mentioned, for successful music making, executive functions are important (Jäncke, 2009b). They are needed for monitoring progress, practicing music in an organized and disciplined way, inhibiting competing motor impulses, and switching to new rules when accidentals are inserted, or switching from playing one instrument to another. Hence, in music lessons, executive functions are trained in an adaptive way as well, which may be a particularly effective way of training according to Klingberg et al. (2002).

Thinking about it in a broader sense, I would put forward that it is the highly adaptive nature of music making that gives rise to general transfer/far transfer. This adaptability ensures that the student is more or less constantly trained in the zone of proximal development, the best *window* for successful learning: This zone is defined as the distance between a child's current developmental level as determined by independent problem solving and the next higher level as determined by problem solving under adult guidance or in collaboration with more capable peers (Wygotski, 1978, p. 86). In the Vygotskian framework, successful learning is slightly ahead of development (adapted to an individual's abilities) and takes place most effectively in the zone of proximal development.

I would like to highlight the factors that make it so likely for musical learning to consistently take place in this developmentally desirable zone. First, new milestones are often reached during music lessons: A good music teacher fits the to-be-learned materials to the learner's current skill level in a way that something new is mastered with the help of the music teacher within the lessons (i.e., a milestone). While for learning this milestone the teacher is needed, the mastery of it will (hopefully) take place at home during independent practice of the new skills. This training routine, when mediated by small group sizes and good teachers, ensures learning in the zone of proximal development. Second, what helps to keep music making in the zone of proximal development is its special form of feedback. Student and teacher receive immediate auditory feedback about their skills. If you like it or not, you hear immediately whether you have mastered a new piece or not. Third, the immediate feedback gives the opportunity to work on an issue directly as it comes up.

Taken together, the adaptability and the immediate feedback that allow for training in the zone of proximal development might be what makes music lessons special. This might be what enables them to affect cognition (e.g., executive functions) in a far-transfer kind of fashion.

It is highly debated if general transfer or general learning exists, and if yes, under what conditions. Green and Bavelier (2008) pointed out that general learning can take place if the learning paradigm is complex and "real world" enough, like video game training (needs to be games that are complex and not predictive, engaging, and difficult enough; so not every game would have this effect), athletic training, or musical training. They proposed that in addition, task difficulty, motivation and arousal, feedback, and variability are important contributors to general learning. I would in general agree to their ideas but argue that adaptability and its consequences might be the most important factor for enhancing effects of training. Adaptability can be directly related to their conditions: It can keep a task complex and difficult (always training in the zone of proximal development), it can produce motivated and aroused learners (oriented on the mastery of the next step), and it is based on feedback as well as variability. Video game training and sports training share an adaptive nature with music training to a certain extent, video gaming probably even more, but music training is more or less unique in its high amount of adaptability.

## AUTHOR CONTRIBUTIONS

The content of the paper was developed by FD. The manuscript was written by FD.

## ACKNOWLEDGMENTS

The author would like to thank Ulrike Frischen and Tina Roeske for helpful comments on earlier versions of this manuscript.

## REFERENCES

Bahr, N., and Christensen, C. A. (2000). Inter-domain transfer between mathematical skills and musicianship. *J. Struct. Learn. Intell. Syst.* 14, 187–197.

Benz, S., Sellaro, R., Hommel, B., and Colzato, L. S. (2015). Music makes the world go round: the impact of musical training on non-musical cognitive functions-a review. *Front. Psychol.* 6:2023. doi: 10.3389/fpsyg.2015.02023

Bialystok, E., and DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 565–574. doi: 10.1037/a0012735

Bugos, J., and DeMarie, D. (2017). The effects of a short-term music program on preschool children's executive functions. *Psychol. Music* 45, 855–867. doi: 10.1177/0305735617692666

Bugos, J., Perlstein, W. M., McCrae, C. S., Brophy, T. S., and Bedenbaugh, P. H. (2007). Individualized piano instruction enhances executive functioning and working memory in older adults. *Aging Mental Health* 11, 464–471. doi: 10.1080/13607860601086504

Butzlaff, R. (2000). Can music be used to teach reading? *J. Aesthetic Educ.* 34, 167–178. doi: 10.2307/3333642

Corrigall, K. A., and Schellenberg, E. G. (2015). Predicting who takes music lessons: parent and child characteristics. *Front. Psychol.* 6, 282. doi: 10.3389/fpsyg.2015.00282

Corrigall, K. A., Schellenberg, E. G., and Misura, N. M. (2013). Music training, cognition, and personality. *Front. Psychol.* 4, 222. doi: 10.3389/fpsyg.2013.00222

Corrigall, K. A., and Trainor, L. J. (2009). Effects of musical training on key and harmony perception. *Ann. N. Y. Acad. Sci.* 1169, 164–168. doi: 10.1111/j.1749-6632.2009.04769.x

Costa-Giomi, E. (2005). Does music instruction improve fine motor abilities? *Ann. N. Y. Acad. Sci.* 1060, 262–264. doi: 10.1196/annals.1360.053

Degé, F., Frischen, U., and Schwarzer, G. (2017). "Der Zusammenhang zwischen Musikunterricht und spezifischen kognitiven Fähigkeiten bei 9- bis 12-jährigen Kindern: Die mediierende Rolle der exekutiven Funktionen," in *Paper Presented at the Jahrestagung der Deutschen Gesellschaft für Musikpsychologie* (Hamburg).

Degé, F., Kubicek, C., and Schwarzer, G. (2011). Music lessons and intelligence: a relation mediated by executive functions. *Music Percept.* 29, 195–201. doi: 10.1525/mp.2011.29.2.195

Degé, F., and Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Front. Psychol.* 2, 124. doi: 10.3389/fpsyg.2011.00124

Degé, F., and Schwarzer, G. (2017). Music lessons and verbal memory in 10- to 12-year-old children: Investigating articulatory rehearsal as mechanism underlying this association. *Psychomusicol. Music Mind Brain* 27, 256–266. doi: 10.1037/pmu0000201

Frischen, U., Schwarzer, G., and Degé, F. (2019). Comparing the effects of rhythm-based music training and pitch-based music training on executive functions in preschoolers. *Front. Integr. Neurosci*. 13:41. doi: 10.3389/fnint.2019.00041

Gaser, C., and Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *J. Neurosci.* 23, 9240–9245. doi: 10.1523/JNEUROSCI.23-27-09240.2003

Gordon, R. L., Fehd, H. M., and McCandliss, B. D. (2015). Does music training enhance literacy skills? A meta-analysis. *Front. Psychol.* 6, 1777. doi: 10.3389/fpsyg.2015.01777

Green, C. S., and Bavelier, D. (2008). Exercising your brain: a review of human brain plasticity and training-induced learning. *Psychol. Aging* 23, 692–701. doi: 10.1037/a0014345

Hannon, E. E., and Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends Cogn. Sci.* 11, 466–470. doi: 10.1016/j.tics.2007.08.008

Hetland, L. (2000a). Learning to make music enhances spatial reasoning. *J. Aesthetic Educ.* 34, 179–238. doi: 10.2307/3333643

Hetland, L. (2000b). Listening to music enhances spatial-temporal reasoning: evidence for the Mozart Effect. *J. Aesthetic Educ.* 34, 105–148. doi: 10.2307/3333640

Ishihara, T., Sugasawa, S., Matsuda, Y., and Mizuno, M. (2017). Improved executive functions in 6-12-year-old children following cognitively engaging tennis lessons. *J. Sports Sci.* 35, 2014–2020. doi: 10.1080/02640414.2016.1250939

Jäncke, L. (2009a). Music drives brain plasticity. *Biol. Rep.* 1:78. doi: 10.3410/B1-78

Jäncke, L. (2009b). "Planung, Handlung und Selbstkontrolle der frontalkortex und die musik," in *Pauken mit Trompeten* (Bonn; Berlin: Bundesministerium für Bildung und Forschung), 88–97.

Jaschke, A. C., Honing, H., and Scherder, E. J. A. (2018). Longitudinal analysis of music education on executive functions in primary school children. *Front. Neurosci.* 12, 103. doi: 10.3389/fnins.2018.00103

Klauer, K. J. (2010). *Transfer des Lernens: Warum Wir Oft Mehr Lernen Als Gelehrt Wird*. Stuttgart: Kohlhammer.

Klingberg, T., Forssberg, H., and Westerberg, H. (2002). Training of working memory in children with ADHD. *J. Clin. Exp. Neuropsychol.* 24, 781–791. doi: 10.1076/jcen.24.6.781.8395

Mähler, C., and Stern, E. (2006). "Transfer," in *Handwörterbuch Pädagogische Psychologie (3. überarb. u. erweit. Aufl. ed.)*, ed D. H. Rost (Weinheim: PVU), 782–793.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., and Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychol. Sci.* 22, 1425–1433. doi: 10.1177/0956797611416999

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb Cortex* 19, 712–723. doi: 10.1093/cercor/bhn120

Norton, A., Winner, E., Cronin, K., Overy, K., Lee, D. J., and Schlaug, G. (2005). Are there pre-existing neural, cognitive, or motoric markers for musical ability? *Brain Cognit.* 59, 124–134. doi: 10.1016/j.bandc.2005.05.009

Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford: University Press.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychol.* 2:142. doi: 10.3389/fpsyg.2011.00142

Rauscher, F. H., and Zupan, M. A. (2000). Classroom keyboard instruction improves kindergarten children's spatial-temporal performance: a field experiment. *Early Childhood Res. Q.* 15, 215–228. doi: 10.1016/S0885-2006(00)00050-8

Roden, I., Grube, D., Bongard, S., and Kreutz, G. (2014). Does music training enhance working memory performance? Findings from a quasi-experimental longitudinal study. *Psychol. Music* 42, 284–298. doi: 10.1177/0305735612471239

Sala, G., and Gobet, F. (2017). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Curr. Dir. Psychol. Sci.* 26, 515–520. doi: 10.1177/0963721417712760

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychol. Sci.* 15, 511–514. doi: 10.1111/j.0956-7976.2004.00711.x

Schellenberg, E. G. (2006a). "Exposure to music: the truth about the consequences," in *The Child as Musician: A Handbook of Musical Development*, ed G. E. McPherson (New York, NY: Oxford University Press), 111–134.

Schellenberg, E. G. (2006b). Long-term positive associations between music lessons and IQ. *J. Educ. Psychol.* 98, 457–468. doi: 10.1037/0022-0663.98.2.457

Schellenberg, E. G. (2011). Examining the association between music lessons and intelligence. *Br. J. Psychol.* 102, 283–302. doi: 10.1111/j.2044-8295.2010.02000.x

Schellenberg, E. G., and Peretz, I. (2008). Music, language and cognition: unresolved issues. *Trends Cogn. Sci.* 12, 45–46. doi: 10.1016/j.tics.2007.11.005

Schellenberg, E. G., and Weiss, M. W. (2013). "Music and cognitive abilities," in *The Psychology of Music*, eds D. Deutsch (Amsterdam: Elsevier), 499–550.

Sergeant, D., and Thatcher, G. (1974). Intelligence, social status and musical abilities. *Psychol. Music* 2, 32–57. doi: 10.1177/030573567422005

Slevc, R. L., Davey, N. S., Buschkuehl, M., and Jaeggi, S. M. (2016). Tuning the mind: exploring the connections between musical ability and executive functions. *Cognition* 152, 199–211. doi: 10.1016/j.cognition.2016.03.017

Standley, J. M. (2008). Does music instruction help children to learn to read? Evidence of a meta-analysis. *Update Appl. Res. Music Educ.* 27, 17–32. doi: 10.1177/8755123308322270

Taatgen, N. A. (2016). "Theoretical Models of Training and Transfer Effects," in *Cognitive Training: An Overview of Features and Applications*, eds T. Strobach and J. Karbach (Springer International Publishing), 19–29.

Wygotski, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 25 January 2021
doi: 10.3389/fnhum.2020.566841

# Influence of Background Musical Emotions on Attention in Congenital Amusia

Natalia B. Fernandez[1,2*], Patrik Vuilleumier[1,2], Nathalie Gosselin[3,4] and Isabelle Peretz[3,4]

[1] Laboratory of Behavioral Neurology and Imaging of Cognition, Department of Fundamental Neuroscience, University of Geneva, Geneva, Switzerland, [2] Swiss Center of Affective Sciences, Department of Psychology, University of Geneva, Geneva, Switzerland, [3] International Laboratory for Brain, Music and Sound Research, University of Montreal, Montreal, QC, Canada, [4] Department of Psychology, University of Montreal, Montreal, QC, Canada

OPEN ACCESS

**Edited by:**
*Franco Delogu, Lawrence Technological University, United States*

**Reviewed by:**
*Anne Caclin, Institut National de la Santé et de la Recherche Médicale (INSERM), France*
*Caicai Zhang, The Hong Kong Polytechnic University, Hong Kong*

***Correspondence:**
*Natalia B. Fernandez*
*Natalia.fernandez@unige.ch*

**Specialty section:**
*This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience*

**Received:** *28 May 2020*
**Accepted:** *30 November 2020*
**Published:** *25 January 2021*

**Citation:**
*Fernandez NB, Vuilleumier P, Gosselin N and Peretz I (2021) Influence of Background Musical Emotions on Attention in Congenital Amusia. Front. Hum. Neurosci. 14:566841. doi: 10.3389/fnhum.2020.566841*

Congenital amusia in its most common form is a disorder characterized by a musical pitch processing deficit. Although pitch is involved in conveying emotion in music, the implications for pitch deficits on musical emotion judgements is still under debate. Relatedly, both limited and spared musical emotion recognition was reported in amusia in conditions where emotion cues were not determined by musical mode or dissonance. Additionally, assumed links between musical abilities and visuo-spatial attention processes need further investigation in congenital amusics. Hence, we here test to what extent musical emotions can influence attentional performance. Fifteen congenital amusic adults and fifteen healthy controls matched for age and education were assessed in three attentional conditions: executive control (distractor inhibition), alerting, and orienting (spatial shift) while music expressing either joy, tenderness, sadness, or tension was presented. Visual target detection was in the normal range for both accuracy and response times in the amusic relative to the control participants. Moreover, in both groups, music exposure produced facilitating effects on selective attention that appeared to be driven by the arousal dimension of musical emotional content, with faster correct target detection during joyful compared to sad music. These findings corroborate the idea that pitch processing deficits related to congenital amusia do not impede other cognitive domains, particularly visual attention. Furthermore, our study uncovers an intact influence of music and its emotional content on the attentional abilities of amusic individuals. The results highlight the domain-selectivity of the pitch disorder in congenital amusia, which largely spares the development of visual attention and affective systems.

**Keywords: congenital amusia, emotion, executive control, music exposure, selective attention**

## INTRODUCTION

Music is prevalent in the modern environments of our daily life. Frequently used in concomitance with mental or motor routine activities such as working out, reading a book, or driving a car, music exposure is also present in many public places, such as restaurants or shops. Although music can be distracting (Kämpfe et al., 2011; Silvestrini et al., 2011), it also has the capacity to enhance cognitive or physical activities in various populations, including Alzheimer's and Parkinson's disease patients, elderly with fall risks (Trombetti et al., 2011; Hars et al., 2013), as well as healthy people (Thompson et al., 2001; Trost et al., 2014; Fernandez et al., 2019b). For instance, music exposure can enhance visuo-spatial attention (Rowe et al., 2007; McConnell and Shore, 2011; Trost et al., 2014; Fernandez et al., 2019b), an essential cognitive ability involving allocating processing resources to specific goal-relevant sensory information.

The beneficial effect of music exposure on visuo-spatial attention may be at least partly related to the emotions music conveys. Previous research has demonstrated the importance of affective states in influencing attention allocation (Mitchell and Phillips, 2007; Vanlessen et al., 2016). Notably, positive affect has frequently been associated with a broader scope of attention (Fredrickson, 2004) which might in turn impair selective attention due to reduced selectivity (Rowe et al., 2007). During music exposure, both positive valence and high arousal may play a similar role in enhancing visuo-spatial information processing (McConnell and Shore, 2011; Trost et al., 2014; Fernandez et al., 2019b).

We (Fernandez et al., 2019b) previously investigated the impact of exposure to music evoking different emotions, including joy, tension, tenderness, and sadness, on the deployment of selective attention processes using a classic visuo-spatial attention network test (ANT) (Fan et al., 2002). In line with other results (Trost et al., 2014), we found that when control participants were exposed to highly arousing background music, especially pleasant music, their performance on the test improved, as revealed by faster target detection in the presence of distractors and greater engagement of fronto-parietal areas (Fernandez et al., 2019b) which is associated with top-down attentional control (Corbetta and Shulman, 2002). These findings are consistent with the notion that rhythmic stimuli can stimulate physiological systems including attentional processes, through their beat structure, especially when the target appears in synchrony with the strong beats of the presented music (Escoffier et al., 2010; Bolger et al., 2013; Trost et al., 2014). The effect of music exposure on selective attention resources is probably mediated by an entrainment of brain rhythms and induced changes in emotional state.

Critically, the above-mentioned studies were conducted using a sample from the general population without regard to individual musical abilities. Notably, people are not equal when it comes to music abilities. Although most people develop normal musical skills, 1.5 to 4% of the population (Peretz and Vuvan, 2017) may suffer from a genetic, music-specific neurodevelopment disorder called congenital amusia (Peretz et al., 2002; Hyde and Peretz, 2004). This musical disorder can be separated into two variants: the most common pitch-based form (also referred as "pitch deafness") and the more recently described time-based form (also referred as "beat deafness") (Phillips-Silver et al., 2013; Peretz and Vuvan, 2017). In the pitch-based form of congenital amusia, the focus of this study, individuals (*amusics* hereafter) present a dysfunction in the fine-grained processing of the pitch structure of music which plays a fundamental role in developing a normal musical system and in fully experiencing music's subtleties (Peretz, 2016). At the cortical level, congenital amusia is associated with neural anomalies affecting both functional and structural connectivity in fronto-temporal networks of the right hemisphere (Hyde et al., 2011; Peretz, 2016). At the behavioral level, congenital amusia is characterized by a selective impairment in the perception and production of very small (<2 semitones) variations in pitch (Hyde and Peretz, 2004). This pitch deficit can affect several musical tasks, such as singing in tune (Dalla Bella et al., 2009), perceiving dissonance (Ayotte et al., 2002; Cousineau et al., 2012), and recognizing familiar melodies without the aid of lyrics (Ayotte et al., 2002). Although this pitch deficit does not usually appear alongside any other psychoacoustic deficits, many amusics also experience difficulties with rhythm (Ayotte et al., 2002), especially in the presence of pitch variation (Foxton et al., 2006; Phillips-Silver et al., 2013).

While the emotional information conveyed in music and speech largely depends on pitch (among other acoustical features), studies investigating the relationship between emotional sensitivity and pitch deficits present in congenital amusia show contrasting results. Previous studies reported that amusics could still correctly recognize the emotion content expressed by music (Ayotte et al., 2002; Gosselin et al., 2015; Jiang et al., 2017) or by speech prosody (Ayotte et al., 2002; Hutchins et al., 2010) despite their musical impairment. In most cases, amusics were found to rely on alternative acoustic features (e.g., tempo, timbre, or roughness) to correctly distinguish musical emotions (Cousineau et al., 2012; Gosselin et al., 2015; Marin et al., 2015). However, other work observed mild impairments in the discrimination of emotions in music or in speech but with preserved intensity judgements (Lévêque et al., 2018; Pralus et al., 2019), or moderately reduced capacities to discriminate emotional prosody in speech (Thompson et al., 2012; Lolli et al., 2015; Lima et al., 2016) as compared with control participants. In addition to suggesting that pitch processing impacts musical emotion perception, the latter findings could also be explained by recent evidence linking congenital amusia to music-specific disturbances in consciousness. Because of these heterogeneous findings on emotion recognition in amusics and their inability to create conscious representations of pitch, it remains unclear whether amusics may still respond to the presence of affective music exposure in other conditions that rely on more implicit processing (Tillmann et al., 2016; Lévêque et al., 2018; Pralus et al., 2019), or whether their perception of emotional content in music may be dampened by their limited musical resources when attentional demands are focused on other stimuli. In the latter case, amusics' perceptual system might be unable to extract relevant affective cues from music due to disrupted pitch processing, and therefore fail to produce the indirect effects of musical emotions on other cognitive functions, such as those observed in attentional processing. Thus, testing attentional effects of musical emotions in amusics would allow us not only to better characterize the extent of musical deficits in this population, but also to clarify the possible role of pitch processing in driving these effects.

Finally, although congenital amusia deficits are thought to occur without any other (cognitive) impairments, several theories have suggested an intimate connection between music and visual-spatial abilities, particularly linking sound frequency representation with spatial codes (e.g., lower pitch is usually represented lower in space than higher pitch) (Rusconi et al., 2006). In line with this assumption, enhanced attentional processing has been reported in musicians compared to non-musicians (Brochard et al., 2004; Sluming et al., 2007) including better executive control performance measured with an ANT paradigm (Medina and Barraza, 2019). However, the relationship between music abilities and visuo-spatial attention is unclear in musically (or visually) impaired people. While Douglas and Bilkey (2007) linked poor performance on a classic mental rotation task in amusic individuals to their deficits in processing contour components, other studies found preserved performance in a similar rotation task in amusics (Tillmann et al., 2010; Williamson et al., 2011). These divergent findings further motivate our aim to better characterize the relationship between different visuo-spatial attention components and musical capacities in amusia.

Given these gaps in the current literature, we here assess to what extent congenital amusia deficit might interfere with the indirect (implicit) effects of musical emotions on selective attentional processes. To address this question, amusic and control participants performed the Attentional Network task (ANT; Fan et al., 2002) mentioned above while they were exposed to music communicating four emotional expressions (differentially organized along both arousal and valence dimensions), in addition to a silent condition. In one single task, the ANT probes three distinct components of selective visuo-spatial processing, namely, executive control, alerting, and orienting (Posner and Petersen, 1990; Petersen and Posner, 2012). Executive control is characterized as the ability to selectively attend to specific information by filtering out concurrent distractors. Alerting is defined as the ability to maintain a highly reactive state toward sensory stimuli, while orienting involves the ability to change the focus of attention and direct it to a specific feature or location of stimuli. To our knowledge, the present study is the first to evaluate the effects of several properties of music exposure on distinct components of attention in individuals presenting the pitch-based form of congenital amusia, particularly concerning the emotional aspects of music and their underlying arousal and valence dimensions, and also more generally the first to probe for any link between musical and visuo-spatial attention abilities in this population.

## MATERIALS AND METHODS

### Participants

Fifteen amusic participants meeting the criteria for the pitch-based form of congenital amusia and fifteen controls matched for education level and musical education duration took part in the study. Participants were mainly right-handed. Participants' characteristics are provided in **Table 1**.

Prior to being selected for participation in this study, the participants were tested on their musical abilities with the online test of amusia (Peretz and Vuvan, 2017), the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003), and the Pitch-Change Detection task (Hyde and Peretz, 2004), all reliable tools to identify amusic individuals (Vuvan et al., 2018). The online test is composed of three tasks, namely the scale test, the off-beat test, and the out-of-key test. Standard testing with the MBEA comprises the same scale test and additional contour, interval, rhythm, meter, and memory tests (Peretz et al., 2003; Vuvan et al., 2018). A melodic composite score was computed by averaging the scale, contour, and interval values measured with the MBEA. Finally, the pitch-change detection task evaluates the severity of the pitch deficit. This task assesses the participant's accuracy in detecting a pitch change of the fourth tone in a five-tone sequence. Here, the pitch-change detection scores represent the detection accuracy for the smallest pitch change in the task, i.e., a pitch change of a quarter semitone (25 cents), which is the most discriminant change (Hyde and Peretz, 2004). All amusic participants included in the present study scored below cut-off scores for both the scale test (22/30) and the melodic composite test (21.4/30), except for one amusic who scored below cut-off on the scale test but slightly above cut-off on the melodic composite test. These cut-off scores (i.e., 2SD below the mean of a normative sample) were chosen in accordance with latest normative data (Peretz and Vuvan, 2017; Vuvan et al., 2018) and used as inclusion criteria. All control participants presented normal music abilities, while all amusic participants scored below the cut-off, indicating the presence of pitch deficits.

All amusic participants had normal non-verbal reasoning and verbal working memory abilities as assessed by the Matrix Reasoning and the Digit Span tests from the WAIS-III (Wechsler Adult Intelligence Scale; Wechsler et al., 1997). All participants

**TABLE 1 |** Participants' characteristics and musical abilities, measured with the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003) and the pitch change detection task (Hyde and Peretz, 2004).

| Participants' characteristics | Amusic group ($n$ = 15) | Control group ($n$ = 15) |
|---|---|---|
| Gender (*n*) | 10F, 5M | 11F, 4M |
| Age (y) | 59.3 ± 19.0 | 57.2 ± 18.7 |
| Handedness (*n*) | 13R, 1L, 1A | 14R, 1A |
| Education (y) | 17.0 ± 3.4 | 16.0 ± 3.2 |
| Musical education (y) | 1.0 ± 1.0 | 2.0 ± 2.0 |
| **Musical abilities** | | |
| MBEA–scale (22[a]/30[b]) | 17.0 ± 2.0 | 28.0 ± 1.2 |
| MBEA–melodic composite score (21.4[a]/30[b]) | 17.8 ± 2.2 | 27.6 ± 1.4 |
| 25-cents pitch-change detection (%) | 32.7 ± 24.7 | 93.0 ± 7.2 |

*For both amusic and control participants separately, the mean group values are presented with the standard deviation, except for gender information (number of female and male participants). N, number of participants; y, years; F, female; M, male; R, right-handed; L, left-handed; A, Ambidextrous; %percentage correct; [a]cut-off value; [b]maximum score.*

had normal or corrected-to-normal vision, had no hearing deficits, and no psychiatric, neurological, or toxicological history.

Only non-musicians were recruited according to the following criteria: (i) no music education/practice before 10 years old, and (ii) no current/past regular music practice for a duration of over 5 years. All participants provided informed written consent in accordance with the regulations of the Research Ethics Council for the Faculty of Arts and Sciences at the Université de Montréal.

## Auditory Material

Twelve pieces of instrumental classical music validated in previous work (Trost et al., 2012; Fernandez et al., 2019b) were used in the current study and categorized into four emotions, namely joy, tension, tenderness, and sadness. These four emotions are organized along orthogonal dimensions of arousal (i.e., *Relaxing–Stimulating*) and valence (i.e., *Unpleasant–Pleasant*) (Trost et al., 2012) and represent the major emotion types identified in the Geneva Emotional Music Scale (Zentner et al., 2008). Hence, both joy and tension are typically associated with highly arousing ratings, while sadness and tenderness are defined as low-arousing. Orthogonally, joy and tenderness are categorized as positively valenced, while tension and sadness are negatively valenced. Our musical excerpts comprised three pieces for each of these four categories, and were presented three times each during the experiments. All musical excerpts had a 45-s duration. Acoustic characteristics of music excerpts are presented for the four emotion categories in **Table 2**.

## Experimental Design

### Attention Network Task

The experimental task took place in a sound-isolated room. Visual stimuli were displayed on a screen at a distance of 50 cm while auditory stimuli were presented binaurally through high-quality headphones (DT 770 pro—250 Ohms, Beyerdynamic) with optimal tolerable loudness determined for each participant. Stimuli presentation and response recording (through a standard keyboard) were controlled using Cogent toolbox (developed by Cogent 2000 and Cogent Graphics) implemented in Matlab 2009b (Mathworks Inc., Natick, MA, USA).

We used a modified Attention Network Test (ANT; Fan et al., 2002), similar to our previous study in young and older healthy individuals (Fernandez et al., 2019b). In this task, participants are asked to judge the direction of a central arrow (leftward or rightward) considered as the target, presented together with either congruent or incongruent flankers (i.e., distractor arrows with the same or different direction, respectively). This visual display (five arrows including central target) is preceded by one of four types of cues (represented by a zero) at different positions on the screen (**Figure 1**). These cues correspond to either a central, double, spatial (valid or invalid), or no-cue condition. Based on previous work by Fan et al. (2002), the executive component of attention was assessed by contrasting congruent vs. incongruent conditions regardless of cue type. Alerting and Orienting components were measured by comparing different cue conditions (i.e., double vs. no-cue conditions for alerting; center vs. valid spatial cue conditions for orienting).

Different blocks of the ANT were performed during exposure to the musical pieces from the four different emotional categories as well as during silence to provide a baseline condition. The set of 12 musical excerpts (three pieces for each emotional category) was repeated three times, while the silence condition was presented nine times, leading to a total of 45 blocks (nine for each emotional category or silence condition) of 45 s duration each, and comprising 10 or 11 trials per block. Cue and target types were presented in a pseudo-randomized order and

**TABLE 2 |** Acoustic characteristics of the music excerpts included in the study for the four emotion categories.

| Dimension | Name | Perceptual characteristics | Audio scores | | | |
|---|---|---|---|---|---|---|
| | | | Joy | Tension | Tend | Sad |
| Rhythm | Tempo | Speed at which a piece of music is played | 129.83 (45.47) | 117.71 (36.35) | 118.72 (37.56) | 121.24 (36.41) |
| | Event density | Complexity of the piece; How many musical events (i.e., average of note occurrence) played in one time unit (sec) | 3.33 (1.01) | 2.71 (1.27) | 2.30 (0.84) | 1.82 (0.84) |
| Beat perception | Pulse clarity | How clearly the beat was detectable in the musical piece | 0.37 (0.12) | 0.34 (0.18) | 0.23 (0.08) | 0.15 (0.04) |
| Timbre | Attack numbers | Number of notes onset or pulses in the piece, related to the expressiveness at which the music piece is played | 23.25 (5.13) | 20.52 (5.78) | 17.61 (4.16) | 18.25 (5.61) |
| | Brightness | Sharpness of the sound | 0.31 (0.10) | 0.33 (0.15) | 0.12 (0.04) | 0.29 (0.07) |
| Frequency/Energy-dominant | Inharmonicity | Amount of energy outside an ideal harmony, supposedly reflect the unpleasantness of the sound | 0.43 (0.03) | 0.41 (0.04) | 0.40 (0.02) | 0.36 (0.05) |
| | Loudness | Information of the intensity of the music piece | 0.08 (0.03) | 0.05 (0.02) | 0.08 (0.04) | 0.07 (0.05) |
| | Dissonance | Roughness, and supposedly the unpleasantness the sound | 291.23 (222.10) | 137.04 (130.12) | 244.41 (216.05) | 199.64 (266.51) |

*Mean audio scores (SD) extracted using MIRtoolbox, are presented for eight acoustic features covering four major dimensions of music in addition to their perceptual characteristics.*

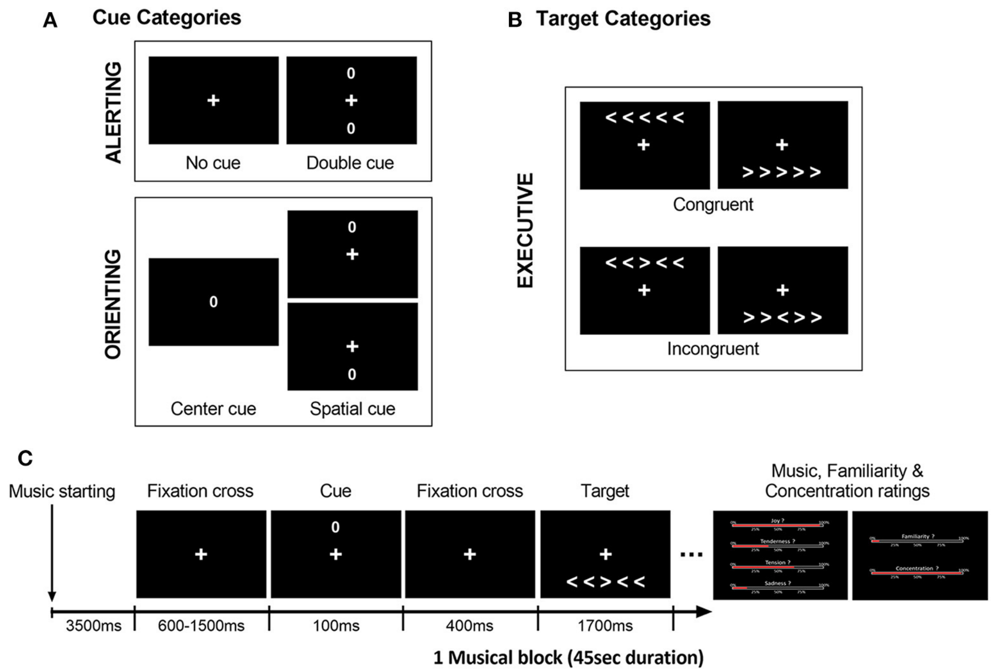



**FIGURE 1 |** Illustration of the modified ANT design, with distinct **(A)** cue categories and **(B)** target categories used to assess the executive, alerting, and orienting components of attention. **(C)** Example of a typical trial time-course at the beginning of a block (here musical block), with an empty 3,500 ms interval prior to the first trials. All trials began with a cue presentation (spatially invalid cue here), followed by the target display (incongruent condition here). Participants had to indicate the direction of the central arrow of the target (right or left) as fast and as accurately as possible after its presentation (within 1,700 ms max). Music, familiarity and concentration ratings for the preceding musical excerpt were presented after the blocks (i. emotion intensity for joy, tenderness, tension and sadness, ii. familiarity, and iii. concentration).

intermixed within the same block. Auditory conditions were also alternated in a pseudo-randomized order between blocks.

The ANT included a total number of 480 trials. As illustrated in **Figure 1**, for each block, the auditory exposure (music or silence) started 3,500 ms prior to the first visual stimulus appearance. Each trial began with a central fixation cross (duration between 600 and 1,500 ms) followed by one of the four possible cues (100 ms duration, visual angle of 0.91°) and lastly the target display with five arrows (1,700 ms duration). Each cue type (i.e., no-cue, center, double, and spatial cue) was presented in a quarter of the total number of 480 trials. In the no-cue condition (120 trials), only the fixation cross was displayed. In the center cue condition (120 trials), only a single cue circle was displayed at screen center. In the double cue condition (120 trials), two circle cues were presented simultaneously above and below the fixation cross, corresponding to the positions of the target stimuli. In the spatial cue condition (120 trials, regardless of validity), only one cue was presented to indicate either the correct location of the upcoming target (i.e., spatially valid, 60 trials) or the opposite location (i.e., spatially invalid, 60 trials). The cue was followed by an empty 400 ms interval during which only the fixation cross was displayed. The target stimuli consisted of a row of five horizontal arrows, among which the central arrow was the target (visual angle of 5.90° and 1.03°, respectively), randomly presented either above or below the central fixation cross (visual angle of 2.31°). In half of the total of 480 trials (240 trials), the central target arrowhead pointed in the same direction as the flanker arrows [i.e., Congruent condition (Con)], while in the other half of the trials (240 trials), it pointed in the opposite direction [i.e., Incongruent condition (Inc)]. Participants were asked to maintain their gaze directed to the fixation cross and to indicate, as fast and as accurately as possible, the direction of the central arrow after its presentation. Responses were given by pressing a corresponding button on the keyboard (i.e., right or left index for indicating right or left direction, respectively). Buttons presses as well as response times were measured throughout the whole experiment.

This experimental design is similar to our prior study (Fernandez et al., 2019b), with the exception of (i) a central fixation cross that was maintained during the presentation of visual stimuli (Fan et al., 2002), (ii) a slightly smaller visual size of the stimuli, and (iii) an additional silence condition used as a baseline condition.

## Emotion, Familiarity, and Concentration Ratings

At the end of a subset of the musical blocks of the ANT (i.e., 12 blocks), participants were asked to rate (i) the emotion intensity (i.e., "To what extent the musical excerpt expresses joy, tenderness, tension, and sadness"); (ii) familiarity (i.e., "To what extent were you familiar with this musical piece?"); and (iii) concentration level during the preceding block (i.e., "To what extent were you concentrated on the task?"). Each musical excerpt received four ratings, one for each emotion category along four distinct scales ranging from 0 (= not at all) to 6 (= extremely). Familiarity and concentration ratings were measured using a numerical scale ranging from 0 (= not at all) to 100 (= extremely), which resulted in one single value per musical excerpt. The order of the requested ratings was pseudo-randomized so that each musical excerpt was evaluated once. After a subset of the silence blocks (i.e., three blocks), only concentration level was assessed.

## Data Analysis

All statistical tests were chosen according to the normality of the residuals distribution and the equality of variances in our data, using R Software version 3.2.4 (R., R Development Core Team). The direction of significant interactions was tested using *t*-tests whose resulting *p*–values were adjusted using Bonferroni corrections.

### Emotion, Familiarity, and Concentration Ratings

First, individual emotion ratings (i.e., emotion intensity) were averaged over the three different musical excerpts for each emotion category (i.e., *joy*, *tenderness*, *tension*, and *sadness*). Second, the ability to correctly discriminate musical emotions (i.e., discrimination ratio) was assessed for each emotion category by subtracting the mean intensity scores of the three other categories from the emotion intensity score of a specific category, as used in previous studies of emotion recognition (e.g., Cristinzio et al., 2010). Familiarity ratings were computed for each emotion category by averaging scores for the three different excerpts associated with this emotion. Finally, concentration levels were assessed by averaging the concentration scores obtained for each emotion category (i.e., *joy*, *tenderness*, *tension*, *sadness, and silence*) and each participant.

Three distinct 4 (emotion category) × 2 (group) mixed-model repeated measures ANOVA analyses were used to separately examine emotion intensity, discrimination ratio, and familiarity ratings. The concentration ratings were entered into a 5 (auditory condition) × 2 (group) mixed-model repeated measures ANOVA. The emotion or auditory conditions were considered as within-subject factors, and the group as a between-subjects factor.

### Attention Network Task

Both the percentage of correct responses (accuracy, AC) and mean reaction times (RT) of correct trials were calculated for each participant and each group separately (*Amusics* and *Controls*), for each of the five auditory conditions (*Joy, Tenderness, Tension, Sadness,* and *Silence*), each target type (*Congruent* and *Incongruent*), and each cue type (*Central*, *Double*, *Spatial* and *No-cue*). A trial was considered accurate when participants correctly indicated the direction of the central arrow within the trial time limit (1,700 ms). The three distinct attentional components were separately analyzed according to the specific cues or stimulus combination (Fan et al., 2002; Fernandez et al., 2020). Executive control was assessed by contrasting incongruent vs. congruent arrow conditions regardless of the cue type (i.e., *Con* and *Inc*), while the alerting and orienting components were determined by contrasting distinct cues regardless the target type (i.e., *Double* vs. *No-cue* conditions for alerting; *Center* vs. *Valid Spatial* conditions for orienting).

Because residuals were not normally distributed (preventing the use of parametric statistical tests), AC analyses were performed for each attentional component using paired Wilcoxon rank tests to determine differences between groups and visual conditions (Con vs. Inc; Double vs. No-Cue; Center vs. Valid Spatial). Close-to-ceiling accuracy rates were found in both groups (>95% correct), and no major effect of musical emotion ($p > 0.3$, Wilcoxon rank tests) was found for any component or groups. Consequently, our main analyses and results concerning the influence of music focused on RTs only.

As RT scores revealed normally distributed residuals and equal variances, the effect of music exposure on RT was assessed using three distinct ANOVAs performed, one for each attentional component. Mean RT measures were entered in a separate 2 × 2 × 5 mixed-model repeated-measure ANOVA with trial type (Con and Inc for executive; Double and No-Cue for alerting; Center and Valid Spatial for orienting) and musical emotion category (Joy, Tenderness, Tension, Sadness, and Silence) as within-subject levels and group (Amusics and Controls) as a between-subject factor. Because arousal and valence dimensions are considered as two essential and independent dimensions of all emotion types (Russell, 2003) and thus also describe the functional organization of music-induced emotions (Trost et al., 2012), emotional effects on attention performance and RTs were further assessed by sorting emotion categories into their corresponding arousal and valence dimensions (in line with the 4 categories used in our study). This was achieved using three distinct 2 × 2 × 2 × 2 mixed-model repeated-measure ANOVAs (one for each attentional component) with trial type (Con and Inc for executive; Double and No-Cue for alerting; Center and Valid Spatial for orienting), valence (high and low), and arousal (high and low) as within-subject levels, plus group as a between-subject factor.

The effect of age was determined by entering the mean RT results in three additional 2 × 2 × 5 × 1 mixed-model repeated measure ANCOVAs (one for each attentional component), with the corresponding trial conditions (Con and Inc for executive; Double and No-Cue for alerting; Center and Valid Spatial for orienting) and auditory conditions (Joy, Tenderness, Tension, Sadness, and Silence) as within-subject levels, groups as a between-subject factor, plus age as a covariate.

Finally, Pearson correlation analyses were performed to assess any relationship between musical abilities scores (MBEA scale, melodic composite or pitch-change detection scores) and indices of the three attentional components, namely the executive cost ($RT_{Inc} - RT_{Con}$), the alerting efficiency ($RT_{No\ Cue} - RT_{Double\ Cue}$), and the orienting efficiency ($RT_{Center\ Cue} - RT_{Spatial\ valid\ cue}$), calculated following previous work (e.g., Jiang et al., 2011).

Of note, the sample size of the current study was determined by a power analysis using G∗Power (version 293 3.1.9.7, Heinrich Heine University). This analysis indicated a probability ≥ 90% to replicate the experimental differences on our attentional task with $n \geq 15$ based on the effect size ($d = 1.83$) observed in a previous study using the same paradigm in both younger and older individuals (Fernandez et al., 2019b).

# RESULTS

## Attention Network Task Results

AC (in percentage, %) and RT results (in milliseconds, ms) for each attentional component measured in the ANT are presented in **Table 3**. Additionally **Figure 2** shows more detailed results for the executive control component. Because AC showed no significant effect for music on performance, our main analysis comparing different emotions and different groups focused on RT data.

### Executive Control Component

Executive control performance was assessed by comparing trials where targets appeared with congruent vs. incongruent flankers. As expected, participants were more accurate (paired Wilcoxon test, $V = 224.5$; $p < 0.001$) and faster [$F_{(1, 28)} = 156.50$; $p < 0.001$] for congruent ($M = 99\%$; SD = 0.6; M = 682 ms; SD = 112.6) compared to incongruent trials ($M = 98\%$; SD = 1.2; $M = 833$ ms; SD = 130.3). Critically, amusic and control participants did not differ for global executive control performance. Both groups showed similar accuracy scores ($V = 34$; $p = 0.71$), and RT data disclosed no main group effect [$F_{(1, 28)} = 0.41$; $p = 0.52$] nor any group by trial type interaction [$F_{(1, 28)} = 0.35$; $p = 0.55$].

More critically, the influence of music on executive control performance was assessed by comparing RTs between the different auditory background conditions. A first ANOVA performed on RTs across all auditory conditions (four emotion types and silence) and trial types (Con vs. Inc) revealed a main effect of auditory condition [$F_{(4, 112)} = 3.20$; $p = 0.01$] with no group effect [$F_{(4, 112)} = 0.59$; $p = 0.66$] nor any group by trial type interaction [$F_{(4, 112)} = 0.63$; $p = 0.63$]. Joyful music ($M = 747$ ms; SD = 116.9) yielded faster visual target detection than sad music across groups [$t_{(29)} = -4.06$; $p = 0.003$; $M = 762$ ms; SD = 120.2]. An additional ANOVA treating music valence and arousal as separate factors revealed a main effect of

**TABLE 3 |** Behavioral scores in the Attention Network Task (ANT) for measures assessing executive control, alerting, and orienting components, for both the amusic and control groups.

| | Accuracy [Percentage (SD)] | | | | | Reaction times [Milliseconds (SD)] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **EXECUTIVE NETWORK** | | | | | | | | | | |
| **A** | **Con** | **Inc** | | | | **Con** | **Inc** | **Cost** | | |
| Amusic | 99% (0.7) | 98% (1.2) | | | | 693 ms (100.0) | 850 ms (139.2) | 157 ms (76.6) | | |
| Control | 99% (0.5) | 99% (1.3) | | | | 672 ms (126.9) | 816 ms (126.9) | 144 ms (55.0) | | |
| **B** | **Joy** | **Tens** | **Tend** | **Sad** | **Silence** | **Joy** | **Tens** | **Tend** | **Sad** | **Silence** |
| Amusic | 99% (1.5) | 99% (1.6) | 99% (1.3) | 99% (1.1) | 99% (1.2) | 757 ms (116.2) | 767 ms (124.7) | 776 ms (117.9) | 774 ms (116.2) | 758 ms (113.9) |
| Control | 99% (1.0) | 99% (1.1) | 99% (1.7) | 99% (1.4) | 99% (1.3) | 737 ms (122.4) | 740 ms (128.9) | 741 ms (120.5) | 750 ms (127.6) | 749 ms (124.7) |
| **ALERTING NETWORK** | | | | | | | | | | |
| **A** | **No cue** | **Double cue** | | | | **No cue** | **Double cue** | **Efficiency** | | |
| Amusic | 99% (0.9) | 99% (1.1) | | | | 793 ms (119.1) | 755 ms (113.2) | 37 ms (25.1) | | |
| Control | 99% (1.0) | 99% (1.2) | | | | 750 ms (125.8) | 728 ms (132.3) | 22 ms (19.8) | | |
| **B** | **Joy** | **Tens** | **Tend** | **Sad** | **Silence** | **Joy** | **Tens** | **Tend** | **Sad** | **Silence** |
| Amusic | 99% (1.7) | 98% (2.0) | 99% (1.5) | 99% (1.3) | 100% (0.8) | 765 ms (121.9) | 764 ms (121.3) | 775 ms (117.6) | 780 ms (116.3) | 786 ms (118.9) |
| Control | 100% (0.9) | 99% (1.5) | 98% (2.5) | 99% (1.3) | 99% (1.3) | 737 ms (131.5) | 739 ms (130.7) | 734 ms (125.0) | 742 ms (126.9) | 743 ms (134.8) |
| **ORIENTING NETWORK** | | | | | | | | | | |
| **A** | **Center cue** | **Sp. valid cue** | **Sp. invalid cue** | | | **Center cue** | **Sp. valid cue** | **Sp. invalid cue** | **Efficiency** | |
| Amusic | 98% (1.3) | 99% (1.7) | 99% (1.7) | | | 769 ms (115.8) | 767 ms (119.1) | 771 ms (127.7) | 2 ms (20.2) | |
| Control | 99% (1.2) | 99% (1.5) | 99% (1.1) | | | 738 ms (120.6) | 740 ms (124.6) | 749 ms (127.1) | −2 ms (25.8) | |
| **B** | **Joy** | **Tens** | **Tend** | **Sad** | **Silence** | **Joy** | **Tens** | **Tend** | **Sad** | **Silence** |
| Amusic | 99% (2.1) | 99% (1.9) | 98% (1.9) | 99% (1.3) | 99% (2.0) | 753 ms (115.4) | 774 ms (133.0) | 773 ms (117.5) | 764 ms (114.7) | 782 ms (109.6) |
| Control | 99% (1.6) | 100% (1.1) | 100% (1.0) | 99% (1.6) | 99% (2.1) | 732 ms (116.6) | 740 ms (132.9) | 743 ms (124.4) | 750 ms (132.2) | 748 ms (119.0) |

*For each attentional component, the upper results (A) display the mean data for each trial type and each cue condition of interest, while the lower results (B) display the mean data for each emotion condition, across the corresponding stimulus and cue conditions. Mean percentage correct [accuracy (%) left panel] and mean reaction times for correct responses [milliseconds (ms), right panel] with standard deviation (SD) values in parentheses. Con, congruent; Inc, incongruent; Sp. Valid Cue, spatial valid cue; Sp. Invalid Cue, spatial invalid cue; Tens, Tension; Tend, Tenderness; Sad, Sadness.*

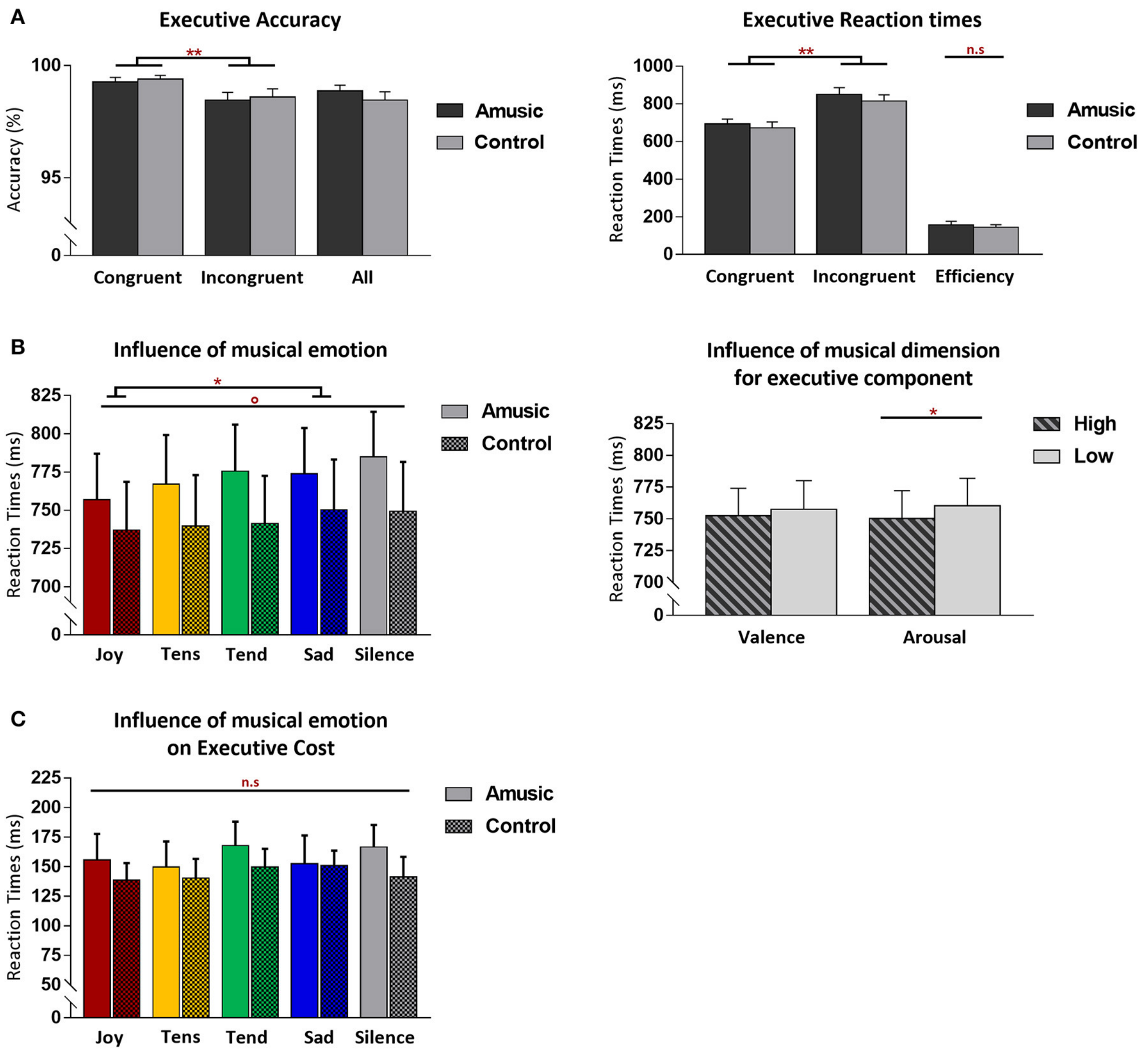



**FIGURE 2 |** Behavioral results from the Attention Network Task (ANT) for the executive control component of attention, shown for amusic and control participants. **(A)** Mean accuracy (%) (left) and RT (in milliseconds, ms) (right) for congruent and incongruent target conditions. Mean accuracy (%) from both target conditions merged together, as well as interference cost (ms) [(mean $RT_{Inc}$-mean $RT_{Con}$)] are also presented. **(B)** Mean RT (ms) regardless of congruence and group conditions, as a function of the discrete emotion category of music, plus a silence condition (left) or as a function of the valence and arousal dimensions of music (right) during the task. **(C)** Executive cost (ms) as a function of the emotion category of music (plus silence), presented without group distinction. Graphs illustrate standard errors of the mean (SEM) and *p*-values (* or °) with the following meaning: *$p < 0.01$; **$p < 0.001$; °$p < 0.05$ without Bonferroni corrections. The n.s. abbreviation indicates non-significant results.

arousal [$F_{(1, 28)} = 8.42$; $p = 0.007$] with faster RT during high-arousing ($M = 750$ ms; SD = 120.6) compared to low-arousing music ($M = 760$ ms; SD = 118.2) in both amusics and control participants taken together. No group by arousal interaction was found [$F_{(1, 28)} = 0.53$; $p = 0.46$]. No main effect of valence [$F_{(1, 28)} = 2.08$; $p = 0.16$], nor group by valence interaction was found [$F_{(1, 28)} = 0.06$; $p = 0.80$]. There was no arousal by valence interaction [$F_{(1, 28)} = 0.25$; $p = 0.61$], and no significant group by arousal by valence interaction [$F_{(1, 28)} = 2.61$; $p = 0.27$].

Finally, a significant main effect of age was demonstrated by our covariate analysis, with longer RTs regardless of congruency/trial type and auditory background [$F_{(1, 26)} = 38.64$; $p < 0.001$], reflecting a general age-related slowing in performance which was similar between amusic and control participants [$F_{(1, 26)} = 2.63$; $p = 0.11$]. Specific correlation analyses revealed no association between the executive cost ($RT_{Inc}$-$RT_{Con}$) and musical abilities in amusics, assessed with MBEA scale [$r_{(13)} = 0.37$, $p = 0.17$], melodic composite [$r_{(13)} = 0.09$, $p = 0.74$], and pitch-change detection scores [$r_{(13)} = -0.07$, $p = 0.78$]. No such association was found either in control participants [MBEA scale: $r_{(12)} = -0.33$, $p = 0.24$; melodic composite: $r_{(12)} = -0.15$, $p = 0.60$; pitch-change detection score: $r_{(12)} = 0.42$, $p = 0.13$].

## Alerting Component

The alerting component of attention was assessed by comparing trials where targets were preceded by double (non-informative) cues vs. no cues. No significant difference in AC was found between the two cue types (paired Wilcoxon test, $V = 106$; $p = 0.51$) or between the two groups ($V = 115$; $p = 0.93$). For RTs, results showed a main effect of cue type [$F_{(1, 28)} = 51.49$; $p < 0.001$] with slower RTs with no-cue ($M = 771$ ms; SD = 122.3) than with double cues ($M = 742$ ms; SD = 121.8). No group effect emerged [$F_{(1, 28)} = 0.61$; $p = 0.43$] nor any group by cue type interaction [$F_{(1, 28)} = 3.46$; $p = 0.07$].

Again, music effects were examined by ANOVAs on RTs measured in different auditory exposure conditions. We found no significant influence of auditory background on alerting for both groups ($p > 0.04$). There was no group by auditory background interaction implicating emotion category [$F_{(1, 28)} = 0.67$; $p = 0.61$], valence [$F_{(1, 28)} = 0.09$; $p = 0.76$], or arousal [$F_{(1, 28)} = 2.70$; $p = 0.11$].

Finally, the age covariate also revealed a main effect on RTs across the different alerting cue types [$F_{(1, 26)} = 28.99$; $p < 0.001$], without group distinction [$F_{(1, 26)} = 2.19$; $p = 0.15$]. None of the two groups showed any correlation of alerting efficiency ($RT_{No\ Cue} - RT_{Double\ Cue}$) with musical abilities, namely MBEA scale [amusic: $r_{(13)} = 0.02$, $p = 0.92$; control: $r_{(12)} = 0.11$, $p = 0.70$], melodic composite [amusic: $r_{(13)} = 0.18$, $p = 0.51$; control: $r_{(12)} = -0.59$, $p = 0.02$, with did not survive multiple comparisons correction], or pitch-change detection scores [amusic: $r_{(13)} = -0.27$, $p = 0.31$; control: $r_{(12)} = -0.12$, $p = 0.65$].

### Orienting Component

Finally, orienting effects were probed by comparing performance on trials where targets were preceded by valid spatial cues vs. center cues. AC showed no difference between the two cue types (paired Wilcoxon test, $V = 67$; $p = 0.16$) or between the two groups ($V = 75$; $p = 0.12$). Likewise, RTs showed no effect of cue type [$F_{(1, 28)} = 0.01$; $p = 0.90$], no effect of group [$F_{(1, 28)} = 0.42$; $p = 0.52$], and no group by cue type interaction [$F_{(1, 28)} = 0.38$; $p = 0.54$].

The effects of auditory exposure on RTs were not significant in terms of either emotion category [$F_{(4, 112)} = 0.61$; $p = 0.65$], valence [$F_{(4, 112)} = 0.47$; $p = 0.49$], or arousal [$F_{(4, 112)} = 0.002$; $p = 0.96$], in both amusics and control participants taken together. Although there was a significant interaction between emotion category and cue type [$F_{(4, 112)} = 2.59$; $p = 0.04$], none of the pairwise *post hoc* comparisons survive multiple-comparison correction. The triple interaction emotion category by cue type by group was not significant [$F_{(4, 112)} = 0.95$; $p = 0.43$].

Finally, the age covariate again showed a main effect on RTs [$F_{(1, 26)} = 41.56$; $p < 0.001$] across the two different orienting conditions, without any group distinction [$F_{(1, 26)} = 1.99$; $p = 0.16$]. No correlation was found between orienting efficiency ($RT_{Center\ Cue} - RT_{Spatial\ valid\ cue}$) and musical abilities for amusics: MBEA scale [$r_{(13)} = -0.08$, $p = 0.76$], melodic composite [$r_{(13)} = -0.12$, $p = 0.64$], or pitch-change detection scores [$r_{(13)} = 1.5$, $p = 0.58$]; or for control participants [MBEA scale: $r_{(12)} = -0.25$, $p = 0.38$; melodic composite: $r_{(12)} = -0.30$, $p = 0.28$; pitch-change detection score: $r_{(12)} = 0.25$, $p = 0.38$].

## Emotion, Familiarity, and Concentration Ratings

The ratings obtained for emotion intensity, emotion discrimination, familiarity, and concentration ratings are presented in **Table 4** and **Figure 3**. As can be seen, amusics generally judged the emotions expressed by music as less intense than controls did [$F_{(1, 28)} = 4.48$; $p = 0.04$]. In both groups, a main effect of emotion category [$F_{(3, 84)} = 15.92$; $p < 0.001$] showed that joy, tense, and tender music pieces were judged as more intense than sad music. No group by emotion interaction was found [$F_{(3, 84)} = 0.46$; $p = 0.71$]. Emotion discrimination (i.e., relative ratio of ratings for the expressed emotion subtracted by the average ratings given to the other three emotions) showed a main effect of emotion category [$F_{(3, 84)} = 21.33$; $p < 0.001$] with a lower discrimination ratio for sad ($M = 16\%$; SD = 23.04) compared to joyful [$t_{(29)} = -7.55$; $p < 0.001$; $M = 54\%$; SD = 19.29], tense [$t_{(29)} = -3.45$; $p = 0.006$; $M = 36\%$; SD = 30.51], and tender music [$t_{(29)} = -6.23$; $p < 0.001$; $M = 45\%$; SD = 21.16]. There was no group effect [$F_{(1, 28)} = 3.25$; $p = 0.08$], nor group by emotion interaction [$F_{(3, 84)} = 0.64$; $p = 0.58$]. Hence, emotion categorization was comparable between amusics and controls.

Familiarity ratings were low for all musical pieces (values $< 50$), but joyful music was generally rated as more familiar than other emotion categories [$F_{(3, 84)} = 19.90$; $p < 0.001$]. Importantly, there was no group effect [$F_{(1, 28)} = 0.04$; $p = 0.83$], nor group by emotion interaction [$F_{(3, 84)} = 1.20$; $p = 0.31$].

**TABLE 4** | Music and Concentration ratings for both amusic and control participants separately.

| | Music and concentration ratings | | | |
|---|---|---|---|---|
| | **Emotion intensity** | **Emotion discrimination** | **Familiarity** | **Concentration** |
| **Joy (V+/A+)** | | | | |
| Amusic | 70% (13.9) | 46% (16.9) | 46% (27.4) | 76% (19.5) |
| Control | 84% (9.1) | 62% (18.4) | 49% (28.2) | 83% (12.6) |
| **Tension (V–/A+)** | | | | |
| Amusic | 59% (20.1) | 28% (27.2) | 26% (21.2) | 72% (21.8) |
| Control | 67% (22.4) | 43% (32.6) | 16% (15.6) | 74% (21.2) |
| **Tenderness (V+/A–)** | | | | |
| Amusic | 68% (14.6) | 42% (18.8) | 37% (24.7) | 76% (16.8) |
| Control | 73% (17.8) | 47% (23.6) | 38% (20.5) | 80% (15.2) |
| **Sadness (V–/A–)** | | | | |
| Amusic | 48% (20.2) | 12% (25.9) | 30% (22.1) | 77% (16.5) |
| Control | 56% (14.4) | 19% (20.1) | 31% (22.0) | 82% (15.2) |
| **Silence** | | | | |
| Amusic | NA | NA | NA | 87% (15.8) |
| Control | NA | NA | NA | 87% (12.7) |

*Music ratings included emotion intensity, emotion discrimination (i.e., discrimination ratio), and familiarity, recorded for each excerpt from the four emotion categories, while concentration ratings were computed for all five auditory conditions (music and silence). Mean values of maximum scores on a continuous scale (from 0 to 100) are presented (in percentage, %) with standard deviation (SD) values in parentheses. NA, Not applicable; $V^{+,}$ positive valence; $V^{-}$, negative valence; $A^{+}$, positive arousal; $A^{-}$, negative arousal.*

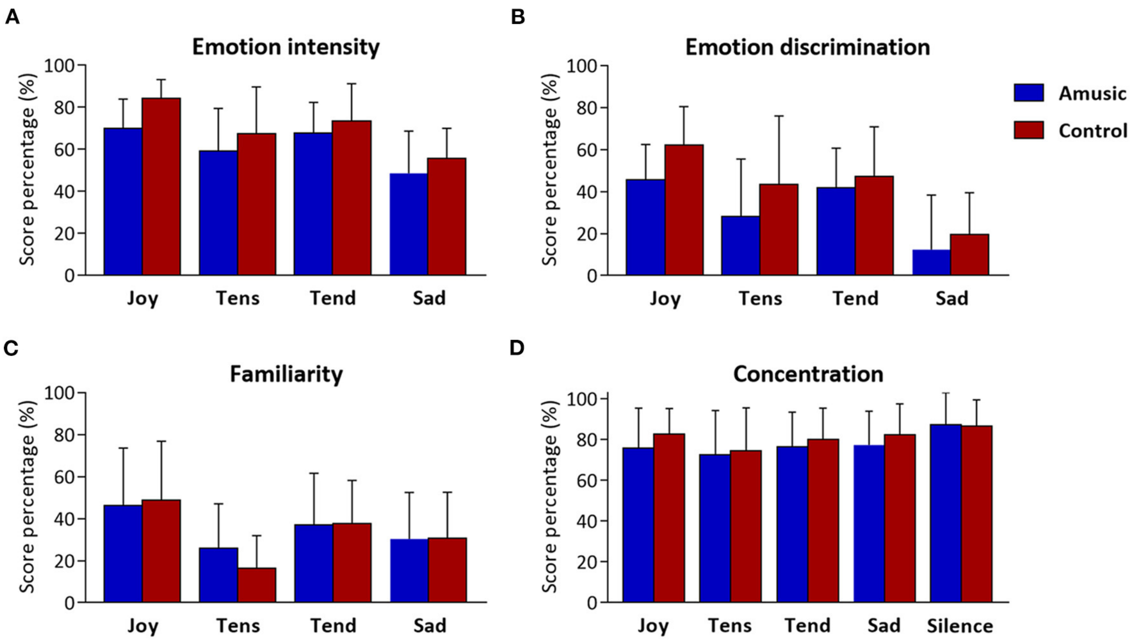



**FIGURE 3 |** Music and concentration ratings obtained at the end of musical blocks. Scores resulting from **(A)** emotion intensity, **(B)** emotion discrimination (i.e., discrimination ratio), and **(C)** familiarity are presented as percentages (%) for each of the four emotion categories (i.e., Joy, Tension, Tenderness, and Sadness), for amusic and control participants separately. Concentration scores **(D)** are presented (%) for each of the five auditory conditions (including silence) for the two groups separately.

Finally, as expected, subjective concentration ratings revealed a main effect of auditory condition [$F_{(4, 112)} = 5.70$; $p < 0.001$], showing higher concentration ratings during silence than during music exposure conditions [$t_{(29)} > 2.78$; $p < 0.03$], but there was no significant difference between groups [$F_{(4, 112)} = 0.44$; $p = 0.51$], nor an auditory condition by group interaction [$F_{(4, 112)} = 0.45$; $p = 0.76$].

## DISCUSSION

The present study investigated to what extent pitch deficits characterizing the most common form of congenital amusia could influence indirect effects of music exposure on attentional performance. Our results show that major visuo-spatial attention components are preserved in individuals with congenital amusia. Furthermore, and more critically, amusics are still sensitive (just as the general population is) to pleasant and arousing music while performing a visuo-spatial attentional task.

### Impact of Amusia on Emotion Processing

Although emotions conveyed by music were evaluated as less intense by the amusic participants, these individuals categorized the emotions expressed by different pieces with similar accuracy to the control participants. This result is in line with the notion that discrimination and intensity perception might be dissociated in emotion processing (Hirel et al., 2014). Our finding is also consistent with previous work that highlighted relatively spared music-related emotional judgements in congenital amusia (Ayotte et al., 2002; Gosselin et al., 2015). However, the results diverge from other work where poorer emotion recognition was found in amusics as compared to controls for both speech (Thompson et al., 2012; Lolli et al., 2015; Lima et al., 2016) and music (Lévêque et al., 2018). In the latter study, the authors suggested that such a discrepancy could have resulted from the use of orchestral music (Lévêque et al., 2018). Specifically, orchestral music might be more sensitive to subtle amusic deficits during an emotion discrimination task in comparison to the piano music excerpts, like those used in our previous studies (Ayotte et al., 2002; Gosselin et al., 2015). However, in this study, we used orchestral music as well as piano and string music excerpts. Interestingly, more recent work reported deficits in explicit emotion processing in amusics for both music and speech when using a forced-choice method (i.e., choosing a specific emotion among given categories) (Thompson et al., 2012; Lévêque et al., 2018; Pralus et al., 2019), while relatively comparable implicit processing abilities of musical emotions and emotional prosody were observed using indirect investigation methods (e.g., intensity ratings of emotions) (Tillmann et al., 2016; Lévêque et al., 2018; Pralus et al., 2019). Here we employed a free scale method for both intensity and discrimination measures, but these judgments could have engaged participants' internal representations of musical emotions differently compared to explicit emotion categorization tasks with forced-choice labels that require conscious retrieval (Cleeremans and Jiménez, 2002). While discrimination judgments may rely on more strategic access to categorical knowledge based on discrete cues, intensity judgements may involve other perceptual abilities operating on more continuous sensory dimensions. Future work should help further disentangle the complex relationships between musical emotion recognition, auditory complexity, and both explicit/implicit investigation methods in the amusic population.

We also note that all participants had more difficulties correctly categorizing sadness compared to other emotions, and had more success categorizing joy, as found previously (Gosselin et al., 2015; Lévêque et al., 2018). In everyday life, sadness is typically associated with unpleasant feelings. However, in music it is often associated either with pleasantness (Kawakami et al., 2013) or with complex emotions, namely simultaneous positive and negative feelings (Juslin et al., 2014; Sachs et al., 2015), as is similarly reported for nostalgia (Barrett et al., 2010; Trost et al., 2012; Schindler et al., 2017). Such expressions of mixed emotions could be a factor behind the lower discrimination ratios and lower intensity scores observed here. Most importantly, all the differential effects of musical emotion on the ratings were similar between our amusic and control participants. Therefore, the results indicate that amusics do have a fairly average capacity to respond to various musical emotions despite a deficit in pitch processing. This is consistent with a possible dissociation between emotional and perceptual components in the processing of music (Gosselin et al., 2015).

## Impact of Amusia on Attentional Processing

Overall, behavioral performance in the attentional task fully accorded with the literature and showed that the ANT paradigm was effective in assessing visual attention. All participants showed sensitivity to conflicting distractor information when detecting a visual target, with more errors and longer RTs on incongruent compared to congruent trials (Casey et al., 2000; Fan et al., 2005; Fernandez et al., 2019b). They were also faster during trials for which they were temporally warned about the imminent target (i.e., double cue) in comparison to no-cue trials, showing benefits of phasic alertness (Fan et al., 2002; Finucane et al., 2010; McConnell and Shore, 2011). Unexpectedly, behavioral results associated with the orienting component of the ANT did not show any modulation of accuracy or RTs (i.e., for center compared to spatially valid cues), indicating no significant beneficial effect of shifting the attentional focus toward the target location prior to its presentation, contrary to what is typically reported in the literature (Fan et al., 2005; Finucane et al., 2010; McConnell and Shore, 2011). This lack of orienting effect may be caused by an inadequate interval between cue and target displays, or by insufficient spatial preparation following the presentation of the visual cue [perhaps due a 50% validity contingency used here, compared to 100% validity contingency in the original Fan's ANT (Fan et al., 2002)]. Nevertheless, other well-established effects were replicated and accompanied by a significant, age-related slowing of attentional performance, as consistently reported in the literature, particularly for executive control (Zhu et al., 2010; Fernandez et al., 2019a,b). Overall, therefore, our findings converge with previous work on attention and validate our modified ANT task ensuring its sensitivity to assessing congenital amusic individuals in the presence or absence of music exposure.

Critically, our findings for the three distinct attentional components showed comparable accuracy and RT performance between amusic and control participants across all conditions. These findings were confirmed by further correlation analyses revealing that the severity of musical deficits, measured with the MBEA scale, melodic composite, and pitch-change detection, did not predict attentional performance in any of the three attentional components. This spared performance in congenital amusia stands in sharp contrast with executive control deficits documented in several populations, including the elderly (Zhu et al., 2010; Fernandez et al., 2019a) and patients with visual attentional developmental disorders (e.g., ADHD) (Johnson et al., 2008; Mogg et al., 2015) who show abnormal distractor susceptibility (i.e., more errors/larger RTs in incongruent trials) in several paradigms, including the ANT version with arrow flankers as used here. Similarly, an attenuation of alerting states has been reported in ADHD (Johnson et al., 2008) and patients with strokes (Spaccavento et al., 2019) compared to the control population, but it was not seen in the amusic group. Finally, the age-related slowing in visual attention was similar in the amusic and control participants who were relatively old but age-matched suggesting that musical deficits have no distinctive impact on attentional performance with increasing age. Thus, the present study highlights preserved attention processing in congenital amusia, in keeping with the notion that their impairment is a selective musical disorder affecting pitch processing (Peretz et al., 2002; Peretz, 2016; Peretz and Vuvan, 2017).

Interestingly, finding normal visuo-spatial attention in congenital amusia does not support earlier claims of an intimate link between musical and visuo-spatial attentional processing abilities (Douglas and Bilkey, 2007), based on the assumption that sound frequency representations may be intertwined with spatial codes (e.g., lower pitch is usually represented lower in space than higher pitch) (Rusconi et al., 2006). Rather, the present study is more in line with prior studies showing preserved mental rotation abilities in amusics with low pitch or contour MBEA scores (Tillmann et al., 2010; Williamson et al., 2011). We found no attenuation of visuo-spatial attentional processing indices in amusic individuals, even in the most musically impaired, by showing no link between three distinct scores measuring musical deficits and attentional performance. These findings suggest that the severity of musical deficits does not impact visuo-spatial attention and further questions the notion of a continuum of visuo-spatial cognition with musical abilities (Douglas and Bilkey, 2007). However, we cannot disregard the possibility that more complex aspects of object representation/manipulation in space might interplay with musical skills, but any such interaction seems unrelated to the attentional processes probed here, again reinforcing the view that congenital amusia occurs independently of any other cognitive deficits.

## Effects of Emotional Music on Attention in Amusia

Our main goal with the present work was to assess whether the influence of music exposure on attentional processes has a similar effect on the congenital amusia population as compared to the general population. In accordance with previous work (Fernandez et al., 2020), we found a reliable modulation of processing speed, regardless of visual flanker congruency, when participants were exposed to joyful (i.e., high-arousing and high-valence) compared to sad music (i.e., low-arousing and low-valence), and this effect was similar in amusics and controls. This attentional enhancement was likely mainly driven by the arousal dimension of music since valence did not appear to modulate performance. Such effects of arousal might result from a greater engagement of the attentional control network in frontal and parietal cortices, as observed in a recent fMRI study using a similar paradigm (Fernandez et al., 2019b).

Unlike in previous work, however, we did not find significantly faster performances when comparing joyful musical exposure to the silence condition (Fernandez et al., 2019b), probably because of the small effect size of this modulation in the current population sample. However, the pattern of absolute RT effects (see **Figure 3**) fully accords with earlier observations, suggesting not only that joyful music produced fastest responses, while sad music and silence produced the slowest, but also that such differences reflect a facilitation of stimulus processing due to joyful (and more generally high-arousing) music rather than a slowing or distracting effect of (negative) emotional music relative to silence (Trost et al., 2014).

Overall, the normal influence of music exposure on executive attentional control in congenital amusia suggests that these individuals still receive the indirect effects of music in spite of their musical deficits. This finding supports the increasingly supported theory that an amusics' brain has the capacity to track subtle musical (pitch) variations without awareness (e.g., Peretz et al., 2009; Tillmann et al., 2012; Zendel et al., 2015). Taken together, our results highlight the powerful and pervasive capacity of music and musical emotions to influence the mind and our behavior through relatively automatic and unconscious pathways, including high-level cognitive functions associated with executive control. Our study yields precious insights into the remarkable relationships between emotional and attentional processing in the human brain through which music can enhance cognitive abilities.

A few possible limitations of the current study should be acknowledged. First of all, our main results and conclusions concerning the preserved attentional effects of musical emotions rely on negative findings, i.e., no significant differences between congenital amusics and controls in critical behavioral effects of interest. We feel that this finding is unlikely to be caused by insufficient power, given that our sample size was validated by previous work (Fernandez et al., 2019b). In addition, although the ANT has been successfully employed in several studies to assess major attentional components across various populations (Wang et al., 2005; Fernandez-Duque and Black, 2006; Mahoney et al., 2010; Park et al., 2019), it may have failed to capture another attentional dimension that is possibly affected by amusia. Nevertheless, the current finding of comparable visuo-spatial attentional performance in this paradigm, despite the congenital impairment associated with amusia, helps to further refine our understanding of this disorder. Further investigations should confirm intact attentional performances in amusics across a wider range of tasks, for instance, by comparing their performance to populations known to present attentional deficits. Finally, another general limitation of our work might be the relatively small sample of individuals with congenital amusia who were included in the study. As this condition has a low prevalence (1.5% up to 4%) in the general population, we deliberately chose strict inclusion criteria (i.e., scores below cut-off scores on two scale tests in accordance with the latest normative data) to ensure an inclusion of individuals presenting clear and substantial musical deficits only (Peretz and Vuvan, 2017; Vuvan et al., 2018), but this strictness inherently limited the size of our sample. We also acknowledge a potential lack of statistical power for some task conditions, particularly the alerting and orienting components whose assessment is made by computing a subset (half) of the total trials (see Fan et al., 2002), unlike the executive control component. This lack of statistical power might account for a failure to demonstrate a main effect of emotion on these components, contrary to our previous study (Fernandez et al., 2019b). In addition, the manipulation of the orienting components might have produced insufficient spatial preparation (50% validity contingency) and allowed participants to ignore the spatial cues, accounting for a lack of a significant validity effect during attentional orienting, unlike the original ANT paradigm (Fan et al., 2002).

In any case, to our knowledge, this study is the first to assess selective attention abilities as well as the influence of music exposure on attentional processes in congenital amusia. We were able to confirm normal emotional processing and cognitive control functions in amusics, notably by demonstrating that they exhibit similar accuracy and reaction time performances compared to the control population in the attentional network task measuring three distinct attentional components. Furthermore, they also exhibit faster reaction times in attention conflict conditions during joyful/high-arousing music compared to sad/low-arousing music, similar to people with normal music perception (Fernandez et al., 2019b). These data reveal that affect-related influences of music on attention control do not depend on the neural system altered in congenital amusia and still operate despite defective pitch processing. Our study yields insights on the remarkable relationships between emotional and attentional processing in the human brain through which music can enhance cognitive abilities.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Council for the Faculty of Arts and Sciences at the Université de Montréal. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

NF designed the project and validated the task. NF performed the data collection and analyses and wrote the initial draft of manuscript with contributions from other authors. PV supervised the research protocol, methodology, and analyses. NG and IP supervised the research protocol and provided testing material and resources for data collection. All authors actively contributed to the revision, corrections of the manuscript, approved the final version for publication, and contributed to the research question formulation.

## FUNDING

This work was supported in parts by a grant awarded to NF from the Swiss National Science Foundation (P1GEP3_161687) as well as by a grant from the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council of Canada, and the Canada Research Chairs program to IP.

## ACKNOWLEDGMENTS

Experimental data were collected during NF collaborative stay in BRAMS at the University of Montréal. We warmly thank Wiebke J. Trost, Ph.D (Cognitive and Affective Neuroscience Lab, University of Zurich, Switzerland) for sharing the auditory material which was valuable for the completion of this project. We are grateful to the staff of the International Laboratory for Brain, Music and Sound Research (BRAMS) of the University of Montreal (Canada) for their technical assistance in the acquisition of the data. In particular, the authors thank Mihaela Felezeu for her invaluable help in the recruitment of congenital amusic and control participants. We thank Caitlyn Trevor, Ph.D (Cognitive and Affective Neuroscience Lab, University of Zurich, Switzerland) for proofreading the manuscript. We would like to thank the reviewers for their careful reading and helpful comments that allowed us to improve our paper.

## REFERENCES

Ayotte, J., Peretz, I., and Hyde, K. (2002). Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain* 125, 238–251. doi: 10.1093/brain/awf028

Barrett, F. S., Grimm, K. J., Robins, R. W., Wildschut, T., Sedikides, C., and Janata, P. (2010). Music-evoked nostalgia: affect, memory, and personality. *Emotion* 10:390. doi: 10.1037/a0019006

Bolger, D., Trost, W., and Schön, D. (2013). Rhythm implicitly affects temporal orienting of attention across modalities. *Acta Psychol.* 142, 238–244. doi: 10.1016/j.actpsy.2012.11.012

Brochard, R., Dufour, A., and Despres, O. (2004). Effect of musical expertise on visuospatial abilities: evidence from reaction times and mental imagery. *Brain Cogn.* 54, 103–109. doi: 10.1016/S0278-2626(03)00264-1

Casey, B., Thomas, K. M., Welsh, T. F., Badgaiyan, R. D., Eccard, C. H., Jennings, J. R., et al. (2000). Dissociation of response conflict, attentional selection, and expectancy with functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8728–8733. doi: 10.1073/pnas.97.15.8728

Cleeremans, A., and Jiménez, L. (2002). Implicit learning and consciousness: a graded, dynamic perspective. *Implicit Learn. Consc.* 2002, 1–40.

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Cousineau, M., McDermott, J. H., and Peretz, I. (2012). The basis of musical consonance as revealed by congenital amusia. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19858–19863. doi: 10.1073/pnas.1207989109

Cristinzio, C., N'diaye, K., Seeck, M., Vuilleumier, P., and Sander, D. (2010). Integration of gaze direction and facial expression in patients with unilateral amygdala damage. *Brain* 133, 248–261. doi: 10.1093/brain/awp255

Dalla Bella, S., Giguère, J.-F., and Peretz, I. (2009). Singing in congenital amusia. *J. Acoust. Soc. Am.* 126, 414–424. doi: 10.1121/1.3132504

Douglas, K. M., and Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nat. Neurosci.* 10, 915–921. doi: 10.1038/nn1925

Escoffier, N., Sheng, D. Y. J., and Schirmer, A. (2010). Unattended musical beats enhance visual processing. *Acta Psychol.* 135, 12–16. doi: 10.1016/j.actpsy.2010.04.005

Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., and Posner, I. M. (2005). The activation of attentional networks. *Neuroimage* 26, 471–479. doi: 10.1016/j.neuroimage.2005.02.004

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., and Posner, I. M. (2002). Testing the efficiency and independence of attentional networks. *J. Cogn. Neurosci.* 14, 340–347. doi: 10.1162/089892902317361886

Fernandez, N. B., Hars, M., Trombetti, A., and Vuilleumier, P. (2019a). Age-related changes in attention control and their relationship with gait performance in older adults with high risk of falls. *Neuroimage* 189, 551–559. doi: 10.1016/j.neuroimage.2019.01.030

Fernandez, N. B., Trost, W. J., and Vuilleumier, P. (2019b). Brain networks mediating the influence of background music on selective attention. *Soc. Cogn. Affect. Neurosci.* 14, 1441–1452.

Fernandez, N. B., Trost, W. J., and Vuilleumier, P. (2020). *Brain networks mediating the influence of background music on selective attention. Soc. Cogn. Affect. Neurosci*. 14, 1441–1452. doi: 10.1093/scan/nsaa004

Fernandez-Duque, D., and Black, S. E. (2006). *Nature reviews neuroscience. Neuropsychology* 20:133. doi: 10.1037/0894-4105.20.2.133

Finucane, A. M., Whiteman, M. C., and Power, J. M. (2010). The effect of happiness and sadness on alerting, orienting, and executive attention. *J. Atten. Disord.* 13, 629–639. doi: 10.1177/1087054709334514

Foxton, J. M., Nandy, R. K., and Griffiths, D. T. (2006). Rhythm deficits in 'tone deafness'. *Brain Cogn.* 62, 24–29. doi: 10.1016/j.bandc.2006.03.005

Fredrickson, B. L. (2004). The broaden-and-build theory of positive emotions. *Philos. Trans. R. Soc. B Biol. Sci.* 359, 1367–1378. doi: 10.1098/rstb.2004.1512

Gosselin, N., Paquette, S., and Peretz, I. (2015). Sensitivity to musical emotions in congenital amusia. *Cortex* 71, 171–182. doi: 10.1016/j.cortex.2015.06.022

Hars, M., Herrmann, F. R., Gold, G., Rizzoli, R., and Trombetti, A. (2013). Effect of music-based multitask training on cognition and mood in older adults. *Age Ageing 43,* 163–200. doi: 10.1093/ageing/aft163

Hirel, C., Lévêque, Y., Deiana, G., Richard, N., Cho, T., Mechtouff, L., et al. (2014). Acquired amusia and musical anhedonia. *Rev. Neurol.* 170, 536–540. doi: 10.1016/j.neurol.2014.03.015

Hutchins, S., Gosselin, N., and Peretz, I. (2010). Identification of changes along a continuum of speech intonation is impaired in congenital amusia. *Front. Psychol.* 1:236. doi: 10.3389/fpsyg.2010.00236

Hyde, K. L., and Peretz, I. (2004). Brains that are out of tune but in time. *Psychol. Sci.* 15, 356–360. doi: 10.1111/j.0956-7976.2004.00683.x

Hyde, K. L., Zatorre, R. J., and Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cereb. Cortex* 21, 292–299. doi: 10.1093/cercor/bhq094

Jiang, C., Liu, F., and Wong, C. P. (2017). Sensitivity to musical emotion is influenced by tonal structure in congenital amusia. *Sci. Rep.* 7:7624. doi: 10.1038/s41598-017-08005-x

Jiang, J., Scolaro, A. J., Bailey, K., and Chen, A. (2011). The effect of music-induced mood on attentional networks. *Int. J. Psychol.* 46, 214–222. doi: 10.1080/00207594.2010.541255

Johnson, K. A., Robertson, I. H., Barry, E., Mulligan, A., Daibhis, A., Daly, M., et al. (2008). Impaired conflict resolution and alerting in children with ADHD: evidence from the Attention Network Task (ANT). *J. Child Psychol. Psychiatry* 49, 1339–1347. doi: 10.1111/j.1469-7610.2008.01936.x

Juslin, P. N., Harmat, L., and Eerola, T. (2014). What makes music emotionally significant? *exploring the underlying mechanisms. Psychol. Music* 42, 599–623 doi: 10.1177/0305735613484548

Kämpfe, J., Sedlmeier, P., and Renkewitz, F. (2011). The impact of background music on adult listeners: a meta-analysis. *Psychol. Music* 39, 424–448. doi: 10.1177/0305735610376261

Kawakami, A., Furukawa, K., Katahira, K., and Okanoya, K. (2013). Sad music induces pleasant emotion. *Front. Psychol.* 4:311. doi: 10.3389/fpsyg.2013.00311

Lévêque, Y., Teyssier, P., Bouchet, P., Bigand, E., Caclin, A., and Tillmann, B. (2018). Musical emotions in congenital amusia: impaired recognition, but preserved emotional intensity. *Neuropsychology* 32, 880–894. doi: 10.1037/neu0000461

Lima, C. F., Brancatisano, O., Fancourt, A., Müllensiefen, D., Scott, S. K., Warren, J. D., et al. (2016). Impaired socio-emotional processing in a developmental music disorder. *Sci. Rep.* 6:34911. doi: 10.1038/srep34911

Lolli, S., Lewenstein, A. D., Basurto, J., Winnik, S., and Loui, P. (2015). Sound frequency affects speech emotion perception: Results from congenital amusia. *Front. Psychol.* 6:1340. doi: 10.3389/fpsyg.2015.01340

Mahoney, J. R., Verghese, J., Goldin, Y., Lipton, R., and Holtzer, R. (2010). Alerting, orienting, and executive attention in older adults. *J. Int. Neuropsychol. Soc.* 16, 877–879. doi: 10.1017/S1355617710000767

Marin, M. M., Thompson, W. F., Gingras, B., and Stewart, L. (2015). Affective evaluation of simultaneous tone combinations in congenital amusia. *Neuropsychologia* 78, 207–220. doi: 10.1016/j.neuropsychologia.2015.10.004

McConnell, M. M., and Shore, D. I. (2011). Upbeat and happy: arousal as an important factor in studying attention. *Cogn. Emot.* 25, 1184–1195. doi: 10.1080/02699931.2010.524396

Medina, D., and Barraza, P. (2019). Efficiency of attentional networks in musicians and non-musicians. *Heliyon* 5:e01315. doi: 10.1016/j.heliyon.2019.e01315

Mitchell, R. L., and Phillips, L. H. (2007). The psychological, neurochemical and functional neuroanatomical mediators of the effects of positive and negative mood on executive functions. *Neuropsychologia* 45, 617–629. doi: 10.1016/j.neuropsychologia.2006.06.030

Mogg, K., Salum, G., Bradley, B., Gadelha, A., Pan, P., Alvarenga, P., et al. (2015). Attention network functioning in children with anxiety disorders, attention-deficit/hyperactivity disorder and non-clinical anxiety. *Psychol. Med.* 45, 2633–2646. doi: 10.1017/S0033291715000586

Park, J., Miller, C. A., Sanjeevan, T., van Hell, J. G., Weiss, D. J., and Mainela-Arnold, E. (2019). Bilingualism and attention in typically developing children and children with developmental language disorder. *J. Speech Language Hear. Res.* 62, 4105–4118. doi: 10.1044/2019_JSLHR-L-18-0341

Peretz, I. (2016). Neurobiology of congenital amusia. *Trends Cogn. Sci.* 20, 857–867. doi: 10.1016/j.tics.2016.09.002

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., et al. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron* 33, 185–191. doi: 10.1016/S0896-6273(01)00580-3

Peretz, I., Brattico, E., Järvenpää M., and Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain* 132, 1277–1286. doi: 10.1093/brain/awp055

Peretz, I., Champod, A. S., and Hyde, K. (2003). Varieties of musical disorders. *Ann. N. Y. Acad. Sci.* 999, 58–75. doi: 10.1196/annals.1284.006

Peretz, I., and Vuvan, D. T. (2017). Prevalence of congenital amusia. *Eur. J. Hum. Genet.* 25, 625–630. doi: 10.1038/ejhg.2017.15

Petersen, S. E., and Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annu. Rev. Neurosci.* 35, 73–89. doi: 10.1146/annurev-neuro-062111-150525

Phillips-Silver, J., Toiviainen, P., Gosselin, N., and Peretz, I. (2013). Amusic does not mean unmusical: beat perception and synchronization ability despite pitch deafness. *Cogn. Neuropsychol.* 30, 311–331. doi: 10.1080/02643294.2013.863183

Posner, M. I., and Petersen, S. E. (1990). The attention system of the human brain. *Annu. Rev. Neurosci.* 13, 25–42. doi: 10.1146/annurev.ne.13.030190.000325

Pralus, A., Fornoni, L., Bouet, R., Gomot, M., Bhatara, A., Tillmann, B., et al. (2019). Emotional prosody in congenital amusia: impaired and spared processes. *Neuropsychologia* 134:107234. doi: 10.1016/j.neuropsychologia.2019.107234

Rowe, G., Hirsh, J. B., and Anderson, K. A. (2007). Positive affect increases the breadth of attentional selection. *Proc. Natl. Acad. Sci. U.S.A.* 104, 383–388. doi: 10.1073/pnas.0605198104

Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C., and Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition* 99, 113–129. doi: 10.1016/j.cognition.2005.01.004

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev.* 110, 145–172. doi: 10.1037/0033-295X.110.1.145

Sachs, M. E., Damasio, A., and Habibi, A. (2015). The pleasures of sad music: a systematic review. *Front. Hum. Neurosci.* 9:404. doi: 10.3389/fnhum.2015.00404

Schindler, I., Hosoya, G., Menninghaus, W., Beermann, U., Wagner, V., Eid, M., et al. (2017). *Measuring aesthetic emotions: a review of the literature and a new assessment tool. PLoS ONE* 12:e0178899. doi: 10.1371/journal.pone.0178899

Silvestrini, N., Piguet, V., Cedraschi, C., and Zentner, R. M. (2011). *Music and auditory distraction reduce pain: emotional or attentional effects? Music Med.* 3, 264–270. doi: 10.1177/1943862111414433

Sluming, V., Brooks, J., Howard, M., Downes, J. J., and Roberts, N. (2007). Broca's area supports enhanced visuospatial cognition in orchestral musicians. *J. Neurosci.* 27, 3799–3806. doi: 10.1523/JNEUROSCI.0147-07.2007

Spaccavento, S., Marinelli, C. V., Nardulli, R., Macchitella, L., Bivona, U., Piccardi, L., et al. (2019). Attention deficits in stroke patients: the role of lesion characteristics, time from stroke, and concomitant neuropsychological deficits. *Behav. Neurol.* 2019:7835710. doi: 10.1155/2019/7835710

Thompson, W. F., Marin, M. M., and Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19027–19032. doi: 10.1073/pnas.1210344109

Thompson, W. F., Schellenberg, E. G., and Husain, G. (2001). Arousal, mood, and the Mozart effect. *Psychol. Sci.* 12, 248–251. doi: 10.1111/1467-9280.00345

Tillmann, B., Gosselin, N., Bigand, E., and Peretz, I. (2012). Priming paradigm reveals harmonic structure processing in congenital amusia. *Cortex* 48, 1073–1078. doi: 10.1016/j.cortex.2012.01.001

Tillmann, B., Jolicoeur, P., Ishihara, M., Gosselin, N., Bertrand, O., Rossetti, Y., et al. (2010). The amusic brain: Lost in music, but not in space. *PLoS ONE* 5:e10173. doi: 10.1371/journal.pone.0010173

Tillmann, B., Lalitte, P., Albouy, P., Caclin, A., and Bigand, E. (2016). Discrimination of tonal and atonal music in congenital amusia: The advantage of implicit tasks. *Neuropsychologia* 85, 10–18. doi: 10.1016/j.neuropsychologia.2016.02.027

Trombetti, A., Hars, M., Herrmann, F. R., Kressig, R. W., Ferrari, S., and Rizzoli, R. (2011). Effect of music-based multitask training on gait, balance, and fall risk in elderly people: a randomized controlled trial. *Arch. Intern. Med.* 171, 525–533. doi: 10.1001/archinternmed.2010.446

Trost, W., Ethofer, T., Zentner, M., and Vuilleumier, P. (2012). Mapping aesthetic musical emotions in the brain. *Cereb. Cortex* 22, 2769–2783. doi: 10.1093/cercor/bhr353

Trost, W., Frühholz, S., Schön, D., Labbé C., Pichon, S., Grandjean, D., et al. (2014). Getting the beat: entrainment of brain activity by musical rhythm and pleasantness. *Neuroimage* 103, 55–64. doi: 10.1016/j.neuroimage.2014.09.009

Vanlessen, N., De Raedt, R., Koster, E. H., and Pourtois, G. (2016). Happy heart, smiling eyes: a systematic review of positive mood effects on broadening of visuospatial attention. *Neurosci. Biobehav. Rev.* 68, 816–837. doi: 10.1016/j.neubiorev.2016.07.001

Vuvan, D., Paquette, S., Goulet, G. M., Royal, I., Felezeu, M., and Peretz, I. (2018). The montreal protocol for identification of amusia. *Behav. Res. Methods* 50, 662–672. doi: 10.3758/s13428-017-0892-8

Wang, K., Fan, J., Dong, Y., Wang, C.-Q., Lee, T. M., and Posner, I. M. (2005). Selective impairment of attentional networks of orienting and executive control in schizophrenia. *Schizophr. Res.* 78, 235–241. doi: 10.1016/j.schres.2005.01.019

Wechsler, D., Coalson, D. L., and Raiford, E. S. (1997). *WAIS-III: Wechsler Adult Intelligence Scale.* San Antonio, TX: Psychological corporation.

Williamson, V. J., Cocchini, G., and Stewart, L. (2011). The relationship between pitch and space in congenital amusia. *Brain Cogn.* 76, 70–76. doi: 10.1016/j.bandc.2011.02.016

Zendel, B. R., Lagrois, M.-É., Robitaille, N., and Peretz, I. (2015). Attending to pitch information inhibits processing of pitch information: the curious case of amusia. *J. Neurosci.* 35, 3815–3824. doi: 10.1523/JNEUROSCI.3766-14.2015

Zentner, M., Grandjean, D., and Scherer, R. K. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* 8, 494–521. doi: 10.1037/1528-3542.8.4.494

Zhu, D. C., Zacks, R. T., and Slade, M. J. (2010). Brain activation during interference resolution in young and older adults: an fMRI study. *Neuroimage* 50, 810–817. doi: 10.1016/j.neuroimage.2009.12.087

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

BRIEF RESEARCH REPORT
published: 27 January 2021
doi: 10.3389/fnins.2021.588914

# Multi-Voiced Music Bypasses Attentional Limitations in the Brain

**Karen Chan Barrett[1]*, Richard Ashley[2], Dana L. Strait[3], Erika Skoe[4], Charles J. Limb[1] and Nina Kraus[5]**

[1] UCSF Sound and Music Perception Lab, Department of Otolaryngology-Head and Neck Surgery, University of California, San Francisco, San Francisco, CA, United States, [2] Program in Music Theory and Cognition, Bienen School of Music, Northwestern University, Evanston, IL, United States, [3] Division of Strategy and Finance, Saint Mary's College, Notre Dame, IN, United States, [4] Department of Speech, Language, and Hearing Sciences, University of Connecticut, Storrs, CT, United States, [5] Auditory Neuroscience Laboratory, Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, United States

Attentional limits make it difficult to comprehend concurrent speech streams. However, multiple musical streams are processed comparatively easily. Coherence may be a key difference between music and stimuli like speech, which does not rely on the integration of multiple streams for comprehension. The musical organization between melodies in a composition may provide a cognitive scaffold to overcome attentional limitations when perceiving multiple lines of music concurrently. We investigated how listeners attend to multi–voiced music, examining biological indices associated with processing structured versus unstructured music. We predicted that musical structure provides coherence across distinct musical lines, allowing listeners to attend to simultaneous melodies, and that a lack of organization causes simultaneous melodies to be heard as separate streams. Musician participants attended to melodies in a Coherent music condition featuring flute duets and a Jumbled condition where those duets were manipulated to eliminate coherence between the parts. Auditory–evoked cortical potentials were collected to a tone probe. Analysis focused on the N100 response which is primarily generated within the auditory cortex and is larger for attended versus ignored stimuli. Results suggest that participants did not attend to one line over the other when listening to Coherent music, instead perceptually integrating the streams. Yet, for the Jumbled music, effects indicate that participants attended to one line while ignoring the other, abandoning their integration. Our findings lend support for the theory that musical organization aids attention when perceiving multi–voiced music.

**Keywords: attention, electroencephalography, N100 response, multivoiced music, counterpoint, polyphony, auditory scene analysis**

OPEN ACCESS

***Edited by:***
*Franco Delogu,*
*Lawrence Technological University, United States*

***Reviewed by:***
*Inyong Choi,*
*The University of Iowa, United States*
*Jufang He,*
*City University of Hong Kong, Hong Kong*

****Correspondence:***
*Karen Chan Barrett*
*Karen.Barrett@ucsf.edu*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *29 July 2020*
***Accepted:*** *06 January 2021*
***Published:*** *27 January 2021*

***Citation:***
*Barrett KC, Ashley R, Strait DL, Skoe E, Limb CJ and Kraus N (2021) Multi-Voiced Music Bypasses Attentional Limitations in the Brain. Front. Neurosci. 15:588914. doi: 10.3389/fnins.2021.588914*

## INTRODUCTION

Humans are limited in their ability to perform two things at once as the performance of simultaneous tasks results in task interference (Neisser and Becklen, 1975; Pashler, 1992; Simons and Chabris, 1999; Herath, 2001; Simons, 2010). Attention is a central bottleneck (Pashler, 1994). Within the auditory domain, it is difficult-to-impossible to attend to multiple independent auditory events simultaneously (Eramudugolla et al., 2005). The "cocktail party problem" refers to the difficulty of paying attention to a single conversation within a crowded room; humans can only pay close attention to one conversation, not several (Cherry, 1953; Hayken et al., 2005).

Speech and music are both complex acoustic stimuli but are fundamentally different. With speech, streams (e.g., a single voice) are heard independently, rarely relying on their integration to convey meaning. Empirical studies of verbal interactions have uncovered certain basic rules of human communication. Within conversations, speakers take turns, leaving little temporal gaps in between turns of speakers, and the overlap between speakers is brief (Sacks et al., 1974; Wooffitt, 2005). There is a relatively orderly process by which conversations unfold (Schegloff, 1987, 1996, 2007). With two people speaking simultaneously, the speakers utterances are not cooperative, but competitive, violating rules of effective communication. With music, however, simultaneous melodies can be both independent and interdependent: possessing their own character (independence), yet also blending to create the composite structure (interdependence). Music and conversation differ markedly in their relative structures. Because of this, the organization of music – its structure – may be one of the key differences between music and speech.

Factors like timbre, rhythm and register aid in the separation of concurrent lines, in what is termed auditory scene analysis (Wright and Bregman, 1987; Bregman, 1990). Comprehending music, however, demands both an appreciation of how individual melodies fit together as well as tracking them individually, a cognitive balance between auditory segregation and integration (Keller, 2008; Ragert et al., 2014; Disbergen et al., 2018). This is particularly true of Baroque polyphony: complex, multi-voiced music written in the 17th and 18th centuries. But how is such polyphony comprehended, given known limitations in human attention?

Experiments investigating attention in music listening have produced opposing theories to explain the perception of polyphony. Gregory (1990) proposed a "divided attention model," which suggested that listeners are capable of simultaneously perceiving multiple musical streams. Sloboda and Edworthy (1981) proposed a different model of attention – the "figure-ground model." In this model, listeners focus on one melody while staying aware of the melody temporarily relegated to the background. By shifting concentration onto different parts, the perception of the music changes; both parts of the percept are processed, but the "figure" or foreground melody receives a different processing compared to the "background." Bigand et al. (2000) contrasted (a) an "Integrative Model of Attention," where listeners integrate two or more voices of polyphony into a single stream, and (b) another model with listeners rapidly shifting their attention between musical lines. This latter attention model, otherwise known as "attentional switching," allows attention to rove between the two melodies, alternately paying attention to one line and then the other.

Multi-voiced music therefore provides an intriguing domain to examine attention and how the auditory system copes with multiple independent inputs (Madsen, 1997; Bigand et al., 2000; Keller, 2001; Ragert et al., 2014). Composers organize the music carefully to highlight individual lines while also promoting a holistic gestalt that integrates multiple lines. Musical structure, or the coherence it yields, may be the prime framework allowing listeners to comprehend multiple melodies in a way that is impossible with multiple independent speech streams.

Electroencephalography (EEG) has emerged as a strategy for studying how listeners parse complex soundscapes. One popular paradigm, the dichotic listening paradigm, asks participants to listen to two simultaneous auditory streams, one directed to each ear. Participants are told to attend to one stream while ignoring the other. Auditory cortical-evoked responses are collected to attended and ignored probes buried within these streams. The N100 response (also known as "N1") is a negative cortical component occurring roughly 100 ms after probe onset. Importantly, the N100 response has been widely used as a neural index of attention: N100 amplitude has been found to be more negative to attended than ignored stimuli (Hillyard et al., 1973), a phenomenon referred to here as the "N100 attention effect." The dichotic paradigm has been employed with simple tone patterns (Snyder et al., 2006) and concurrent speech streams (Coch et al., 2005; Sanders et al., 2006; Stevens et al., 2006, 2009).

Prior experiments have investigated the neural networks underlying perception of multi-voiced music (Satoh et al., 2001; Janata et al., 2002; Uhlig et al., 2013). To our knowledge, however, no experiment has specifically adapted the dichotic listening paradigm to study neural indices of attention when listening to real-world melodies. It was hypothesized that: (1) musical structure provides coherence across different musical lines, creating a cognitive framework for listeners to attend to co-occurring melodies and that (2) lack of coherence-creating structure would cause melodies to be heard as separate streams rather than as an integrated percept. To test these hypotheses, two musical conditions were created: (1) a *Coherent* music condition using Baroque counterpoint duets and (2) a *Jumbled* condition in which the lines of the duets were scrambled, and the melodies played asynchronously with one another, no longer forming coherent compositions (**Figure 1**). In both conditions, listeners were instructed to selectively attend to one melody while ignoring the other. By using the dichotic listening paradigm, this experiment examined how musical structure affects attention to multi-voiced music at a biological level in trained musician participants. Since we propose that musical structure facilitates attention to multiple melodies, we hypothesized that the two musical conditions would elicit different "N100 attention effects." It was predicted that smaller (if any) N100 attention effects would be seen for the normal Coherent music condition compared to the Jumbled. In normal music, melodies in a composition are meant to be integrated into a single gestalt, and this, we argue, strongly interferes with the ability to selectively attend to one stream and ignore the other, hence small if any, N100 selective attention effects will be seen. By contrast, the Jumbled condition is predicted to yield larger N100 attention effects, because the broken coherence helps to focus attention onto a single stream. In this cacophonous Jumbled music, it is relatively easy to selectively attend to one line while ignoring the other since the two lines are not meant to sound like a cohesive unit. By testing trained musicians well-versed in navigating the attentional demands of multipart musical


**FIGURE 1 |** Sample portion of Musical Excerpts. **(A)** shows sample of Coherent excerpt. Music in Coherent excerpts were presented as written by the composers. **(B)** shows sample of corresponding Jumbled excerpt. The score was scrambled and pulled apart by taking large portions of one line (see highlighted box for example) and inserting it in a different location. Additionally, the *tempo* (i.e., speed) of bottom line was slower than *tempo* of top line. Coherence between the two lines has been destroyed, although each line sounded relatively intact when played alone because chunks were excerpted at natural stopping points (ex. cadences or other structural pauses). Excerpt sounds as if the two melodies are playing completely independently and separately.

structures, this experiment aims to illuminate the biological, attentional mechanisms that support the comprehension of sophisticated musical soundscapes. Studying expert listeners may shed light on how and why humans can perceive polyphony – a process that should be challenging and yet occurs relatively effortlessly among listeners from a variety of musical cultures.

# METHODS

## Participants

All experimental procedures were approved by the Northwestern University Institutional Review Board. 21 young adult musicians (11 females), ages 18–37 (mean = 23.81 ± S.D 4.80 years) recruited from Northwestern University participated in this study. All participants had normal hearing (<20 dB HL pure tone thresholds at octave frequencies from 125 to 8000 Hz), no reported history of neurological or learning disorders, and normal IQ (mean = 125.76 ± SD 9.71) as measured by the 2-subtest Wechsler Abbreviated Scale of Intelligence (WASI) (Harcourt Assessment, San Antonio, TX, United States) thereby passing our screening procedures. Participants were all actively practicing their instrument at time of testing and had a mean of 12 years of musical training (SD = 5.22 years). For this experiment, only highly trained-musicians were tested because pilot testing revealed that the paradigm was challenging (see Methods below). Oboists and clarinetists were excluded from participating as the sounds of those instruments were used in the study, and previous research has found that musicians show enhanced neural activity and preferential attention to the sound of their instrument (Pantev et al., 2001a; Shahin et al., 2003, 2008; Margulis et al., 2007; Strait et al., 2012). All participants had significant small ensemble experience or played a multi-line instrument, ensuring familiarity with hearing multi-voice music.

## Musical Stimuli

There were two experimental music conditions: Coherent and Jumbled. Stimuli consisted of four excerpts from Johann Philippe Quantz's *Six Duets for Two Flutes*, Op. 2 and George Telemann's *Six Sonatas for Two Flutes*, TWV 40, No. 103 and 107. The Coherent music condition featured the compositions as written by the composers. In the Jumbled condition, one line was played at a slower *tempo* than the other and the score was pulled apart and subsequently rearranged so that musical figures in one line no longer coordinated with those in the other line. Thus, in the Jumbled condition the music lacked structural coherence *between* the two lines because of the scrambling of the music (see **Figure 1** for details).

All musical excerpts were transcribed into Sibelius 4 (Sibelius, Daly City, CA, United States) and MIDI files were exported into Logic Pro 8 (Apple Inc., Cupertino, CA, United States). Stock oboe and bass clarinet sounds were replaced with high-quality, realistic-sounding instruments from the EastWest Sound Library (EastWest Studios/Quantum Leap, Hollywood, CA, United States). Note that this resulted in two lines which were in different timbres and registers. Although the Baroque duets were written for two flutes, instruments and registers were altered here; pilot testing revealed that selective attention to one line was too difficult even for expert listeners when melodies were played by the same instrument in the same register. To ensure and verify that participants were paying attention during the experiment, a behavioral, perceptual task was given. Since it was difficult to ask concrete questions about the music (i.e., simple questions easily yielding objective answers), participants were told to count "target tones" – infrequently occurring, quarter-tone flat mistunings inserted into the line. Versions of sound files were made where the intended "attended" line featured 3–7 randomly chosen, quarter-tone flat mistunings during each 3-min excerpt. These target tone mistunings were made using the Pitch Bend Plugin in Logic Pro 8. At the end of the excerpt, participants were quizzed on the total number of mistunings detected as a screening method for ensuring they were paying attention and following experimental directions (see Methods below). The melodies were normalized for volume using the Level16 sound editing program (Tom Carrell and Bob Tice, University of Nebraska).

## Electrophysiology

### N100 Evoking (Probe) Stimulus

The evoking stimulus was a Steinway piano sound ($G_1$, F0 = 100 Hz, 200 ms) synthesized using Logic Pro (Apple Inc) that was superimposed into the musical lines, occurring at randomized interstimulus intervals (ISIs) of 600, 900, or 1200 ms. This sound possesses a sharp onset that is conducive to recording robust N100 responses.

### Recording Parameters and Protocol

Auditory-evoked potentials were recorded to the probe using a 32-channel silver electrode cap (Electrocap International, Eaton, OH, United States) in NeuroScan Acquire 4.3 (Compumedics) while participants were seated in a sound-attenuated booth. A single electrode was placed on each of the right and left earlobes; right ear acted as reference during the online recording and the recordings were re-referenced to linked earlobes offline. Single electrodes were placed on the medial canthus of the right eye and on the lower eyelid of the left eye to act as eye-blink monitors, so that trials containing eyeblink artifacts could be rejected from the average. Contact impedance for all electrodes was under 10 kΩ (Ferree et al., 2001; Kappenman and Luck, 2010). Neural recordings were off-line filtered from 0.1 to 100 Hz and digitally sampled at a rate of 500 Hz.

The probe was presented with the contrapuntal melodies played through two wall-mounted speakers located exactly 1 meter to the left and right of the participant at 180° apart from one another. The melodies were played dichotically, one to each speaker. Participants were asked to attend to one of the two simultaneously presented melodies which differed in location (left/right speaker), instrument (bass clarinet or oboe), and musical content. This procedure was adapted from previous experiments that used dichotic speech streams (Coch et al., 2005; Sanders et al., 2006; Stevens et al., 2009). Participants were initially told which instrument to attend to and were directed to which speaker the instrument would be presented from. The attended instrument and its initial location (right/left) were randomized across participants, as was the order of the Jumbled and Coherent conditions. The probe was presented randomly to the left or right (i.e., attended or ignored) sides of the head. The musical melodies were played at 55 dB SPL and the probe was played at 65 dB SPL, creating a 10 dB difference in keeping with protocol used in other dichotic listening paradigms (Coch et al., 2005; Sanders et al., 2006; Stevens et al., 2009). The recording took place in four three-minute blocks. After each 3-min block, participants were quizzed on the number of mistunings they heard in the attended melody to ensure active engagement and listening. To control for any ear advantages, the melody then switched sides and participants were asked to change their attended side (left/right) in order to continue the task with the same melody and timbre. For example, if the bass clarinet melody originally came out of the left speaker, in the next 3-min segment it originated from the right speaker. Participants were told to switch their attention to the other speaker.

All participants were able to perform the selective attention task as indicated by their performance in tracking the number of target tone mistunings in the attended line (average percent correct = 80%).

## Data Processing and Analysis

The continuous recordings were bandpass filtered offline from 0.1 to 40 Hz (12 dB/octave, zero phase shift) and subjected to a spatial filtering algorithm in Neuroscan Edit 4.3 (Compumedics) to reduce the influence eye blinks on the recordings. For each musical condition, recordings were epoched over a window of −100 to 500 ms, using the onset of the probe stimulus to define 0 ms. Epochs containing muscle artifact that exceeded a $\pm$ 100 $\mu$V threshold were removed using a spatial algorithm where the algorithm computes the degree of similarity between each epoch and the average of all epochs using Pearson's correlations. Individual responses were ranked according to their Pearson's $r$-values and the most poorly correlated 30% were discarded. The remaining 70% were averaged, making up the final averaged evoked response for each subject in each condition (Abrams et al., 2008); these 500 artifact free responses from each participant were subsequently used for statistical analysis.

### Statistical Analysis

Mean amplitudes for the N100 cortical response (occurring from 100 to 150 ms following stimulus onset) were calculated for each electrode channel. This time range aligns with well-described characteristics of the N100. Mean Amplitudes during this time window were then averaged into scalp regions of interest (ROI, see **Figure 2**) by mathematically averaging the mean amplitudes from individual channels; these ROI blocks were based on precedent set from a previous paper (Strait et al., 2014). This resulted in a single amplitude for each ROI for each participant in both the attend and ignore conditions. Visual inspection of waveforms revealed that responses at the Frontal, Parietal, and Occipital ROI were not robust (i.e., did not resemble a clean auditory-evoked response) and therefore excluded from analysis. Differences in N100 amplitudes were compared across musical conditions using a 2 $\times$ 2 $\times$ 4 RMANOVA with musical condition (Jumbled vs. Coherent), attention (Attend vs. Ignore), and scalp ROI (Prefrontal, Central, Left, Right, see **Figure 2**) as within-subject factors. Data was normally distributed as assessed by the Kolmogorov Smirnov test of normality. Mauchly's Test of Sphericity showed that the assumption of sphericity was violated, therefore RMANOVA statistics are reported below using the Greenhouse-Geisser correction (see Results).

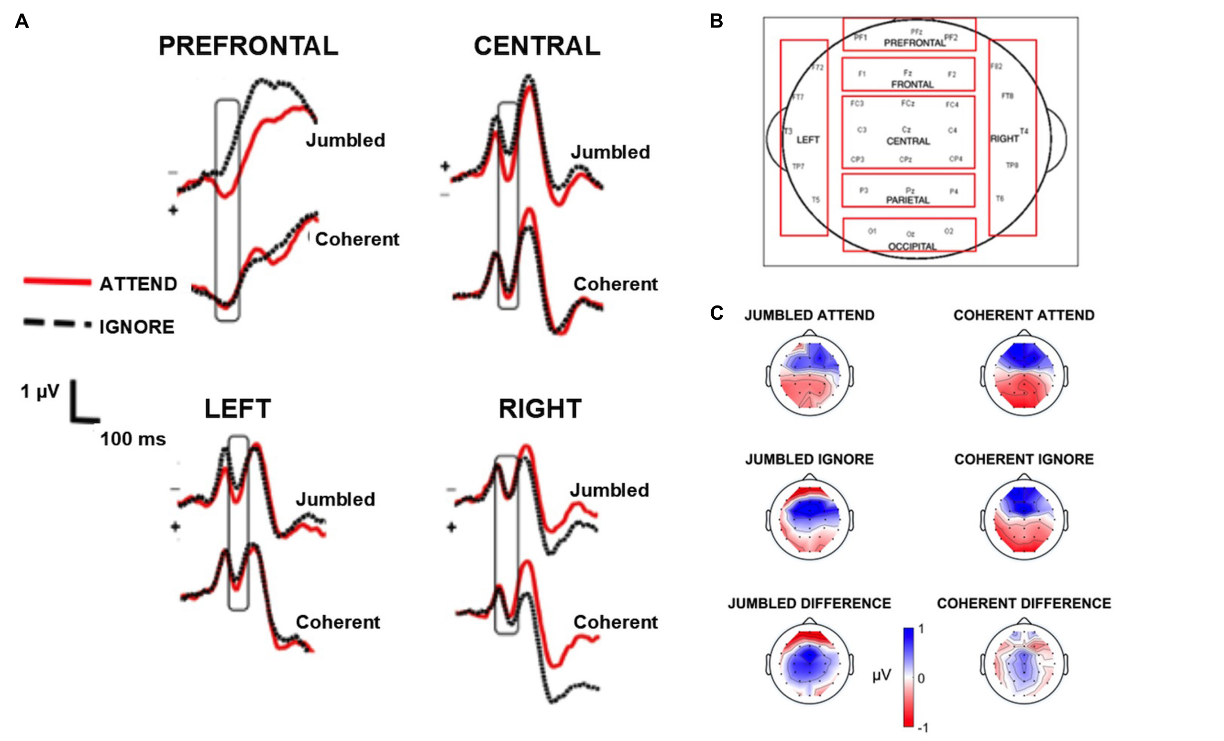



**FIGURE 2 |** EEG waveforms. **(A)** shows the grand average waveforms highlighting the N100 component (box) for the four regions of interest [prefrontal, central, left, and right regions, see **(B)**]. **(C)** depicts the topographic maps of the N100 component for the four conditions as well as the difference between attend and ignore conditions, generated by subtracting the topographic map of the attend from the ignore conditions. Topographic maps zoom in at 115 ms where the greatest negativity occurs in the grand average waveforms.

# RESULTS

## Summary of Results

Participants demonstrated significant selective attention effects in the Jumbled Condition but not the Coherent Condition, particularly in prefrontal and central scalp regions (see below and **Table 1**).

## N100 Attention Effect Seen in Jumbled Condition but Not Coherent Condition

Analysis revealed a non-significant but trending effect of attention $F(1,20) = 3.49$, $p = 0.077$ and a significant main effect of condition $F(1,20) = 5.23$, $p = 0.033$. Non-significant but trending two-way interaction effects were seen both (1) between attention and condition $F(1,20) = 3.39$, $p = 0.080$, suggesting that attention might be impacting the two musical conditions differently, and (2) between condition and scalp region $F(1.05,21) = 3.47$, $p = 0.075$ suggesting that the cortical regions were eliciting different responses in the two musical conditions. Finally, a non-significant but trending three-way interaction between attention, condition, and scalp region was observed $F(1.32,26.35) = 3.06$, $p = 0.082$.

**TABLE 1 |** Within-music comparisons for N100 amplitudes.

| Scalp region | JUMBLED Attend vs. Ignore | COHERENT Attend vs. Ignore |
|---|---|---|
| Prefrontal | $\mathbf{t = -2.13, p = 0.046^*}$ | $t = -0.029$, $p = 0.98$ |
| Central | $\mathbf{t = -2.45, p = 0.024^*}$ | $t = -0.32$, $p = 0.75$ |
| Left | $t = -0.273$, $p = 0.79$ | $t = -1.58$, $p = 0.13$ |
| Right | $t = 1.12$, $p = 0.28$ | $t = 0.32$, $p = 0.75$ |

*Table shows post hoc paired t-tests comparing Attend to Ignore N100 amplitudes over each scalp region of interest and within each musical condition. The Jumbled condition shows significant N100 attention effects (i.e., significantly more negative Attend amplitudes compared to Ignore amplitudes) in prefrontal and central sites (see bolded values), while the Coherent music condition shows no effect of attention *p < 0.05.*

Following a main effect of condition as well as trending interaction effects, *post hoc* paired sample *t*-tests were employed comparing Attend vs. Ignore amplitudes within each musical condition (**Table 1**).

*Post hoc* paired-sample *t*-tests were performed comparing Attend to Ignore (**Table 1** and **Figure 3**). In the Jumbled condition, Attend amplitudes were significantly more negative than Ignore amplitudes at the Prefrontal region [$t(20) = -2.13$, $p = 0.046$] and Central region [$t(20) = -2.45$, $p = 0.024$]. Thus, in the Jumbled condition a selective "N100 attention effect" was seen; the amplitude of the N100 response was more negative to attended rather than ignored stimuli (**Figure 3**). This suggests that participants were able to selectively attend to one line over another. No significant differences between Attend and Ignore amplitudes, however, were found for the Coherent music condition (**Figure 3**).

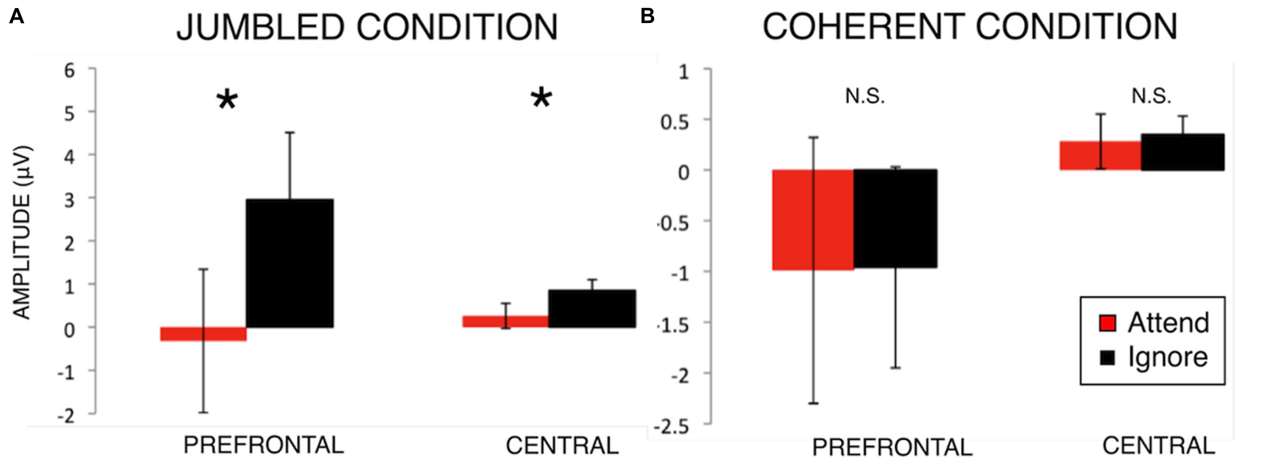



**FIGURE 3 |** N100 amplitudes for Attend and Ignore responses in the Jumbled Condition **(A)** and Coherent Condition **(B)**. Bar graphs comparing N100 group-mean amplitudes in both music conditions (mean ± S.E.) for Prefrontal and Central scalp regions of interest. An effect of attention on the N100 response magnitude (i.e., N100 attention effect) is seen in the Jumbled music condition (cacophonous compositions, **A**) but not the Coherent music condition (normal music compositions, **B**). $^*p < 0.05$.

## DISCUSSION

The effect of musical structure on attention was examined to evaluate the hypothesis that musical coherence aids the processing of simultaneous musical melodies. This hypothesis was investigated through neural indices of attention when musicians listened to incoherent versus coherent musical stimuli. Results showed that attentional responses are modulated by whether the musical lines were coherent or disjunct: (1) No N100 selective attention effect was found when participants heard normally structured polyphonic compositions (Coherent condition) in that the attend and ignore conditions did not differ from each other; (2) An N100 selective attention effect, however, was found when listening to incoherent music derived from the coherent excerpts (Jumbled condition). These results suggest that musician participants were capable of selectively attending to one line while ignoring the other when the lines no longer blended together to form an intact composition. When listening to coherent music, however, they integrated the two melodies into one percept.

These results provide biological support for the "integrative model of attention" that suggested that listeners integrate two or more voices of polyphony into a single perceptual stream (Bigand et al., 2000). Our results cannot rule out the "selective attention model" for the Coherent condition as it is possible that participants' attention flickered between streams so rapidly as to not affect the N100 cortical response. While possible, the results, especially the fact that Ignored inputs were similar to Attend inputs when listening to coherent music (**Table 1**), seem to more plausibly support an "integrative model of attention." Note that the Jumbled music condition with incoherent musical lines supported the selective attention model. Taken together, results support our hypothesis that musical organization facilitates attention to multi-voiced music. The Coherent musical condition resulted in listeners using integrative listening strategies while the Jumbled condition yielded selective attention effects. Musical structure may compensate for attentional limitations typically experienced when two auditory signals are presented concurrently.

Here, "musical structure" loosely describes coherence, or the global organization of a composition. The key difference between the Coherent and Jumbled music conditions is that the melodies in the Jumbled condition eliminated correspondence between the two lines and they moved asynchronously with each other, disrupting meter and tonality. Periodic rhythmic structures and metric frameworks may play a key part in attention and integrative listening to music (Keller, 2001; Hurley et al., 2018), freeing attentional resources to efficiently process multiple parts (Keller, 1999; Keller and Burnham, 2005). Furthermore, factors like unstable tonality have been predicted to interfere with integrative listening as well (Keller, 2001). These ideas support the concept that in the Jumbled music condition, by virtue of its disorganization, it was difficult to fuse the two lines.

This "disorganization" may account for why the Jumbled music condition yielded N100 results similar to those found in simultaneous speech conditions used in other studies (Coch et al., 2005; Sanders et al., 2006; Stevens et al., 2009). A parallel exists between competing conversational speech effects, where two people are talking over each other, and the cacophonous music used in the Jumbled music condition. The organization underlying "normal" music may be the vital distinguishing feature that sets music apart from other auditory inputs like speech. Multiple speakers are heard as competing when they talk at the same time, but simultaneous melodies are composed to reinforce and complement one another.

To our knowledge, this is the first study using naturalistic musical materials in a dichotic listening paradigm to investigate neural indices of attention with concurrent musical lines. Even so, the polyphonic duets used featured relatively simple counterpoint. Future experiments may want to employ more complex polyphonic compositions (e.g., fugues, Baroque trios). Additional information about the neural activation patterns behind segregation and integration may be gained through the incorporation of neural imaging techniques. It is thought that the planum temporale (PT) is involved in stream segregation while the inferior parietal cortex (IPC) is involved in auditory integration (see Gutschalk and Dykstra, 2014; Ragert et al., 2014).

In this experiment, only the effect of musical structure, an exogenous factor of attention, was investigated when listening. Different types of musical training, conceivably considered as endogenous factors of attention, may also influence indices of attention when perceiving polyphony. For example, adult musicians demonstrate enhanced cortical and subcortical responses to the timbre of their own instruments, showing a preference for their major instrument (Pantev et al., 2001b; Margulis et al., 2007; Shahin et al., 2008; Trainor et al., 2009; Strait et al., 2012; Barrett et al., 2013; Shahin et al., 2003). This experiment specifically excluded musicians who played oboe or clarinet, the instruments used in the stimuli, to control for any possible timbre preference. Future experiments may want to investigate the interaction between endogenous, top-down (extramusical or training-related) and exogenous, bottom-up (music structure-related) factors that affect attention. Would musicians' preference for their own timbre override or change their responses to the musical structure and its effect on attention?

Aside from examining musicians playing different instruments, one could also divide subject groups according to various aspects of musical training. Participants here were experienced in listening to simultaneous musical lines. Future research might investigate whether neural responses differ between instrumentalists who play single-line instruments (ex. flute, oboe) as opposed to those who play multi-line instruments (e.g., organ, piano, keyboard instruments). Additionally, neural responses might differ between musicians with minimal vs. substantial amounts of aural skills training since ear training aims to develop an ability to hear multiple concurrent lines. Moreover, non-musicians and those struggling with acoustic scene analysis, such as cochlear implant users who receive degraded auditory input, may show different results as well.

In conclusion, this experiment demonstrated that musical organization facilitates attention to the broader musical context when trained musicians listened to multi-voiced music as evidenced through auditory cortical-evoked potentials. Musical structure may help humans process simultaneous melodies as a way to cope with the attentional limitations that one would have for other auditory stimuli (such as speech). This organization may allow listeners to integrate musical melodies into one percept, thereby aiding in the comprehension of polyphonic music.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Northwestern University Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

KB designed the experiment, collected and analyzed the data, wrote the manuscript with feedback from co-authors. RA advised KB in design of the experiment and contextualization of scientific work. DS co-designed the experiment, aided in technical support in EEG collection, and co-analyzed the data. ES wrote the scripts that allowed for data analysis and aided in making figures. CL co-wrote the manuscript. NK oversaw the entire experiment from experimental design to data collection to manuscript preparation. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was conducted as part of KB's dissertation research at Northwestern University and was supported by a Northwestern University Cognitive Science fellowship awarded to KB. Publication fees for this article were partially supported by the Open Access Publishing Fund at University of California, San Francisco (UCSF).

## ACKNOWLEDGMENTS

We would like to thank Rafael Escobedo-Quiroz for his technical advice on EEG collection and Shivani Bhatia for her *ad hoc* assistance on this project. We would also like to thank Jen Krizman and Trent Nicol for comments on revised versions of this manuscript.

## REFERENCES

Abrams, D. A., Nicol, T., Zecker, S., and Kraus, N. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J. Neurosci.* 28, 3958–3965. doi: 10.1523/Jneurosci.0187-08.2008

Barrett, K. C., Ashley, R., Strait, D. L., and Kraus, N. (2013). Art and science: how musical training shapes the brain. *Front. Psychol.* 4:713. doi: 10.3389/fpsyg.2013.00713

Bigand, E., McAdams, S., and Foret, S. (2000). Divided attention in music. *Int. J. Psychol.* 35, 270–278.

Bregman, A. S. (1990). *Auditory Scene Analysis?: The Perceptual Organization of Sound*. Cambridge, Mass: MIT Press.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979.

Coch, D., Sanders, L. D., and Neville, H. J. (2005). An event-related potential study of selective auditory attention in children and adults. *J. Cogn. Neurosci.* 17, 605–622.

Disbergen, N. R., Valente, G., Formisano, E., and Zatorre, R. J. (2018). Assessing top-down and bottom-up contributions to auditory stream segregation and integration with polyphonic music. *Front. Neurosci.* 12:121. doi: 10.3389/fnins.2018.00121

Eramudugolla, R., Irvine, D. R. F., McAnally, K. I., Martin, R. L., and Mattingley, J. B. (2005). Directed attention eliminates change deafness in complex auditory scenes. *Curr. Biol.* 15, 1108–1113.

Ferree, T. C., Luu, P., Russell, G. S., and Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clin. Neurophysiol.* 112, 536–544. doi: 10.1016/S1388-2457(00)00533-2

Gutschalk, A., and Dykstra, A. R. (2014). Functional imaging of auditory scene analysis. *Hear. Res.* 307, 98–110. doi: 10.1016/j.heares.2013.08.003

Hayken, S., Chen, Z., Haykin, S., and Chen, Z. (2005). The cocktail party problem. *Neural Comput.* 17, 1875–1902. doi: 10.1162/0899766054322964

Herath, P. (2001). Neural correlates of dual task interference can be dissociated from those of divided attention: an fMRI study. *Cereb. Cortex* 11, 796–805. doi: 10.1093/cercor/11.9.796

Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science* 182:177.

Hurley, B. K., Fink, L. K., and Janata, P. (2018). Mapping the dynamic allocation of temporal attention in musical patterns. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 1694–1711. doi: 10.1037/xhp0000563

Janata, P., Tillmann, B., and Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cogn. Affect. Behav. Neurosci.* 2:121.

Kappenman, E. S., and Luck, S. J. (2010). The effects of electrode impedance on data quality and statistical significance in ERP recordings. *Psychophysiology* 47, 888–904. doi: 10.1111/j.1469-8986.2010.01009.x

Keller, P. (1999). Attending in complex musical interactions: the adaptive dual role of meter. *Aust. J. Psychol.* 51, 166–175. doi: 10.1080/00049539908255354

Keller, P. E. (2001). Attention resource allocation in musical ensemble performance. *Psychol. Music* 29, 20–38. doi: 10.1177/0305735601291003

Keller, P. E. (2008). "Joint action in music performance," in *Enacting Intersubjectivity: A Cognitive and Social Perspective on the Study of Interactions Emerging communication: Studies on new technologies and practices in communication*, Vol. 10, eds F. Morganti, A. Carassa, and G. Riva (Amsterdam: IOS Press), 1–18.

Keller, P. E., and Burnham, D. K. (2005). Musical meter in attention to multipart rhythm. *Music Percept.* 22, 629–661. doi: 10.1525/Mp.2005.22.4.629

Madsen, C. K. (1997). Focus of attention and aesthetic response. *J. Res. Music Educ.* 45, 80–89. doi: 10.2307/3345467

Margulis, E. H., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., and Wong, P. C. M. (2007). Selective neurophysiologic responses to music in instrumentalists with different listening biographies. *Hum. Brain Mapp.* 30, 267–275.

Neisser, U., and Becklen, R. (1975). Selective looking: attending to visually specified events. *Cogn. Psychol.* 7, 480–494. doi: 10.1016/0010-0285(75)90019-5

Pantev, C., Engelien, A., Candia, V., and Elbert, T. (2001a). Representational cortex in musicians. Plastic alterations in response to musical practice. *Ann. N. Y. Acad. Sci.* 930, 300–314.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., and Ross, B. (2001b). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport* 12:169.

Pashler, H. (1992). Attentional limitations in doing two tasks at the same time. *Curr. Dir. Psychol. Sci.* 1, 44–48.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychol. Bull.* 116, 220–244. doi: 10.1037/0033-2909.116.2.220

Ragert, M., Fairhurst, M. T., and Keller, P. E. (2014). Segregation and integration of auditory streams when listening to multi-part music. *PLos One* 9:e84085. doi: 10.1371/journal.pone.0084085

Sacks, H., Schegloff, E., and Jefferson, G. (1974). A simplest systematics for organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.2307/412243

Sanders, L. D., Stevens, C., Coch, D., and Neville, H. J. (2006). Selective auditory attention in 3-to 5-year-old children: an event-related potential study. *Neuropsychologia* 44, 2126–2138.

Satoh, M., Takeda, K., Nagata, K., Hatazawa, J., and Kuzuhara, S. (2001). Activated brain regions in musicians during an ensemble: a PET study. *Cogn. Brain Res.* 12, 101–108. doi: 10.1016/S0926-6410(01)00044-1

Schegloff, E. A. (1987). Analyzing single episodes of interaction – an exercise in conversation analysis. *Soc. Psychol. Q.* 50, 101–114. doi: 10.2307/2786745

Schegloff, E. A. (1996). "Turn organization: one intersection of grammar and interaction," in *Interaction and Grammar*, eds E. Schegloff, E. Ochs, and S. Thompson (Cambridge: Cambridge University Press), doi: 10.1017/cbo9780511620874.002

Schegloff, E. A. (2007). "Sequence organization in interaction: a primer in conversation analysis I," in *Sequence Organization in Interaction: A Primer in Conversation Analysis I*, ed. E. Schegloff (Cambridge: Cambridge University Press), doi: 10.1017/CBO9780511791208

Shahin, A., Bosnyak, D. J., Trainor, L. J., and Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J. Neurosci.* 23, 5545–5552.

Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., and Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *NeuroImage* 41, 113–122.

Simons, D. J. (2010). Monkeying around with the gorillas in our midst: familiarity with an inattentional-blindness task does not improve the detection of unexpected events. *IPerception.* 1, 3–6. doi: 10.1068/i0386

Simons, D. J., and Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception* 28, 1059–1074. doi: 10.1068/p281059

Snyder, J. S., Alain, C., and Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cogn. Neurosci.* 18, 1–13. doi: 10.1162/089892906775250021

Sloboda, J., and Edworthy, J. (1981). Attending to two melodies at once: the of key relatedness. *Psychol. Music* 9, 39–43. doi: 10.1177/03057356810090010701

Stevens, C., Lauinger, B., and Neville, H. (2009). Differences in the neural mechanisms of selective attention in children from different socioeconomic backgrounds: an event-related brain potential study. *Dev. Sci.* 12, 634–646. doi: 10.1111/j.1467-7687.2009.00807.x

Stevens, C., Sanders, L., and Neville, H. (2006). Neurophysiological evidence for selective auditory attention deficits in children with specific language impairment. *Brain Res.* 1111, 143–152. doi: 10.1016/j.brainres.2006.06.114

Strait, D. L., Chan, K., Ashley, R., and Kraus, N. (2012). Specialization among the specialized: auditory brainstem function is tuned in to timbre. *Cortex* 48, 360–362.

Strait, D. L., Slater, J., Abecassis, V., and Kraus, N. (2014). Cortical response variability as a developmental index of selective auditory attention. *Dev. Sci.* 17, 175–186. doi: 10.1111/desc.12107

Trainor, L. J., Shahin, A. J., and Roberts, L. E. (2009). Understanding the benefits of musical training: effects on oscillatory brain activity. *Ann. N. Y. Acad. Sci.* 1169, 133–142. doi: 10.1111/j.1749-6632.2009.04589.x

Uhlig, M., Fairhurst, M. T., and Keller, P. E. (2013). The importance of integration and top-down salience when listening to complex multi-part musical stimuli. *NeuroImage* 77, 52–61. doi: 10.1016/j.neuroimage.2013.03.051

Wooffitt, R. (2005). *Conversation Analysis and Discourse Analysis?: a Comparative and Critical Introduction*. Thousand Oaks, CA: SAGE.

Wright, J. K., and Bregman, A. S. (1987). Auditory stream segregation and the control of dissonance in polyphonic music. *Contemp. Music Rev.* 2, 63–92. doi: 10.1080/07494468708567054

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 23 February 2021
doi: 10.3389/fnhum.2021.578594

# Eros, Beauty, and Phon-Aesthetic Judgements of Language Sound. We Like It Flat and Fast, but Not Melodious. Comparing Phonetic and Acoustic Features of 16 European Languages

*Vita V. Kogan*[1] *and Susanne M. Reiterer*[2,3]*

*[1] School of European Culture and Languages, University of Kent, Kent, United Kingdom, [2] Department of Linguistics, University of Vienna, Vienna, Austria, [3] Teacher Education Centre, University of Vienna, Vienna, Austria*

OPEN ACCESS

***Edited by:***
*Elvira Brattico,*
*Aarhus University, Denmark*

***Reviewed by:***
*Francesca M. M. Citron,*
*Lancaster University, United Kingdom*
*Li Su,*
*Academia Sinica, Taiwan*

****Correspondence:***
*Susanne M. Reiterer*
*Susanne.Reiterer@univie.ac.at*

***Specialty section:***
*This article was submitted to Speech and Language, a section of the journal Frontiers in Human Neuroscience*

***Received:*** *17 July 2020*
***Accepted:*** *12 January 2021*
***Published:*** *23 February 2021*

***Citation:***
*Kogan VV and Reiterer SM (2021) Eros, Beauty, and Phon-Aesthetic Judgements of Language Sound. We Like It Flat and Fast, but Not Melodious. Comparing Phonetic and Acoustic Features of 16 European Languages. Front. Hum. Neurosci. 15:578594. doi: 10.3389/fnhum.2021.578594*

This article concerns sound aesthetic preferences for European foreign languages. We investigated the phonetic-acoustic dimension of the linguistic aesthetic pleasure to describe the "music" found in European languages. The Romance languages, French, Italian, and Spanish, take a lead when people talk about melodious language – the music-like effects in the language (a.k.a., phonetic chill). On the other end of the melodiousness spectrum are German and Arabic that are often considered sounding harsh and un-attractive. Despite the public interest, limited research has been conducted on the topic of phonaesthetics, i.e., the subfield of phonetics that is concerned with the aesthetic properties of speech sounds (Crystal, 2008). Our goal is to fill the existing research gap by identifying the acoustic features that drive the auditory perception of language sound beauty. What is so music-like in the language that makes people say "it is music in my ears"? We had 45 central European participants listening to 16 auditorily presented European languages and rating each language in terms of 22 binary characteristics (e.g., beautiful – ugly and funny – boring) plus indicating their language familiarities, L2 backgrounds, speaker voice liking, demographics, and musicality levels. Findings revealed that all factors in complex interplay explain a certain percentage of variance: familiarity and expertise in foreign languages, speaker voice characteristics, phonetic complexity, musical acoustic properties, and finally musical expertise of the listener. The most important discovery was the trade-off between speech tempo and so-called linguistic melody (pitch variance): the faster the language, the flatter/more atonal it is in terms of the pitch (speech melody), making it highly appealing acoustically (sounding beautiful and sexy), but not so melodious in a "musical" sense.

**Keywords: phon-aesthetics, language attitudes and ideologies, speech melody, speech rate, language perception, crosslinguistic comparison, rhythm in language, prosody and intonation perception**

## INTRODUCTION

There is almost universal agreement that Italian, Spanish, and French are appealing and melodious languages to the human ear. Italian, it is often said, is the language of opera, and only a rare singer does not have it in their linguistic repertoire. There are several reasons why Italian might be so pleasant to hear. Dr. Patti Adank, professor of Speech, Hearing, and Phonetic Science at the University College London, says that an open syllabic structure and a high vocalic share make Italian the optimal language for singing (as cited in Kerr, 2017). Matteo Dalle Fratte, a musicologist and founder of *Melofonetica.com*, says that "Italian is the language built to be sung," that the alternation of short and long consonants in Italian (gemination or double consonants) produces "the agogic accent" and "incredible expression and dramatic tension to the text" (Dalle Fratte, 2018). Opera lovers seem to agree. Online forums for singers are replete with comments such as "As for the best sounding language when sung, I feel that it's [*sic*] Finnish, and Italian and Spanish sound good too. All are heavily vowel-y languages which is pretty much essential for a good singing language" (Guest123456, 2010).

Whereas the Romance languages are frequently described as melodious, beautiful and sexy, German and Arabic, on the other side of the likeability spectrum, are for some too harsh or vocally unpleasant due to their consonant clusters, and to many Western ears, tonal languages, such as Cantonese or Mandarin, sound whiny (Science Chat Forum, 2011; Quora, 2015). Certainly, it is hard to separate the effects of phonetic features from the influence of the languages' socio-cultural aura – the speakers and the history behind a language. French might sound lovely to one's ear because it has a high vocalic index (every other sound in French is a vowel) but also because listening to French brings memories of Les Champs-Élysées, fragrant wines, and Duma's novels. In our previous study (Reiterer et al., 2020), we found that pre-existing socio-cultural factors, like, second language experience, as well as the speaker's voice, explained most (two thirds) of the variation in listeners' aesthetic judgments. Yet, the phonetic properties of languages also played a significant role. In the present study, we focus on the phonetic-acoustic dimension of the linguistic aesthetic pleasure and try to quantify the "music" found in European languages. Despite the public interest, there has been little research into phonaesthetics, a subfield of phonetics that is concerned with the aesthetic properties of speech sounds (Crystal, 2008). This is surprising, given the success of aesthetic research in other fields: e.g., the aesthetics of objects (Jacobsen et al., 2004), the experience of music (Brattico et al., 2013; Reuter and Siddiq, 2017), and art (Zaidel et al., 2013; Leder et al., 2014), as well as the appreciation of mathematical beauty (Zeki et al., 2014).

The idea that some languages sound like music is not counterintuitive. According to the *musical protolanguage hypothesis*, speech and music originate from the same source; i.e., they come from the imitation and modification of environmental sounds to express basic emotions such as love, anger, pity, and sadness (Kirby, 2011; Fitch, 2013; Ma et al., 2019). Because of the vocal tract constraints associated with speaking and singing, similar emotions are conveyed by similar acoustic features in both domains. Previous studies have shown that the speech rate or tempo and F0/pitch function more or less the same way in speech and music in terms of their effects on listeners' ratings (Juslin, 1997, 2000).

Speech rate (tempo in music) is a temporal aspect that typically signifies a certain number of units per duration; e.g., syllables per second. In both speech and music, rate or tempo increases with high-arousal or "active" emotions such as anger, fear, and happiness. The opposite is true for low-arousal or "passive" emotions such as sadness and tenderness (Juslin and Laukka, 2003; Ma and Thompson, 2015). Fundamental frequency F0 (analogous to pitch in music and the acoustic correlate of the main portion of the perceived pitch in the speaker's voice) is characterized by the rate at which the vocal folds open and close across the glottis. In both music and speech, a low pitch is associated with sadness and a high pitch – with happiness. A rising F0 contour evokes active emotions, whereas a falling F0 contour is associated with passive emotions (Cordes, 2000). The same holds true for pitch variation: happy, angry, and frightened responses increase with higher pitch variation and the perception of sad and angry stimuli is influenced by lower pitch variation (Breitenstein et al., 2001).

Not all studies demonstrate overlapping emotional ratings for speech and music. Ilie and Thompson (2006) manipulated several acoustic cues in both domains, and even though they observed similarities for some of the cues – e.g., fast speech and fast music were perceived as more energetic and tense compared to slower speech and music, – the effects were not consistent across all acoustic dimensions. For example, participants found a high-pitched speech (though not music) more pleasant. The authors concluded that even though the same circuitry might be involved in connecting acoustic events and their corresponding affective meanings in both speech and music, different attentional strategies might be used for the two types of stimuli; that is, listeners pay greater attention to prominent aesthetic properties of music than language, where verbal information is probably the primary attention attractor. Some studies find that speech and music have domain-specific cues to emotions because they have different structural features and functions (e.g., Krumhansl, 1990). For example, Quinto et al. (2013), after measuring pitch variability and rhythmic properties of speech and music, concluded that whereas a changing pitch conveys emotional intentions in speech, it does not behave the same way in music. One explanation for this, they say, relates to cross-cultural differences in emotional verbal communication – the so-called "pull-effects" – (Scherer et al., 2003), and even though emotional decoding across cultures is relatively good, people within a culture are still better able to identify emotions than outsiders (Mesquita, 2003). However, the cultural component, which is inseparable from language and emotional events, makes comparisons between verbal and musical stimuli even more challenging.

It should be noted though, that since there are no direct acoustic analogs between speech and music, it is problematic to make conclusive comparisons. Chow and Brown (2018) attempted to solve this problem by using the musical notation in their analysis of speech melody, converting the fundamental-frequency trajectories of the recorded words and utterances from hertz into semitones and transcribing them into musical scores of a relative pitch. They reported that, compared to music, speech is atonal and characterized by a weak type of chromaticism. Nevertheless, even within the compressed pitch range of standard speech production, language-specific melodic patterns can be found. Font-Rotchés and Torregrosa-Azor (2015) compared the intonation of yes-no questions in Spanish and German using the Melodic Analysis of Speech, a method of analysis that is based on the principle of phonic hierarchy and the measures of the F0 of the tonal segments, the vowels. Although the authors observed a close correspondence between some Spanish and German melodic patterns, they also reported substantial differences in the tonal range: to the point that a statement produced in one language might sound like a question in another language. Mennen et al. (2012) confirmed the existence of significant cross-linguistic differences in the F0 range. Yet, the aesthetic value of the melodic patterns and how they might contribute (or not) to the overall music-like effect in some languages remain unknown.

Although our knowledge of the overlap between music and language is notable (Sammler, 2020), their aesthetic acoustic properties are still poorly understood. In compariosn to speech, the aesthetic investigations of music are more promising. Several studies describe rhythm and its perceptual attribute, the beat, and tonal harmony as the most prominent contributors to auditory pleasure (e.g., Brattico et al., 2013). Rhythm perception arise from a grouping mechanism that creates patterns of prominence recurring in time (Arvaniti, 2009; Falk et al., 2014). While it is quite likely that the aesthetic value of rhythm is shaped by biological constraints – e.g., the limitations of working memory (Ravignani et al., 2016) or even by the inner pacemaker: heart beat (Chahal et al., 2017) – the formal structure of a musical piece is also important: tonality, harmony, and meter create specific temporal regularities, which, when they meet the expectations, build an emotional aesthetic response. In this sense, anticipation or expectancy, which also depends on musical knowledge, is the essential mechanism for a pleasurable experience (Steinbeis et al., 2006; Vuust and Kringelbach, 2010). Some listeners experience "chills," an intense physical sensation such as goosebumps or trembling, in response to a favorite tune or melody (Reuter and Oehler, 2011; Starcke et al., 2019), and while musical events that elicit such reactions vary from person to person, there are a few patterns that might be connected to chills: the onset of vocals, the beginning of a structurally new part, and contrasting voices are strong acoustic triggers (Grewe et al., 2007; Guhn et al., 2007).

Compared to music, little is known about the phonetic chill or the auditory pleasure that arises from listening to languages. First of all, it is hard to evaluate the aesthetic value of rhythm in language because speech rhythmic patterns are not beat-based (or metrical), unlike the rhythmic patterns of music. Ozernov-Palchik and Patel (2018) observe that "the temporal patterning of linguistic units is highly structured, but is not based on an underlying grid of equal time intervals" (p. 166). Beat-based processing and speech processing may be cognitively related at a more abstract level concerned with prediction in structured sequences: after all, listeners routinely predict upcoming linguistic material, although the prediction is not based on temporal periodicity but rather on phonological, semantic, and syntactic structures. Yet, there is something in the speech that contributes to the perceptual experience of a beat. Infant et al. (2013) suggest that the beat distribution patterns in speech are cued by stressed syllables and p-centers, a psychological phenomenon that coincides with syllabic nuclei and vowel onsets (Lin and Rathcke, 2020). Arvaniti (2012) also connects rhythm with syllabic prominence.

On closer examination, we can see that a more regular syllabic structure (e.g., CVCV, where "C" stands for a consonant and "V" for a vowel) might produce a similar pleasurable anticipatory effect that is fundamental in music (Steinbeis et al., 2006; Vuust and Kringelbach, 2010). For that reason, it might have a higher aesthetic value than a more complex unpredictable syllabic structure (e.g., CVCCVCC). There is, indeed, a language-universal preference for the CV structure that, at times, overrides a preference for native-specific structures (Greenberg, 1965; Blevins, 1995). Either way, it is reasonable to assume that languages that predominantly use the CV structure also sound more pleasant. Unlike Italian, which often uses the CV structure, German, its northern neighbor, displays a structural variety with complex syllabic combinations, including heavy consonant clusters (Rabanus, 2003).

Another reason for the difficulty in quantifying the timing characteristics of the linguistic rhythm is that it is not a unidimensional phenomenon and influenced by several factors, including F0 movements (Tilsen and Arvaniti, 2013). So, unlike in the case of music, the linguistic rhythm is the product of various phonological phenomena, each interacting with others. It does not mean that linguistic rhythm is unconnected to musical rhythm: several studies have shown the influence of speech rhythms on non-linguistic rhythmic grouping preferences (e.g., when the composer's native language influences the composition) (Patel et al., 2006; Jekiel, 2014).

Remarkably few empirical studies have investigated the music-like effects in language. Even though much is known about the intonation and the rhythmic architecture of speech, it is still unclear what aesthetic value these elements have and how they contribute (or not) to a pleasurable auditory experience derived from listening to the spoken word. While prior studies compared the sounds of language and music (Patel et al., 2006; Chow and Brown, 2018), none looked at the aesthetic value of the acoustic parameters responsible for music-like effects in language. In this regard, the goal of the present study is *multi*-disciplinary (as opposed to just inter-disciplinary) since we aim at quantifying a phenomenon at the juncture of three domains: linguistics/language sciences, musicology, and aesthetics. Although this approach is advantageous in many senses as it allows for new questions to be formed in a new way, it also has a number of limitations. For example, we are not aware of the previous research that would employ a similar design and, therefore, provide important guidance for our methodology and data analysis. Thus, we had to rely solely on our own knowledge of the question and scientific intuition, which makes the present study highly exploratory in nature. Another challenge for any inter- and multidisciplinary research is developing a common language across the disciplines to describe a complex phenomenon in the most comprehensive way. Despite these and other limitations, we believe it is important to begin a conversation on this topic and open the line of research that future scholars can confirm, refute, or finesse at their leisure.

Our immediate goal is to fill an existing gap in the research by identifying those acoustic features that drive the auditory perception of language beauty to create a musical effect in the sound of the language. What is so musical in a language that makes people say “It’s music in my ears”? Why are we so mesmerized by the sounds of the Romance languages, especially Italian, French and Spanish (the Latin lover effect), and less enthusiastic about the Germanic and Slavic families? Is it true that Italian is a language built to be sung? Can we explain at least part of the charm or ‘sound pleasure’ (“Ohrenschmaus”) by characteristics derived from acoustic-phonetic measures? Here, we are primarily interested in the auditory allure of the Romance languages, which have been consistently marked across various surveys as the most melodious languages in the world (Burchette, 2014). In our previous study (Reiterer et al., 2020), the Romance languages were described as “pleasant to listen to,” “melodic,” and the “languages of music and songs” (e.g., one participant observed: "French sounds to me very soft and “round.” It is often the language of love and in many songs, some phrases are in French. Lady Gaga: Bad Romance/Christina Aguilera: Lady Marmalade/ABBA: Voulez-vous"). However, here we look at the acoustic parameters that are responsible for pleasurable aesthetic effects in music – rhythm and melody – and explore the higher-order linguistic phenomena we seem to perceive, the phonetic chill. This is the first in a series of studies that, like any step into the unknown, is highly exploratory. For this reason, it is peppered with caveats and limitations that we discuss at the end of this article.

To summarize, our research questions are:

1. What acoustic-phonetic features are responsible for the music-like or phon-aesthetically pleasing effects in languages?
2. How are these features distributed across language families? Do the Romance languages lead the list in this sense?

Apart from shedding light, we hope, on the nature of the hedonic pleasure derived from the architecture of language sounds, this study has pedagogical implications for foreign language learning. Appreciating the acoustic makeup of a target language might activate additional affective learning pathways in the learner’s brain and support auditory memory. For example, neuropsychological studies show that emotional events are remembered better than neutral events, thanks to the amygdalae – two almond-shaped nuclei in the brain that enhance the function of the medial temporal lobe memory system (Dolcos et al., 2004; Koelsch et al., 2006). Approaching language as a song also helps to alleviate speaking anxiety and produces an overall relaxing effect that is essential for successful learning (Fonseca-Mora et al., 2011]BR110). Teachers can use the acoustic properties of the language-to-be-learned and complement the classroom work with synesthetic activities that emphasize specific phonetic features (Wrembel, 2010).

# MATERIALS AND METHODS

## Participants

The participants ($N$ = 45) were students or young academics with the following native languages (L1s = first languages): Slovenian (22), German (11), English (5), Serbo-Croatian (3), Finish (1), Italian (1), Kazak (1), and Portuguese (1). Participants’ ages ranged between 22 and 49. On average, the participants reported being able to speak 2.9 foreign languages ($SD$ = 1.6, min = 0, max = 8). The following languages were mentioned as participants’ foreign languages or L2s (from most to least common): English (71%), German, Italian, French, Croatian, and Russian. Other languages mentioned as L3/LX (in the alphabetic order): Arabic, Chinese, Esperanto, Hungarian, Finnish, Japanese, Ladin, Latin, Portuguese, Romanian, Russian, Slovene, Spanish, Swedish, Turkish, and Welsh.

## Materials

The recordings of the 16 European languages were presented as auditory stimuli recorded by native speakers, half of which were females (**Supplementary Table 2S**). Each stimulus featured a reading of a translation of Aesop’s fable *The North Wind and the Sun*. The 16 languages were distributed over four language families: (1) Romance (French, Italian, Spanish, and Catalan). (2) Germanic (German, English, Icelandic, and Danish). (3) Slavic (Russian, Polish, Serbo-Croatian, and Ukrainian). (4) Other smaller languages or isolates (Hungarian, Greek, Basque, and Welsh) – see also **Figure 1** with a map of the languages.

## Phonetic Measures

***Speech rate*** equated to the number of syllables per second (see Coupé et al. (2019) for a discussion on how to calculate speech rate). A researcher with linguistic-phonetic training auditorily calculated speech rate for each language with the aid of a digital audio workstation (Adobe Audition [Computer software], 2018, Version 11.1.0) and visual control of the scripts. Silences longer than 50 milliseconds were excluded and only stretches with a continuous speech signal of at least one second were considered. Ten such speech streams were investigated for their syllable counts and the mean over those ten individual 1-s streams formed the final syllable rate per language. For quality control, speech rate for a second set of independent voice/language recordings of *The North Wind and the Sun* in the 16 European languages of the experiment was enumerated in the same way as described above and compared with the experimental set (set 1). The second set of recordings (henceforth called *the second set)*, was produced at a laboratory at the University of Vienna and consisted of all-female samples. There was a strong positive correlation between the experimental and the second set: $r$ = 0.8, $p$ = 0.000). For further quality control the speech rate data was compared to the values reported by an earlier publication (Coupé et al., 2019) resulting in a strong pos. correlation again ($r$ = 0.8, $p$ = 0.006) between our first experimental set and the sample published by Coupe

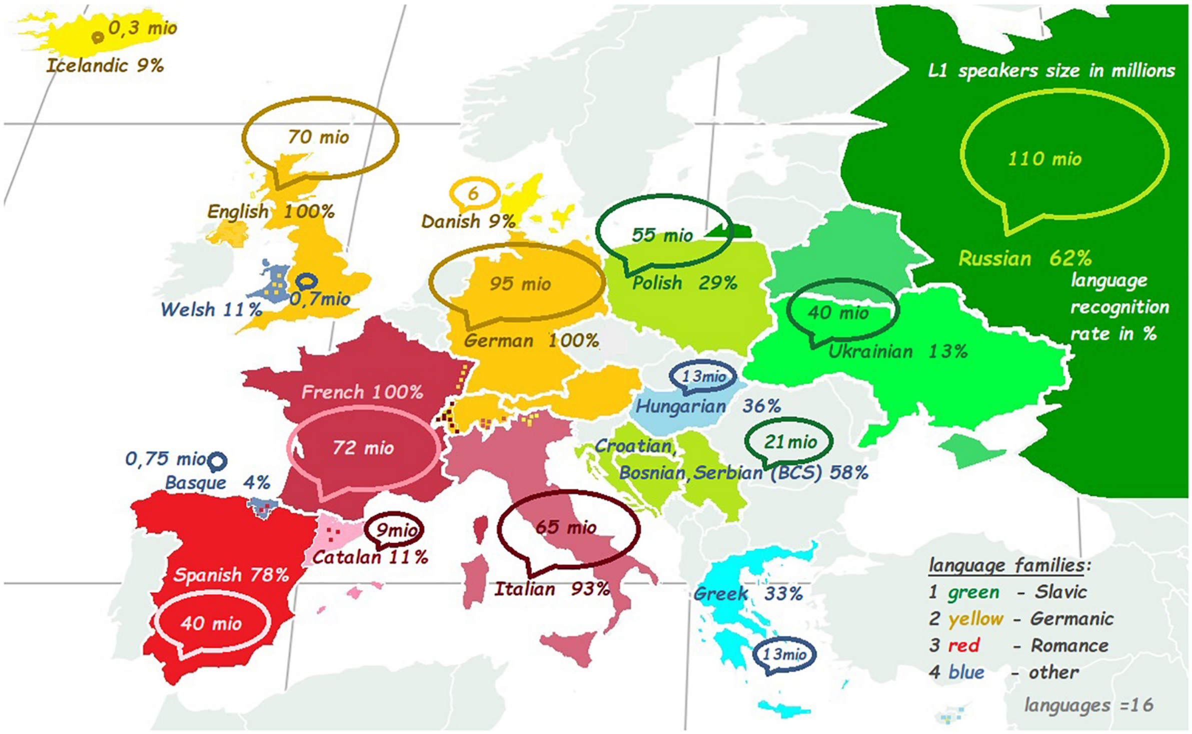



**FIGURE 1 |** The number of speakers in millions and the recognition rate in percentage. The green color refers to Slavic languages, yellow – Germanic languages, red – Romance languages, and blue indicates other language families or languages-isolates.

et al., and another strong pos. correlation of $r = 0.76$ ($p = 0.017$) between our second set of recorded languages and the sample reported by Coupè.et al. (Note: the speech rate was not correlated with the speaker's gender, $r = 0.03$; $p = 0.8$).

***The mean F0*** (fundamental frequency in hertz) of the voice recordings representing the 16 languages was extracted in Praat (Boersma and Weenink, 2017), a software package for speech analysis. Half of the recordings were spoken by male voices and half by female voices ($N = 8$). A continuous variable F0 was used to introduce a more neutral acoustic measure for the speaker's gender. The F0 vector of the speech samples was also used to measure pitch modulations, melody, prosody.

## F0-Trajectory Pitch Variation (Prosody) Measurement

The F0 pitch trajectories (pitch lists or vectors) were extracted by using Praat. Both sets of 16 voice recordings (first/main set of 16 languages, 50% female voices, and a second set of own voice recordings of the same 16 languages – all-female voices) were manually, visually and auditorily, checked and screened for pitch artifacts. An individual pitch range (see **Table 1**) was determined for every voice according to the artifact removal strategies in voice recordings as in Mayer (2019). Individual ranges were determined as cut-off frequencies and checked again for remaining artifacts that occurred due to hissing or creaky voices (high or low frequencies).

**TABLE 1 |** Frequency ranges across languages in the first and the second set.

| Languages | The first (experiment) set | | The second (control) set |
|---|---|---|---|
| | Female speakers | Male speakers | Female speakers |
| Basque | 125–300 Hz | | 175–350 Hz |
| Catalan | | 75–200 Hz | 165–350 Hz |
| Croatian | 120–260 Hz | | 145–350 Hz |
| Danish | 115–280 Hz | | 145–350 Hz |
| English | 115–300 Hz | | 170–350 Hz |
| French | 130–250 Hz | | 140–350 Hz |
| German | | 80–230 Hz | 135–350 Hz |
| Greek | | 80–200 Hz | 155–400 Hz |
| Hungarian | | 75–300 Hz | 135–270 Hz |
| Icelandic | | 75–200 Hz | 165–350 Hz |
| Italian | 120–370 Hz | | 120–290 Hz |
| Polish | 90–400 Hz | | 155–330 Hz |
| Russian | | 75–200 Hz | 140–350 Hz |
| Spanish | 120–300 Hz | | 125–350 Hz |
| Ukrainian | | 75–250 Hz | 155–400 Hz |
| Welsh | 75–250 Hz | | 110–410 Hz |

The extracted and artifact-controlled pitch trajectories were then converted into the international music cents scale with 55 Hz as reference frequency according to the formula [1,200 * log2 (pitch vector in Hz/55)] to neutralize pitch effects between male and female speakers by transferring them into a normed reference frame (cents). From cents we converted further into semitones [1,200 cents or 12 semitones = 1 octave; 100

cents = 1 semitone]. Finally, we were interested in the pitch variation (melody of speech, voice modulations, and prosody) and calculated the variance or SD (standard deviation) of the music cents and semitones as ***Pitch variation***.

## Other Measures

***Lexical distance*** between languages (hereinafter the ***Serva-Petroni distance***) was estimated to control the distance between L1 of the perceivers (participants) and the 16 experimental languages. The ***Serva-Petroni distance*** based on a new automated method that uses the normalized Levenshtein distance developed by Serva and Petroni (2008), ranged from the minimal distance of 0 to the maximal distance of 1.

***Percentage learned as Lx*** refers to how much or how frequently a certain language of the experimental set (of those 16) was known or had previously been learned as a second/foreign language (Lx) by the participants. E.g. English has a score of 95% percent learned as Lx (foreign language) by our participants, whereas Greek, Basque, Catalan, Polish, Danish, Icelandic, and Ukrainian all got 0% because they were not learned by a single person of the sample as Lx.

Participants also self-reported on their ***musicality, singing ability***, and the number of and expertise in any musical instruments they play. Previous studies have shown that self-assessment serves as a reliable measure and is comparable to expert assessments (Christiner and Reiterer, 2013, 2015; Christiner et al., 2018).

## Procedure

We recruited primarily central European participants ($N$ = 45). Following an online link, the participants listened to 16 European languages and evaluated them using opposite descriptors (e.g., beautiful vs ugly – see **Table 2**). The participants were instructed to use headphones. For the evaluations, an intuitive scale between 0 and 100 points with a slider was provided for each of the 22 adjective descriptors. We asked participants to estimate how familiar they were with the languages (self-perceived familiarity) and how much they liked the speakers' voices. Furthermore, to control self-perceived familiarity, the participants were asked to name or guess the names of the languages they heard (if they were unsure, they could write whatever they associated with the language or name a language family). The later measure constituted ***Recognition rate***. A final comment box collected optional comments about the task and appeared after each language evaluation.

All participants were informed about the main purpose of the research, its procedures, risks, and potential benefits. They volunteered to participate, signed a consent form, and received monetary compensation for their participation. The experiment was administered online using Gorilla Experiment Builder (Anwyl-Irvine et al., 2020), which generated a unique URL for each participant. All instructions and other texts featured in the experiment were presented in English. First, participants were asked to fill in a personal background questionnaire about their demographics and a language background questionnaire. Next, they were presented with a rating task featuring 17 recordings of languages in a randomized order (16 European languages of the study and 1 language doublet for control purposes). Participants were asked to rate each recording according to 22 aesthetic descriptors, available as sliding scales, and to provide further information about their impressions. They were required to use speakers or headphones and to complete the tasks using Mozilla Firefox as a default browser.

The instructions emphasized the importance of focusing on the sounds of the languages and not on the meaning of the presented text. Participants were told that there were no correct answers and were encouraged to use both extremes of the scale to their liking. They were allowed to listen to each recording as often as they wanted to. By the end of each trial, each language was evaluated in terms of its familiarity to a participant: the same sliding scale from 0 ("I don't know this language") to 100 ("I recognized this language") was employed for this purpose. Participants could also guess which language they thought they have heard. Lastly, the speaker's voice was evaluated following the same principle: 0 for "Very unpleasant voice" and 100 for "Very pleasant voice." The optional subsection "Other impressions" allowed the participants to comment further on their experiences.

**TABLE 2 |** Aesthetic descriptors.

| Scale | Negative descriptor | Positive descriptor | Source |
|---|---|---|---|
| *Beauty* | Ugly | Beautiful | Giles and Niedzielski (1998) |
| *Coolness* | Uncool | Cool | Reiterer |
| *Culture* | Uneducated | Cultured | Giles and Niedzielski (1998) |
| *Elegance* | Inelegant | Elegant | Giles and Niedzielski (1998) |
| *Eroticism* | Unerotic | Erotic | Reiterer |
| *Fashion* | Unfashionable | Fashionable | Giles and Niedzielski (1998) |
| *Fun* | Boring | Fun | Giles and Niedzielski (1998) |
| *Generosity* | Stingy | Generous | Giles and Niedzielski (1998) |
| *Importance* | Marginal | Influential | Giles and Niedzielski (1998) |
| *Intelligence* | Stupid | Intelligent | Giles and Niedzielski (1998) |
| *Melody* | Tuneless | Melodic | Reiterer |
| *Memorability* | Unmemorable | Memorable | Reiterer |
| *Orderliness* | Chaotic | Orderly | Giles and Niedzielski (1998) |
| *Pleasantness* | Unpleasant | Pleasant | Giles and Niedzielski (1998) |
| *Romanticism* | Unromantic | Romantic | Reiterer |
| *Seductiveness* | Unseductive | Seductive | Reiterer |
| *Sexiness* | Unsexy | Sexy | Reiterer |
| *Softness* | Hard | Soft | Reiterer |
| *Status* | Low status | High status | Giles and Niedzielski (1998) |
| *Sweetness* | Harsh | Sweet | Reiterer |
| *Wealth* | Poor | Wealthy | Giles and Niedzielski (1998) |
| *Welcomingness* | Repellant | Welcoming | Giles and Niedzielski (1998) |

# RESULTS

## Linguistic Variables and Aesthetic Ratings

As a part of the study, we measured linguistic background or control variables and the aesthetic ratings (see **Table 3** for descriptive statistics).

**TABLE 3** | Descriptive statistics with the linguistic background variables and aesthetic ratings (0–100) with voice.

| | | L1 speakers in mio | Ls learned as Lx in % | Recognition rate in % | Lexical distance | Overall mean score | BEAUTY | STATUS | EROS | SOFTNESS | ORDERLINESS | Melody | Voice |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N | Valid | 16 | 16 | 16 | 14 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| | missing | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mean | | 38,1725 | 15,8638 | 46,6263 | ,770087 | 53,7174 | 56,6732 | 56,2097 | 45,0611 | 50,3681 | 54,4125 | 59,7556 | 59,8222 |
| Median | | 30,5000 | 3,3100 | 34,4000 | ,790846 | 50,9086 | 53,7825 | 54,7986 | 41,2806 | 47,5611 | 52,6222 | 57,2333 | 59,6000 |
| Standard Deviation SD | | 35,81612 | 25,06462 | 37,37447 | ,0854314 | 8,72251 | 8,29062 | 9,25013 | 12,12394 | 12,66038 | 9,14027 | 10,82261 | 9,42046 |
| Range | | 109,69 | 95,00 | 95,60 | ,2602 | 29,90 | 25,94 | 30,35 | 39,86 | 49,69 | 32,33 | 35,76 | 34,24 |
| Minimum | | ,31 | ,00 | 4,40 | ,6236 | 41,93 | 44,31 | 45,62 | 31,68 | 25,44 | 41,82 | 43,56 | 47,00 |
| Maximum | | 110,00 | 95,00 | 100,00 | ,8838 | 71,83 | 70,24 | 75,97 | 71,54 | 75,13 | 74,16 | 79,31 | 81,24 |

*Here the 22 aesthetic descriptors are collapsed into five underlying factors (see the description of the principle component analysis applied below).*

All correlation coefficients of zero-order correlations are reported in Spearman's Rho; we used the classical symbol "r" for these coefficients. Two-tailed testing was the default, and the results are reported at the significance level of $p < 0.05$, if not stated otherwise. For an overview correlation matrix of all variables see **Figure 1**).

The factors ***Beauty, Eros, Softness, Status,*** **and** ***Orderliness*** were based on an exploratory factor analysis (EFA: described in Reiterer et al., 2020; see **Table 4** below and **Supplementary Figures 1S + 2S**) which resulted in the reduction of 22 ratings to five factors based on $N = 45$. One of the original ratings was ***Melody*** (how melodious the language is). This adjective was subsumed under the factor ***Beauty*** by the EFA. However, since this adjective is of particular interest in studying the musical aspects of the ratings, we present the results of the melody ratings as well. Nevertheless, it should be noted that ***Melody*** is comprised as one of the eight scales under the factor ***Beauty***. ***Melody*** results that are reported below will always follow after the main five factors. The ratings of the five main factors are depicted in **Figures 2**, **3** where the ***Overall*** rating scores are color-coded on the geography map. For further details see **Supplementary Tables 1S, 3S**.

The languages with the highest ***Beauty*** ratings were French (70.24), English (70), and Italian (67.5). The lowest ***Beauty*** ratings were Welsh (44.3), German (47.4), and Polish (49.3). The languages with the highest ***Eros*** ratings were French (71.5), Spanish (59.7), and Italian (59.2) – a not surprising "Latin Lover effect." The least erotic dubbed languages were Welsh (31.6), German (32.4), and Danish (32.7).

**TABLE 4** | The 22 aesthetic ratings collapsed into five factors (Beauty-yellow, Status-blue, Eros-red, Softness-orange, and Orderliness-green).

| Aesthetic rating | Beauty | Eros | Softness | Status | Orderliness |
|---|---|---|---|---|---|
| Beauty | X | | | | |
| Melody | X | | | | |
| Fun | X | | | | |
| Memorability | x | | | | |
| Welcoming | x | | | | |
| Generosity | x | | | | |
| Pleasantness | x | | | | |
| Culture | | | | x | |
| Status | | | | x | |
| Wealth | | | | x | |
| Importance | | | | x | |
| Intelligence | | | | x | |
| Fashion | | | | x | |
| Elegance | | | | x | |
| Coolness | | | | x | |
| Sexiness | | x | | | |
| Eroticism | | x | | | |
| Seductiveness | | x | | | |
| Romantic | | x | | | |
| Softness | | | x | | |
| Sweetness | | | x | | |
| Orderliness | | | | | x |

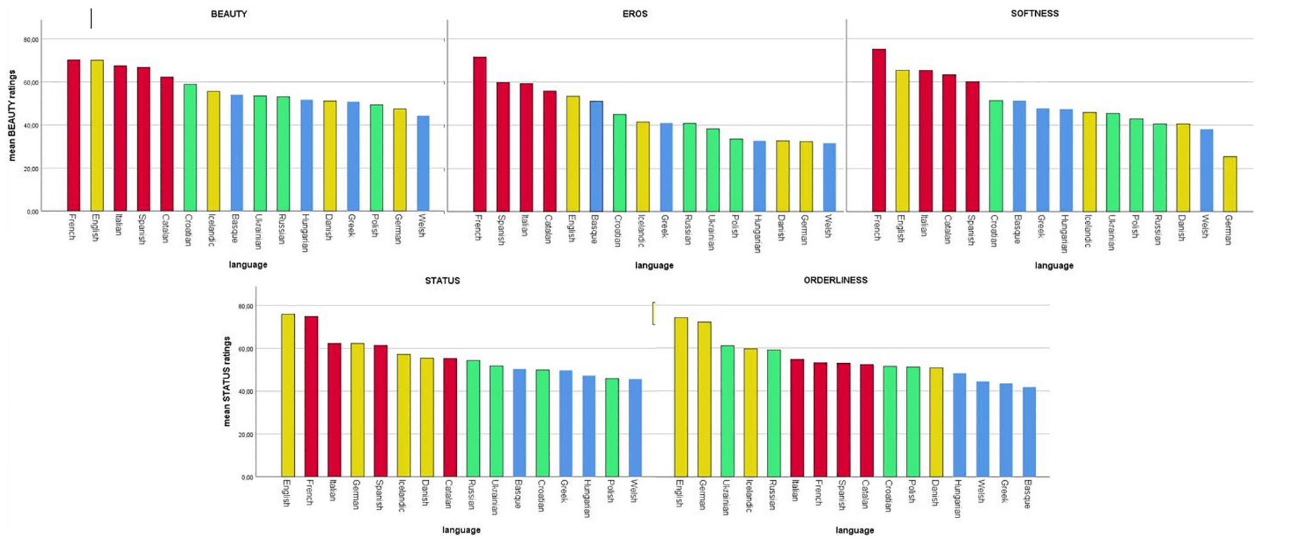



**FIGURE 2 |** Aesthetic ratings, grouped by factors (Beauty, Eros, Softness, Status, and Orderliness), for all 16 languages, colored by language families (red = Romance, yellow = Germanic, green = Slavic, and blue = other).

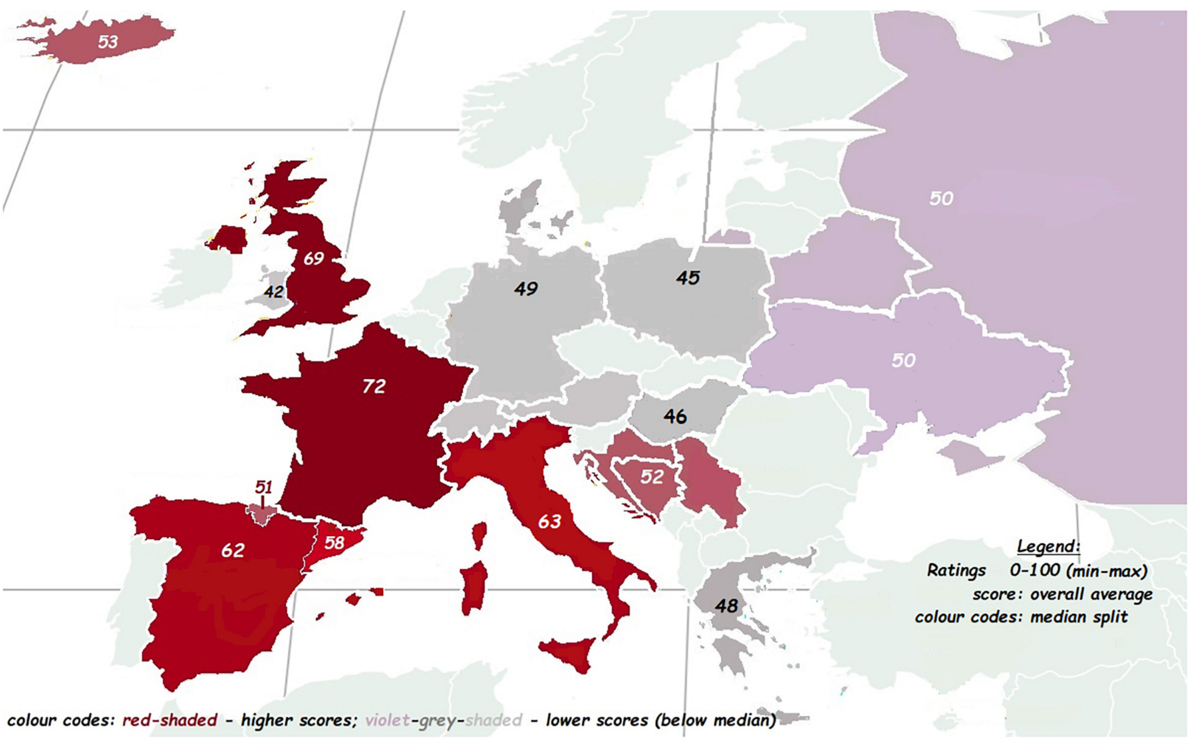



**FIGURE 3 |** Mean (overall) aesthetic rating per language displayed on a geographical map. Color codes reflect the median split of the overall aesthetic ratings per language; red colors reflect higher ratings and violet or gray shaded reflect the lower ratings (0–100 min-max).

On the ***Softness/Sweetness*** scale, the highest languages were French (75), English (65.3) and Italian (65.3). The lowest scores were German (25.4), Welsh (38), and Danish (40.5).

On the ***Culture-Status*** scale, the highest languages were English (75.9), French (74.8), and Italian (62.3). The lowest scores were Welsh (45.6), Polish (45.8), and Hungarian (47). Note that, these ratings (unlike the first three) also correlate moderately and positively with ***L1 community size***.

The last factor found by EFA was ***Orderliness***: the highest languages were English (74), German (72), and Ukrainian (61).

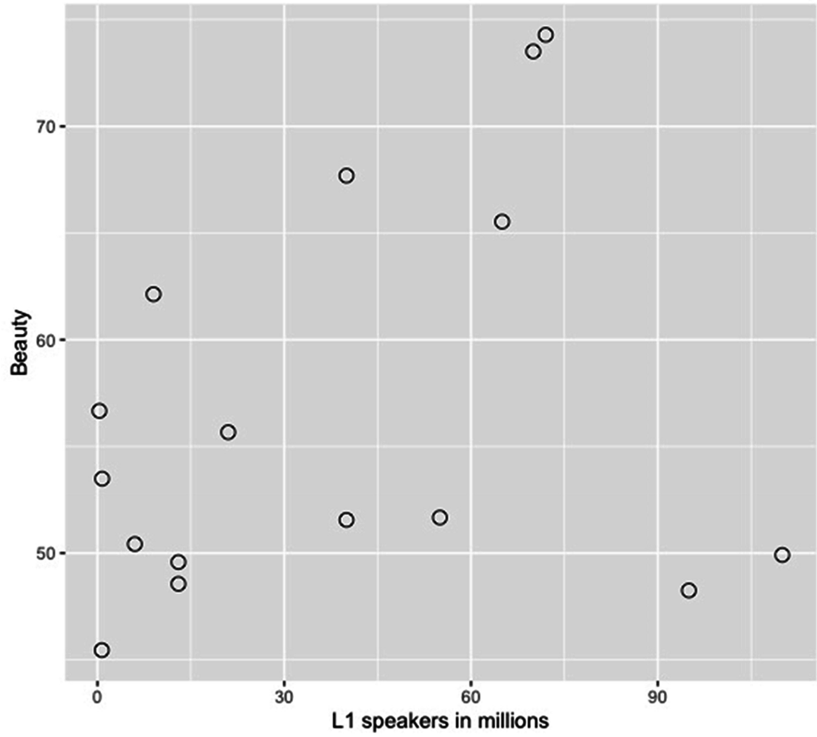



**FIGURE 4 |** Scatterplot of the correlation between Beauty ratings and the number of speakers in Millions (community size).

The lowest languages were Basque (41.8), Greek (43.5), and Welsh (44.4). Recall that we found a highly significant correlation between ***Orderliness***: and ***L1 community size*** with the more orderly languages having larger native speaker communities (German, English, and Ukrainian).

***Melody***, the single rating subsumed already under the factor ***Beauty***, yielded similar results to ***Eros***. The languages perceived as most melodious were French (79.3), Italian (76), and Spanish (72.2) – followed by English and Catalan. In the lowest range were German (43.5), Welsh (44.8), and Greek (51.6) – followed by Polish and Russian. Again, this picture resembles closely the ratings of the factors presented above. Perceived ***Melody*** ratings did not reflect the Hertz-based ***Melody*** measurements as measured by ***F0 pitch variation*** (see also ***Melody cents variance*** below).

***L1 community size*** (or a number of L1 speakers in millions) operationalized social-linguistic power relations. The smallest language community was Icelandic with only 31,000 speakers (0.31 mio) and the biggest community was Russian with 110 mio speakers. This variable was introduced to see the effects of cultural-political power on the aesthetic ratings. The results showed that there were no significant relationships between ***L1 community size*** and the ratings ***Eros***, ***Beauty*** and ***Sweetness*** [Spearmans's $r = 0.2$, $p = 0.5$ (Eros); $r = 0.2$, $p = 0.4$ (Beauty), $0 = 0.1$, $p = 0.6$ (Softness)]. However, L1 community size did affect the ratings ***Orderliness*** and marginally – ***Status*** [($r = 0.6$, $p = 0.01$ (Orderliness), $r = 0.5$, $p = 0.07$ (Status)]. It is interesting to see that the greater ***L1 community size***, the higher the perception of a language status or "orderliness" – a clear sign of the impact of socio-political power on a perceived language status. However, this power does not transfer to the concepts that are more "emotional" or "aesthetic" in nature, such as language beauty, eroticity, and softness/sweetness. Further analysis revealed a strong and positive correlation with ***Recognition rate*** ($r = 0.9$, $p = 0.000^{**}$) and with ***Percentage learned as Lx*** ($r = 0.7$, $p = 0.003^{**}$). These results came as no surprise since they demonstrate the power effect, namely that the languages of the large communities (dominant languages) are familiar and recognized. These are also the languages that are typically acquired as foreign languages at mainstream educational institutions.

Further, in **Figure 1**, ***Recognition rate*** is expressed as a percentage and ***L1 community size*** as a circle with a corresponding size. **Figure 4** shows the (low) correlation between ***Beauty*** ratings and ***L1 community size***.

The next variable was ***Percentage learned as Lx*** or ***Percentage Lx***, the percentage of the participants that learned a given language as a foreign (L2, L3, and Lx) language. The share was rather low in total. Only 16% of all languages used in the experiment were learned as Lx by the participants. At the same time, the average number of foreign languages learned was three indicating a rather multilingual sample of participants. The experimental set consisted of many rarer languages that represented smaller language families, and that could be the reason why the overall share of Lx was rather small. Seven out of the 16 languages were not learned as a foreign language by

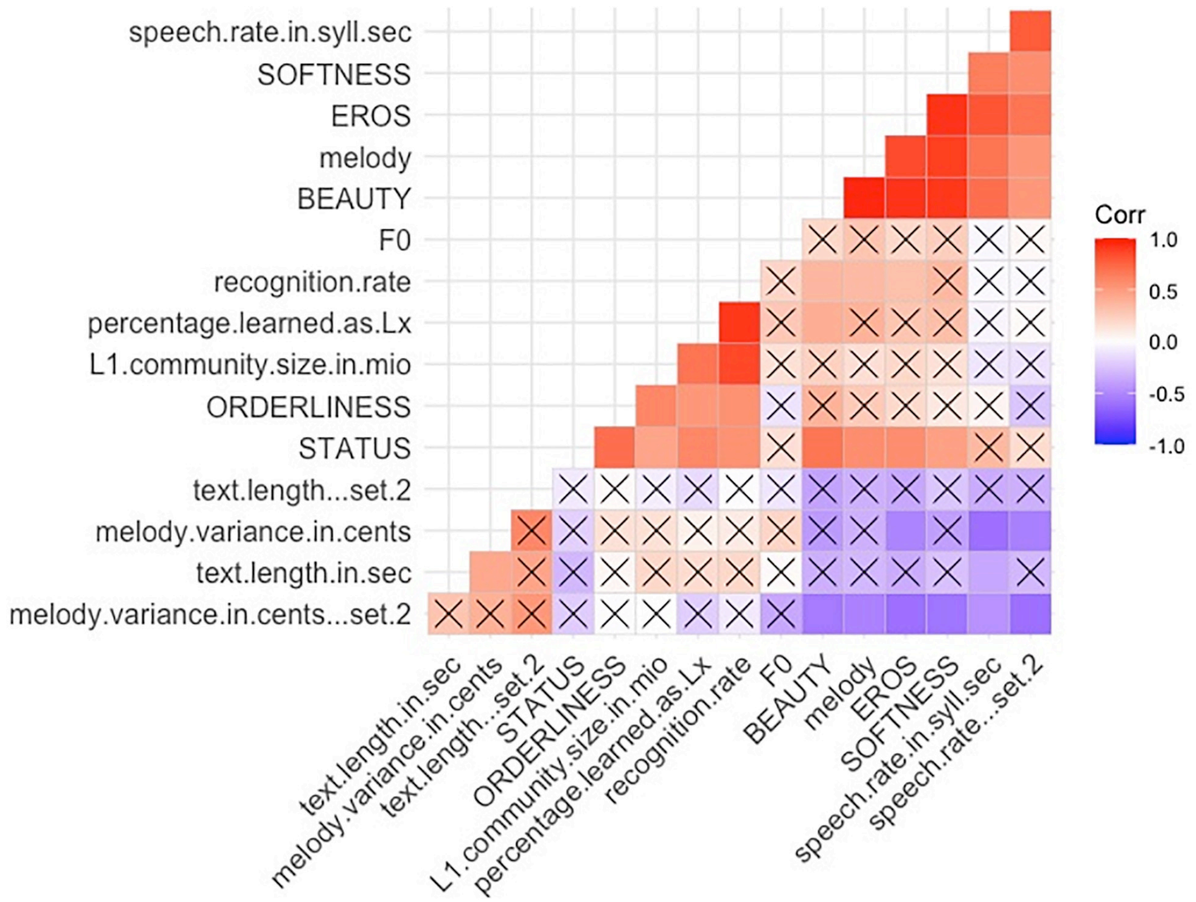



**FIGURE 5 |** Zero Order (Spearman) Correlations Matrix Overview of all variables, $p < 0.1$ ($p$ level was set to include trend-level correlations. For the chart displaying all correlation coefficients at $p < 0.1^{trend}$, $p < 0.05^{*}$, and $p < 0.01^{**}$ level see **Supplementary Material Table S2**).

anyone: Basque, Catalan, Danish, Greek, Icelandic, Polish, and Ukrainian. On the other hand, one language reached a 95% share of Lx learning and that was English, followed by German with 42% and French with 30%. Regression analysis showed that foreign language knowledge influenced ***Recognition rate***, and with that – the aesthetic ratings: ***Percentage Lx*** explained 40% variance in the aesthetic ratings (see a detailed discussion on familiarity and foreign language knowledge in Reiterer et al., 2020). Thus, in terms of language socio-political power, the **power of second language education** (foreign languages that are traditionally acquired in schools) is more influential than the size of the L1 community. We found that ***Percentage Lx*** correlated highly with ***Recognition rate*** ($r = 0.9$, $p = 0.000^{**}$), i.e., the more language is recognized, the more widely it has been learned as Lx. ***Percentage Lx*** also correlated positively with ***Status*** ($r = 0.6$, $p = 0.01^{*}$) and ***Orderliness*** ($r = 0.5$, $p = 0.04^{*}$): higher ***Status*** and ***Orderliness*** were ascribed to well-known languages. Concerning ***Beauty***, ***Eros***, and ***Softness***, there were non-significant relationships between these ratings and ***Percentage Lx***, as well as other acoustic variables.

***Recognition rate*** showed the percentage of languages that were identified correctly and reflected the familiarity with the languages. The most recognized languages were English, French, and German (all 100%). The least recognized languages were Basque (4%), Danish (9%), and Icelandic (9%). ***Recognition rate***, other than correlating to ***L1 community size*** and ***Percentage Lx***, correlated positively with ***Status*** and ***Orderliness*** ($r = 0.55$, $p = 0.03^{*}$; see also **Figures 1**, **5**).

In the case of Basque, two-thirds of the participants when asked to identify the language commented alongside these lines: "some kind of Romance language either Portuguese, Romanian, or another one." Thirteen participants could not provide any answer suggesting that it could be "an Indo-European language." To check whether associating Basque with the Romance language had effects on aesthetic ratings, the independent samples t-test was performed (for the ratings of ***Beauty***, ***Eros***, ***Softness***, ***Status***, and **Orderliness**). The results showed no differences ($p = 0.6$ to.9) between the group that associated Basque with a Romance language and the group that did not.

The same picture emerged for the other two barely recognized languages. In the case of Danish 35 participants believed it to be either Dutch (the majority), German or "some kind of Germanic or Northern language." Six participants could not identify it at all, with one participant suggesting it was Swahili. There was again no distinction in the group mean ratings as measured by the independent samples median test. In the case of Icelandic, a North Germanic language, the results were again similar with no significant group differences (t-tests) between those ($N = 18$) who thought it was "some kind of a Northern or Germanic or Scandinavian language" and those who ($N = 23$) though it was something else (e.g., a Finno-Ugric language with many participants suggesting Finnish and a few – a Romance or Slavic language). With all three unrecognized languages, we found no evidence of language family influence on the aesthetic ratings. This is particularly important for the case of Basque, where there is a danger to acoustically (and wrongly) classify it as a Romance language due to its century-long phonetic co-habitat with other Romance languages on the Iberian Peninsula.

One of the most diverse guesses was evoked by Welsh (the Celtic language family, Welsh was identified by 11% of the participants). Fourteen different languages and families were mentioned (other than "I don't know") in relation to Welsh, most pertaining to the Northern European regions, from the Germanic to Finno-Ugric language families, including guesses such as Lithuanian, Estonian, or Scandinavian. Some participants went as far as Arabic or Hebrew, with one particularly interesting answer: "at some point, it sounded like English with an accent. . .but it might be Bulgarian or Hungarian. . .not sure." Many comments were linguistically interesting, such as "an English creole?" or "something like Gaelic – an Englishman who speaks Celtic." For further details of the qualitative results (note, the guesses were obligatory, but the comments were optional) please see **Supplementary Material**.

The next variable was ***Lexical distance*** that referred to the distance between participants' L1 and the languages of the study. This variable was used to control for typological influences of the mother tongue. ***Lexical distance*** was based on Levenshtein distances (Serva and Petroni, 2008; Petroni and Serva, 2010) and ranged from 0 (no distance between participant's L1 and a given language) to 1 (maximal distance). The most distant language (to all participants) in the experimental set was Welsh with a coefficient of 0.88. The coefficients for Basque and Hungarian were not computed since no scores could be obtained for these language isolates (both languages do not belong to the Indo-European family). The closest languages for all participants on average were Croatian with a coefficient of 0.62 and German with 0.63. ***Lexical distance*** yielded no relationships (non-significant correlation coefficients, $N = 14$) to the aesthetic ratings ($r = 0.04$, $p = 0.8$ for ***Beauty***; $r = 0.2$, $p = 0.5$ for ***Eros***; $r = 0.2$, $p = 0.5$ for ***Softness***; $r = -0.1$, $p = 0.7$ for ***Status***, and $r = -0.5$, $p = 0.07$ for ***Orderliness***), reflecting no influence of L1 on the ratings. Only in the case of ***Orderliness***, there was a negative trend (yet non-significant): the more distant the language was, the less orderly it was perceived.

***Fundamental frequency (F0)*** was introduced to quantify acoustical differences between the languages voiced by male and female speakers. The male speakers ($N = 8$) had a F0 mean of 123 Hz, while female speakers had significantly higher values – 184 Hz on average. Such difference confirms previous research findings that describe gender-specific acoustic profiles. The overall mean of F0 in the sample ($N = 16$) was 153 Hz. The lowest F0 was for the Russian male voice and equated to 108 Hz, and the highest F0 was for the Italian and Polish female voices and both equated to 208 Hz. The mean F0 in the comparison recordings set (all female speakers) was 208 Hz (SD $\pm$ 24.7, range 165–243 Hz), with the highest pitched voice being the Ukrainian speaker and the lowest pitched voice – the Italian speaker. While ***F0*** correlated highly with ***Gender*** (the higher, the more female the voice of the recordings), $r = 0.86$, $p = 0.000$), the continuous variable **F0** in Hertz did not yield significant correlations with the likability ratings ($r = 0.2$ or below). However, a slightly different picture emerged when the calculations were carried out with the traditional binary male/female category.

Here, differences (the Mann Whitney U test for independent samples) between the genders emerged tendentially ($p = 0.065$) with the languages voiced by female speakers receiving higher likability ratings. The medians were: The ***Overall*** likeability score of 57 (female-voiced languages) vs 49 (male-voiced languages); ***Eros*** – 52 (female-voiced languages) vs 39 (male-voiced languages); ***Beauty*** – 63 (female-voiced languages) vs 54 (male-voiced languages); ***Softness*** – 56 (female-voiced languages) vs. 47 (male-voiced languages); ***Melody*** – 67 (female-voiced languages) vs. 55 (male-voiced languages); ***Status*** – 58 (female-voiced languages) vs 53 (male-voiced languages); ***Orderliness*** – 52 (female-voiced languages) vs 56 (male-voiced languages). Note that the differences for ***Status*** and ***Orderliness*** are not that discrepant. The gender effect is discussed extensively in Reiterer et al. (2020). While this might reflect a stereotyped well-known scenario, it is not clear why the continuous variable F0 – as a variable that belongs to a higher-order data level – does not confirm the same trend. As we know, gender can be predicted by F0 quite reliably.

## Acoustic-Phonetic Variables

Our variables of interest were acoustic-phonetic variables **Table 5**. Here, we included: ***Melody variance in cents*** (variance of F0 contour, so-called "melody of speech," measured on the music cents scale), ***Melody variance/SD in semitones*** (same as above, just converted into a semitone scale), ***Speech rate*** (speech rate measured in syllables per second with both language sets), and ***Text length*** (the duration of the audio recording in seconds; see **Table 3** for the summary).

***Melody ñents variance*** referred to the F0 fluctuations or trajectories that were translated into music cents and later into the semitone scale. To track speech melody in terms of musical melody (as opposed to the linguistic melody), we measured the original F0 speech contours on the basis of Hertz of the fundamental frequency in both language sets. The mean speech melody fluctuations expressed in standard deviations of semitones were $\pm$3.2 semitones (SD of $\pm$0.76) in the first set and $\pm$2.9 semitones (SD of $\pm$0.8) in the second set. The largest deviations (high variance, more F0 modulations) were $\pm$4.6 or $\pm$5 semitones, whereas the lowest fluctuations were $\pm$2 or $\pm$1.9 semitones, respectively, on average. The larger modulation ranges, such as $\pm$5 semitones (considering $\pm$) this means a range or tone interval of 9.2–10 semitones, which corresponds to almost one "sixth" of the octave range. This is a rather large range for speech that was produced by reading aloud neutral

**TABLE 5** | Descriptive statistics with the phonetic-acoustic variables for both language sets ($N$ = 16).

| | | F0 mean (set 1) | F0 mean (set 2) | Pitch contour in cents variance (set 1) | Pitch contour in semitones SD (set 1) | Pitch contour in cents variance (set 2) | Pitch contour in semitones SD (set 2) | Syllables / sec (set 1) | Syllables / sec (set 2) | syll / sec compared to (Coupé et al 2019) | Text recording in sec – (set 1) | Text recording in sec – (set 2) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N | Valid | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 9 | 16 | 16 |
| | missing | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 0 |
| Mean | | 153,6804 | 208,1914 | 107855,5334 | 3,2006 | 87289,7104 | 2,8507 | 5,9063 | 5,8312 | 6,8691 | 37,3281 | 37,4931 |
| Median | | 152,0234 | 212,3133 | 102388,8365 | 3,1867 | 67992,7190 | 2,6065 | 5,9000 | 5,7500 | 7,0650 | 39,0900 | 37,8700 |
| Standard Deviation SD | | 34,40028 | 24,70695 | 50018,54414 | ,76029 | 55450,10240 | ,80150 | ,67845 | ,61937 | ,64028 | 4,42692 | 2,80527 |
| Range | | 102,22 | 77,58 | 171114,42 | 2,61 | 221171,15 | 3,16 | 2,35 | 1,90 | 1,86 | 14,33 | 12,14 |
| Minimum | | 107,88 | 165,71 | 38734,19 | 1,97 | 37026,79 | 1,92 | 4,85 | 4,90 | 5,87 | 29,73 | 33,06 |
| Maximum | | 210,10 | 243,29 | 209848,60 | 4,58 | 258197,94 | 5,08 | 7,20 | 6,80 | 7,73 | 44,06 | 45,20 |

texts (containing no expressions of extreme emotions or cries). When compared to a song, one could realize a restricted variety of songs with a pitch range of one sixth. The other extreme (4 semitones = more or less equivalent to a "second" pitch interval in music) is suboptimal for realizing a tune or a song melody. Thus, we could not think of any tune or song which suffices only on a pitch range of one "second" (only the very beginning of the Beatles song "*Yesterday*" could be started on that interval). The least melodious languages in our set were comparable to such a "song." The least melody/pitch variance [i.e., flattest melody) was found for French (2), Croatian (2.4), and Catalan (2.4), followed by Basque (2.5). In the second language set, similarly, these languages were Croatian (1.9), Basque (2.2), and Spanish (2.2). The languages with the highest pitch modulations in the first set were Polish (4.6), Italian (4.3), and Welsh (4), followed by Hungarian (3.8). In the second set, we had Welsh (5), Ukrainian (3.9), and Greek (3.7) with this range (see also **Figures 6**, **7**). With the pitch intervals found for Italian, Polish, and Welsh (a fourth, fifth, and a sixth pitch interval in terms of F0 variance), one could try and sing children's songs such as *All my little ducklings* or *Twinkle, twinkle little star. Even* these rather wide speech pitch ranges are not comparable to the much more variant pitch ranges/intervals used in songs, typically performed in most musical styles. Thus, to summarize, we found that music uses far wider pitch ranges than language expressed by speech melody. Speech is rather flat when compared to music.

***Melody cents variance*** showed a series of significant correlations. First and foremost, it correlated negatively with ***Speech rate*** ($r$ = −0.6, $p$ < 0.01**), meaning higher ***Melody ñents variance*** went hand in hand with lower numbers of syllables per second (see also **Figures 7A,B**). The same result ($r$ = −0.6, $p$ < 0.01**) – the speed-melody trade-off – we observed in the totally independent second set of language recordings (**Figure 7B**). Thus, the slower the speech was, the more pitch modulations were found; the opposite was also true – the faster the speech, the fewer (F0) modulations (flatter, fewer semitones up and down) were present in the recording. In other words, the speakers who "sang" (modulated more in terms of pitch) slowed down; and the speakers who kept their F0 stable spoke faster. Interestingly, a significant negative correlation was found between ***Melody ñents variance*** and ***Eros*** ($r$ = -0.5; $p$ = 0.04). Thus, higher ***Eros*** ratings were in line with **lower** and not higher **speech melody**, as it would be suspected. It was the faster speech and not the melody that participants found erotic (see **Figure 8**). There were no significant correlations (below or around 0.2) between F0 distribution as a marker of an acoustic ***Gender*** and ***Melody cents variance*** in the first (gender-mixed) language set. This finding did not confirm the common belief that female speakers display more speech melody or prosody. However, larger recording sets are needed to investigate the gender aspect thoroughly.

Furthermore, in terms of ***Melody ñents variance*** the two language sets showed a trend toward a weak or moderate positive correlation, but this was not significant ($r$ = 0.4, $p$ = 0.12). This indicates that pitch ranges and speech melodies are subject to both, language inherent features and individual differences in voice, and more research is necessary to show how stable

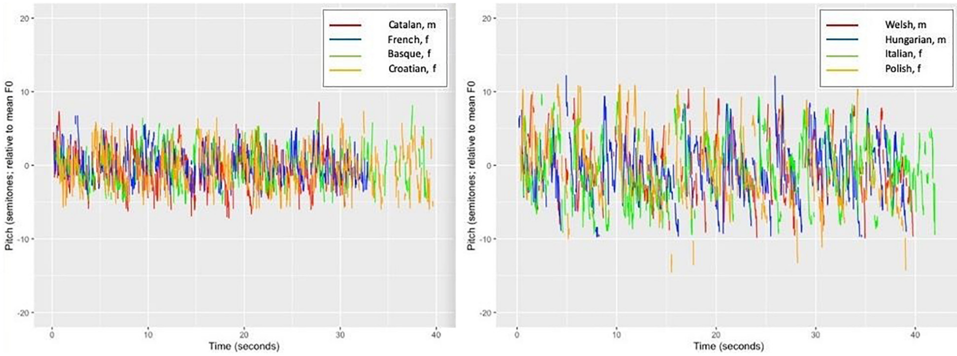



**FIGURE 6 |** The pitch range (F0 fluctuations) for the four flattest (Catalan, French, Basque, and Croatian) and the four most variable/melodic (Welsh, Hungarian, Italian, and Polish) languages. F0 fluctuations are shown in terms of ±standard deviations of musical semitones.

the prosody characteristics are across languages (inherent to languages). We discuss this question further in *Limitations*.

***Speech rate*** in syllable per second was measured as another acoustic-phonetic feature of speech. ***Speech rate*** was compared to the second set of recordings matched for language and text type (*The Northwind and the Sun*) to check the reproducibility and reliability. The sets correlated highly and positively ($r = 0.8$, $p = 0.000^{**}$). The paired samples t-test showed no significant difference between both language sets: the mean for the first set was 5.9 syllables per second (±0.6) and the mean for the second set was 5.8 syllables per second (±0.6). Next, we compared our ***Speech rate*** to the same variable in the study conducted by Coupé et al. (2019). In their study, ***Speech rate*** was measured across a variety of world languages and based on numerous text genres and speakers. We had nine overlapping languages with their dataset: Basque, Catalan, Serbo-Croatian, English, French, German, Hungarian, Italian, and Spanish. We found that both our language sets (first and second) correlated highly with ***Speech rate*** Coupè and colleagues reported (the first set: $\mathbf{r = 0.72}$; $\mathbf{p = 0.03}^*$; the second set: $\mathbf{r = 0.68}$, $\mathbf{p = 0.04}^*$).

The languages with the highest ***Speech rate*** in the first set were Spanish (7.2 syll/sec), Catalan (7 syll/sec), and Italian (6.5 syll/sec). The lowest ***Speech rate*** were found for Welsh (4.8 syll/sec), German (5.1 syll/sec), and Polish (5.2 syll/sec). Furthermore, ***Speech rate*** yielded strong positive correlations to all the aesthetic ratings, except for ***Status*** and ***Orderliness***: with ***Eros*** ($r = 0.8$, $p = 0.000^{**}$), with ***Beauty*** ($r = 0.7$, $p = 0.002^{**}$), with ***Melody*** $r = 0.7$, $p = 003^{**}$), and with ***Softness*** $r = 0.6$, $p = 0.007^{**}$) – the faster, the better. These perceptions unanimously show that the faster the spoken speech, the higher the likability ratings. The example of this result is illustrated in **Figures 8A,B** with ***Eros*** ratings. Two "big" languages form outliers to the correlation: English and French. French reaches a rather high ***Eros*** rating in proportion to its speed, and so does English. However, since both languages are amongst the most popular foreign languages learned and institutionalized this finding should not be that surprising.

Since it is still unclear how pitch ranges and F0 variance fluctuate across languages (and how much they vary across individuals as personal voice/prosody traits), we think that the pitch variations found here reflect both: language-inherent prosodic characteristics and voice-individual characteristics (see more notes on voice in the Discussion).

***Text length*** (refers to the length of the recordings in seconds) was used to control for the influences of the inter-individual speaker or language-related (lexicon, translation, word-formation) differences of the text. Although the storyline of *The Northwind and the Sun* is the same across languages, every language uses its own morpho-syntactic and lexico-semantic rules to realize the text, and moreover many different versions of the translations exist. Notwithstanding, despite all the idiosyncrasy, we found no significant correlations with all the variables, but two: in the first set and the second set of independently collected voice recordings' text length showed a marginally significant moderate correlation ($r = 0.46$; p.076) and there is a significant positive high correlation between ***Melody cents variance*** **and** ***Text length*** ($r = 0.6$, $p = 0.01^*$) in the second set and there was a similar one at trend level in the first set of recordings ($r = 0.44$, $p = 0.085$). The positive correlation means that the more speech melody is modulated along the pitch trajectory, the longer it takes to voice the text. Lower ***Melody cents variance*** or flatter voices read the text quicker. This hints toward an articulatory trade-off between time and vocal space (e.g., to "sing" a sentence takes longer).

## Further Correlations: Musicality and Singing Ability

We asked participants ($N = 45$) to self-report on their musicality (the scale of 0–10), the number of instruments played, and singing ability. As a result, ***Singing ability*** correlated with the likability ratings, but also with ***Recognition rate***, i.e., the familiarity with the languages (in fact, all music-related scores correlated significantly with the recognition rate: Pearson's r ranged between $0.38^{**}$ and $0.47^{**}$). The higher the music expertise/practice, the higher also the foreign language expertise. Musical people were more successful in guessing the languages they heard. Even though it did not necessarily imply that they spoke these languages, it seemed that they had a better feeling

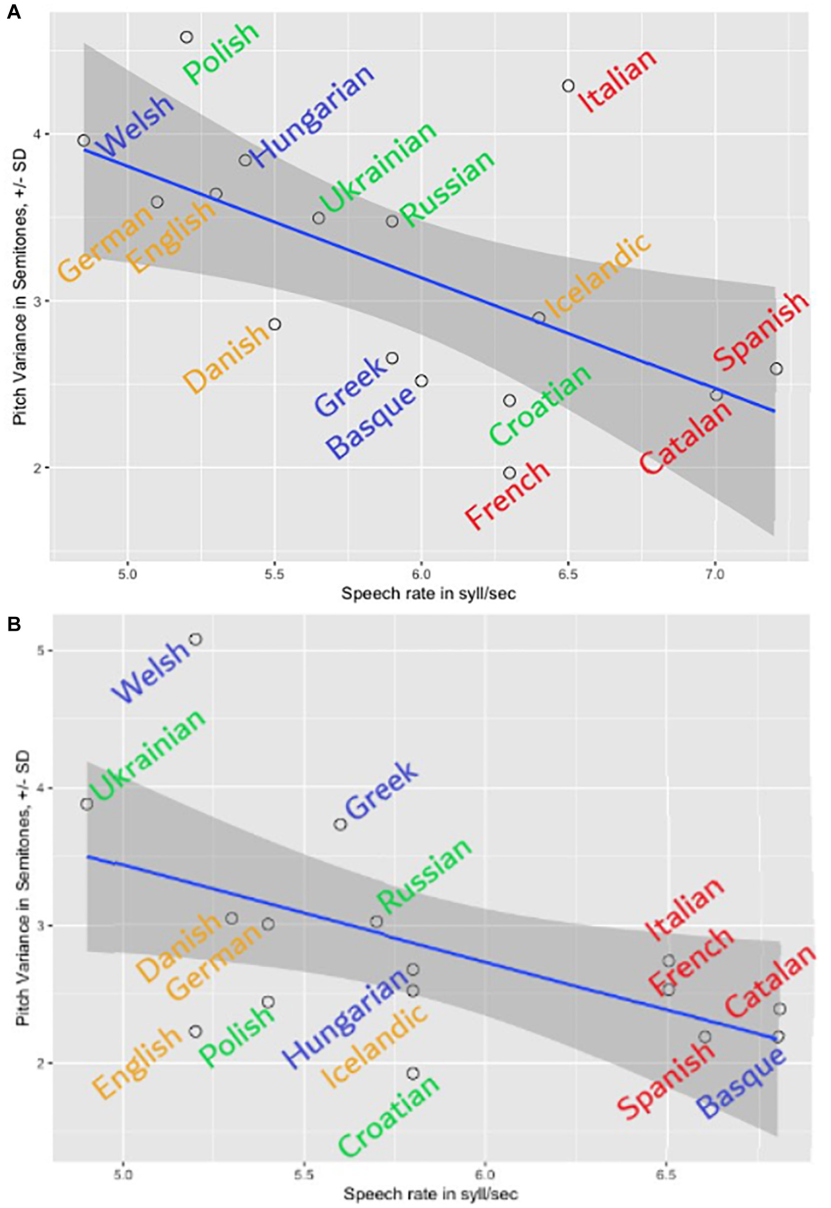



**FIGURE 7 |** The speed-melody trade-off in the first **(A)** and the second **(B)** sets of language recordings: the slower the speech, the more pitch modulations are found.

or "ear" for the languages overall. There was a trend between ***Singing ability*** and the number of foreign languages spoken ($r = 0.3$, $p = 0.052$). ***Singing ability*** also correlated positively with ***Beauty*** ($r = 0.3$, $p = 0.04$), and ***Eros*** ($r = 0.3$, $p = 0.02$), and not significantly – with ***Softness***, ***Melody***, ***Status***, ***Orderliness***, or ***Voice*** ratings. ***Musicality*** correlated highly ($r = 0.75^{**}$, $p = 0.000$) with ***Singing ability***, but also with ***Beauty*** ($r = 0.3$, $p = 0.02$) and ***Status*** ($r = 0.3$, $p = 0.04$) and not with the other factors (***Melody***, ***Orderliness***, ***Softness***, ***Voice***, and ***Eros***). In sum, a sensitivity toward music and singing was connected to the enhanced perception of language beauty, eroticity, and status. Musical people were also better language guessers: it was enough for some of them to get exposure to the sound shape of the language sample once to identify the language correctly. It could be that having a good ear for music works for language as well.

## Voice as a Nuisance Co-variate

***Voice*** judgments correlated highly with the following aesthetic ratings: ***Beauty***, ***Eros***, ***Softness***, and ***Melody*** ($r > 0.8$, $p = 0.000^{**}$) and ***Status*** ($r = 0.5$, $p = 0.04$), but not with ***Orderliness*** ($r = 0.2$, $p = 0.36$). Furthermore, ***Voice*** correlated significantly with ***Speech rate*** ($r = 0.6$, $p = 0.01$). See Discussion for the relevant comments on this trend.

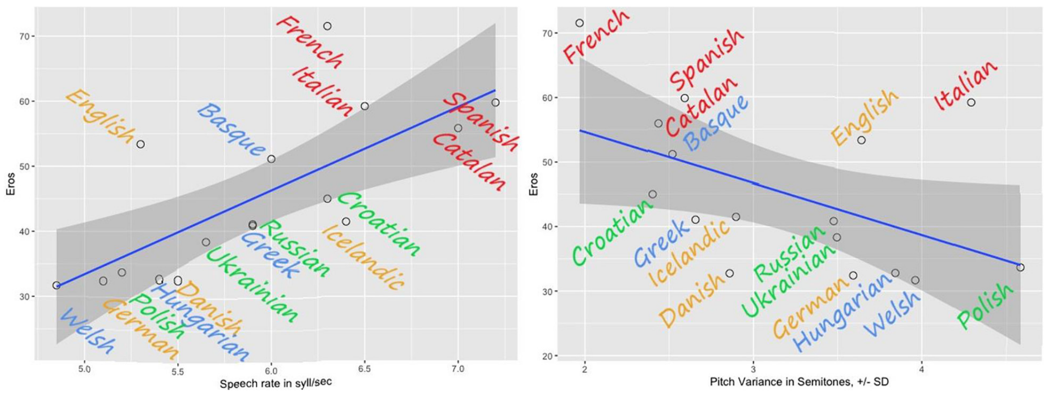



**FIGURE 8 |** The relationships between the Eros factor and the speech rate (on the **left**) and the Eros factor and the pitch variance (on the **right**). The highest Eros ratings are assigned to faster flatter languages.

# DISCUSSION

It is widely believed that some languages sound more musical than others, and attempts to identify the music-in-the-language have been around for at least several centuries. Charles V (1500 – 1558), Holy Roman Emperor, dryly remarked, "I speak Spanish to God, Italian to women, French to men, and German to my horse" (Brunner, 2014). Prompted by comments such as this, the present article aims to capture the quantitative sources behind the aesthetic judgments of cross-linguistic stimuli and explains the music-like effect that some languages, particularly the Romance languages, convey to listeners.

## Familiarity

Participants in this study were familiar with many of the languages we used: 100% recognized German, French, and English, 93% recognized Italian, and 78% recognized Spanish. The less recognized languages were Icelandic, Danish, and Basque, with a recognition rate of 9, 9, and 4%, respectively. Welsh evoked the most guesses.

Overall, participants derived more pleasure from listening to the languages they recognized (except German; on the contrary, Basque or Icelandic were highly rated but barely recognized) because familiar languages seem more beautiful, erotic and of a higher cultural status. Decades of research into the connection between familiarity and liking have shown a robust positive correlation between the two (e.g., Birch and Marlin, 1982), the gist of which is summarized by an oft-quoted German proverb: "What a farmer does not know, he does not eat." Evolutionary psychologists explain this preference by saying that familiar objects are perceived as favorable because they have been proven harmless after the initial exposure (Bornstein, 1989). More recent socio-cognitive explanations point to a facilitation effect, which occurs when a stimulus is processed on a second or third-occasion (Winkielman and Cacioppo, 2001; Reber et al., 2004). Thus, familiar languages, requiring less effort to process, might be perceived as sounding more pleasant because they are easily recognized. In this sense, the auditory pleasure derived from listening to the sounds of language resembles the pleasure derived from music: anticipation or expectancy (previous knowledge of language or music) is the essential mechanism for a pleasurable experience (Steinbeis et al., 2006; Vuust and Kringelbach, 2010).

Nevertheless, our participants did not prefer languages that were their native languages (L1s) or close to L1s, the same linguistic family they usually spoke. Rather, they were influenced by the familiarity associated with a foreign or second language-learning experience (the so-called "exotic touch"); the more languages they spoke, the more they enjoyed the sound of foreign languages. In this regard, the Serva-Petroni lexical distance (Serva and Petroni, 2008) mattered: distant but familiar languages were more welcome than less distant but familiar languages. This balance between sounding exotic and being familiar influenced (i.e., significantly correlated with) most of the aesthetic ratings, namely, culture-status, sweet-softness, and eroticity, which again echoed the findings of similar research in music (Eisentraut, 2012).

Despite the positive effect of facility by familiarity, several languages did not conform to this pattern: e.g., German was recognized by 100% of participants but received unfavorable aesthetic ratings (only in orderliness and culture-status), whereas the unrecognizable Basque and Icelandic languages enjoyed favorable ratings. On the other hand, exotic, and barely recognizable Basque came sixth in eroticity rating and seventh in sweetness and softness. Thus, with Basque and Icelandic, it appears to be that the language's sound shape and not its socio-cultural associations evokes pleasurable responses.

Certainly, when dealing with natural (and not artificial) languages and aesthetic judgments, it is almost impossible to separate the socio-cultural influences (e.g., enjoying the sounds of Italian because Italy is beautiful) from the language's acoustic contributors (e.g., Italian is spoken rapidly with a high sonority index). In our previous study (Reiterer et al., 2020), we concluded that it is the complex interplay of socio-cultural factors, idiosyncratic properties of voice, and language-specific acoustic features that account for aesthetic judgments about language. In the present study, the goal was to focus on language-specific features exclusively. Nevertheless, we looked at one socio-cultural aspect – the size of the L1 speaker's community and observed a connection between it and the higher ratings for orderliness and culture-status. While we might plausibly assume that "big" languages are empowered ecologically, as they have greater prestige in comparison to minority languages, we might also expect participants to rate such languages (e.g., German and Russian) as highly ordered with a high culture status. However, we found no relationships between community size and beauty, eroticity, or melody ratings. It seems that our participants' linguistic aesthetic judgments were unaffected by this soci-cultural factor, i.e., they did not find high-status languages to sound more pleasant.

## Musical Abilities and Individual Differences

Not all participants were affected by the acoustic features of the language recordings to the same degree. In our study, participants with higher self-perceived musical abilities (particularly in singing) tended to find languages more beautiful and erotic and assigned a higher social status to them. Such individual differences might be explained by neural mechanisms that underlie language and music. Taking congenital amusia (a neurodevelopmental disorder commonly known as "tone-deafness") as one extreme example and an exceptional musical talent as another, individuals vary in the ways they process musical stimuli. Previous research has shown that emotional prosody in speech, when other linguistic information is absent, is processed similarly to music (Patel et al., 2005; Hutchins et al., 2010; Zhishuai et al., 2017), so that a non-musical person might be unaffected by the emotional valence of the acoustic features that language conveys. However, the opposite might also be true: sensitivity to emotional speech prosody or sound properties of language might be enhanced in individuals who can process and interpret music well (Thompson et al., 2012).

Investigating a massed repetition effect inducing a perceptual transformation from speech to a song, Falk et al. (2014) found individual differences in sensitivity to underlying acoustic cues. Thus, rhythmically sensitive participants experienced the speech-to-music transformation more often compared to other less sensitive participants. In their study, a surprising finding was that professional musicians are not always high-perceivers, perhaps because they might have higher criteria for auditory signals to be interpreted as musical. On the other hand, amateurs tend to evaluate speech as music much more readily. In this regard, musicality that is connected to "phonetic chill," the intense pleasure one experiences when listening to the sounds of language, should be interpreted not only in terms of professional training but also as a stand-alone individual perceptual and productive ability and strategy (Christiner and Reiterer, 2015). Assaneo et al. (2019) similarly observed that formal musical training does not explain an enhanced ability to synchronize motor output to auditory input (e.g., tapping to music). Synchronization type is a consistent individual trait supported by functional and/or structural brain differences and influences, among other tasks, language learning outcomes. Assaneo et al. (2019) found that high synchronizers were also better language learners, an association confirmed by the present study in which participants with musical abilities spoke, on average, more foreign languages than their non- or less-musical peers. The connection between musicality, especially singing, and foreign language aptitude appears in children as early as kindergarten, as several studies show (e.g., Nardo and Reiterer, 2009; Christiner and Reiterer, 2018; Christiner et al., 2018).

## The Heartbeat of Language Is Rhythm

While some languages are spoken faster than others, yet, the rate of information they convey is much the same. Coupé et al. (2019) analyzed 17 languages from nine language families and reported comparable average rates at which information is emitted across languages (the channel capacity). In their sample, the fastest languages in the sample were Japanese, Spanish and Basque; the slowest were Yeu Chinese/Cantonese, Vietnamese and Thai. Because speech rate does not influence encoding efficiency, it might nevertheless have aesthetic value. Our participants found faster languages (Spanish, Catalan, Italian, and French) more beautiful, erotic and 'melodious' but less orderly and more chaotic. Participants characterized Spanish as sounding "strong," "convincing," and "decisive." Catalan sounded "too fast." For Italian, one participant said: "pleasant flow" and for French – "wunderschön gesprochen" and "in this case I experienced slight ASMR" (autonomous sensory meridian response, i.e. chill). English was also one of the slowest languages in our sample, yet, it was difficult for participants to evaluate this language objectively due to a high proficiency level: "Initially I couldn't help not to judge the language including the meaning. The experience of the sound immediately gave me visuals and the atmosphere." Some of our participants labeled slower languages as sounding "harsh, hard to pronounce" (German), "middle age/older" (Polish/Danish), and "the pronunciation seems very hard" (Welsh).

Previous research that analyzed how speech rate and tempo in music relate to emotions demonstrated that increased speed in both domains is associated with high-arousal or "active" emotions (Juslin and Laukka, 2003; Ma and Thompson, 2015), such as happiness, the anger and fear. While we are unaware of research connecting speech rate with aesthetic judgments, one way to interpret our findings is to look at the connection between the rate of speech and syllabic structure. The most likable languages (in our study) employ primarily the CV structure: e.g., 58% of syllables in Italian are CV syllables, compared to 31% in German (Rabanus, 2003). Such structures imply not only a greater articulatory economy (Janson, 1986) but also produce a simple rhythmic pattern with regular syllabic pulses that might boost the speech rate. Slower languages (e.g., German and Polish) have more elaborate syllabic structures that might be perceived as syncopated rhythms – complex and ambiguous rhythms that sound "off-beat." Such rhythms stress weak positions in the metrical structure while leaving nearby strong positions "empty," or without stress. Fitch and Rosenfeld (2007) investigated syncopation in music and concluded that its rhythms are difficult to process and remember since they require listeners to reset their internal pulse representations. In the case of languages, syncopated rhythms might also call for greater cognitive and articulatory effort, leading to slower speech rate, and negative aesthetic ratings (However, since syncopation in language is rarely explored, this must be said with caution).

Dellwo and Wagner (2003) found that the consonantal interval measure correlates negatively with speech rate, meaning that languages with complex consonant clusters (e.g., German, Polish, or Russian) would have a slower speech rate compared to languages that lack this feature (e.g., Italian, French, and Spanish). They also found a positive correlation between speech rate and sonority that is tightly connected to vocalic share – vowels are the most sonorous sounds in most languages. One would expect that languages with the predominant CV structure normally have more vowels, and therefore, a higher sonority index. In our previous study (Reiterer et al., 2020) we found a positive correlation between sonority and beauty ratings: more sonorous languages were also perceived as sounding more beautiful and erotic.

The CV syllabic structure is faster to perceive, process and produce, and when the structure is repeated it might produce a music-like effect. Falk et al. (2014) looked at the repetition effects that induce a perceptual transformation from speech to song – the situation, in which a spoken sentence repeated several times begins to resemble a song. They explained the emergence of musical percepts by a more detailed acoustic encoding that is facilitated as a result of repetition. In other words, when the same verbal structure is repeated its content fails to become of primary importance (in our study, most languages were not understood semantically) and the acoustic characteristics (e.g., melody, rhythm) that did not matter before become more salient. Even though this study was conducted with utterances and not syllables, we can assume that a similar perceptual strategy might be applied to a repeated syllable. If a syllabic structure is the same or similar every time (e.g., "banana"), cognitive resources, freed from attending to a varying syllabic structure and associating with its rhythmic (syncopated) complexity, are used for the enhanced perception of the acoustic properties creating a music-like effect.

Finally, a recent preliminary investigation by Lin and Rathcke (2020) into the anchor of sensorimotor synchronization suggests that the moments of local maximal energy increase (maxD) as well as vowel onsets, are prominent acoustic landmarks for rhythmic attention and synchronization with speech. Based on these findings, languages with more vowels could be perceived as possessing a faster rhythm (in our sample, Basque and French had the highest vocalic share, German and Polish the lowest; Reiterer et al., 2020). Infant et al. (2013) measured rhythm perception based on the amplitude modulation structure of the speech envelope, which is also connected to syllabic nuclei and vowels, and observed that syllable-timed languages (e.g., Spanish or French) indeed have faster rates than stress-timed languages (e.g., German or Russian).

## The Melody Paradox

When we speak of melody, we traditionally mean music and music-like stimuli (environmental sounds, the melody of the voice, etc.). However, in linguistics melody refers to an organized pitch pattern in speech (Patel, 2010). In the present study, we employ 'speech melody' in its musical sense – how melodiously a language sounds, the music-in-the-language. Our results showed that languages with fewer pitch variations (F0 contour as measured in hertz/cents/semitones) are perceived as more erotic and melodious. Except for Italian, the Romance languages (Catalan, Spanish and French leading the list) show the flattest intonation contour. At the opposite end were Welsh, English, and Polish (and Italian), which used an impressive range of pitch variation. Compared to music semitones, Polish would sound like *My bonnie lies over the ocean//Nobody knows the trouble I've seen,* and, Catalan, the most atonal, would sound like the first bar of two bars in *Strangers in the night* or *Yesterday*. This surprising finding can be summarized thus: the most "melodious" languages are those without melody!

It has been observed previously that linguistic intonation indeed lacks the complexity and aesthetic potency of musical pitch, which has an elaborate system of intervals with a rich network of pitch relations. Patel (2010) commented that: "Intonation contours are aesthetically inert, as evidenced by the fact that people rarely hum intonation contours or find themselves captivated by the pitch patterns of speech" (p. 184). Indeed, pitch has a different function in speech. In music, it is an aesthetic object, but in speech, pitch has a practical purpose – to convey structural information. Since pitch is not the only linguistic feature responsible for this function, spoken language, compared to music, is rather atonal (Chow and Brown, 2018), making it very hard to derive an aesthetic pleasure from linguistic intonation. Even when pitch varies minimally it does not affect comprehension significantly.

A recent study by Albouy et al. (2020) demonstrated that perception of speech is most affected by the degradation of information in the temporal dimension (rhythm), whereas perception of music is most affected by degradation in the spectral dimension (pitch). The authors propose that these two domains employ opposite extremes of the spectro-temporal continuum, which is reflected neurally as two specialized complementary systems, one in each hemisphere. In the light of these new findings, a wide range of F0 variation throughout an utterance is not essential for comprehension. The temporal as well as timbral (Reiterer et al., 2008) information, on the other hand, is a more crucial dimension for speech. In this regard, flat and fast Romance languages might be processed by naive listeners with greater ease and generate positive judgments for that reason. This means that the 'melodiousness' of French and Spanish is rooted in the rhythm and socio-cultural stereotypes attached to these languages and not in the actual speech melody, i.e., pitch variation.

In our study, the tradeoff between pitch and rate is surprisingly consistent across languages: the more pitch is modulated, the slower the speech rate. The opposite is also true: the less pitch is modulated, the faster the speech rate. The idea that auditory cognition depends on the processing of spectro-temporal energy patterns and that these features often trade-off against one another is not new and has been demonstrated before (e.g., Elliott and Theunissen, 2009; Flinker et al., 2019). Speech can be understood with either very coarse spectral information or very coarse temporal information (Arai and Greenberg, 1998). In the present study, the languages display various ratios between pitch variation and rate, with the Romance languages being the flattest and the fastest, a combination that can hardly be said to be melodious unless the claimant is a fan of hip-hop. Interestingly, that Spanish has been often called a "machine gun" language because of its fast and flat sound. On the other hand, the Germanic and Slavic language families show greater variety in terms of pitch, and therefore, are more melodious in the linguistic sense of the term. At the same time, these languages are slower, thus, producing an opera-like effect – the prolonged intervals of "singing" that take time and allow for more pitch variations.

Just as languages can be classified loosely into tempo-dominant and pitch-dominant, listeners might be sensitive to one domain more than the other. In a recent study, Christiner et al. (2018) found that listeners with superior rhythmic discrimination ability as measured by a subtest of IMMA (a test of musicality; Gordon, 1982) imitated an unfamiliar language, Tagalog, better than they did unfamiliar Chinese. Because Tagalog is a non-tonal language, therefore, its rhythmical organization can be predominantly recognized by naïve listeners. In marked contrast, a singing ability appeared to be key to imitating Chinese, a tonal language where pitch plays an important meaning-bearing function. The authors concluded that people vary in the type of specific acoustic features they rely on when processing utterances in an unfamiliar language. These individual differences do not have to be structured by a native language alone; they can be determined by the type of musical training – e.g., the type of instrument played (Schneider et al., 2002) – or innate abilities; e.g., singing (Oikkonen et al., 2015). To shed light on the nature of aesthetic preferences and to explain phonetic aptitude in some language learners, we recommend that the interplay between language typology and individual differences in auditory processing should be researched further (Kogan and Mora, 2017).

## The Case of Italian

Italian was the only language that did not demonstrate a clear trade-off between speech rate and intonation (at least in the first set of recordings): both fast and varied in pitch, it remains a linguistic enigma. More research – and certainly more recordings coming from a variety of speakers – is needed to understand how Italian speakers (as it was one the outlier cases here) manage to be fast and varied in pitch at the same time. However, as the second set of independent recordings showed, Italian was a perfect fit to the trade-off between speed and melody.

Another interpretation underlying this trade-off phenomenon in general could be that it reflects a more basic physiological mechanism, namely a sort of "Heisenbergian" trade-off between space and time resolution. It is imaginable that our vocal articulators are not optimized for producing melodious (pitch-intense) but precise and highly intelligible speech, e.g. by spanning octaves (like in arias), and being very fast at the same time (like in casual conversation). Articulatory movements (larynx/pharynx, lungs, and mouth cavity) have to be coordinated in space and time. Usually, song is slower than speech, for if we sang everything we wanted to say, we would need a multiple of time for information exchange.

## Limitations

One of this study's principal limitations is that the languages it used are represented by single speakers (one speaker per language). This methodological decision was dictated by the study's preliminary nature: to build a larger repository of language recordings (in terms of speakers per language and the number of languages) at a later stage. Although this is not an ideal scenario, previous studies indicated that while there is wide interspeaker variation in an acoustic parameter within a language, this variation is also structured by language (Coupé et al., 2019). In other words, individual speech behavior is not due to individual characteristics alone but is further defined and guided by the language being spoken. Yet, it would be optimal to have several speakers, preferably of the same gender, voicing the same language. In fact, the second set of language recordings was collected after the present study was conducted to confirm some of the acoustic-phonetic phenomena observed in the first set. The *speed-melody tradeoff* was, in fact, exactly replicated in the second set, as well as the ordering of languages in terms of the *speech rate*. However, in terms of the pitch variance (melody), the two sets only showed a weak positive trend indicating that the pitch-related results should be considered as reflecting both, individual (speaker inherent) and more global linguistic-typological (language inherent) traits.

In regard to gender, in our previous study (Reiterer et al., 2020) we found that languages presented by female voices were rated significantly higher on average than those presented by male voices. This finding confirms previous research that addressed this methodological issue (McMinn et al., 1993; Wilding and Cook, 2000; Whipple and McManamon, 2002; Edworthy et al., 2003). In this study we used a continuous variable for F0 rather than a binary male/female variable, and the previously observed connection between the speaker's gender and the aesthetic ratings was not present this time. Also, we did not find a connection between female speakers and the average pitch variance or speech rate.

As well as inter-speaker voice differences, there are also individual differences in how a voice is processed by listeners. Some listeners have an enhanced ability to extract, evaluate, and categorize non-linguistic information available in voices (Brück et al., 2011). Such voice-sensitive individuals might experience any spoken stimuli more intensely, and therefore, might be more susceptible to phonetic chill. Studies employing brain imaging techniques would be particularly helpful to account for this variable. The voice as a nuisance variable should be addressed in more detail in future studies.

Finally, the question of familiarity with some of the languages in the study is an important methodological decision that one has to make. We carefully measured the familiarity aspect not only by asking participants to guess the language they were listening to but also by including additional questions about participants' second language education and languages spoken (at which proficiency levels) besides their mother tongue. This information was included in the analysis (polyglot factor) and reported accordingly (Reiterer et al., 2020). By no means do we deny the influence of familiarity on participant's judgments – on the contrary, we would like to endorse that it is a complex interplay between the familiarity, speaker/listener's unique characteristics (individual differences), and language's sound shape that influences the aesthetic rating. We deliberately wanted to address the question of the musical allure of the widely spoken languages that are prototypically (through public surveys and on online forums) declared the most melodious (e.g., Italian). Unfortunately, these languages are also common languages learned in schools and they enjoy public familiarity. In our previous study (Reiterer et al., 2020) we measured the share that familiarity contributes to the overall aesthetic judgments and came to the conclusion that although it plays a prominent role (about 40% of variance), there is more to the story. The present study was dedicated to the acoustic-phonetic dimension only without an explicit and further investigation of familiarity. For future studies, the strategy could be to use only unfamiliar languages or artificial languages or focus on non-European listeners.

Also, it would be beneficial to incorporate a variety of languages in future studies, preferably of languages that belong to different families and show a range of phonetic and acoustic features. The present study was a first exploration into aesthetic preferences for the sound of only 16 European languages.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

SR contributed to conception and experimental design of the scientific work. SR and VK involved in processing of data collection, contributed to data analysis and interpretation, drafted, and wrote and revised the article critically. Both authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

We want to explicitly thank Gasper Pesek and Ziga Bogataj from the University of Ljubljana, and Dr. Jörg Mühlhans, the head of the Media Lab of the Faculty of Philological and Cultural Studies at the University of Vienna for initial help with programming and voice recordings. Special thanks for valuable discussions go to Dr. Isabella Czedik-Eysenberg, University of Vienna. Many thanks also to George A. James for final reads and thoughtful editing as well as to Markus for critical remarks. Furthermore, we are indebted to international volunteers who were involved in the realization of some of the voice recordings: Amanda Christiansen; Màrta Czire, Margherita de Gregorio: University of Vienna, Anna Louvrou, Itsaso Garaikoetxea, Gloria, Petra Kartela, Megan, Sandrine Laplanche, Erla Ragnheiður, Karina Rurarz, Susan Gabriel-Dennis, and Alina Yanni.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.578594/full#supplementary-material

## REFERENCES

Abercombie, D. (2013). The North Wind and the Sun, 1951-1978 [sound]. University of Edinburgh, School of Philosophy, Psychology, and Language Sciences. Department of Linguistics and English Language. Edinburgh: University of Edinburgh.

Adobe Audition [Computer software] (2018). Available. at: http://www.adobe.com (accessed February 02, 2020).

Albouy, P., Benjamin, L., Morillon, B., and Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science* 367, 1043–1047. doi: 10.1126/science.aaz3468

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J. K. (2020). Gorilla in our midst: an online behavioral experiment builder. *Behav. Res. Methods* 52, 388–407. doi: 10.3758/s13428-019-01237-x

Arai, T., and Greenberg, S. (1998). "Speech intelligibility in the presence of cross-channel spectral asynchrony," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, New York, NY.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica* 66, 46–63. doi: 10.1159/000208930

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *J. Phonet.* 40, 351–373. doi: 10.1016/j.wocn.2012.02.003

Assaneo, F. M., Ripolles, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., and Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nat. Neurosci.* 22, 627–632. doi: 10.1038/s41593-019-0353-z

Birch, L. L., and Marlin, D. W. (1982). I don't like it; I never tried it: effects of exposure on two-year-old children's food preferences. *Appetite* 3, 353–360. doi: 10.1016/S0195-6663(82)80053-6

Blevins, J. (1995). "The syllable in phonological theory," in *Handbook of Phonological Theory*, Vol. 1, ed. J. Goldsmith (London: Blackwell), 206–244.

Boersma, P., and Weenink, D. (2017). *Praat: Doing Phonetics by Computer [Computer Program]. Version 5.3. 51*. Available online at: https://www.fon.hum.uva.nl/praat/ (accessed February 15, 2020).

Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968-1987. *Psychol. Bull.* 106, 265–289. doi: 10.1037/0033-2909.106.2.265

Brattico, E., Bogert, B., and Jacobsen, T. (2013). Toward a neural chronometry for the aesthetic experience of music. *Front. Psychol.* 4:206. doi: 10.3389/fpsyg.2013.00206

Breitenstein, C., Lancker, D. V., and Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cogn. Emot.* 15, 57–79. doi: 10.1080/0269993004200114

Brück, C., Kreifelts, B., and Wildgruber, D. (2011). Emotional voices in context: a neurobiological model of multimodal affective information processing. *Phys. Life Rev.* 8, 383–403. doi: 10.1016/j.plrev.2011.10.002

Brunner, B. (2014). *The Sound of Difference. Why We Find Some Languages More Beautiful than Others*. New York, NY: The Smart Set.

Burchette, J. (2014). *Sexiest Accents Poll: Where do People Have the Voice of Seduction?*. Atlanta, GA: CNN International.

Carbonell, J. F., and Llisterri, J. (1992). Catalan. *J. Intern. Phonet. Assoc.* 22, 53–56.

Chahal, A., Baghla, S., Kaur, R., and Kaushal, G. (2017). Heart rate monitoring using human speech features extraction: a review. *J. Eng. Technol. Sci. Res.* 4, 444–449.

Chow, I., and Brown, S. (2018). A musical approach to speech melody. *Front. Psychol.* 9:247. doi: 10.3389/fpsyg.2018.00247

Christiner, M., and Reiterer, S. M. (2013). Song and speech: examining the link between singing talent and speech imitation ability. *Front. Psychol.* 4:874. doi: 10.3389/fpsyg.2013.00874

Christiner, M., and Reiterer, S. M. (2015). A mozart is not a pavarotti: singers outperform instrumentalists on foreign accent imitation. *Front. Hum. Neurosci.* 9:482. doi: 10.3389/fnhum.2015.00482

Christiner, M., and Reiterer, S. M. (2018). Early influence of musical abilities and working memory on speech imitation abilities: study with pre-school children. *Brain Sci.* 8:169. doi: 10.3390/brainsci8090169

Christiner, M., Rüdegger, S., and Reiterer, S. M. (2018). Sing Chinese and tap Tagalog? Predicting individual differences in musical and phonetic aptitude using language families differing by sound-typology. *Intern. J. Multilingual.* 15, 455–471. doi: 10.1080/14790718.2018.1424171

Cordes, I. (2000). "Communicability of emotion through music rooted in early human vocal patterns," in *Proceedings of the Sixth International Conference on Music Perception and Cognition*, New York, NY.

Coupé, C., Oh, Y. M., Dediu, D., and Pellegrino, F. (2019). Different languages, similar encoding efficiency: comparable information rates across the human communicative niche. *Sci. Adv.* 5:eaaw2594. doi: 10.1126/sciadv.aaw2594

Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics*, 6th Edn, Oxford: Blackwell.

Dalle Fratte, M. (2018). *Italian is the language built to be sung. Melafonica*. Available online at: https://www.melofonetica.com/italian-is-a-language-built-to-be-sung/ (accessed April 14, 2020).

Dellwo, V., and Wagner, P. (2003). "Relationships between speech rate and rhythm," in *Proceedings of the ICPhS*, London.

Dolcos, F., LaBar, K. S., and Cabeza, R. (2004). Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fMRI study. *Neuroimage* 23, 64–74. doi: 10.1016/j.neuroimage.2004.05.015

Edworthy, J., Hellier, E., Walters, K., Clift-Mathews, W., and Crowther, M. (2003). Acoustic, semantic and phonetic influences in spoken warning signal words. *Appl. Cogn. Psychol.* 17, 915–933. doi: 10.1002/acp.927

Eisentraut, J. (2012). *The Accessibility of Music: Participation, Reception, and Contact*. Cambridge: Cambridge University Press.

Elliott, T. M., and Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Computat. Biol.* 5:302. doi: 10.1371/journal.pcbi.1000302

Falk, S., Rathcke, T., and Dalla Bella, S. (2014). When speech sounds like music. *J. Exper. Psychol.* 40:1491. doi: 10.1037/a0036858

Fitch, W. T. (2013). "Musical protolanguage: darwin's theory of language evolution revisited," in *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain*, eds J. J. Bolhuis and M. B. H. Everaert (Cambridge, MA: MIT Press), 489–503.

Fitch, W. T., and Rosenfeld, A. J. (2007). Perception and production of syncopated rhythms. *Music Percept.* 25, 43–58. doi: 10.1525/mp.2007.25.1.43

Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O., and Poeppel, D. (2019). Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nat. Hum. Behav.* 3, 393–405. doi: 10.1038/s41562-019-0548-z

Font-Rotchés, D., and Torregrosa-Azor, J. (2015). Melodic analysis of speech: a methodological approach for contrastive analysis of the intonation of Yes-No questions in Spanish and German, *Paper Presented at XXXIV Romanistentag*, Mannheim.

Fonseca-Mora, C., Toscano-Fuentes, C., and Wermke, K. (2011). Melodies that help: The relation between language aptitude and musical intelligence. *Int. J. English Stud.* 22, 101–118.

Fougeron, C., and Smith, C. L. (1993). French. *J. Intern. Phonet. Assoc.* 23, 73–76. doi: 10.1017/S0025100300004874

Giles, H., and Niedzielski, N. (1998). "Italian is beautiful, German is ugly," in *Language Myths*, eds L. Bauer and P. Trudgill (London: Penguin), 85–93.

Gordon, E. E. (1982). *Intermediate Measures of Music Audiation*. Chicago, IL: GIA.

Greenberg, J. (1965). Some generalizations concerning initial and final consonant clusters. *Linguistics* 18, 5–34. doi: 10.1515/ling.1965.18.1.5

Grewe, O., Nagel, F., Kopiez, R., and Altenmüüller, E. (2007). Listening to music as a re-creative process: physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Percept.* 24, 297–314. doi: 10.1525/mp.2007.24.3.297

Grønnum, N. (1998). Danish. *J. Intern. Phonet. Assoc.* 28, 99–105.

Guest123456 (2010). *Worst and Best Languages for Singing? February 9 [Online forum post]*. Available online at: http://www.antimoon.com/forum/t15861.htm (accessed January 12, 2019).

Guhn, M., Hamm, A., and Zentner, M. (2007). Physiological and musico-acoustic correlates of the chill response. *Music Percept.* 24, 473–484. doi: 10.1525/mp.2007.24.5.473

Hualde, J. I., Lujanbio, O., and Zubiri, J. J. (2010). Goizueta basque. *J. Intern. Phonet. Assoc.* 40, 113–127. doi: 10.1017/S0025100309990260

Hutchins, S., Gosselin, N., and Peretz, I. (2010). Identification of changes along a continuum of speech intonation is impaired in congenital amusia. *Front. Psychol.* 1:236. doi: 10.3389/fpsyg.2010.00236

Ilie, G., and Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Percept.* 23, 319–330. doi: 10.1525/mp.2006.23.4.319

Infant, E. E. G., Goswami, U., and Leong, V. (2013). Speech rhythm and temporal structure: converging perspectives? *Lab. Phonol.* 4, 67–92. doi: 10.1515/lp-2013-0004

Jacobsen, T., Buchta, K., Köhler, M., and Schröger, E. (2004). The primacy of beauty in judging the aesthetics of objects. *Psychol. Rep.* 94, 1253–1260. doi: 10.2466/pr0.94.3c.1253-1260

Janson, T. (1986). Cross-linguistic trends in CV sequences. *Phonol. Yearbook* 5, 179–196.

Jassem, W. (2003). Polish. *J. Intern. Phonet. Assoc.* 33, 103–107. doi: 10.1017/S0025100303001191

Jekiel, M. (2014). "Comparing rhythm in speech and music: the case of English and Polish," in *Proceedings of the Yearbook of the Poznan Linguistic Meeting*, Merseburg. doi: 10.1515/yplm-2015-0003

Juslin, P. N. (1997). Perceived emotional expression in synthesized performances of a short melody: capturing the listener's judgment policy. *Music. Sci.* 1, 225–256. doi: 10.1177/102986499700100205

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: relating performance to perception. *J. Exper. Psychol.* 26, 1797–1813. doi: 10.1037/0096-1523.26.6.1797

Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129:770. doi: 10.1037/0033-2909.129.5.770

Kerr, B. (2017). *How Italian Became the Language of love. BBC*. Available online at: http://www.bbc.com/travel/story/20170621-how-italian-became-the-language-of-love (accessed November 22, 2019).

Kirby, S. (2011). "Darwin's musical protolanguage: an increasingly compelling picture," in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohmeier, J. A. Hawkins, and I. Cross (Oxford: Oxford University Press), 96–102. doi: 10.1093/acprof:oso/9780199553426.003.0010

Koelsch, S., Fritz, T. V., Cramon, D. Y., Müller, K., and Friederici, A. D. (2006). Investigating emotion with music: an fMRI study. *Hum. Brain Mapp.* 27, 239–250. doi: 10.1002/hbm.20180

Kogan, V. V., and Mora, J. C. (2017). "Which language sounds good to you? researching sources of individual differences in processing unfamiliar non-native vowel contrasts," in *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2017*, New York, NY.

Krumhansl, C. L. (1990). *Oxford Psychology Series. Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.

Landau, E., Lonéariæ, M., Horga, D., and Škariæ, I. (1995). Illustrations of the IPA: Croatian. *J. Intern. Phonet. Assoc.* 25, 83–86.

Leder, H., Gerger, G., Brieber, D., and Schwarz, N. (2014). What makes an art expert? emotion and evaluation in art appreciation. *Cogn. Emot.* 28, 1137–1147. doi: 10.1080/02699931.2013.870132

Lin, C. Y., and Rathcke, T. (2020). "How to hit that beat: testing acoustic anchors of rhythmic movement with speech," in *Proceedings of 10th International Conference on Speech Prosody 2020*, Cham. doi: 10.21437/SpeechProsody.2020-1

Ma, W., Fiveash, A., and Thompson, W. F. (2019). Spontaneous emergence of language-like and music-like vocalizations from an artificial protolanguage. *Semiotica* 2019, 1–23. doi: 10.1515/sem-2018-0139

Ma, W., and Thompson, W. F. (2015). Human emotions track changes in the acoustic envi- ronment. *Proc. Natl. Acad. Sci. U.S.A.* 112, 14563–14568. doi: 10.1073/pnas.1515087112

Martínez-Celdrán, E., Fernández-Planas, A. M., and Carrera-Sabaté, J. (2003). Castilian Spanish. *J. Intern. Phonet. Assoc.* 33, 255–259. doi: 10.1017/S0025100303001373

Mayer, R. E. (2019). "How multimedia can improve learning and instruction," in *The Cambridge Handbook of Cognition and Education*, eds J. Dunlosky and K. A. Rawson (Cambridge: Cambridge University Press), 460–479. doi: 10.1017/9781108235631.019

McMinn, M. R., Brooks, S. D., Triplett, M. A., Hoffman, W. E., and Huizinga, P. G. (1993). The effects of God language on perceived attributes of God. *J. Psychol. Theol.* 21, 309–321.

Mennen, I., Schaeffler, F., and Docherty, G. (2012). Cross-language differences in fundamental frequency range: a comparison of English and German. *J. Acoust. Soc. Am.* 131, 2249–2260. doi: 10.1121/1.3681950

Mesquita, B. (2003). "Emotions as dynamic cultural phenomena," in *The Handbook of the Affective Sciences*, eds R. Davidson, H. Gold-smith, and K. R. Scherer (New York, NY: Oxford University Press), 871–890.

Nardo, D., and Reiterer, S. (2009). "Musicality and phonetic language aptitude," in *Language Talent and Brain Activity*, eds G. Dogil and S. Reiterer (Berlin: Mouton de Gruyter), 213–256.

Oikkonen, J., Huang, Y., Onkamo, P., Ukkola-Vuoti, L., Raijas, P., Karma, K., et al. (2015). A genome-wide linkage and association study of musical aptitude identifies loci containing genes related to inner ear development and neurocognitive functions. *Mol. Psychiatry* 20, 275–282. doi: 10.1038/mp.2014.8

Ozernov-Palchik, O., and Patel, A. D. (2018). Musical rhythm and reading development: does beat processing matter. *Annu. NY Acad. Sci.* 1423, 166–175. doi: 10.1111/nyas.13853

Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford: Oxford university press.

Patel, A. D., Foxton, J. M., and Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain Cogn.* 59, 310–313. doi: 10.1016/j.bandc.2004.10.003

Patel, A. D., Iversen, J. R., and Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: the case of British English and French. *J. Acoust. Soc. Am.* 119, 3034–3047.

Petroni, F., and Serva, M. (2010). Measures of lexical distance between languages. *Phys. A Statist. Mech. Appl.* 389, 2280–2283. doi: 10.1016/j.physa.2010.02.004

Pompino-Marschall, B., Steriopolo, E., and ??ygis, M. (2017). Ukrainian. *J. Intern. Phonet. Assoc.* 47, 349–357.

Quinto, L. R., Thompson, W. F., and Keating, F. L. (2013). Emotional communication in speech and music: the role of melodic and rhythmic contrasts. *Front. Psychol.* 4:184. doi: 10.3389/fpsyg.2013.00184

Quora (2015). *What Makes a Language Sound Beautiful From Others? [Quora answer]*. Available online at: https://www.quora.com/What-makes-a-language-sound-beautiful-from-others (accessed September 18, 2018).

Rabanus, S. (2003). Intonation and syllable structure: a cross-linguistic study of German and Italian conversations. *Zeitschrift Sprachwissenschaft* 22, 86–122. doi: 10.1515/zfsw.2003.22.1.86

Ravignani, A., Delgado, T., and Kirby, S. (2016). Musical evolution in the lab exhibits rhythmic universals. *Nat. Hum. Behav.* 1, 1–7. doi: 10.1038/s41562-016-0007

Reber, R., Schwarz, N., and Winkielman, P. (2004). Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? *Pers. Soc. Psychol. Rev.* 8, 364–382. doi: 10.1207/s15327957pspr0804_3

Reiterer, S., Erb, M., Grodd, W., and Wildgruber, D. (2008). Cerebral processing of timbre and loudness: FMRI evidence for a contribution of Broca's area to basic auditory discrimination. *Brain Imaging Behav.* 2, 1–10. doi: 10.1007/s11682-007-9010-3

Reiterer, S. M., Kogan, V., Seither-Preisler, A., and Pesek, G. (2020). Foreign language learning motivation: phonetic chill or latin lover effect? Does sound structure or social stereotyping drive FLL? *Psychol. Learn. Motiv.* 72, 165–205. doi: 10.1016/bs.plm.2020.02.003

Reuter, C., and Oehler, M. (2011). Psychoacoustics of chalkboard squeaking. *J. Acoust. Soc. Am.* 130:2545. doi: 10.1121/1.3655174

Reuter, C., and Siddiq, S. (2017). "The colourful life of timbre spaces: Timbre concepts from early ideas to meta-timbre space and beyond," in *Body, Sound and Space in Music and Beyond: Multimodal Explorations*, ed. C. Wöllner (London: Routledge), 150–167. doi: 10.4324/9781315569628-9

Roach, P. (2004). British English: received pronunciation. *J. Internat. Phonet. Assoc.* 34, 239–245. doi: 10.1017/S0025100304001768

Sammler, D. (2020). Splitting speech and music. *Science* 367, 974–976. doi: 10.1126/science.aba7913

Scherer, K. R., Johnstone, T., and Klasmeyer, G. (2003). *Vocal Expression of Emotion*. Oxford: Oxford University Press.

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* 5, 688–694.

Science Chat Forum (2011). *Do Some Languages Sound Better than Others? [Online Forum Thread]*. Available online at: http://www.sciencechatforum.com/viewtopic.php?f1/412&t1/41804 (accessed November 20, 2019).

Serva, M., and Petroni, F. (2008). Indo-European languages tree by Levenshtein distance. *EPL* 81:68005. doi: 10.1209/0295-5075/81/68005

Starcke, K., von Georgi, R., Tiihonen, T. M., Laczika, K. F., and Reuter, C. (2019). Don't drink and chill: effects of alcohol on subjective and physiological reactions during music listening and their relationships with personality and listening habits. *Intern. J. Psychophysiol.* 142, 25–32. doi: 10.1016/j.ijpsycho.2019.06.001

Steinbeis, N., Koelsch, S., and Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological, and neural responses. *J. Cogn. Neurosci.* 18, 1380–1393. doi: 10.1162/jocn.2006.18.8.1380

Szende, T. (1994). Hungarian. *J. Internat. Phonet. Assoc.* 24, 91–94. doi: 10.1017/S0025100300005090

Thompson, W. F., Marin, M. M., and Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19027–19032. doi: 10.1073/pnas.1210344109

Tilsen, S., and Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *Proc. Natl. Acad. Sci. U.S.A.* 134, 628–639. doi: 10.1121/1.4807565

Verhoeven, J. (2019). *The North Wind and the Sun in Phonetics*. Available online at: https://www.phonetics.expert/north-wind-and-the-sun (accessed January 10).

Vuust, P., and Kringelbach, M. L. (2010). The pleasure of music. *Pleasur. Brain* 45, 255–269.

Whipple, T. W., and McManamon, M. K. (2002). Implications of using male and female voices in commercials: an exploratory study. *J. Advert.* 31, 79–91. doi: 10.1080/00913367.2002.10673668

Wilding, J., and Cook, S. (2000). Sex differences and individual consistency in voice identification. *Percept. Motor Skills* 91, 535–538. doi: 10.2466/pms.2000.91.2.535

Winkielman, P., and Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: psychophysiological evidence that processing facilitation elicits positive affect. *J. Pers. Soc. Psychol.* 81, 989–1000. doi: 10.1037/0022-3514.81.6.989

Wrembel, M. (2010). L2-accented speech in L3 production. *Intern. J. Multilingual.* 7, 75–90. doi: 10.1080/14790710902972263

Yanushevskaya, I., and Bunéiæ, D. (2015). Russian. *J. Intern. Phonet. Assoc.* 45, 221–228. doi: 10.1017/S0025100314000395

Zaidel, D. W., Nadal, M., Flexas, A., and Munar, E. (2013). An evolutionary approach to art and aesthetic experience. *Psychol. Aesthet. Creat. Arts* 7:100.

Zeki, S., Romaya, J. P., Benincasa, D. M., and Atiyah, M. F. (2014). The experience of mathematical beauty and its neural correlates. *Front. Hum. Neurosci.* 8:68. doi: 10.3389/fnhum.2014.00068

Zhishuai, J., Hong, L., Daxing, W., Pin, Z., and Xuejing, L. (2017). Processing of emotional faces in congenital amusia: an emotional music priming event-related potential study. *Neuroimage Clin.* 14, 602–609. doi: 10.1016/j.nicl.2017.02.024

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 09 April 2021
doi: 10.3389/fnins.2021.596636

# Can Music Influence Patients With Disorders of Consciousness? An Event-Related Potential Study

*Yajuan Hu[1†], Fengqiong Yu[2,3,4†], Changqing Wang[1], Xiaoxiang Yan[1] and Kai Wang[1,3,4*]*

[1] *Department of Neurology, The First Affiliated Hospital of Anhui Medical University, Hefei, China,* [2] *Department of Medical Psychology, Chaohu Clinical Medical College, Anhui Medical University, Hefei, China,* [3] *Collaborative Innovation Center of Neuropsychiatric Disorders and Mental Health, Hefei, China,* [4] *Anhui Province Key Laboratory of Cognition and Neuropsychiatric Disorders, Hefei, China*

**Background:** Long-term disorders of consciousness (DOC) are a huge burden on both patients and their families. Previously, music intervention has been attempted as a potential therapy in DOC, with results indicating an enhancement of arousal and awareness; yet, to date, there are limited studies on music interventions in DOC with electroencephalogram monitoring. Meanwhile, prediction of awareness recovery is a challenge facing clinicians. The predictive value mismatch negativity (MMN), as a classical cognitive component in event-related potential, is still controversial. In this study, we use auditory event-related potential to probe the effect of music in DOC, and investigate whether music may improve the predictive value of MMN in awareness recovery.

**Methods:** Fourteen DOC patients were included in the prospective study. Auditory oddball electroencephalogram data were recorded twice with each patient, before and after 5 min of listening to a Chinese symphony that has joyful associations. The outcome was assessed 6 months later.

**Results:** Significant differences of MMN amplitude were found between healthy controls and pre-music DOC patients ($p < 0.001$), but no significant differences were found between healthy controls and post-music DOC patients. The presence of MMN before music was not correlated with favorable outcome, and 50% of patients with MMN did not recover awareness. When MMN was absent, 50% of patients awoke. After listening to music, among the 11 patients who showed MMN, seven patients recovered awareness. When MMN was absent, no one recovered awareness.

**Conclusions:** Some DOC patients, even those in a minimal consciousness state and those with unresponsive wakefulness *syndrome (UWS),* were affected by music. The MMN amplitude was elevated by the music to some extent. A single test of MMN did not have a good prognostic value of our study; however, retesting of MMN after stimulation with familiar music that has joyful associations might be valuable for observation and detection of possible recovery. The musical processing in DOC patients and the effect of musical therapeutic practices need further investigations.

**Keywords: awareness, coma, minimal consciousness state, music, unresponsive wakefulness syndrome (UWS), mismatch negativity (MMN), prediction**

OPEN ACCESS

**Edited by:**
*Franco Delogu, Lawrence Technological University, United States*

**Reviewed by:**
*Wendy L. Magee, Temple University, United States*
*Dandan Zhang, Shenzhen University, China*
*Li Hu, Chinese Academy of Sciences, China*
*Ruolei Gu, Chinese Academy of Sciences, China*
*Rong Ye, University of Cambridge, United Kingdom*

***Correspondence:**
*Kai Wang*
*wangkai1964@126.com*

*†These authors have contributed equally to this work*

**Specialty section:**
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

**Received:** *19 August 2020*
**Accepted:** *19 March 2021*
**Published:** *09 April 2021*

**Citation:**
*Hu Y, Yu F, Wang C, Yan X and Wang K (2021) Can Music Influence Patients With Disorders of Consciousness? An Event-Related Potential Study. Front. Neurosci. 15:596636. doi: 10.3389/fnins.2021.596636*

## BACKGROUND

Craniocerebral trauma, encephalitis, ischemic-hypoxic encephalopathy, and cerebrovascular disease can lead to disorders of consciousness (DOC). The prognosis of these unconscious patients is varied. Some recovered awareness, while others moved into a minimally consciousness state (MCS) (Giacino et al., 2002), even remained in a long-term coma or vegetative state (VS) until death, a term known as unresponsive wakefulness *syndrome* (UWS). Whether and how these individuals are able to regain consciousness is the most important concerns of the patients' families and clinicians. Evaluation by clinical manifestations alone is usually not sufficient to diagnose consciousness, and the rate of misdiagnosis has been estimated at 37–43% (Hirschberg and Giacino, 2011). In particular, neuroelectrophysiology and neuroimaging technology has shown that clinically diagnosed VS patients retain cortical processing ability. Some studies observed that several MCS or VS patients could answer 'yes' and 'no' by using activation of different brain regions with the aid of functional magnetic resonance imaging (fMRI) (Owen et al., 2006).

It is noted that the auditory network is one of the reliably observed networks, help to differentiate MCS from VS (Demertzi et al., 2015). Since language disorders were prevalent in people with DOC, language-based assessments and treatments for recovering consciousness and cognition were not sufficient (Schnakers et al., 2014). Acoustic stimuli with characteristics of self-related, personal preferences and emotional valence showed more likely to change with neural activity and behavior (di Stefano et al., 2012; O'Kelly et al., 2013). Thus, the optimal auditory stimuli for assessment and treatment in patients with DOC tend to hold these properties (Magee, 2018). Among the auditory stimuli, music possesses both self-referential properties and emotional valence and can also bypass language (Magee, 2018).

Listening to music has been shown in neuroimaging studies to induce a vast bilateral network of brain activation associated with reward systems, emotions, semantic processing, motor function, and attention in normal subjects (Warren, 2008; Koelsch, 2014) and to influence mood and arousal (Gabriela Husain et al., 2002). Brain activation due to music has also been detected by fMRI in DOC patients (Okumura et al., 2014; Heine et al., 2015), even when no behavioral evidence of language recognition is present (Edlow et al., 2017).

It has been noted that familiar patient preferred music could enhance functional connectivity in people with DOC compared with the white noise condition (Perrin et al., 2015). Brain fMRI data revealed familiar music triggered broad emotion-related areas and reward circuit, showing significant more active than induced by unfamiliar music (Pereira et al., 2011). With familiar objects in an enriched environment, a wider range of behavioral responses was elicited in people with DOC (di Stefano et al., 2012). When some DOC patients listened to familiar preferred music, the EEG power spectrum of frontal midline theta and frontal alpha, as well as respiratory rate and eye blink rate, increased to some extent indicating that there were arousal and selective attention responses (O'Kelly et al., 2013).

Listening to music can also induce long-term plastic changes for the recovery of early processing (Särkämö et al., 2010). After 2 months period of listening to patient preferred music daily, cognitive recovery and mood were enhanced after middle cerebral artery stroke (Särkämö et al., 2008). EEG regional coherence was enhanced on mesocircuit model and thalamocortical synchronization after a 6–18 months music therapy in a DOC case study, which may be an indicator of an increase in consciousness (Lord and Opacka-Juffry, 2016). Thus, listening to familiar patient preferred music is proposed as a beneficial way for people with DOC to improve their perceptual and cognitive abilities and may also be used for prognostic purposes (Magee and O'Kelly, 2015; Perrin et al., 2015). A music therapy assessment tool for awareness in disorders of consciousness (MATADOC) is developed for detecting the awareness and behavior responsiveness to live musical stimuli (Magee et al., 2014). The protocol was a live music based measure delivered by trained board music therapist (Magee et al., 2015). Based on the MATADOC, O'Kelly et al. (2013) reported a VS patient diagnosed by non-music measure improved the assessing score of MCS scale. It seemed that the DOC patients who presented more favorably with music condition would make a better recovery (Magee, 2018). A study with 13 DOC patients found that patients who showed discriminatory event-related potential (ERP) responses to their names after listening to preferred music showed a favorable outcome six months later; meanwhile, no such favorable outcome was found in patients who showed the absence of a discriminative response (Castro et al., 2015).

From studies, music may be effective in assessment and treatment with people with DOC, possibly due to a range of responses on arousal, attention and emotion triggered by music stimuli, which regulates reward pathways and improves neural plasticity (Rollnik and Altenmüller, 2014), thus facilitating awareness of self and environment, irrespective of language, visual, and motor disabilities. To date, however, most of the researches in this area were by neurophysiological and behavior assessment and neuroimaging methods. There is little research on music interventions in DOC patients using EEG. Whether the music can enhance consciousness or cognitive processing even in the absence of behavioral signs in people with DOC need further evidence from EEG.

Event related potentials (ERP) have been used to investigate cognitive processing and cortical learning in DOC patients. It has many advantages, including millisecond (ms) time resolution, high event sensitivity, bedside manipulation, and non-invasive, and had become a valuable clinical investigation tool. Mismatch Negativity (MMN) is an important component of ERP, and a marker of perceptual processing of deviant auditory stimuli representing a form of "primitive intelligence." In auditory stimuli, MMN appears in the fronto-central area after an infrequent change in a repetitive sequence of sounds (Sams et al., 1985). Auditory oddball paradigms were shown to elicit MMN in patients with DOC (Rodriguez et al., 2014; Wang et al., 2018). MMN can also be used to predict the outcome of comatose patients (Daltrozzo et al., 2007; Faugeras et al., 2011). Since how music effects the MMN in DOC patients is still not clear, here, we used auditory ERP to probe the effect of music on the stimulation of specific areas of the brain in patients with DOC.

It is proposed to use music stimuli familiar to DOC patients for evaluation (Royal College of Physicians, 2013), as well as the patient's preferred music which have shown more likely to activate arousal and attention relative to patient disliked music (O'Kelly et al., 2013; Pool and Magee, 2016). Meanwhile, compositional features of music stimuli (i.e., tempo, rhythm, melody, harmony, timbre, and loudness) should be considered (Robb et al., 2011). In this study, we chose a Chinese symphony "Spring Festival Prelude" composed by Huanzhi Li in 1956 as the music stimulus. This piece of music was played repeatedly and widely during the Spring Festival in China, which is one the most familiar music to Chinese people and preferred by them. The orchestral music is in C major and F major, allegro, and is associated with scenes of people's excitement, jubilation, beating drums, singing and dancing in traditional festivals.

We propose that listening to this familiar music with joyful associations can increase attention and arousal in DOC patients, and that the MMN would be enhanced after the music stimuli. The aim of this study was to evaluate MMN performance in these unconscious patients after music exposure and to observe the prognosis.

# PARTICIPANTS AND METHODS

## Participants

### DOC Group

From Feb 2016 to Oct 2019, DOC patients admitted to the Neurological Intensive Care Unit of the First Affiliated Hospital of Anhui Medical University were enrolled in this study. The inclusion criteria were: (1) DOC $\geq$ 14 days of onset and (2) age $\geq$ 18 years. The exclusion criteria were: (1) history of drug abuse; (2) severe coexisting disease (e.g., acute myocardial infarction, liver and kidney failure, heart failure, or terminal cancer) with a limited likelihood of survival; (3) shock (systolic blood pressure $<$ 80 mmHg); (4) abnormal body temperature; (5) epilepsy; (6) damaged or missing cranial bones; (7) sedatives administered in the previous 24 h; and (8) use of an invasive ventilator or non-invasive ventilator. Patients were recorded without sedation for at least 48 h. Among the 17 recordings three were discarded due to low electroencephalogram (EEG) quality. We included 14 DOC patients (six women), aged from 18–70 years (mean 49.6 years) in this study (**Table 1**). The etiology of the DOC was ischemic stroke ($n$ = 5), anoxic encephalopathy ($n$ = 2), acute disseminated encephalomyelitis (ADEM) ($n$ = 2), intracranial hemorrhage ($n$ = 2), Japanese encephalitis ($n$ = 1), hypoglycemic encephalopathy ($n$ = 1), and traumatic brain injury ($n$ = 1). Twelve patients were diagnosed as an MCS, and two patients were defined as in a VS (**Table 1**).

### Control Group

Healthy adults (20 in total: eight women, aged 40.0 $\pm$ 11.6 years) with no history of drug abuse, mental disorders, or history of brain injury were used as controls.

## Methods

### Behavior

Neurologic examinations were conducted immediately before ERP. Two trained neurologists evaluated the patients' state consciousness based on the Coma-Recovery Scale-Revision (CRS-R) (Giacino et al., 2004). A patient who consistently followed instructions correctly, used objects functionally, and used gestures to express "yes" or "no" were defined as aware. MCS patients who showed only minimal levels of behavioral interaction with non-reflex movements, such as visual pursuit or fixation, localization of noxious stimulation, and appropriate emotional response were defined MCS−. A patient who showed reproducible movement to commands, reached for objects, and displayed automatic motor response was defined as MCS+. If the patient showed only clinical signs of unresponsiveness, such as auditory or visual startle, localization to sound or noxious stimulation, flexion withdrawal, abnormal posturing, or no response to the noxious stimulation, that patient was considered to be suffering from UWS, or even worse, was in a coma (Bruno et al., 2011).

We followed each patient for 6 months. The outcomes were categorized as follows: (1) conscious awareness, (2) MCS, (3) UWS or comatose, or (4) deceased. We defined category 1 as favorable outcome and categories 2, 3, and 4 as unfavorable outcome.

## Procedure

This study used a modified oddball task. Standard and deviation trials were included in each block (**Figure 1**), of which 85% were standard trials and 15% were deviation trials (Wijnen et al., 2007). There were three experimental blocks of 200 trials each. The standard stimuli had a tone of 800 Hz and the deviation stimuli had a tone of 1200 Hz. The time difference between the two ears was 0 ms for each stimulus presentation. The procedure used pseudorandom design with at least four standard stimuli between two deviation stimuli. The inter-trial interval was 600 ms. The entire session required about 25 min to complete. The standard trial following the deviation trial was removed to keep the sound group balanced and to ensure habituation of processing standard stimuli. The procedure was presented using E-prime 2.0 software (Psychology *Soft- ware* Tools Inc., Pittsburgh PA, United States).

Electroencephalogram data were recorded twice for each patient. Between the two oddball tasks, patients were allowed to rest for 5 min, and then listened to a Chinese symphony with joyful associations called "Spring Festival Prelude," which is well known and preferred by Chinese people, for 5 min (**Figure 1**). The version is included in the album "Golden China" (2008), and played by the China National Symphony Orchestra. The total duration is 5minutes and 2 seconds. The auditory stimuli were delivered by a binaural headphone with 90 dB sound pressure in oddball task and 60–70 dB in music. Through the inquiry of the patients' close family members, all these patients were familiar with this music, and liked it.

**TABLE 1 |** Patients' characteristics and outcomes.

| Patient | Gender | Age, years | Etiology | Disease duration, days | CRS-R (A-V-M-O-C-Ar) | MMN | | N100 | | 6 months follow-up, CRS-R (A-V-M-O-C-Ar) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Pre-music | Post-music | Pre-music | Post-music | |
| 1 | F | 44 | Stroke | 15 | MCS+ (1, 1, 3, 3, 1, 1) | + | +↑ | + | + | Awake (4, 5, 6, 3, 2, 3) |
| 2 | M | 58 | Acute disseminated encephalomyelitis | 19 | MCS− (0, 2, 2, 0, 0, 1) | - | +↑ | + | + | Awake (4, 5, 6, 3, 2, 3) |
| 3 | M | 43 | Anoxicencephalopathy | 180 | MCS− (2, 2, 1,1, 0, 2) | + | +↑ | + | + | MCS− (2, 2, 1,1, 0, 2) |
| 4 | F | 18 | Japanese encephalitis | 19 | MCS+ (3, 3, 2, 0, 1, 3) | + | +↑ | + | + | Awake (4, 5, 6, 0, 2, 3) |
| 5 | M | 70 | Stroke | 24 | MCS+ (3, 3, 0, 0, 1, 3) | + | +↑ | + | + | Awake (4, 5, 5, 0, 2, 3) |
| 6 | F | 68 | Stroke | 16 | MCS− (2, 3, 2, 0, 1, 2) | + | + | + | + | Awake (3, 5, 6, 1, 2, 3) |
| 7 | M | 30 | Acute disseminated encephalomyelitis | 25 | MCS− (1, 3, 0, 0, 0, 2) | - | +↑ | + | + | Awake (4, 5, 6, 3, 2, 3) |
| 8 | F | 48 | Hypoglycemic encephalopathy | 15 | UWS (1, 1, 1, 0, 0, 2) | - | +↑ | - | + | UWS (1, 1, 1, 0, 0, 2) |
| 9 | M | 30 | Anoxic encephalopathy | 31 | UWS (1, 1, 1, 0, 0, 2) | - | - | - | - | UWS (1, 1, 1, 0, 0, 2) |
| 10 | F | 25 | Intracranial hemorrhage | 150 | MCS− (1, 2, 0, 0, 0, 2) | + | - | + | + | MCS− (1, 3, 2, 0, 0, 2) |
| 11 | M | 60 | Traumatic brain injury | 180 | MCS− (2, 3, 2, 0, 0, 2) | + | +↑ | + | + | MCS− (2, 3, 2, 0, 0, 2) |
| 12 | F | 67 | Stroke | 15 | MCS+ (3, 1, 3, 0, 0, 1) | + | +↑ | + | + | Awake (4, 5, 6, 1, 2, 3) |
| 13 | M | 49 | Anoxic encephalopathy | 530 | MCS− (2, 2, 1,1, 0, 2) | + | - | + | + | MCS− (2, 2, 1, 1, 0, 2) |
| 14 | M | 63 | Stroke | 15 | MCS− (2, 0,3, 1, 0, 1) | + | +↑ | + | + | MCS+ (3, 4, 5, 0, 1, 3) |

*CRS-R, Coma Recovery Scale-Revised; CRS-R subscales: A, auditory function; V, visual function; M, motor function; O, oromotor; C, communication; Ar, arousal; ↑, an increase in MMN amplitude; ↓, a decrease in MMN amplitude.*

## EEG Data Recording

Electroencephalogram data were recorded from 64 tin electrodes placed on the scalp according to the extended International 10/20 system using a Neuroscan recording system (Neuro Scan, Sterling, VA, United States). EEG signals were recorded using a left mastoid electrode as the online reference. All electrode impedances were maintained below 10 kΩ. EEG activities were amplified with 0.01–100 Hz band-pass filtering and continuously sampled at 500 Hz/channel.

MATLAB scripts using functions from the EEGLAB environment were adopted to process and analyze the EEG data (Delorme and Makeig, 2004). The collected data were re-referenced to the average of the left and right mastoids and were down-sampled to 250 Hz. Then, the data were subjected to a high-pass filter at 1 Hz (FIR filter conducted with pop_eegnewfilt with a default parameters and a cutoff frequency of 0.5 Hz and 26 dB, respectively, to remove baseline drift, thereby ensuring reliable results for independent component analysis) (Delorme et al., 2012). Artifactual channels and non-brain electrodes were rejected by the clean_raw data plugin in EEGLAB, leaving an average of 58.52 (95%, [29, 61]) clean channels per participant. Continuous data were filtered and segmented from 1000 ms before the go and stop signal to 2000 ms after the stimulus. Artifactual epochs were identified and removed based on: (a) abnormal spectral characteristics of high frequency noise (rejspec; 20–40; $< -35$ or $> 35$ dB); (b) abnormal trends (rejtrend; slope $> 200$ μV with R2 $> 0.3$); (c) abnormal amplitude (threshold $-500$ μV or $+ 500$ μV); (d) improbable data using joint probability (jointprob, 8 standard deviations (SD) for single channel and 4 SD for all channels); or (e) abnormal distributions (rejkurt; 8 SD for single channel and 4 SD for all channels). Data from electrodes responsible for more than 10% of rejected epochs were discarded. Subsequently, epoch data were decomposed into maximally independent components using an extended infomax algorithm implemented by the runica function with default parameters. Artifactual components electrocardiogram and electromyogram were identified and removed by the EEG_SASICA plugin and visual inspection in EEGLAB (Chaumon et al., 2015). On average, there were 53.56 (95%, [52.8, 54.33]) components left per participant. The mean proportion of rejected epochs was 4.66 (95%, [3.22, 6.09]) in the healthy control group, 5.0 (95%, [2.12, 7.88]) in the pre-DOC group (the DOC group prior to listening to music), and 5.92 (95%, [1.64, 10.21]) in the post-DOC group (the DOC group after listening to music). Rejection rates did not differ significantly among the groups ($F_{2,47}$ = 0.26, $p$ = 0.77).The cleaned ERP waveforms were time-locked to stimulus onset and epoched to 200 ms pre-stimulus and 1000 ms post-stimulus. The ERPs were averaged separately for standard and deviation trials. N100 was defined as a negative wave between 70–130ms, preceded by a positive deflection P50.

When standard and deviant N100 were detected, the MMN was defined as a negative difference with an average amplitude

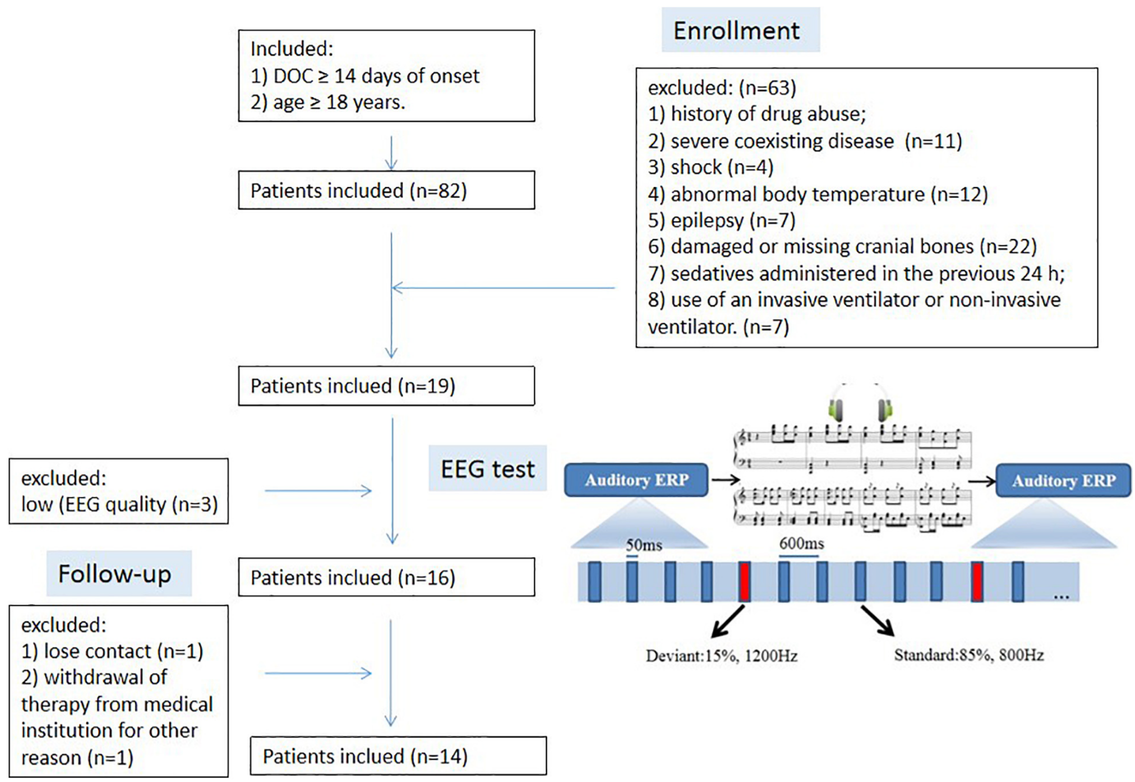



**FIGURE 1** | Flow chart of the study.

between 100 and 200 ms. The discriminative MMN wave was accepted when the amplitude was greater than 0.75 uV (Fischer et al., 1999). P3 was defined as the average amplitude between 250-350ms. ERPs data were extracted from the frontal area with F3, FZ, F4, FC3, FCZ, and FC4 electrodes, central area with C3, CZ, and C4 electrodes, and central-parietal area with CP3, CPZ, CP4, P3, PZ and P4 electrodes.

## Statistical Analysis

Quantitative data were presented as mean ± standard errors. Multiple comparisons of MMN amplitude among the groups (control, pre-music DOC, and post-music DOC) were conducted using one-way ANOVA test followed by *post hoc* Dunnett's T3 test. Multiple repeated ANOVA was conducted with scalp area as within-subject factor, group as between-subject factor. Spearman's rank test was calculated to test the correlation between CRS-R total scores and MMN amplitudes. $\chi^2$ test or Fisher's exact test was applied for categorical variables. Data were analyzed using SPSS software, version 17.0 (SPSS Inc., Chicago, IL, United States).

# RESULTS

## Control Group

In healthy controls, discriminative MMN was present in 20/20 (100%) of subjects. The amplitude elicited by deviation trials was significantly higher ($p < 0.05$) than that elicited by standard trials at 100–200 ms in frontal and central areas. In the frontal area, the MMN measured −5.1 ± 1.8 uV with a latency of 123.8 ± 14.2 ms. In central area, the MMN measured −4.5 ± 1.7 uV with a latency of 120.6 ± 14.9 ms. P3 were elicited with the average amplitude 3.19 ± 2.56 uV at frontal area, 3.256 ± 2.05 uV at central-parietal area.

## DOC Group

In the 14 patients (four MCS+, eight MCS− and two with UWS), discriminative MMNs were evoked in ten subjects (four MCS+ and six MCS−) prior to listening to music, and 11 subjects (four MCS+, six MCS−, and one with UWS) after listening to music listening.

Prior to listening to music, the MMN measured −1.3 ± 2.6 uV with a latency of 122.6 ± 24.6 ms in frontal area, and −1.6 ± 2.1 uV with a latency of 125.4 ± 17.7 ms in central area. Meanwhile, the average P3 measured was −0.95 ± 2.29 uV at frontal area and −0.93 ± 2.24 uV at central-parietal area. After listening to music, the MMN measured was −3.6 ± 3.1 uV with a latency of 117.1 ± 23.7 ms in frontal area and −3.6 ± 3.8 uV with a latency of 127.4 ± 26.9 ms in central area. The average P3 was −0.24 ± 2.22 uV at the frontal area and −0.36 ± 3.5 uV at the central-parietal area.

Multiple comparisons of MMN amplitude among the groups (control, pre-music DOC, and post-music DOC) were conducted using one-way ANOVA. Significant differences were found among the groups ($p < 0.001$). A *post hoc* Dunnett's T3 test showed significant differences between healthy controls and pre-music DOC patients ($p < 0.001$); however, no statistically significant differences was found between healthy controls and post-music DOC patients, or between pre-music DOC patients and post-music DOC patients (**Figure 2**). Multiple repeated ANOVA was conducted with scalp area (frontal or central) as within-subject factor, group (before or after music) as between-subject factor. There was no within-subjects effect ($F = 0.19$, $p > 0.05$), but the effect between groups was significant ($F = 7.97$, $p < 0.05$), which indicated that music intervention has an influence on MMN in DOC patients. Using the same methods for P3 amplitude analysis, significant difference was found between healthy control and pre-music DOC patients ($p < 0.05$), as well as between healthy control and post-music DOC patients ($p < 0.05$). The P3 amplitude had no significant difference between pre-music DOC patients and post-music DOC patients ($p > 0.05$). No within-subject effect or between-group effect was found by multiple repeated ANOVA ($p > 0.05$).

There was no association between total CRS-R scores and the amplitude of MMN, both prior to and after listening to music ($p > 0.05$). The correlation between the changes of MMN amplitudes (before and after music) with CRS-R total scores did not reach the significance ($p = 0.15$). The presence of MMN before music was not linked to favorable patient outcome at the 6 months follow-up (Fisher exact test, $p > 0.05$), with sensitivity of 71%, specificity of 28.6%, a positive predictive value of 50%, and negative predictive value of 50%. Before music, when MMN was present, 5/10 awoke. When MMN was absent, 2/4 awoke. After listening to music, among the 11 patients who showed MMN after listening to music, seven recovered awareness. When the MMN was absent, 3/3 (100%) did not recover awareness. The total predictive value of MMN after the music did not reach to the statistical significance (Fisher's exact test, $p > 0.05$), but with a sensitivity of 100%, and specificity of 42.9%. Positive predictive value was 63.6% and a negative predictive value was 100%. We further used chi-square to evaluate the value of the presence of N100, as well as the combination of MMN and N100 for prognosis. However, both of them did not reach to significant value in the prognosis (Fisher's exact test, $p > 0.05$). Meanwhile, no statistical relationship was established between the presence of increased MMN and a favorable outcome that recovered conscious awareness (Fisher's exact test, $p > 0.05$). In this study, nine patients (9/14) had increased MMN after listening to music, six of which (6/9) recovered conscious awareness; however, five patients (5/14) showed no increases in MMN, four of which had an unfavorable outcome.

## DISCUSSION

Mismatch negativity is the earliest cognitive component in an ERP trace. MMN presents a negative deflection on frontal and central areas, with a latency of between 100 and 250 ms after deviance onset (Sams et al., 1985). MMN links to fresh-afferent neuronal activity, reflecting the brain's ability to determine automatic comparisons (Garrido et al., 2009). The generation of MMN might involve an index of memory and an involuntary attention switching process (Garrido et al., 2009). Furthermore, MMN can be elicited below the level of attention, even in sleep and coma.

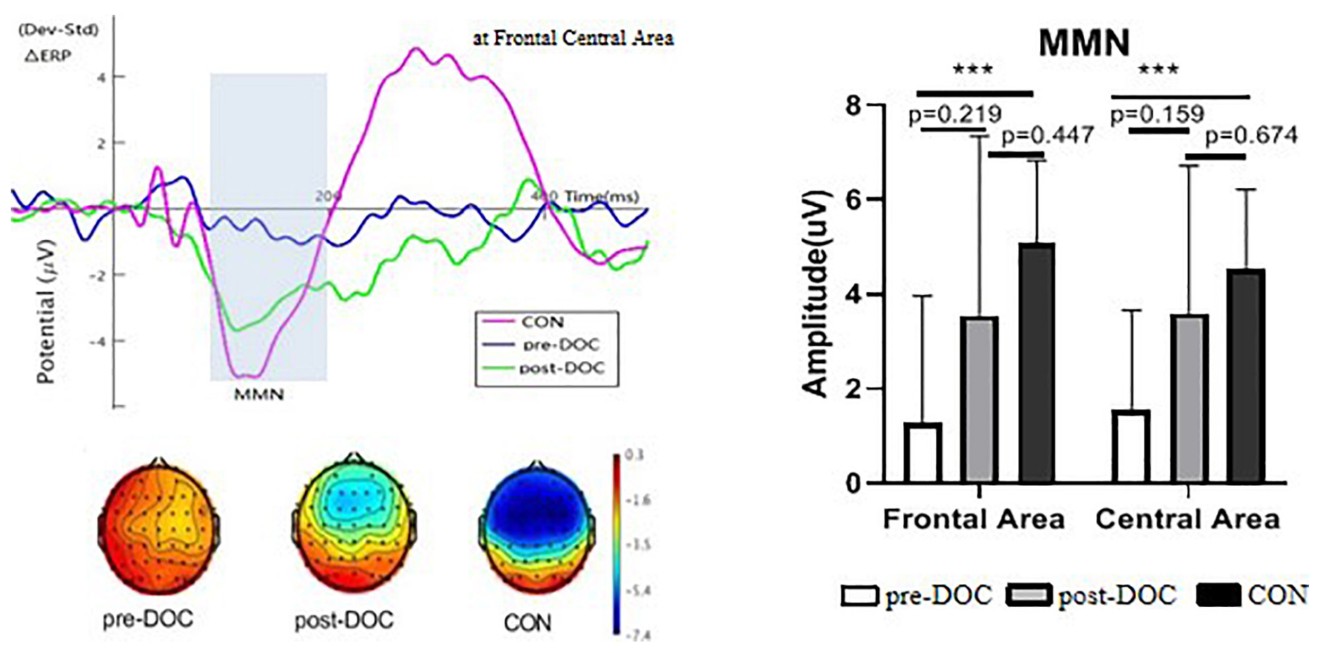



**FIGURE 2 |** Mismatch negativity response with groups. Data (mean ± SEW) are shown separately for DOC patients, before and after music and controls. ***$p < 0.005$. Pre-DOC, DOC patents before music; post-DOC, DOC patients after music; CON, healthy controls.

It has been shown that the frontal cortex and parietal cortex are involved in conscious perception (Del Cul et al., 2007; Koch et al., 2016). DOC patients who elicit an MNN are considered to be in a semi-conscious state, and deficits in MMN are associated with DOC (Faugeras et al., 2011). In our study, MMN was found to be elicited in the frontal and central areas. Meanwhile, the average amplitude of MMN in DOC patients was lower than healthy controls, although no association was found between total CRS-R scores and the amplitude of MMN.

Several studies have indicated that music has been shown to elicit neural responses, altering cortical activity and connectivity both in healthy subjects and clinically "unresponsive" patients (Wilkins et al., 2014). Music therapy appears to be a promising clinical tool for rehabilitation of DOC patients, as well as dementia patients (Sihvonen et al., 2017). Music's features with patients, such as autobiographical re-experiencing, familiarity, and preference, would have beneficial effects (O'Kelly et al., 2013; Castro et al., 2015). In this study, we chose a Chinese symphony with known joyful associations called "Spring Festival Prelude" which was familiar to the subjects and liked by them. The musical excerpt was selected based on the musical stimuli using in prior studies with properties of familiar, preferred, and emotional valence. And it also meets the characteristics of frequent changes in tempo, dynamic, musically coherent and representative of the whole musical piece (Castro et al., 2015), such as Aaron Copland's "Rodeo – Four Dance Episodes" and the first 16 measures of "Les Toreador" from "Carmen" Suite No.1 by Bizet that used in previous studies (Danielsen et al., 2014; Okumura et al., 2014).

The MMN amplitude was also elevated by music to some extent. Nine of the 14 (64%) patients had an increased MMN amplitude after listening to music. There was a statistically significant difference of MMN amplitude between DOC patients before listening to music and healthy controls. But after listening to music, no statistical difference existed between DOC patients and healthy controls. Analysis by multiple repeated ANOVA also indicated that music intervention has an influence on MMN in DOC patients. We therefore considered that many DOC patients, even those who were MCS− and those suffering from UWS, were affected by listening to the music, which in turn improved MMN—a factor considered to be a marker of consciousness involved in acoustic discrimination and sensory memory. As for the improved, MMN is a basis of the emerging conscious perception and awareness, it is rational to use music-intervention for promoting consciousness recovery.

P300 is an endogenous component of ERP, which is related to the cognitive process, attention, memory, intelligence and mental state of the brain. But the sensitivity of the P300 elicited in a short two-tone oddball paradigm was too low in VS and MCS patients for application in a clinical setting (Real et al., 2016). Similar to the previous study, P300 did not elicit well in our patients with DOC when using the two pure tones oddball paradigm, which limited the use of P300 for prognosis. While complex sensory tones, such as subject's own name, familiar sound could induce larger P300 and had predictive value in DOC patients (Daltrozzo et al., 2007) and could be used in a further study.

Mismatch negativity had been used in the prognosis of awareness. Despite having good specificity, MMN was found to have a poor sensitivity and detection rate. A meta-analysis of MMN for prediction of awareness showed a sensitivity of 38%, specificity of 91%, positive predictive value of 88%, negative predictive value of 46% (Daltrozzo et al., 2007). Fisher et al. studied 128 comatose patients, and on the 8th day after average coma, only 33 patients elicited MMN. Thirty of the 95 patients who recovered awareness showed an MMN (31.6% sensitivity) (Fischer et al., 1999). A study of 346 comatose patients found that when MMN was present (25.4%), 88.6% patients awoke, and when it was absent, 62.4% patients did not (Fischer et al., 2004).

In our study, 10 patients who elicited an MMN, half regained awareness prior to the 6 months follow-up, while two of the four the patients that did not elicit an MMN recovered their awareness. In regard to prognostic value, the presence of MMN before listening to music was not correlated with favorable patient outcome, with sensitivity of 71%, specificity of 28.6%, a positive predictive value of 50%, and negative predictive value of 50%. The specificity of the baseline MMN is relatively low compared with the high specificity validated in previous studies in literature. We noted that the duration of disease and the etiology to the coma may confound the results. Within the five patients who did not recover awareness but had MMN, four of them had the EEG test exceeding 6 months after the onset of the coma. Both of the two patients who had no MMN but then regained consciousness were not old and diagnosed as ADEM with predominant white matter damages, which is generally better in prognosis than other causes. Meanwhile, the time of the EEG test in these two ADEM patients was less than one month after onset.

After listening to music, among the 11 patients who elicited an MMN, seven patients recovered awareness prior to the 6 months follow-up. However, three patients who did not elicit MMN after listening to music remained unconscious at 6 months follow-up. We also noted that there were three patients who recovered awareness that did not initially elicit MMN, but did so after listening to music. There were still two patients who showed MMN first but did not induce MMN after the music, and they did not regain consciousness. As a result, the prognostic value of MMN after music seemed to increase than before music, with a sensitivity of 100% vs. 71%, specificity of 42.9% vs. 28.6%, positive predictive value of 63.6% vs. 50%, and a negative predictive value 100% vs. 50%. Thus, retesting of MMN after stimulation with familiar music that has joyful associations may be a good attempt for the observation and detection of possible recovery. However, increased MMN after the music did not prove to be linked with favorable outcome ($p = 0.19$).

There are some limitations of our study. The sample size used in the present study was small, which limited generalization of the conclusions. Many factors may be the confounds of the study, such as the cause of the initial disease, the injured region, a patient's age, disease duration, the delays between the EEG test and the onset of the coma, which should be taken into account in future studies. The MMN we used is far from being a proof of the subclinical consciousness, and we did not elicit P300 well. The auditory oddball paradigm we used need to be modified, and other paradigms such as novelty P3 elicited by the subject's own name, and motor imagery tasks can be used further.

## CONCLUSION

In conclusion, we considered that many DOC patients, even those who were found to be MCS− and who suffered from UWS, could be positively affected by music. In particular, the MMN amplitude was elevated by the music to some extent, so it is reasonable to use music-intervention for promoting the recovery of consciousness. A single test of MMN did not have a good prognostic value in our study; however, retesting of MMN after stimulation with familiar music that has joyful associations may be valuable for observation and detection of possible recovery.

The musical processing in DOC patients and the effect of musical therapeutic practices need further investigation. Larger samples are needed to prove the predictive value of MMN after music with specific etiology of DOC in a long-term clinical follow-up.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The First Affiliated Hospital of Anhui Medical University, Hefei. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

YH and FY designed the experiments. YH, XY, and CW carried out the experiments. FY and YH analyzed the experimental results. YH and FY wrote the manuscript. KW contributed to the manuscript writing, review, and supervision. All authors contributed to the article and approved the submitted version.

## FUNDING

This study was supported by the Key University Natural Science Research Project of Anhui Province (KJ2019A0291) and the National Science & Technology Fundamental Resources Investigation Program of China (2018FY100900).

## ACKNOWLEDGMENTS

We express our gratitude to the patients, controls, and their families for the participation, and thank the staff of the Department of Neurology of Anhui Medical University for their collaboration. We thank International Science Editing (http://www.internationalscienceediting.com) for editing this manuscript.

## REFERENCES

Bruno, M. A., Vanhaudenhuyse, A., Thibaut, A., Moonen, G., and Laureys, S. (2011). From unresponsive wakefulness to minimally conscious PLUS and functional locked-in syndromes: recent advances in our understanding of disorders of consciousness. *J. Neurol.* 258, 1373–1384. doi: 10.1007/s00415-011-6114-x

Castro, M., Tillmann, B., Luauté, J., Corneyllie, A., Dailler, F., André-Obadia, N., et al. (2015). Boosting cognition with music in patients with disorders of consciousness. *Neurorehabil. Neural Repair* 29, 734–742. doi: 10.1177/1545968314565464

Chaumon, M., Bishop, D. V. M., and Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *J. Neurosci. Methods* 250, 47–63. doi: 10.1016/j.jneumeth.2015.02.025

Daltrozzo, J., Wioland, N., Mutschler, V., and Kotchoubey, B. (2007). Predicting coma and other low responsive patients outcome using event-related brain potentials: a meta-analysis. *Clin. Neurophysiol.* 118, 606–614. doi: 10.1016/j.clinph.2006.11.019

Danielsen, A., Otnæss, M. K., Jensen, J., Williams, S. C. R., and Ostberg, B. C. (2014). Investigating repetition and change in musical rhythm by functional MRI. *Neuroscience* 275, 469–476. doi: 10.1016/j.neuroscience.2014.06.029

Del Cul, A., Baillet, S., and Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol.* 5:e260. doi: 10.1371/journal.pbio.0050260

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., and Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS One* 7:e30135. doi: 10.1371/journal.pone.0030135

Demertzi, A., Antonopoulos, G., Heine, L., Voss, H. U., Crone, J. S., De Los Angeles, C., et al. (2015). Intrinsic functional connectivity differentiates minimally conscious from unresponsive patients. *Brain* 138, 2619–2631. doi: 10.1093/brain/awv169

di Stefano, C., Cortesi, A., Masotti, S., Simoncini, L., and Piperno, R. (2012). Increased behavioural responsiveness with complex stimulation in VS and MCS: preliminary results. *Brain Inj.* 26, 1250–1256. doi: 10.3109/02699052.2012.667588

Edlow, B. L., Chatelle, C., Spencer, C. A., Chu, C. J., and Wu, O. (2017). Early detection of consciousness in patients with acute severe traumatic brain injury. *Brain* 140, 2399–2414. doi: 10.1093/brain/awx176

Faugeras, F., Rohaut, B., Weiss, N., Bekinschtein, T. A., Galanaud, D., Puybasset, L., et al. (2011). Probing consciousness with event-related potentials in the vegetative state. *Neurology* 77, 264–268. doi: 10.1212/wnl.0b013e3182217ee8

Fischer, C., Luauté, J., Adeleine, P., and Morlet, D. (2004). Predictive value of sensory and cognitive evoked potentials for awakening from coma. *Neurology* 63, 669–673. doi: 10.1212/01.wnl.0000134670.10384.e2

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., and Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clin. Neurophysiol.* 110, 1601–1610. doi: 10.1016/s1388-2457(99)00131-5

Garrido, M. I., Kilner, J. M., Stephan, K. E., and Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* 120, 453–463. doi: 10.1016/j.clinph.2008.11.029

Giacino, J. T., Ashwal, S., Childs, N., Cranford, R., Jennett, B., Katz, D. I., et al. (2002). The minimally conscious state: definition and diagnostic criteria. *Neurology* 58, 349–353. doi: 10.1212/wnl.58.3.349

Giacino, J. T., Kathleen, K., and John, W. (2004). The JFK coma recovery scale-revised: measurement characteristics and diagnostic utility. *Arch. Phys. Med. Rehabil.* 85, 2020–2029. doi: 10.1016/j.apmr.2004.02.033

Heine, L., Castro, M., Martial, C., Tillmann, B., Laureys, S., and Perrin, F. (2015). Exploration of functional connectivity during preferred music stimulation in patients with disorders of consciousness. *Front. Psychol.* 6:1704. doi: 10.3389/fpsyg.2015.01704

Hirschberg, R., and Giacino, J. T. (2011). The vegetative and minimally conscious states: diagnosis, prognosis and treatment. *Neurol. Clin.* 29, 773–786. doi: 10.1016/j.ncl.2011.07.009

Husain, G., Thompson, W. F., and Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Percept.* 20, 151–171. doi: 10.1525/mp.2002.20.2.151

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321.

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666

Lord, V., and Opacka-Juffry, J. (2016). Electroencephalography (EEG) measures of neural connectivity in the assessment of brain responses to salient auditory stimuli in patients with disorders of consciousness. *Front. Psychol.* 7:397. doi: 10.3389/fpsyg.2016.00397

Magee, W. L. (2018). Music in the diagnosis, treatment and prognosis of people with prolonged disorders of consciousness. *Neuropsychol. Rehabil.* 28, 1331–1339. doi: 10.1080/09602011.2018.1494003

Magee, W. L., Ghetti, C. M., and Moyer, A. (2015). Feasibility of the music therapy assessment tool for awareness in disorders of consciousness (MATADOC) for use with pediatric populations. *Front. Psychol.* 6:698. doi: 10.3389/fpsyg.2015.00698

Magee, W. L., and O'Kelly, J. (2015). Music therapy with disorders of consciousness: current evidence and emergent evidence-based practice. *Ann. N. Y. Acad. Sci.* 1337, 256–262. doi: 10.1111/nyas.12633

Magee, W. L., Siegert, R. J., Daveson, B. A., Lenton-Smith, G., and Taylor, S. M. (2014). Music therapy assessment tool for awareness in disorders of consciousness (MATADOC): standardisation of the principal subscale to assess awareness in patients with disorders of consciousness. *Neuropsychol. Rehabil.* 24, 101–124. doi: 10.1080/09602011.2013.844174

O'Kelly, J., James, L., Palaniappan, R., Taborin, J., Fachner, J., and Magee, W. L. (2013). Neurophysiological and behavioral responses to music therapy in vegetative and minimally conscious states. *Front. Hum. Neurosci.* 7:884. doi: 10.3389/fnhum.2013.00884

Okumura, Y., Asano, Y., Takenaka, S., Fukuyama, S., Yonezawa, S., Kasuya, Y., et al. (2014). Brain activation by music in patients in a vegetative or minimally conscious state following diffuse brain injury. *Brain Inj.* 28, 944–950. doi: 10.3109/02699052.2014.888477

Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., and Pickard, J. D. (2006). Detecting awareness in the vegetative state. *Science* 313:1402. doi: 10.1126/science.1130197

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., and Brattico, E. (2011). Music and emotions in the brain: familiarity matters. *PLoS One* 6:e27241. doi: 10.1371/journal.pone.0027241

Perrin, F., Castro, M., Tillmann, B., and Luauté, J. (2015). Promoting the use of personally relevant stimuli for investigating patients with disorders of consciousness. *Front. Psychol.* 6:1102. doi: 10.3389/fpsyg.2015.01102

Pool, J., and Magee, W. L. (2016). Music in the treatment of children and youth with prolonged disorders of consciousness. *Front. Psychol.* 7:202. doi: 10.3389/fpsyg.2016.00202

Real, R. G. L., Veser, S., Erlbeck, H., Risetti, M., Vogel, D., Müller, F., et al. (2016). Information processing in patients in vegetative and minimally conscious states. *Clin. Neurophysiol.* 127, 1395–1402.

Robb, S. L., Burns, D. S., and Carpenter, J. S. (2011). Reporting guidelines for music-based interventions. *J. Health Psychol.* 16, 342–352. doi: 10.1177/1359105310374781

Rodriguez, R. A., Bussière, M., Froeschl, M., and Nathan, H. J. (2014). Auditory-evoked potentials during coma: do they improve our prediction of awakening in comatose patients? *J. Crit. Care* 29, 93–100. doi: 10.1016/j.jcrc.2013.08.020

Rollnik, J. D., and Altenmüller, E. (2014). Music in disorders of consciousness. *Front. Neurosci.* 8:190. doi: 10.3389/fnins.2014.00190

Royal College of Physicians (2013). *Prolonged Disorders of Consciousness. National Clinical Guidelines*. London: RCP.

Sams, M., Paavilainen, P., Alho, K., and Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalogr. Clin. Neurophysiol.* 62, 437–448. doi: 10.1016/0168-5597(85)90054-1

Särkämö, T., Pihko, E., Laitinen, S., Forsblom, A., Soinila, S., Mikkonen, M., et al. (2010). Music and speech listening enhance the recovery of early sensory processing after stroke. *J. Cogn. Neurosci.* 22, 2716–2727. doi: 10.1162/jocn.2009.21376

Särkämö, T., Tervaniemi, M., Laitinen, S., Forsblom, A., Soinila, S., Mikkonen, M., et al. (2008). Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain* 131, 866–876. doi: 10.1093/brain/awn013

Schnakers, C., Bessou, H., Rubi-Fessen, I., Hartmann, A., Fink, G. R., Meister, I., et al. (2014). Impact of aphasia on consciousness assessment: a cross-sectional study. *Neurorehabil. Neural Repair* 29, 41–47. doi: 10.1177/1545968314528067

Sihvonen, A. J., Särkämö, T., Leo, V., Tervaniemi, M., Altenmüller, E., and Soinila, S. (2017). Music-based interventions in neurological rehabilitation. *Lancet Neurol.* 16, 648–660. doi: 10.1016/s1474-4422(17)30168-0

Wang, X., Fu, R., Xia, X., Chen, X., Wu, H., Landi, N., et al. (2018). Spatial properties of mismatch negativity in patients with disorders of consciousness. *Neurosci. Bull.* 34, 700–708. doi: 10.1007/s12264-018-0260-4

Warren, J. (2008). How does the brain process music? *Clin. Med. (Lond)* 8, 32–36. doi: 10.7861/clinmedicine.8-1-32

Wijnen, V. J. M., Van Boxtel, G., Eilander, H., and Gelder, B. (2007). Mismatch negativity predicts recovery from the vegetative state. *Clin. Neurophysiol.* 118, 597–605. doi: 10.1016/j.clinph.2006.11.020

Wilkins, R. W., Hodges, D. A., Laurienti, P. J., Steen, M., and Burdette, J. H. (2014). Network science and the effects of music preference on functional brain connectivity: from Beethoven to Eminem. *Sci. Rep.* 4, 6130.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 21 July 2021
doi: 10.3389/fnins.2021.663015

# Chinese and Western Musical Training Impacts the Circuit in Auditory and Reward Systems

**Sijia Guo[1,2†], Ke Peng[3†], Rui Ding[2], Junchen Zhou[2], Yan Liu[2], Yao He[2], Yuhong Liu[2], Ke Li[4], Pei Liu[5], Cheng Luo[1,2], Jing Lu[1,2*] and Dezhong Yao[1,2*]**

[1] The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Laboratory for Neuroinformation, University of Electronic Science and Technology of China, Chengdu, China, [2] Center for Information in Medicine, School of Life Sciences and Technology, University of Electronic Science and Technology of China, Chengdu, China, [3] School of Music Education, Xinghai Conservatory of Music, Guangzhou, China, [4] Department of Imaging, The 306th Hospital of the People's Liberation Army, Beijing, China, [5] Department of Music Education, China Conservatory of Music, Beijing, China

OPEN ACCESS

**Edited by:**
Cunmei Jiang,
Shanghai Normal University, China

**Reviewed by:**
Marina Kliuchko,
Technical University of Denmark,
Denmark
Yin Tian,
Chongqing University of Posts
and Telecommunications, China

***Correspondence:**
Jing Lu
lujing@uestc.edu.cn
Dezhong Yao
dyao@uestc.edu.cn

† These authors have contributed equally to this work

**Specialty section:**
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

**Received:** 02 February 2021
**Accepted:** 29 June 2021
**Published:** 21 July 2021

**Citation:**
Guo S, Peng K, Ding R, Zhou J, Liu Y, He Y, Liu Y, Li K, Liu P, Luo C, Lu J and Yao D (2021) Chinese and Western Musical Training Impacts the Circuit in Auditory and Reward Systems. Front. Neurosci. 15:663015. doi: 10.3389/fnins.2021.663015

Previous studies have provided evidence about the brain plasticity effects of musical training, however, the issue of how expertise in music styles induced by Chinese or Western musical training affects neuroplasticity and reward responses has been less considered, especially for subjects of Chinese origin. In this work, 16 musicians who trained in the Western music style (Western-trained musicians) and 18 musicians who trained in the Chinese music style (Chinese-trained musicians) were recruited as the musician group for the experiment, while 15 non-musicians were recruited as the control group. Using a paradigm that consisted of listening to Chinese and Western music and measurements using functional magnetic resonance imaging (fMRI) technology, we found that Chinese-trained musicians activated the bilateral superior temporal gyrus (STG) when listening to music, while Western-trained musicians activated the left STG. In addition, under the condition of listening to music with Chinese style, Chinese-trained musicians have a stronger functional connection in the circuit of the auditory and reward system than Western-trained musicians. The finding is opposite under the condition of listening to music with Western style. Interestingly, it seems that the circuit of Chinese-trained musicians is partial to the right STG, while Western-trained musicians show the opposite, i.e., a tendency toward the left STG. The influence of different music styles on experienced musicians is reflected by the functional activities and connections between the auditory system and the reward system. This outcome indicates that training in Chinese music style or Western music style affects the strategies of musicians when listening to music. Musical characteristics such as rhythm, melody and cultural attributes play an important role in this process. These findings, which provide evidence for functional neuroplasticity based on musical training, can enrich our insights into the musical brain.

**Keywords: musical training, expertise, musician, music styles, auditory system, reward system, fMRI**

## INTRODUCTION

Brain plasticity under the influence of music has been a focus of scholars due to its great potential influence on shaping our cognition. Musical training during childhood and adolescence can affect neural development and enhance abilities of speech-in-noise processing, language skills, etc (Rüsseler et al., 2001; Tervaniemi et al., 2001; Strait et al., 2012; Tierney et al., 2015; Lu et al., 2017). The benefits of musical training have effects across lifespans (Strait and Kraus, 2014). Musical training experience affects neural development (Elbert et al., 1995; Lu et al., 2017; Li et al., 2018), with influencing factors that include the age at which training begins (Ohnishi et al., 2001) and the instruments musicians study (Hund-Georgiadis and von Cramon, 1999; Pantev et al., 2001). Scientists have found that experienced musicians exhibit enhanced auditory cortical representations for musical training associated with their principal instrument (Pantev et al., 2001). Therefore, different instrumental trainings have different influences on musical processing, which influence our cognition and perception of musical components (Drake and El Heni, 2003; Bouwer et al., 2018). In addition, another reason for the difference in neural plasticity between musicians is that musicians listen to music with different styles when they are training different musical instruments. For example, Chinese music style and Western music style; the musical rhythms of Western classical music are considered highly regular and predictable (Levitin et al., 2012). Specifically, majority of Western music has regular beat levels, which include even time divisions and time intervals composed of simple ratios of 1:1 and 2:1 (Stevens, 2012). Beat is a fundamental feature of Western music and that the prominent characteristic of Western music is the regular subdivision of the beat. Typical Western music's primary beat level has one or two levels of subdivision (London, 2012). Different from Western music style, music from Asian regions has irregular beat attributes (Stevens, 2012). The duration of one beat of Chinese music is sometimes based on one monosyllabic Chinese character. Under this circumstance, the instrumental performance of Chinese music usually corresponds to an irregular stress duration (Cao and Li, 2017). Chinese music style pays more attention to melody compared with Western music style. The application of melodic techniques such as atonality, microtones, pantonality and polytonality can be seen in its composition (Jiang, 1991).

Study indicates that music is culturally universal and culture-specific (Morrison and Demorest, 2009). Exposure to the particular melodic, harmonic, rhythmic, or timbral features of the surrounding musical culture leads to the enhancement of musical abilities, similar to children learning language (Morrison et al., 2003). People who have received musical instrument training since childhood can further enhance their musical abilities corresponding to the characteristics (rhythm, melody, etc.) of music style that they are exposed to and practice during daily training. Experienced musicians who have received Chinese or Western musical instrument training can distinguish the difference between Chinese and Western music styles based on their professional knowledge and abilities. Therefore, we suggest that both Chinese-trained musicians and Western-trained musicians have a greater understanding of the musical styles they are exposed to during their daily training than non-musicians and musicians who are trained in other types of music.

Previous neurological and pharmacological studies have shown that the reward system activates when subjects listen to music, and dopamine is the physiological basis of the reward response (Salimpoor et al., 2011; Raghanti et al., 2016; Ferreri et al., 2019). This process is related to the pleasure experienced when listening to music. Research has indicated that expectations can be generated by veridical (characteristic-specific) knowledge about a familiar genre. The predictability of musical characteristics possibly contributes to our enjoyment of music (Levitin et al., 2012). Furthermore, the reward responses are different when subjects are familiar or unfamiliar with the music (Freitas et al., 2018). Thus, the understanding of musical styles probably relates to the pleasure experienced in reward processing and the relatively dominant music characteristics in the music style may have a greater contribution in this process. The brain regions involved in this process include midbrain, abdominal and back lateral striatum, amygdala, prefrontal lobe, cingulate gyrus, etc. (Pantev et al., 1998; Koelsch et al., 1999; Blood and Zatorre, 2001).

Researchers have explored the difference in the neural activation among subjects from different cultures when they encounter cultural familiar/unfamiliar music. For example, Western musicians mainly activated bilateral motor regions in the culturally familiar condition and activated a right lateralized network of angular gyrus and the middle frontal gyrus in the culturally unfamiliar condition (Nan et al., 2006). However, musical interaction related to culture affects human responses to music during infancy (Morrison and Demorest, 2009). In this period, musical abilities are related to attention, memory, and the acquisition of musical information develop (Trehub, 2001). In addition, language learning has a transfer effect to music learning, and native speakers of different languages have different musical abilities (Zhang et al., 2020). Thus, recruiting subjects from different cultures may be disadvantageous to our study of the influence of training in Chinese and Western music styles on neural plasticity. At present, majority of the research on neural plasticity under musical training has been conducted on subjects with Western cultural backgrounds. It is necessary to study neural plasticity under musical training which conducted on subjects coming from other musical traditions as well. Therefore, the current study places its focus on musicians of Chinese origin, who were trained in either Western or Chinese music style.

In summary, we assume that different musical styles which musicians are exposed to during their daily training, have different influences on brain plasticity. Perhaps music styles affect the strategies of musicians when they listen to music. Additionally, in the process of listening to music with different styles, their brain areas related to the processing of musical characteristics have different connections with the reward system. Herein, we plan to recruit Chinese subjects to study the neural plasticity of musicians exposed to the corresponding musical styles due to long-term musical training.

**TABLE 1** | Characteristics of the participants.

| Participants | Age (mean ± SD) | Age of musical acquisition (mean ± SD) | Musical training hours per week (h) | Cultural background |
|---|---|---|---|---|
| Non-musician group | 21.3 ± 2.0 | No musical learning experience | Not applicable | Chinese |
| Western-trained musician group | 20.3 ± 4.1 | 7.7 ± 2.9 | 21.4 | Chinese |
| Chinese-trained musician group | 20.5 ± 3.1 | 6.3 ± 2.1 | 17.2 | Chinese |

# MATERIALS AND METHODS

## Participants

In this experiment, 49 subjects of Chinese origin were recruited among undergraduate and Master's degree candidates at different universities. Background information, such as starting age, training years, weekly training hours, etc., were obtained using the Montreal Music History Questionnaire (MMHQ) (Coffey et al., 2011). Eighteen musicians who played the guqin or erhu were selected as the Chinese-trained musician group; sixteen musicians who played the piano were selected as the Western-trained musician group. Fifteen non-musicians who had no formal musical education and had never played a musical instrument were selected as the control group. All of the musicians were from the Chinese Conservatory of Music or the Central Conservatory of Music. The non-musicians were recruited from other universities. The age of initial musical training was approximately 7 years among the musicians. They trained several hours every day and did not study other instruments during their daily non-training hours. Information about the participants is provided in **Table 1**. The comparisons of background information between Chinese-trained musicians and Western-trained musicians are provided in **Figure 1**.

All of the subjects were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971) and had normal hearing abilities. They were paid for participating in the study. The study was performed with the approval of the Ethics Committee of the School of Life Sciences and Technology at the University of Electronic Science and Technology of China. All procedures were conducted in accordance with approved guidelines. Before the experiment, all of the subjects were fully informed about the nature and procedures of the study, and informed written consent was obtained from participants prior to enrollment in the study.

## Experimental Tasks

The musical stimuli in the experiment consisted of 20 pieces, each of which was 10 s in length. Considering the daily training courses of Chinese-trained and Western-trained musicians, ten pieces of Chinese music style were excerpted from traditional Chinese music (Ai Ying, Da Mo Shu Huai, etc.), and ten pieces of Western music style were excerpted from the Baroque period to the 20th century (Concerto Grosso in D Major, Memories of Constantinople, etc.). The musical stimuli were downloaded from the International Music Score Library Project[1] and Netease Cloud Music[2]. For each musical stimulus, a 10-s excerpt that represented the most characteristic and recognizable segment of the music was selected. The musical features of the selected musical stimuli remained intact and true to the original version. The loudness of the musical pieces was fixed and comfortable for the subjects. All of the subjects had never listened to the musical stimuli used in the experiment before. The degree of familiarity with music mentioned later refers to whether the subjects are familiar with the style of the music they listen to. To eliminate irrelevant variables, purely instrumental music, e.g., that which did not contain vocal sections, was used during the entire experiment. In addition to the musical stimuli, a 10-s blank period was used as a control condition during which a blank screen without any auditory stimulation was presented. The subjects who participated in the experiment did not know these details in advance, ensuring they would not have a deliberated response to the music. Before the experiment, the subjects were instructed to avoid strong head movements and to carefully listen to the presented music. One practice trial was presented on the screen to confirm that the participants understood the task instructions (**Figure 2**). Each participant completed 10 trials of Chinese music, 10 trials of Western music and 5 trials of the control condition. Each trial consisted of presentation of a fixation cross (of 4–6 s; jitter was used to cause the brain-blood oxygen signal to return to baseline and to separate the stimulation time), 10 s for the stimuli to appear randomly and 4 s for the subjects to select one of the three response possibilities indicating different stages of familiarity with the musical styles (familiar, ordinary, and unfamiliar). Responses were accomplished by pressing a button on an MRI-compatible response box (if subjects could not distinguish the familiarity of the musical styles, they were instructed to respond with "ordinary"). The performance of the task was judged according to the responses of the subjects. The total duration of the task for each subject was 450–500 s. The tasks were designed and presented using the E-prime software, version 2.0 (Psychology Software Tools, Inc., United States) (**Figure 2**).

## fMRI Scans

Functional MRI scans were acquired using a 3T magnetic resonance imaging (MRI) scanner (Siemens MAGNETOM Trio 3T, Germany) with a standard GE whole head coil performed at the MRI Research Center of the 306th Hospital of the People's Liberation Army. During scanning, foam padding and earplugs were used to reduce head motion and scanning

[1]http://cn.imslp.org/

[2]http://music.163.com/

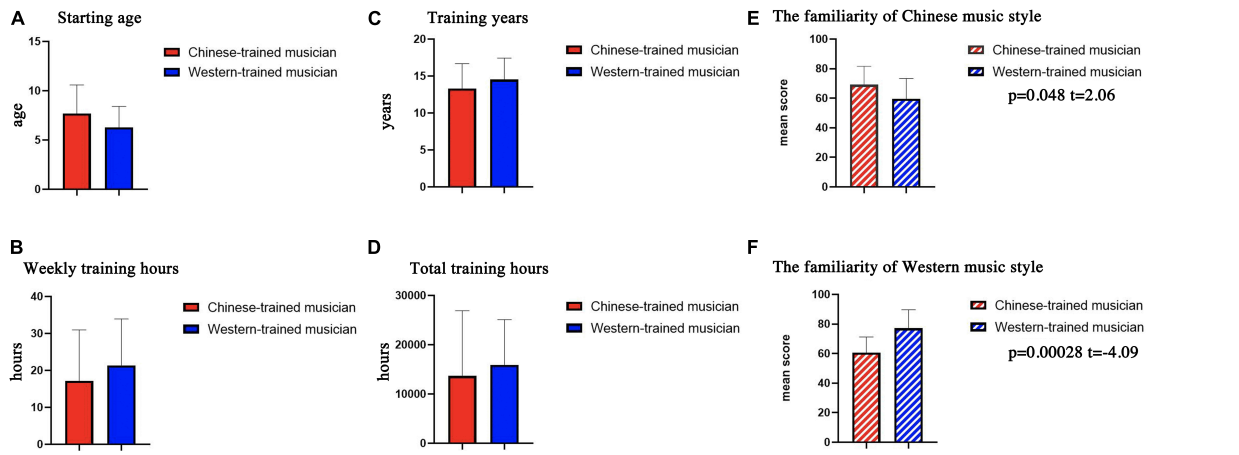



**FIGURE 1 |** The comparisons of background information between Chinese-trained musicians and Western-trained musicians. **(A)** Starting age, i.e., the mean age at which participants started to learn how to play the instrument. **(B)** Weekly training hours, i.e., the mean hours they spend practicing every week reported by the participants. **(C)** Training years, i.e., the mean total number of years since they started to learn the instrument. **(D)** Total training hours, which is calculated by weekly training hours × the number of weeks in a year × training years. There was no significant difference between the two types of musicians in these results. **(E,F)** Musicians have significant differences in familiarity with Chinese and Western music styles. Specifically, compared with Western-trained musicians, Chinese-trained musicians are more familiar with Chinese music style. In addition, compared with Chinese-trained musicians, Western-trained musicians are more familiar with Western music style.

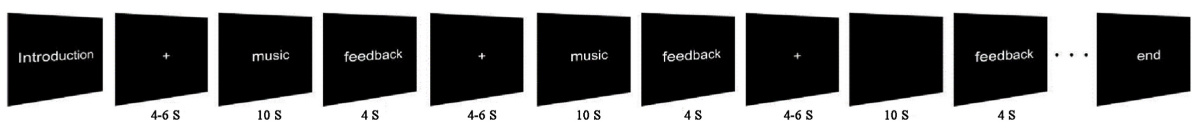



**FIGURE 2 |** Experimental task. In each trial, the musical stimuli (10 pieces from Chinese musical pieces and 10 from Western musical pieces) and the control stimuli (10-s blank period) were presented in a pseudorandom order after presentation of a fixation cross for 4–6 s. The participants were instructed to provide feedback on their level of familiarity with music styles (familiar, ordinary, and unfamiliar) by using a keyboard within 4 s.

noise, respectively. The functional images were acquired using gradient echo-planar imaging (EPI) sequences [slices = 30, averages/measurements = 1/244, echo time (TE) = 30 msec, repetition time (TR) = 2,000 msec, flip angle (FA) = 90°, field of view [FOV] = 210 mm × 210 mm, matrix = 64 × 64, acquisition voxel size = 3.3 mm × 3.3 mm × 4.0 mm, reconstructed voxel size = 3 mm × 3 mm × 3 mm, multislice mode/series: interleaved/descending, bandwidth = 2,232, and slice thickness/gap = 4 mm/0.8 mm] with an eight channel phased array head coil. To ensure steady-state longitudinal magnetization, the first five volumes were discarded. Subsequently, high-resolution T1-weighted images were acquired using a 3-dimensional fast spoiled gradient echo (T1-3D FSPGR) sequence (TR = 2,300 msec, TE = 2.98 msec, FA = 9°, matrix = 256 × 256, FOV = 240 mm × 256 mm, slice thickness/gap = 1 mm/0.5 mm, and slices = 176).

## Functional Imaging Analysis

### Processing

The preprocessing and statistical analysis were performed using SPM8 software (Statistical Parametric Mapping)[3]. We conducted slice time correction, 3D motion detection and correction, spatial normalization to the Montreal Neurological Institute (MNI) template supplied by SPM, and spatial smoothing using an isotropic Gaussian kernel (8 mm full width at half maximum). To avoid MRI machine field effects and to eliminate the head movements of the participants, a series of preprocessing steps, including discarding the first five volumes, normalizing the images with an echo planar imaging template to the MNI atlas space (Evans et al., 1993), and resampling to 3 mm × 3 mm × 3 mm, were performed. Temporal bandpass filtering (pass band 0.01–0.08 Hz) was conducted using a phase-insensitive filter, which was used to reduce the effects of low-frequency drift and high-frequency noise. The time series was further corrected for the effect of six head motion parameters obtained in the realigning step.

### Statistical Testing

We compared the background information differences between Chinese-trained musicians and Western-trained musicians, including starting age, training years, weekly training hours, total training hours and familiarity with Chinese and Western music styles (**Figure 1**). For fMRI data, the second-level analysis embedded in SPM8 software was used. For the first level analyses, data were analyzed using a generalized linear model (GLM) for the three conditions (Western music style,

[3]http://www.fil.ion.ucl.ac.uk/spm

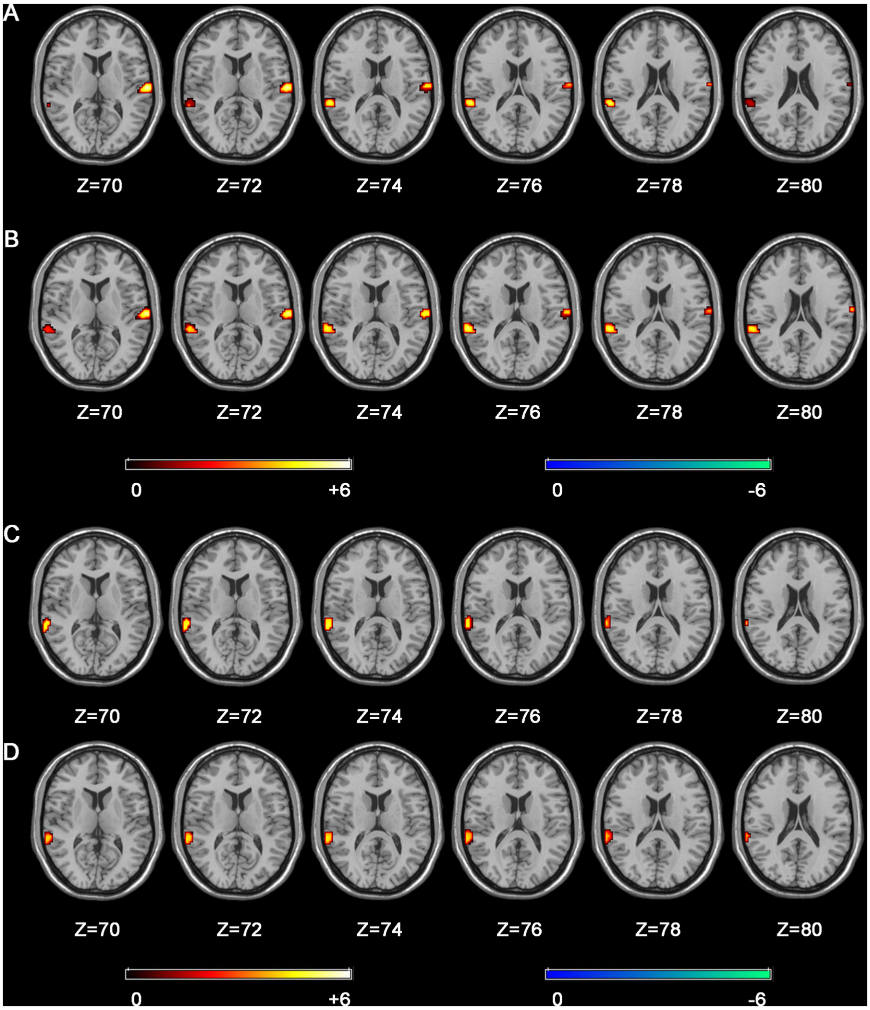



**FIGURE 3 |** Comparison of brain activity during the experimental task when Chinese and Western-trained musicians listened to music with familiar and unfamiliar styles compared to non-musicians. Compared to non-musicians, Chinese-trained musicians have greater activation of the bilateral superior temporal gyrus, and Western-trained musicians have greater activation of the left superior temporal gyrus. **(A)** Chinese-trained musicians – music with Chinese style. **(B)** Chinese-trained musicians – music with Western style. **(C)** Western-trained musicians – music with Chinese style. **(D)** Western-trained musicians – music with Western style.

Chinese music style and the control condition) for each subject. These conditions were modeled by a function convolved with the hemodynamic response function, which was used as the regressor to analyze the task. The six head motion parameters were included in the design matrix to control for movement-related artifacts. For the second level analyses, the contrast images resulted in the first level analyses were entered into group analyses for each of the corresponding contrasts. To further evaluate the differences of listening to music with Chinese or Western styles, comparisons were then made between Western-trained musicians vs. non-musicians and between Chinese-trained musicians vs. non-musicians under the conditions of listening to music with Chinese or Western style in order to ensure basic task-related activation by using a two-sample $t$-test ($p < 0.05$, false discovery rate (FDR)-corrected, and cluster size $\geq$621 $mm^3$). The MNI space displayed the coordinates of the voxel with the maximal $T$-value corrected for the cluster size. Clusters whose values were below 23 were rejected. After analyzing the data, areas that were more activated during the experimental task were identified (**Figure 3**).

To study the relationship between the auditory system and reward system, we defined 6-mm radius spherical ROIs conclude left and right STG, ventral and dorsal striatum, medial prefrontal cortex, midbrain, amygdala, hippocampus, and orbitofrontal cortex in the functional connectivity analysis (Schott et al., 2007; Furukawa et al., 2014; Kruse et al., 2018; Zou et al., 2018). Then, we extracted the signals of ROIs during the task section and calculated the functional connections by using code, to further investigate four conditions (Chinese-trained musicians and Western-trained musicians listened to music with Chinese or Western style). In addition, we compared the connections between the conditions (Chinese music style: Chinese-trained musicians vs. Western-trained musicians and Western music

**TABLE 2 |** Stereotaxic locations of significant clusters of activation. Stereotaxic locations were used to locate significant clusters of activation (Talairach and Pa, 1988).

**Chinese-trained musicians listening to music with Western style**

| Region (AAL) | BA | MNI coordinates (mm) | | | T score | Cluster (voxels) |
|---|---|---|---|---|---|---|
| | | x | y | z | | |
| Temporal_Sup_L | 22 | −63 | −39 | 15 | 5.2762 | 54 |
| Temporal_Sup_R | 42 | 66 | −18 | 9 | 5.4844 | 67 |

*L, left; R, right; AAL, anatomical automatic labeling; BA, Brodmann area; FDR-corrected $p < 0.05$; $p = 0.00078572$; and cluster size $\geq 621$ mm$^3$.*

**Chinese-trained musicians listening to music with Chinese style**

| Region (AAL) | BA | MNI coordinates (mm) | | | T score | Cluster (voxels) |
|---|---|---|---|---|---|---|
| | | x | y | z | | |
| Temporal_Sup_L | 22 | −60 | −39 | 15 | 4.2404 | 46 |
| Temporal_Sup_R | 42 | 63 | −18 | 9 | 5.6102 | 85 |

*L, left; R, right; AAL, anatomical automatic labeling; BA, Brodmann area; FDR-corrected $p < 0.05$; $p = 0.00086851$; and cluster size $\geq 621$ mm$^3$.*

**Western-trained musicians listening to music with Chinese style**

| Region (AAL) | BA | MNI coordinates (mm) | | | T score | Cluster (voxels) |
|---|---|---|---|---|---|---|
| | | x | y | z | | |
| Temporal_Sup_L | 22 | −66 | −42 | 12 | 4.9514 | 91 |

*L, left; R, right; AAL, anatomical automatic labeling; BA, Brodmann area; FDR-corrected $p < 0.05$; $p = 0.0006865$; and cluster size $\geq 621$ mm$^3$.*

**Western-trained musicians listening to music with Western style**

| Region (AAL) | BA | MNI coordinates (mm) | | | T score | Cluster (voxels) |
|---|---|---|---|---|---|---|
| | | x | y | z | | |
| Temporal_Sup_L | 22 | −66 | −42 | 12 | 4.4412 | 74 |

*L, left; R, right; AAL, anatomical automatic labeling; BA, Brodmann area; FDR-uncorrected $p < 0.05$; $p = 0.005$; and cluster size $\geq 621$ mm$^3$.*

style: Western-trained musicians vs. Chinese-trained musicians, **Figure 4**).

## RESULTS

We find that there is no significant difference in terms of starting age, training years, weekly training hours, and total training hours (weekly training hours × the number of weeks in a year × training years) (these results are shown in **Figures 1A–D**). However, musicians have a significant difference in their familiarity with Chinese and Western music styles. Specifically, Chinese-trained musicians are more familiar with Chinese music style than Western-trained musicians. In addition, Western-trained musicians are more familiar with Western music style than Chinese-trained musicians (**Figures 1E,F**).

The activation of the left and right superior temporal gyrus of Chinese-trained musicians is stronger than that of non-musicians and the activation of the left superior temporal gyrus of Western-trained musicians is stronger than that of non-musicians during this experimental task. The three comparisons, which were Chinese-trained musicians vs. non-musicians under the Chinese music style condition, Chinese-trained musicians vs. non-musicians under the Western music style condition and Western-trained musicians vs. non-musicians under the Chinese music style condition passed FDR-correction, but the comparison of Western-trained musicians vs. non-musicians under the Western music style condition did not survive the correction (**Figure 3** and **Table 2**). By comparing the connections between the superior temporal gyrus and the reward system of Chinese and Western-trained musicians, we find that the connection between the right superior temporal gyrus and reward system when Chinese-trained musicians listen to music with Chinese style is stronger than that of Western-trained musicians listening to it. For Western-trained musicians, the connection between the left superior temporal gyrus and reward system when they listen to music with Western style is stronger compared with Chinese-trained musicians (**Figure 4** and **Table 3**, **Table 4**).

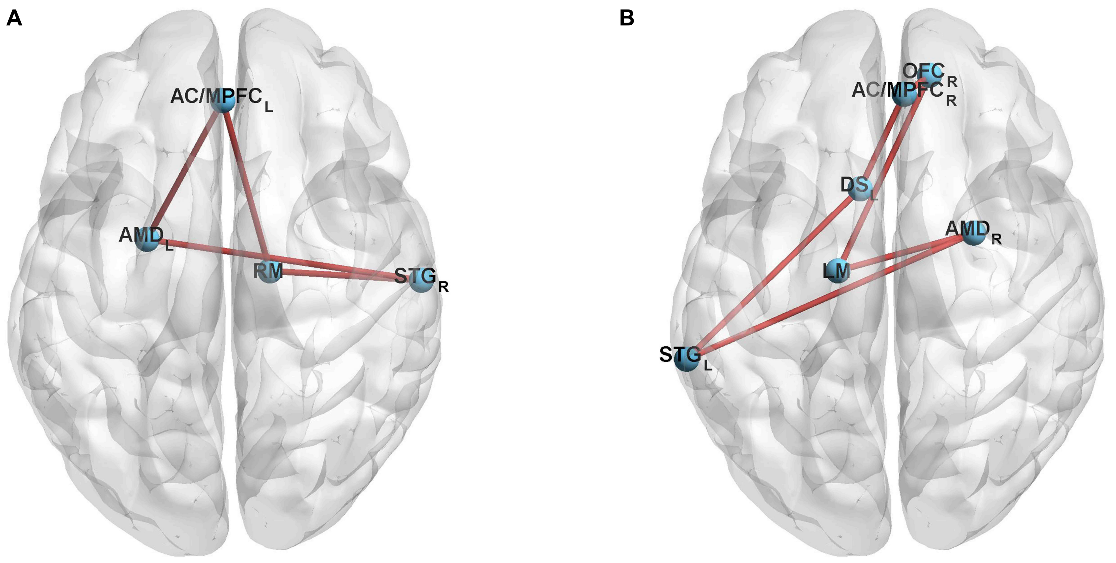



**FIGURE 4 |** Comparison of the connections between the superior temporal gyrus and the reward system of musicians when musicians listen to music with familiar and unfamiliar styles. **(A)** Music with Chinese style: Chinese-trained vs. Western-trained musicians. Stronger connections concluded edges among the right superior temporal gyrus, right midbrain, left amygdala, and left medial prefrontal cortex. **(B)** Music with Western style: Western-trained vs. Chinese-trained musicians. Stronger connections concluded edges among the left superior temporal gyrus, left midbrain, right amygdala, right medial prefrontal cortex, left dorsal striatum and right orbitofrontal cortex. The red lines represent the enhanced connections. The statistical results are presented in **Tables 3**, **4**.

## DISCUSSION

The familiarity of music styles caused by instrument training is an important factor in music recognition. We examined three groups of subjects, namely, Western-trained musicians, Chinese-trained musicians and non-musicians, to study brain plasticity under the influence of musical stimuli. Previous studies have found that different abilities in musical processing, such as reading musical scores (Bouhali et al., 2020), recognition of timbre (Mazzucchi et al., 1982), speech in-noise processing (Strait et al., 2012), and perception of pitch and rhythm (Liégeois-Chauvel et al., 1998; Matsui et al., 2013; Yu et al., 2017), are associated with the temporal lobe. Thus, the temporal lobe plays an important role in the perception of music (Satoh et al., 2001, 2003) and is involved in processing musical features (Patterson et al., 2002; Grahn and Rowe, 2009; Samson et al., 2011).

**Table 3 |** The connection difference between the superior temporal gyrus and the reward system of different musicians when listening to music with Chinese style.

| Connection degree | *t* | *p* |
|---|---|---|
| Superior temporal gyrus_R – Midbrain_R | 2.57 | 0.015 |
| Superior temporal gyrus_R – Amygdala_L | 2.96 | 0.006 |
| Medial prefrontal cortex_L – Midbrain_R | 2.28 | 0.029 |
| Medial prefrontal cortex_L – Amygdala_L | 2.19 | 0.036 |

*Chinese-trained musicians vs. Western-trained musicians.*

**Table 4 |** The connection difference between the superior temporal gyrus and the reward system of different musicians when listening to music with Western style.

| Connection degree | *t* | *p* |
|---|---|---|
| Superior temporal gyrus_L – Amygdala_R | 2.15 | 0.039 |
| Amygdala_R – Midbrain_L | 2.37 | 0.024 |
| Midbrain_L – Orbitofrontal cortex_R | 1.61 | 0.116 |
| Dorsal striatum_L – Superior temporal gyrus_L | 2.31 | 0.027 |
| Orbitofrontal cortex_R – Medial prefrontal cortex_R | 2.85 | 0.008 |
| Medial prefrontal cortex_R – Dorsal striatum_L | 2.13 | 0.041 |

*Western-trained musicians vs. Chinese-trained musicians.*

In this study, greater activation of the left superior temporal gyrus was found during Western-trained musicians listening to musical stimuli compared with non-musicians. Whereas Chinese-trained musicians had greater activation in the bilateral superior temporal gyrus than non-musicians. These results indicate that different music styles which musicians are exposed to during their daily training will affect their strategies of listening to music. Specifically speaking, exposure to Western music style due to long-term musical training will lead Western-trained musicians rely more on the processing of rhythmic features when they listen to music, which is reflected in the activation of the left STG. Rhythm recognition relies on temporal prediction in temporal processing. There is evidence that the left auditory cortex is mainly responsible for temporal processing (LaCroix et al., 2015) and that it has a higher degree of temporal sensitivity (Zatorre and Belin, 2001; Zatorre et al., 2002). Additionally, there is greater left lateralization in the brain activity of musicians compared to non-musicians during rhythm perception, particularly within the superior temporal gyrus (Limb et al., 2006). In some ways, regular rhythm and stress are considered fundamental features of Western music style, which provides evidence for the above statement (Levitin et al., 2012;

London, 2012; Stevens, 2012; Ding et al., 2017). For Chinese-trained musicians, exposure to Chinese music style due to long-term musical training will lead them to rely more on the processing of melodic features when they listen to music, which is reflected in the bilateral STG activation. Rogalsky et al. have found that listening to the melody would lead to the activation of the bilateral STG (Rogalsky et al., 2011). Study on patients with the bilateral temporal lobe lesions also reported that these patients could not perceive melody (Peretz et al., 1994). Other studies found that activation of bilateral temporal lobes is correlated with familiarity with melodies (Groussard et al., 2010; Agustus et al., 2018). Additionally, researchers have found melody processing with more activity in the right auditory system (Peretz et al., 1994; Zatorre et al., 2002; Harms et al., 2014), which may be due to it is mainly involved in spectral processing and has higher spectral sensitivity (Zatorre and Belin, 2001; Zatorre et al., 2002; LaCroix et al., 2015). Part of Chinese traditional music has irregular beat attributes and melody is a prominent feature, which provides evidence for the above statement (Jiang, 1991; Stevens, 2012; Chen, 2016).

The results also show the differences in the functional connections between the superior temporal gyrus and the reward system when Chinese and Western-trained musicians listen to musical stimuli. These connections are stronger when musicians listen to the musical pieces from their familiar musical style as compared to listening to music from unfamiliar style. Studies have shown that implicit expectations of music are based on the history of listening to music and the implicit understanding of music rules. For example, the rhythm is considered highly regular and predictable and has been thought to contribute to our enjoyment of music (Levitin et al., 2012). Whereas the prediction of melody is an important aspect of music listening (Thiessen and Saffran, 2009; Egermann et al., 2013; Lappe et al., 2013). Meanwhile, exposure effects support the contention that accurate prediction is rewarding, familiar music will make musicians more enjoyable, which has been confirmed its presence for melodies (Wilson, 1979; Zald and Zatorre, 2011). Hence, just like prior exposure to a piece of music makes it more predictable. Musical training contributes to predicting upcoming musical events by raising the accuracy of the discrimination of musical material (Lappe et al., 2013). Familiarity with the music style contribute to musical pleasure, rhythm characteristics in Western music style as well as melody characteristics in Chinese music style may play an important role in this process.

There is evidence supporting a deep understanding and clear familiarity with the rules of musical structure have contributed to the anticipation of upcoming musical events, listeners will pay attention to whether it meets their expectations, in turn leading to emotional arousal (Salimpoor et al., 2011). The activity in the STG is correlated with the reward system when subjects listen to personally pleasurable music (Martínez-Molina et al., 2016; Gold et al., 2019). Therefore, although we do not obtain musicians' rating of pleasure from the musical pieces, but we suggest that when the implicit expectation is not met (i.e., when Western-trained musicians listen to Chinese music or Chinese-trained musicians listen to Western music), the connection between the STG and reward system is low because the musical characteristics in the music style do not meet their predictions. This outcome is possibly caused by music that violates expectations at the beginning and will be considered unattractive because it prevents accurate predictions (Zald and Zatorre, 2011). These findings support that when musicians listen to music in their field of expertise, their clear familiarity and deep understanding of the musical characteristics in the musical style will lead to enhanced connectivity between the STG and reward system.

In general, we think that musicians will have a deeper understanding of the music style that they often hear during daily training, and familiar musical style can help musicians predict upcoming musical events. Acquired musical expertise of musicians can raise the prediction accuracy, which makes upcoming musical events meet their expectations. This process leads to musicians' enjoyment of music, which is related to the increased connection between the STG and reward system. Furthermore, this result is consistent with our background information results. That is, Chinese-trained musicians are familiar with Chinese music style, which means they are more familiar with the rules and features of Chinese music than Western-trained musicians. The same holds for Western-trained musicians (**Figures 1E,F**). Considering that the musicians recruited in this study are typical Chinese and Western instrumentalists and that the experimental stimuli also contain typical cultural characteristics of Chinese and Western music styles, we think that musicians' familiarity with culture attribute is also one of the reasons that leads to the differences in functional connectivity when they listen to Chinese and Western music. This helps explain the differences of the connections between the STG and the reward system of Chinese and Western-trained musicians when listening to music with Chinese or Western styles.

The results also show the laterality of the connections between the STG and the reward system. The connections of Chinese-trained musicians are partial to the right STG, and the connections of Western-trained musicians are partial to the left STG. Comparatively speaking, Western music style emphasizes rhythm, while Chinese music style emphasizes melody. The left hemisphere of the brain is mainly involved in temporal processing and rhythm recognition whereas the right hemisphere principally participates in spectral processing and melody recognition. These results explain the laterality of the circuit between the STG and the reward system in Chinese-trained and Western-trained musicians. In our results, there is no difference in connections of the STG and the reward system between musicians and non-musicians. We suggest that this finding occurred because non-musicians have less understanding of the rules of Chinese or Western music than experienced musicians. Thus, they hardly have implicit expectations based on the musical features in music style and just enjoy the music; therefore, they can always produce a reward response.

## CONCLUSION

This investigation researched trained musicians' brain activities under condition of listening to music of the style in which they are experts or not experts. The results showed that Chinese-trained musicians activated the bilateral superior temporal gyrus and Western-trained musicians activated the left superior temporal gyrus during the task. When listening to the music in which they are experts, Chinese-trained musicians showed enhanced connections between the right superior temporal gyrus and the reward system compared with Western-trained musicians. However, Western-trained musicians showed enhanced connections between the left superior temporal gyrus and the reward system. This outcome indicates that the circuit that connects the superior temporal gyrus and reward system has different patterns between Chinese-trained musicians and Western-trained musicians when they are listening to music in which they were experts or not experts. The activities and connections in the task are related to the characteristics of the music style that musicians are exposed to in their daily training. Chinese-trained musicians and Western-trained musicians have laterality in processing the characteristics of music style they expertise. One of the reasons for the difference may also be the familiarity with culture attribute.

## LIMITATIONS

This study has some limitations. Although we mentioned that the cultural attributes of music style may lead to differences between Chinese and Western musicians in listening to familiar music style, the influence of musical instruments needs to be considered in the further exploration. Research on the cultural attributes suggests that musical instruments are potent cultural phenomena. They are embodiments of culturally based belief and value systems (Dawe, 2011). In this work, the instruments (guqin, erhu, and piano) trained by musicians are considered as typically traditional Chinese and Western instruments, and which are culturally specific. In addition, these musicians have been selected and tested. Thus, we believe that they are almost fully influenced by musical culture and can be the professional model for this study. Professional scales for familiarity of musical culture are necessary for us to distinguish the influence of musical instruments and culture. However, at present, there is no authoritative scale to measure the familiarity of musical culture. When the authoritative scale is developed, it will be helpful for enriching this hypothesis. Furthermore, the comparison of Western-trained musicians vs. non-musicians under the Western music style condition did not survive the correction, even though it showed a tendency similar to that in the Chinese music style condition. This may be due to the small sample size caused by the recruitment of subjects (we try our best to recruit the professional guqin and erhu players from conservatory of music, but guqin and erhu, the typically traditional Chinese instruments, are learnt by a small crowd). Thus, sample sizes should be increased in future studies.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The study was done with the approval of the Ethics Committee of the School of Life Sciences and Technology at the University of Electronic Science and Technology of China. All procedures were carried out in accordance with approved guidelines. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

JL, DY, and PL contributed for ideas and design of experiment. SG, RD, JZ, and YL contributed for literature research and review. KL contributed for providing experimental sites and equipment. KP and PL contributed for selecting experimental stimuli. KP, JZ, YL, and YH contributed for recruiting subjects and collecting experimental data. SG, RD, YH, and YhL contributed for processing data. SG, RD, and JL contributed for making charts and writing manuscript. JL, CL, and DY contributed for reviewing manuscript and providing suggestions. All authors contributed to the article and approved the submitted version.

## FUNDING

This study was supported by grants from the National Natural Science Foundation of China (Nos. 61761166001, 31600798, 82001906, and 61527815), Project of Science and Technology Department of Sichuan Province (No. 2021YFS0135), and 111 project (B12027).

## ACKNOWLEDGMENTS

We thank Weixia Wang for the help of collecting some information of participants. We also thank Hui He for helpful discussion and Zhuo Chen for helping with partial data processing.

## REFERENCES

Agustus, J. L., Golden, H. L., Callaghan, M. F., Bond, R. L., Benhamou, E., Hailstone, J. C., et al. (2018). Melody Processing Characterizes Functional Neuroanatomy in the Aging Brain. *Front. Neurosci.* 12:815. doi: 10.3389/fnins.2018.00815

Blood, A. J., and Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11818–11823. doi: 10.1073/pnas.191355898

Bouhali, F., Mongelli, V., Thiebaut de Schotten, M., and Cohen, L. (2020). Reading music and words: The anatomical connectivity of musicians' visual cortex. *Neuroimage* 212:116666. doi: 10.1016/j.neuroimage.2020.116666

Bouwer, F. L., Burgoyne, J. A., Odijk, D., Honing, H., and Grahn, J. A. (2018). What makes a rhythm complex? The influence of musical training and accent type on beat perception. *PLoS One* 13:e0190322. doi: 10.1371/journal.pone.0190322

Cao, G., and Li, D. (2017). Rigidity and Flexibility: a Comparative Study of Traditional Chinese and Western Music from the Perspective of Complex Information System Theory. *Proceedings* 1, 12–16. doi: 10.3390/is4si-2017-03971

Chen, C. H. (2016). "The Comparison between Chinese and western music education" in *2016 International Conference on Education, Sports, Arts and Management Engineering*. (Paris: Atlantis Press). doi: 10.2991/icesame-16.2016.212

Coffey, E. B. J., Herholz, S. C., Scala, S., and Zatorre, R. J. (2011). "Montreal Music History Questionnaire: a tool for the assessment of music-related experience in music cognition research" in *The Neurosciences and Music IV: learning and Memory Conference*. (UK: Edinburgh).

Dawe, K. (2011). *The Cultural Study of Musical Instruments*, 2nd Edn. England: Routledge.

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., and Poeppel, D. (2017). Temporal modulations in speech and music. *Neurosci. Biobehav. Rev.* 81, 181–187. doi: 10.1016/j.neubiorev.2017.02.011

Drake, C., and El Heni, J. B. (2003). Synchronizing with Music: intercultural Differences. *Ann. N. Y. Acad. Sci.* 999, 429–437. doi: 10.1196/annals.1284.053

Egermann, H., Pearce, M. T., Wiggins, G. A., and McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cogn. Affect Behav. Neurosci.* 13:533—553. doi: 10.3758/s13415-013-0161-y

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., and Taub, E. (1995). Increased Cortical Representation of the Fingers of the Left Hand in String Players. *Science* 270:305. doi: 10.1126/science.270.5234.305

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., and Peters, T. M. (1993). *3D Statistical Neuroanatomical Models from 305 MRI Volumes*. United States: IEEE

Ferreri, L., Mas-Herrero, E., Zatorre, R. J., Ripollés, P., Gomez-Andres, A., and Alicart, H. (2019). Dopamine modulates the reward experiences elicited by music. *Proc. Natl. Acad. Sci. U. S. A.* 116, 3793–3798. doi: 10.1073/pnas.1811878116

Freitas, C., Manzato, E., Burini, A., Taylor, M. J., Lerch, J. P., and Anagnostou, E. (2018). Neural Correlates of Familiarity in Music Listening: a Systematic Review and a Neuroimaging Meta-Analysis. *Front. Neurosci.* 12:1–14. doi: 10.3389/fnins.2018.00686

Furukawa, E., Bado, P., Tripp, G., Mattos, P., Wickens, J. R., and Bramati, I. E. (2014). Abnormal Striatal BOLD Responses to Reward Anticipation and Reward Delivery in ADHD. *PLoS One* 9:e89129. doi: 10.1371/journal.pone.0089129

Gold, B. P., Mas-Herrero, E., Zeighami, Y., Benovoy, M., Dagher, A., and Zatorre, R. J. (2019). Musical reward prediction errors engage the nucleus accumbens and motivate learning. *Proc. Natl. Acad. Sci. U. S. A.* 116, 3310–3315.

Grahn, J. A., and Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J. Neurosci.* 29, 7540–7548. doi: 10.1523/JNEUROSCI.2018-08.2009

Groussard, M., Rauchs, G., Landeau, B., Viader, F., Desgranges, B., Eustache, F., et al. (2010). The neural substrates of musical memory revealed by fMRI and two semantic tasks. *Neuroimage* 53, 1301–1309. doi: 10.1016/j.neuroimage.2010.07.013

Harms, V. L., Cochran, C., and Elias, L. J. (2014). Melody and Language: an Examination of the Relationship Between Complementary Processes. *Open Psychol. J.* 7, 1–8. doi: 10.2174/1874350101407010001

Hund-Georgiadis, M., and von Cramon, D. Y. (1999). Motor-learning-related changes in piano players and non-musicians revealed by functional magnetic-resonance signals. *Exp. Brain Res.* 125, 417–425. doi: 10.1007/s002210050698

Jiang, J. (1991). The Influence of Traditional Chinese Music on Professional Instrumental Composition. *Asian Music* 22, 83–96. doi: 10.2307/834308

Koelsch, S., Schroger, E., and Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport* 10, 1309–1313. doi: 10.1097/00001756-199904260-00029

Kruse, O., Leon, T. I., Stalder, T., Stark, R., and Klucken, T. (2018). Altered reward learning and hippocampal connectivity following psychosocial stress. *Neuroimage* 171, 15–25. doi: 10.1016/j.neuroimage.2017.12.076

LaCroix, A. N., Diaz, A. F., and Rogalsky, C. (2015). The relationship between the neural computations for speech and music perception is context-dependent: an activation likelihood estimate study. *Front. Psychol.* 6:1138. doi: 10.3389/fpsyg.2015.01138

Lappe, C., Steinsträter, O., and Pantev, C. (2013). Rhythmic and melodic deviations in musical sequences recruit different cortical areas for mismatch detection. *Front. Hum. Neurosci.* 7:260. doi: 10.3389/fnhum.2013.00260

Levitin, D. J., Chordia, P., and Menon, V. (2012). Musical rhythm spectra from Bach to Joplin obey a 1/f power law. *Proc. Natl. Acad. Sci. U. S. A.* 109, 3716–3720. doi: 10.1073/pnas.1113828109

Li, Q., Wang, X., Wang, S., Xie, Y., Li, X., Xie, Y., et al. (2018). Musical training induces functional and structural auditory-motor network plasticity in young adults. *Hum. Brain Mapp.* 39, 2098–2110. doi: 10.1002/hbm.23989

Liégeois-Chauvel, C., Peretz, I., Babaï, M., Laguitton, V., and Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain* 121, 1853–1867. doi: 10.1093/brain/121.10.1853

Limb, C. J., Kemeny, S., Ortigoza, E. B., Rouhani, S., and Braun, A. R. (2006). Left hemispheric lateralization of brain activity during passive rhythm perception in musicians. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* 288, 382–389. doi: 10.1002/ar.a.20298

London, J. (2012). *Hearing in Time*. New York: Oxford University Press.

Lu, J., Yang, H., He, H., Jeon, S., Hou, C., and Evans, A. C. (2017). The Multiple-Demand System in the Novelty of Musical Improvisation: evidence from an MRI Study on Composers. *Front. Neurosci.* 11:1–13. doi: 10.3389/fnins.2017.00695

Martínez-Molina, N., Mas-Herrero, E., Rodríguez-Fornells, A., Zatorre, R. J., and Marco-Pallarés, J. (2016). Neural correlates of specific musical anhedonia. *Proc. Natl. Acad. Sci. U. S. A.* 113, E7337–E7345. doi: 10.1073/pnas.1611211113

Matsui, T., Tanaka, S., Kazai, K., Tsuzaki, M., and Katayose, H. (2013). Activation of the left superior temporal gyrus of musicians by music-derived sounds. *Neuroreport* 24, 41–45. doi: 10.1097/wnr.0b013e32835c1e02

Mazzucchi, A., Marchini, C., Budai, R., and Parma, M. (1982). A case of receptive amusia with prominent timbre perception defect. *J. Neurol. Neurosurg. Psychiatry* 45, 644–647. doi: 10.1136/jnnp.45.7.644

Morrison, S. J., and Demorest, S. M. (2009). Cultural constraints on music perception and cognition. *Prog. Brain Res.* 178, 67–77. doi: 10.1016/S0079-6123(09)17805-6

Morrison, S. J., Demorest, S. M., Aylward, E. H., Cramer, S. C., and Maravilla, K. R. (2003). FMRI investigation of cross-cultural music comprehension. *Neuroimage* 20, 378–384. doi: 10.1016/s1053-8119(03)00300-8

Nan, Y., Knösche, T. R., and Friederici, A. D. (2006). The perception of musical phrase structure: a cross-cultural ERP study. *Brain Res.* 1094, 179–191. doi: 10.1016/j.brainres.2006.03.115

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., et al. (2001). Functional anatomy of musical perception in musicians. *Cereb. Cortex* 11, 754–760. doi: 10.1093/cercor/11.8.754

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature* 392, 811–814. doi: 10.1038/33918

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., and Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport* 12, 169–174. doi: 10.1097/00001756-200101220-00041

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36, 767–776. doi: 10.1016/s0896-6273(02)01060-7

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., et al. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain* 117:1283. doi: 10.1093/brain/117.6.1283

Raghanti, M. A., Edler, M. K., Stephenson, A. R., Wilson, L. J., Hopkins, W. D., Ely, J. J., et al. (2016). Human-specific increase of dopaminergic innervation in a striatal region associated with speech and language: a comparative analysis of the primate basal ganglia. *J. Comp. Neurol.* 524, 2117–2129. doi: 10.1002/cne.23937

Rogalsky, C., Rong, F., Saberi, K., and Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J. Neurosci.* 31, 3843–3852. doi: 10.1523/JNEUROSCI.4515-10.2011

Rüsseler, J., Altenmüller, E., Nager, W., Kohlmetz, C., and Münte, T. F. (2001). Event-related brain potentials to sound omissions differ in musicians and non-musicians. *Neurosci. Lett.* 308, 33–36. doi: 10.1016/S0304-3940(01)01977

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat. Neurosci.* 14, 257–262. doi: 10.1038/nn.2726

Samson, F., Zeffiro, A. T., Toussaint, A., and Belin, P. (2011). Stimulus complexity and categorical effects in human auditory cortex: an activation likelihood estimation meta-analysis. *Front. Psychol.* 1:241. doi: 10.3389/fpsyg.2010.00241

Satoh, M., Takeda, K., Nagata, K., Hatazawa, J., and Kuzuhara, S. (2001). Activated brain regions in musicians during an ensemble: a PET study. *Cogn. Brain Res.* 12:101. doi: 10.1016/s0926-6410(01)00044-1

Satoh, M., Takeda, K., Nagata, K., Hatazawa, J., and Kuzuhara, S. (2003). The anterior portion of the bilateral temporal lobes participates in music perception: a positron emission tomography study. *Am. J. Neuroradiol.* 24, 1843–1848.

Schott, B. H., Niehaus, L., Wittmann, B. C., Schütze, H., Seidenbecher, C. I., and Heinze, H. (2007). Ageing and early-stage Parkinson's disease affect separable neural mechanisms of mesolimbic reward processing. *Brain* 130, 2412–2424. doi: 10.1093/brain/awm147

Stevens, C. J. (2012). Music perception and cognition: a review of recent cross-cultural research. *Top. Cogn. Sci.* 4, 653–667. doi: 10.1111/j.1756-8765.2012.01215.x

Strait, D. L., and Kraus, N. (2014). Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hear. Res.* 308, 109–121. doi: 10.1016/j.heares.2013.08.004

Strait, D. L., Parbery-Clark, A., Hittner, E., and Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain Lang.* 123, 191–201. doi: 10.1016/j.bandl.2012.09.001

Talairach, J., and Pa, T. (1988). *Co-Planar Stereotaxic Atlas of The Human Brain. 3-D proportional System: an Approach to Cerebral Imaging*. New York: Thieme Medical Publishers, Inc.

Tervaniemi, M., Rytkonen, M., Schroger, E., Ilmoniemi, R. J., and Naatanen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learn. Mem.* 8, 295–300. doi: 10.1101/lm.39501

Thiessen, E. D., and Saffran, J. R. (2009). How the melody facilitates the message and vice versa in infant learning and memory. *Ann. N.Y. Acad. Sci.* 1169, 225–233. doi: 10.1111/j.1749-6632.2009.04547.x

Tierney, A. T., Krizman, J., and Kraus, N. (2015). Music training alters the course of adolescent auditory development. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10062–10067. doi: 10.1073/pnas.1505114112

Trehub, S. E. (2001). Musical predispositions in infancy. *Ann. N. Y. Acad. Sci.* 930, 1–16. doi: 10.1111/j.1749-6632.2001.tb05721.x

Wilson, W. R. (1979). Feeling more than we can know: exposure effects without learning. *J. Personal. Soc. Psychol.* 37, 811–821. doi: 10.1037/0022-3514.37.6.811

Yu, M., Xu, M., Li, X., Chen, Z., Song, Y., Liu, J., et al. (2017). The Shared Neural Basis of Music and Language. *Neuroscience* 357, 208–219. doi: 10.1016/j.neuroscience.2017.06.003

Zald, D. H., and Zatorre, R. J. (2011). *Neurobiology of Sensation and Reward*. United States: CRC Press/Taylor & Francis.

Zatorre, R. J., and Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11, 946–953. doi: 10.1093/cercor/11.10.946

Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46. doi: 10.1016/s1364-6613(00)01816-7

Zhang, L., Xie, S., Li, Y., Shu, H., and Zhang, Y. (2020). Perception of musical melody and rhythm as influenced by native language experience. *J. Acoust. Soc. Am.* 147:EL385. doi: 10.1121/10.0001179

Zou, L., Zhou, H., Zhuang, Y., van Hartevelt, T. J., Lui, S. S. Y., and Cheung, E. F. C. (2018). Neural responses during the anticipation and receipt of olfactory reward and punishment in human. *Neuropsychologia* 111, 172–179. doi: 10.1016/j.neuropsychologia.2018.02.003

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

REVIEW
published: 16 September 2021
doi: 10.3389/fnins.2021.662087

# Use of Music Therapy as an Audiological Rehabilitation Tool in the Elderly Population: A Mini-Review

*Anne Sophie Grenier[1], Louise Lafontaine[2] and Andréanne Sharp[1]**

[1] CERVO Research Center, Université Laval, Québec City, QC, Canada, [2] École d'Orthophonie et d'Audiologie, Université de Montréal, Montreal, QC, Canada

It is well known and documented that sensory perception decreases with age. In the elderly population, hearing loss and reduced vestibular function are among the most prevalently affected senses. Two important side effects of sensory deprivation are cognitive decline and decrease in social participation. Hearing loss, vestibular function impairment, and cognitive decline all lead to a decrease in social participation. Altogether, these problems have a great impact on the quality of life of the elderly. This is why a rehabilitation program covering all of these aspects would therefore be useful for clinicians. It is well known that long-term music training can lead to cortical plasticity. Behavioral improvements have been measured for cognitive abilities and sensory modalities (auditory, motor, tactile, and visual) in healthy young adults. Based on these findings, it is possible to wonder if this kind of multisensory training would be an interesting therapy to not only improve communication but also help with posture and balance, cognitive abilities, and social participation. The aim of this review is to assess and validate the impact of music therapy in the context of hearing rehabilitation in older adults. Musical therapy seems to have a positive impact on auditory perception, posture and balance, social integration, and cognition. While the benefits seem obvious, the evidence in the literature is scarce. However, there is no reason not to recommend the use of music therapy as an adjunct to audiological rehabilitation in the elderly when possible. Further investigations are needed to conclude on the extent of the benefits that music therapy could bring to older adults. More data are needed to confirm which hearing abilities can be improved based on the many characteristics of hearing loss. There is also a need to provide a clear protocol for clinicians on how this therapy should be administered to offer the greatest possible benefits.

**Keywords: music, cognition, movement, social participation, presbycusis**

OPEN ACCESS

***Edited by:***
*Cunmei Jiang,*
*Shanghai Normal University, China*

***Reviewed by:***
*Peter Schneider,*
*Heidelberg University, Germany*
*Jakub Ryszard Matyja,*
*Polish Academy of Sciences, Poland*

***Correspondence:***
*Andréanne Sharp*
*andreanne.sharp@fmed.ulaval.ca*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *01 February 2021*
***Accepted:*** *13 August 2021*
***Published:*** *16 September 2021*

***Citation:***
*Grenier AS, Lafontaine L and Sharp A (2021) Use of Music Therapy as an Audiological Rehabilitation Tool in the Elderly Population: A Mini-Review. Front. Neurosci. 15:662087. doi: 10.3389/fnins.2021.662087*

## INTRODUCTION

Disabling hearing loss is an important problem to address among older adults. It is known to affect one-third of individuals over 65 years old (World Health Organization, 2020). The term used for age-related hearing loss in the literature is presbycusis, which includes degradation of structures in the middle ear, inner ear, and central auditory pathways due to aging (Gates and Mills, 2005). McDermott (2009) explains that all the factors that make it difficult to analyze the auditory scene in normal-hearing people are even more problematic in people with hearing loss. This is why it is not surprising that presbycusis leads not only to reduced hearing thresholds but also to difficulties to understand speech in noise, degradation of central auditory processing, and impaired localization of sounds in the environment.

The auditory image sent to the auditory cortex must be very robust so that the signal analysis (e.g., segregation, categorization) will be carried out correctly. However, in hearing-impaired individuals, this signal is degraded by hearing loss (e.g., hair cell breakage, reduced ossicle transfer function). The signal strength will therefore be reduced and distorted. In the end, the signal received at the cortex is always degraded. This results in reduced temporal and spectral resolution (Shinn-Cunningham and Best, 2008). The reduction of auditory filters makes it more difficult to analyze the auditory scene and to segregate it into its components. In addition, the harmonic structure and attack or release times appear to be less well perceived in individuals with hearing loss. These components are important for the formation of auditory objects in a short time scale (e.g., the formation of syllables during the presence of several conversations at the same time). The fine spectro-temporal structure is also less well encoded. This point can make auditory unmasking difficult, a skill necessary to isolate auditory objects from the auditory scene, for example, when it is needed to understand speech in noise.

Hearing loss in the elderly can have a variety of causes that will mostly lead to sensorineural or sometimes mixed hearing loss (Gates and Mills, 2005). There is a small number of elderly will present bone-conduction gaps in their audiogram (Nondahl et al., 2013). One way to improve hearing thresholds, for both cases, would be to find a way to directly repair the affected structures, i.e., the inner ear for the former, and the middle ear for the latter (as for example some studies try to do with stem cells, see for example Pauley et al., 2008, for a review). However, these new technologies are still in development and are not commercially available. This is why, in both situations, the most common way to improve hearing perception is by the use of hearing aids. The gain provided by hearing aids does not fully restore hearing skills. It will rather amplify environmental sounds based on hearing loss to improve audibility. However, despite sufficient hearing amplification, difficulties persist, and these are linked to persistent problems at the level of intelligibility. Several studies have shown that speech-in-noise (Marrone et al., 2008; Abrams and Kihm, 2015) and musical perception (Feldmann and Kumpf, 1988; Chasin and Russo, 2004) difficulties are recurrent complaints among hearing aid users. A way to improve auditory abilities is to combine the use of hearing aids with audiological rehabilitation. Auditory training of any kind will not be able to improve hearing thresholds due to anatomical and physiological damage limitations but will try to improve auditory processing abilities.

A lot of audiological rehabilitation programs to date are focusing mostly on auditory training (for a review, see Sweetow and Palmer, 2005). However, most audiologists are faced with patients present multiple sensory declines with vestibular disorders due to the normal aging process (dizziness, loss of balance, falls, etc.) (Hülse et al., 2019). Presbyvestibulopathy is a phenomenon of growing interest because it interferes with the maintenance of posture, balance, and movement of the elderly (Jahn, 2019). The increased risks of falls due to loss of balance in the elderly are considered a significant burden on the health system and the health of the population (World Health Organization, 2008). This other sensory loss combined with hearing loss needs to be addressed in rehabilitation approaches with the hearing-impaired elderly to enable them to have a better quality of life.

Other side effects of these sensory deprivations must be addressed in audiological rehabilitation. More and more studies are suggesting a link between hearing loss due to age, cognitive decline, and even dementia when no assistive listening devices are used (Nadhimi and Llano, 2020; Slade et al., 2020). Many leading medical organizations around the world are suggesting that age-related hearing loss can be the largest potentially preventable risk factor for dementia (Livingston et al., 2017, 2020). Once put together, all of these elements illustrate a global portrait of the limitations caused by aging of the auditory and vestibular system. These elements should be central in hearing rehabilitation programs to obtain a global portrait of the elderly and promote more accurately healthy aging.

A second important side effect of presbycusis and presbyvestibulopathy is the impact on social life of older adults. When both communication and ease of movements are altered, this can lead to impaired social participation daily to many activities (e.g., conversations, family reunions, watching television, or listening to music). Therefore, it is not surprising that another huge impact of hearing loss is on social and emotional wellbeing. As stated by World Health Organization (2020): “Exclusion from communication can have a significant impact on everyday life, causing feelings of loneliness, isolation, and frustration, particularly among older people with hearing loss.” When hearing loss is not taken care of, it will contribute to social isolation, depression, loss of self-esteem, decrease in psychological wellbeing, reduction of the ability to function in social situations, and a reduction in quality of life in general (Gates and Mills, 2005). If we add to this portrait the difficulties in maintaining posture and balance leading to limited movement, it then becomes evident that the normal aging of the inner ear can have a significant impact on the social participation of the elderly and lead to long-term psychological problems that are important to address to ensure a better quality of life for this part of the population.

Here, we question whether musical therapy could be a useful tool in audiological rehabilitation to help reverse the maladaptive plasticity linked to this sensory deprivation. It is well established that musical training can lead to functional and structural brain changes in people with normal hearing. Imaging studies have revealed that several areas of the brain, including the planum temporale, corpus callosum, the main motor area of the hand, and the cerebellum, differ in structure and size between normal musicians and non-musicians (e.g., Schlaug et al., 1995; Münte et al., 2002; Schneider et al., 2002, 2009; Herholz and Zatorre, 2012; White-Schwoch et al., 2013). These anatomical changes are associated with behavioral improvements, mainly for hearing skills. For example, improved performance has been measured for discrimination of timbre, pitch, rhythm, etc. (for a review, see Kraus and Chandrasekaran, 2010). Basic auditory processes are also improved in musicians. For example, a study by Spiegel and Watson (1984) demonstrated that musicians have better frequency discrimination compared to matched non-musician controls. Moreover, this effect seems to be correlated with the number of years of musical expertise (Kishon-Rabin et al., 2001). It is also known that musicians have better central hearing skills and that these are preserved over time (for a review, see Alain et al., 2014). In addition, it is possible to hypothesize that music not only can impact auditory abilities but also can help with social aspects and cognitive aspects. The multisensory perspective of musical training suggests that it may also lead to improvements in posture, balance, and movements. What is less clear, however, is whether learning a musical instrument at an advanced age is an effective method of rehabilitation. This question is of high relevance considering that not all individuals with hearing loss have learned a musical instrument in their lifetime. It is therefore this portion of the population that is the interest of this current literature review. The main goal of this review is to establish the impact of music therapy in the context of hearing rehabilitation of older adults. To do so, we will examine the impact of music therapy on (1) communication, (2) movements, (3) social participation, and (4) cognitive abilities of non-musician healthy old adults to determine if it could be a useful tool for audiological rehabilitation.

## IMPACT ON COMMUNICATION

It is well known that learning to play a musical instrument can lead to improvement of central auditory abilities. This is why it should be a useful way to improve central auditory abilities following hearing decline due to age. A recent study by Dubinsky et al. (2019) investigated if a group of older adults with age-related hearing loss improved their speech-in-noise and pitch discrimination abilities following musical training. Thirty-four adults with age-related hearing loss between 54 and 76 years old took part in choir lessons for 2 h per week for 10 weeks. They were compared to a do-nothing group of 29 adults matched for age and hearing loss degree. Following this training, results suggested improvement for the choir-singing group for speech-in-noise and pitch discrimination. Also, the neural representation of the pitch discrimination was measured *via* frequency following responses, an auditory evoked potential that represents the phase-locked neural activity that is synchronized to periodic and transient aspects of sound. The neural representation of pitch was also improved following musical training.

Zendel et al. (2019) also used electrophysiology to measure the speech performance in noise of 13 individuals above 65 years old participated in a 6-month piano training. They were compared to two control groups with participants matched for age, cognition, and education levels. The first group had no specific therapy (do-nothing group) ($n = 13$), and the second group ($n = 8$) followed a similar training but using video games instead of piano lessons. Participants were asked to repeat words in three conditions: No-Noise, Quiet [15 dB signal-to-noise ratio (SNR)], and Loud (15 dB SNR). Results were similar to those of Dubinsky et al. (2019): accuracy to repeat words in the loud condition was improved post-training. In addition, encephalographic measurements (EEG) showed an increase in positive cerebral electrical activity at the level of the fronto-left electrodes 200 to 1,000 ms after the presentation of a word in noise linked with the behavioral improvement found in the music group. Source analysis suggests that this activity was due to regions involved in the motor speech system. These results support the idea that musical training provides a causal benefit to hearing abilities.

With the same participants as the last study discussed, Fleming et al. (2019) performed a speech-in-noise task and made functional magnetic resonance imaging (fMRI) measurements during this behavioral task. In contradiction to their previous study, there was no improvement in measured behavior. The authors explain, however, that the nature of the task may have caused this difference. They used the hearing in noise test, which is a standard test used in audiological evaluation that consists of repeating sentences in noise. However, in this study, the participants were asked to identify the picture most related to the sentence between four pictures. This task is easier than the previous one. Also, during the EEG study, the SNR was less good and made the task more difficult, which would have avoided the ceiling effect found during the task in the fMRI. However, notable differences were found in fMRI measurements following music training. The training has led to an increase in activation of the following cortical areas when perceiving speech: the frontal cortices bilaterally, the left parietal cortex, and the right temporal bone. These areas would be associated with speech encoding. Overall, these findings support the hypothesis that musical training improve speech processing skills even when the training is started late in life.

Recently, Worschech et al. (2021) compared the impact of learning piano against listening to music as an intervention to improve speech-in-noise perception. They did their investigation using a randomized controlled trial paradigm. A total of 156 elderly with normal hearing thresholds and non-musicians were randomized into two groups: one playing piano ($n = 74$) and the other listened to cultural music ($n = 82$). Participants of the first group took part in 20 piano lesson sessions for 6 months, while the other group attended to lessons focused on listening to music for the same time period. At the end of each session, both groups completed the assigned homework for about 30 min per day. The authors measured speech reception threshold (SRT) before and after the 6-month training with the International Matrix Test (Jansen et al., 2012) that is a common clinical test used by audiologists. Globally, the results after 6 months of musical training suggested an improvement for binaural SRTs in both groups by an average of −0.14 dB. Furthermore, the piano playing group improved their left SRT by −0.46 dB in comparison to the music listening group that did not show any improvement.

Taken together, these four studies support the hypothesis that musical training started at an advanced age would be beneficial in improving speech-in-noise processing skills. Also, the first study supports that pitch discrimination threshold seems to be improved. Further investigation is needed to measure if auditory scene analysis skills in general are improved following musical training that started at an advanced age. Results for speech in noise are positive, but the amount of evidence indicates that there is a lack of research on the subject. To conclude formally would be premature, even only for that auditory ability.

## IMPACT ON POSTURE AND BALANCE

As discussed earlier, the vestibular system is also a sensory organ in the inner ear affected by normal aging. Presbyvestibulopathy can lead to problems such as difficulties to maintain balance or to perform movements that ultimately lead to increased risk of falls. Considering that learning a musical instrument is a multisensory activity and that it is easily possible to create multisensory workouts based on music, the use of music in rehabilitation could be an interesting avenue to compensate the balance and movement disorders related to normal aging. However, only two studies have looked at the question.

Auditory input has an impact on postural control and balance (e.g., Houde et al., 2016; Maheu et al., 2017a,b, 2019). It has been suggested that the auditory input can be used to compensate for vestibular loss (e.g., Maheu et al., 2019), which is a common problem among older adults (Baloh et al., 2001). By improving hearing skills, as suggested by many studies (for a review, see Kraus and Chandrasekaran, 2010), music therapy could therefore have an indirect impact on postural control and balance.

Hamburg and Clair (2003) formed a group of 16 people aged between 65 and 78 years to participate in 14 weeks of training. It consisted of a 1-h-per-week movement sequence set to music to reflect the dynamics, rhythm, timing, and phrasing of music. Balance, gait speed, and functional reach measurements were taken before training, at 5 weeks, and at the end of training. After 5 weeks, significant improvements were noted for measurements of one foot balance, walking speed, and functional reach. These results therefore suggest the potential use of music as a workout to improve balance in the elderly.

Similarly, Maclean et al. (2014) recruited 45 participants over the age of 65 years. They divided them into three distinct groups. The first group did a short workout before doing the task; they had to practice walking in rhythm to music until they were comfortable doing so. The second group also had music in the background during the walking task to accompany them but was not instructed to walk to the music. Finally, the third group did not have any training or background music during the walking task. All groups were asked to walk two times a distance of 15 m with or without background music depending on the group. They were also asked to subtract out loud from an initial number concurrently to reflect their cognitive performance during one of the two trials (single task vs. dual tasks). During the task, mean step time, velocity, number of steps, and step-time variability were measured. The results suggest that compared to the other two groups, the music trained group participants showed no dual task deficit in speed. Furthermore, after the music training, their gait stability for both the single and dual tasks improved. No improvement post-training was measured for cognitive performance demonstrating neither decline nor improvement as their gait became steadier. Furthermore, it was only the condition with training that showed this improvement. Music in the background was not sufficient to improve stability of walk.

Overall, the two studies that have looked at gait and movement following music training in older adults suggest potential improvements. The second study even suggested that a short training with music in the background may be sufficient to improve gait stability during walk in the elderly. Due to the limited amount of study and the fact that none of these has investigated whether their participants were hard of hearing or had vestibular disorders, these conclusions are, however, uncertain. When combined with evidence suggesting potential hearing improvements, these results suggest that music training could be a powerful tool for vestibular rehabilitation, but there is still a need to investigate further to provide a formal recommendation.

## IMPACT ON COGNITIVE ABILITIES

As mentioned earlier, the aging of the world population leads to a larger number of individuals presenting cognitive decline. The prevalence of dementia is projected to double every 20 years (Lin, 2011). An increasing number of studies are suggesting a direct link between hearing loss and cognitive decline (e.g., Lin et al., 2013). A few studies have investigated musical therapy as a way to improve cognitive function in older adults without dementia.

One recent study used a category fluency task (verbal fluency ability) to investigate the impact of listening to music on cognition. Thompson et al. (2005) investigated the effect of listening to an excerpt of Vivaldi's *The Four Seasons* on Category fluency in healthy older adults (control group) and Alzheimer's disease patients. Thompson et al. (2005) showed that Vivaldi's *The Four Seasons: Winter* had the largest effect on cognitive performance, which is why it was chosen for this research. For the purpose of this review, only the control group will be described. Sixteen healthy and older volunteers (5 males, 11 females, mean of age 74.94 years) were recruited for this study. Four fluency categories were used (fruits, colors, vehicles, and furniture). Participants completed two 1-min category fluency tasks while listening to music and two 1-min category fluency tasks without music. They had to name as many examples as they could among these categories. Each participant was tested individually in a quiet room. In the music condition, participants were told they would hear music. The results showed that the facilitating effect of music was small but considered significant because category fluency performance in the music condition exceeded the performance in the no-music condition.

Another study investigated the impact of music listening on verbal fluency. Mammarella et al. (2007) investigated phonemic fluency and short-term auditory memory in their study. They used Vivaldi's music to investigate the impact of music listening on cognitive abilities during two tasks: (1) number sequence repetition (digit span, maximum of 8) and (2) to say spontaneously as much words as possible in 60 s, starting with a specific letter (phonemic fluency, maximum of 34 words). There were 24 adults, aged between 73 and 86 years (mean of 81). The participants were non-musicians and considered in good health. No training was performed before the testing. The study adopted a repeated-measures design with type of background. There were three types of background (Vivaldi's music vs. white noise vs. no music) randomized between participants. Half of the participants were presented with the digit span first, and the other half took the fluency test first. The results for the digit span (verbal working memory) showed that the background noise had a significant effect. The mean digit span scores in the music condition were 5.25 (SD = 1.03), 4.67 (SD = 0.8) with the white noise and 4.33 (SD = 1.20) in the no-music condition. Planned comparison showed that there was a significant advantage of the music condition over the white noise and the no music condition. No difference was observed between white noise condition and the no-music condition. Concerning the phonemic fluency (verbal fluency), the results also revealed a significant effect of background noise. The mean phonemic fluency scores were 25.83 (SD = 8.23) in the music condition, 20.50 (SD = 8.63) with the white noise, and 19.62 (SD = 7.59) in the no-music condition. Planned comparisons showed the significant advantage of music over white noise and no music.

Likewise, another study used long-term training to investigate the impact of musical therapy on cognitive function. The digit span task was also used as well as many other tasks looking at the hearing/speaking dimension and visual and executive functions. Bugos et al. (2007) investigated if individualized piano instruction (IPI) can be a potential cognitive intervention to mitigate normal age-related cognitive decline in older adults. Thirty-one non-musicians aged between 60 and 85 years participated in this 9-month study. They were divided into an experimental group ($n$ = 16) and a control group ($n$ = 15). First, they passed a test of overall cognitive and music abilities as well as a test for working memory and executive functions in order to confirm that both groups had similar mental abilities. The experimental group received a 6-month IPI, and the control group did not. The IPI was a half-hour lesson each week, and the participants had to practice at least 3 h per week. Following the 6 months, they had 3 months without formal practice. Four subtests [digit symbol (speed, visual working memory, visuospatial processing, and visual attention), digit span (verbal working memory), block design (spatial visualization ability and motor skill), and letter number sequencing (verbal working memory)] were repeated at all three time points (before the IPI, after the 6 months, and after the 3 months of no formal practice). Also, the Trail Making Test (TMT) that assesses visual attention and task switching (Reitan and Wolfson, 1985) was administered at each of the three points. The results demonstrated that the experimental group improved over time for TMT (part B) (passing from 98.4 to 72.1 s) and digit symbols (raw score passing from 50.3 to 72.6) but not the control group. No other differences were measured between groups, suggesting improvements only for visual processing (visual working memory, visual attention, and visuospatial processing). No improvements were measured for verbal working memory in contradiction to those of Mammarella et al. (2007).

Finally, Hirokawa (2004) investigated the effect of listening to music and relaxation on arousal and verbal working memory. For this study, 15 volunteer women aged between 66 and 80 years (mean 72.7) with no dementia and no history of musicianship participated in one preliminary session in order to make sure they understood the task. They were then scheduled for three other sessions (on three different days), which took approximately 20 min each. Those four sessions were conducted in a quiet room in each subject's respective residence. Subjects first started with an Activation–Deactivation Adjective Check List (AD ACL by Thayer, 1978 for arousal level), which is a multidimensional self-rating test constructed and extensively validated for rapid assessments of momentary activation or arousal states. Following the test, the participants were subjected to a 10-min experimental intervention in a random order. These interventions were as follows: subject preferred (a) music chosen over a specific list, (b) relaxation instructions (the modified autogenic training phrases adopted by Alice Green were used as the relaxation instructions), or (c) silence. The subjects adjusted the loudness for their comfort for music and relaxation sessions. Following the 10 min of experimental intervention, the subjects again completed the AD ACL, followed by the reading span test (Daneman and Carpenter, 1980). This test requires participants to read series of unconnected sentences aloud and to remember the final word of each sentence in a series. The subjects were presented increasingly longer sets of sentences until they failed. The results showed that music and relaxation had a positive effect on arousal, but no differences were observed on the verbal working memory performance for any of the three conditions.

Globally, the studies concerning the effect of music on older persons' cognition abilities tend to show that there is an improvement only for verbal fluency, visual processing, and arousal. This conclusion is, however, weak because the connection between each of these studies is hard to establish because the cognitive tasks used in each study were mostly different. Additionally, the assessing conditions sometimes differed. In some of the studies, the participant was able to choose the intensity of the music, which can modify the task's condition. There is also no information concerning the hearing status of the participants. Being in a quiet environment does not necessarily mean that the participant will perfectly understand speech, which was necessary for a lot of the tasks presented in this section. It is clear that the scientific community is keen on pursuing investigations to evaluate the impact of musical training on cognitive abilities. For example, a study protocol was recently published by James et al. (2020). Their study will follow 155 healthy elderly for 12 months to measure the impact of music training vs. musical listening on cognitive and perceptual–motor aptitudes with behavioral and fMRI tasks. Perhaps this new investigation will answer some remaining questions that will reveal the power of musical training as a rehabilitative tool.

## IMPACT ON SOCIAL PARTICIPATION

Social and emotional spheres of life of older adults are affected by hearing loss due to age (World Health Organization, 2020).

Music therapy, which is a multisensory training, could be a beneficial way to improve side effects of sensory loss such as social limitations. A few studies did large-scale qualitative investigation in the population to understand how music was improving or decreasing social aspects of older adults' life. For example, Cohen et al. (2002) integrated a questionnaire in the second phase of the Canadian Study of Health and Aging (CSHA2) to evaluate the importance of music in an elder's life. Based on 300 respondents, results revealed that most seniors evaluated the importance of music as high. Furthermore, Hays and Minichiello (2005) measured the significance of music to elders and the contribution of music to their quality of life. After questioning 19 elders, their qualitative results suggest that music provides a connection between the elders and their memories in a way that it brings a sense of spirituality to life. These qualitative investigations suggest that music is seen as important by older adults and that it influences positively their motivation to participate in music therapy. It is well known that patient's motivation can greatly influence the outcome of a rehabilitation process (Maclean et al., 2000).

Furthermore, some studies did an investigation on social participation in groups of older adults participated weekly to musical activities such as choirs and bands. Wise et al. (1992) studied the relationship between the dimensions of successful aging and singing in a choir. Forty-nine choir participants were compared to 49 non-musician elders matched for gender and marital status. Two main answers were obtained when asking members why they participated in a choir. Firstly, participants mentioned that they loved to sing and make music (musical component). Secondly, participants enjoyed associating with other members of the chorus (non-musical component). In accordance with these results, Hillman (2002) looked at the perceived benefit of participatory singing for seniors. A questionnaire formed of 33 items was sent to members of a choir. The answers given by participants suggest that they perceived several benefits from joining the choir. In fact, 24% of the elders mentioned an improvement of their musical abilities, and 22% reported benefits in their social life. Moreover, increased emotional wellbeing (14%), improved interest in art events (12%), and enhanced physical health (2%) were other benefits reported by the chorus members. These findings suggest a global improvement in quality of life. Furthermore, the elders rated an improvement in their social relationships. These two studies support weakly the fact that singing in a choir can have a positive impact on social life of the elderly in addition to the musical benefits and can improve globally the quality of life of older adults.

Furthermore, Coffman and Adamek (1999) asked 52 volunteer members of a wind band several questions about their quality of life and their perception of the benefits gained from participating in musical activities. First, when they were asked why they joined the band, their motivations were mostly improving their musical skills but also, in second place, to improve their social relationships. They were also asked to describe their perceived benefits from joining the band. The principal outcome was the improvement of musical skills, but the second most given answer was related to social benefits. Taken together, these three studies suggest a similar conclusion that musical activities in groups (band or choir) lead to improvement in social participation. Their findings remain limited because none of these studies compared their participants to a group of older adults doing another leisure activity or to a do-nothing group, which can bias the results.

Finally, only four studies took elders were not already participating in musical activities and introduced them to different kinds of music therapy. Vanderark et al. (1983) looked at the influence of music participation in two groups of elders following a 5-week program of music therapy. Those entered the program were at least 60 years old, and they had to be able to hear. Members were from two different private nursing homes. In the first center, 20 elders constituted the experimental group, received two 45-min sessions of the music program per week. In the second nursing home, 23 elders constituted the control group, did not receive music therapy. The association of the nursing home to either condition was random. Vanderark et al. (1983) administered three self-created questionnaires measuring five variables: self-concept, life satisfaction, quality of life, self-concept in music, and attitudes toward music. Results suggest that life satisfaction of the music program participants was improved in comparison to the do-nothing group. Also, gains for music attitude scores were higher in the experimental group than the control group. Globally, the experimental group had significant improvements for the five previous cited variables compared to the control group.

Solé et al. (2010) evaluated how participating in a choir, a music appreciation class, or a preventive music therapy session can influence the elders' quality of life. Eighty-three elders joined one of these three activities for 9 months. Questionnaires were administrated pre-training and post-training. Globally, results from all questionnaires suggest that the three programs seemed to have positively changed the life of the participants. Also, results suggest that all music programs enhanced the quality of life of the participants.

Chan et al. (2012) conducted a randomized controlled trial to measure the effects of music on depression symptoms in elders. The research team enrolled 50 participants in their trial. In the experimental condition, the elders did 30-min listening sessions ($n = 24$). The control group ($n = 26$) did not receive any music intervention. The data were collected once a week after the music session for the experimental group (weeks 2–8) and after the rest period for the control group (weeks 1–8). The scale used to measure the level of depression symptoms was the Geriatric Depression Scale (GDS-15). The results from this study suggested that the elders with music intervention had significantly lower scores on the depression scale compared to the control group every week of the trial.

Ahessy (2016) investigated the impact of music therapy to help reduce depression symptoms and improve quality of life of seniors in a randomized controlled trial. Forty participants from a long-term residential unit and a day care center in Dubai were assigned to the control or the experimental group. The participants received a musical treatment participated in the choir for 12 weeks at a rate of 1 h per week. In opposition, the control group received the normal nursing care and they were offered, after their participation in the study, four choral sessions. Several tests were administered at the beginning and at the end of the 12-week intervention. The first questionnaire, the Cornell Scale for Depression in Dementia (CSDD), evaluated the depression level in patients with or without dementia. The Cornell Brown Scale (CBS) was the second questionnaire used to evaluate the quality of life. Also, the cognitive functioning was rated with the Mini Mental State Examination (MMSE). Finally, the experimental group could answer the choir evaluation questionnaire (CEQ) to express their thoughts about the intervention. To reduce the possible effects of the circadian rhythm and enhance the reliability of the measures, tests were administered at the same time of the day. The findings revealed that the musical intervention reduced significantly the depressive symptom scores on the CSDD for participants of the choir compared to the control group. Also, the quality of life of the seniors was significantly enhanced after the 12-week program. This improvement was not present in the control group; on the contrary, there was a decrease in their quality of life. The post-intervention results were significantly higher than (a) the results obtained before the beginning of the music therapy and (b) the results obtained for the control group after 12 weeks of standard nursing care. The cognitive functioning was also significantly improved in the experimental group compared to the control group. Regarding the benefits gained from participating in the choir (CEQ), the most answered reasons were firstly to learn new skills and secondly to maintain social relationships. These results support the importance of implementing music therapy to help reduce depression symptoms and improve quality of life of the elders.

In conclusion, a limited number of studies suggest quality of life benefits (perceived and objective) from implementing musical therapy in the life of the elders. In fact, maintaining social relationships has often been perceived as an advantage gained from participating in several music programs. Another benefit perceived from music therapy is that it helps to reduce depression symptoms in seniors. None of these studies have considered if their participants had a hearing loss, which normally accompanies aging. One study did a systematic review to evaluate the benefits of music therapy in patients were diagnosed with a chronic disease (Quach and Lee, 2017). Their results suggest that music intervention helps these patients to reduce depression symptoms. In fact, hearing loss is considered a chronic disease, but the reality of presbycusis is far from diseases like Parkinson, cancer, and diabetes considered in this study. This reflects the importance of conducting studies on (a) the impact of hearing loss on social aspects in the elderly quality of life and (b) the influence of music therapy on seniors coping with hearing loss.

## DISCUSSION

In this review, our goal was to investigate the impact of musical training as an audiological rehabilitative tool for older adults to compensate for the difficulties linked with aging auditory and vestibular system decline and their indirect consequences leading to social and cognitive difficulties. Benefits of musical training on perceptual and cognitive skills are well documented (for a review, see Herholz and Zatorre, 2012). Furthermore, these benefits seem to persist despite the aging process (for a review, see Alain et al., 2014). Engagement in music performance over a lifetime seems to slow age-related decline for certain motor tasks (e.g., Krampe and Ericsson, 1996) and central auditory hearing abilities (e.g., Parbery-Clark et al., 2011; Zendel and Alain, 2012). Whether learning to play a musical instrument or follow a music-based training later in life can lead to similar improvements needs to be examined more closely. The main goal of this review was to establish the impact of music therapy in the context of hearing rehabilitation in older adults. We examined the literature to find studies that selected older non-musicians to participate in any form of music-based training to improve their auditory skills, gait, balance and movement, social participation, and cognitive abilities. Altogether, the big picture suggests that musical training started later in life can improve several abilities that can lead to better quality of life for the elderly. First, the three studies investigating central auditory abilities following a musical training suggest improvements for speech-in-noise perception and pitch representation (Dubinsky et al., 2019; Fleming et al., 2019; Zendel et al., 2019). Secondly, the two studies investigating gait and movements also suggest improvements for maintaining balance on one foot, walk speed, gait stability, and functional reach (Hamburg and Clair, 2003; Maclean et al., 2014). Furthermore, most studies investigating cognitive aspects following musical training suggest improvements for verbal fluency, visual processing, and arousal (Hirokawa, 2004; Thompson et al., 2005; Bugos et al., 2007; Mammarella et al., 2007). Also, social aspects of life seem to be positively influenced by participating in musical group activities for older adults. Finally, there are some major limitations to these studies that need to be addressed before making any recommendation.

One main limitation of all these studies is the lack of quantification of hearing loss. Only three studies measuring speech-in-noise abilities (Dubinsky et al., 2019; Fleming et al., 2019; Zendel et al., 2019) measured hearing thresholds of their participants. Only Dubinsky et al. (2019) used participants with typical presbycusis. The other two studies had participants were mostly normal hearing individuals or with mild hearing loss. Worschech et al. (2021) did not measure hearing thresholds of their participants; this is a major limitation of a study investigating speech-in-noise perception. All studies described above concerning gait and posture, cognitive abilities, and social abilities did not measure properly hearing threshold to quantify if their participants had hearing loss. Some are suggesting that their participants self-reported hearing well, but it is well known that it is not a good estimate of hearing loss. Hearing loss is most of the time unknown and highly stigmatized, and it takes around 10 years for an individual to realize his or her hearing difficulties (Davis L.J., 1995; Davis A., 1995). It is well known that hearing loss impacts social participation and heightens the risk of cognitive decline, and a high proportion of individuals suffering from hearing loss presents vestibular symptoms (e.g., Tien and Linthicum, 2002; Kaga et al., 2008; Cushing et al., 2013; Xu et al., 2015). This is why it should be a standard procedure to evaluate hearing of older participants when investigating any sort of abilities. Furthermore, hearing loss measurement will help provide evidence for the use of musical training as a rehabilitative tool for auditory abilities and to improve general quality of life of older adults. To do so, there is a need to correlate the degrees of hearing loss with the efficiency of musical training to guide clinicians in using this tool. For now, it is not possible to determine if only certain degrees of hearing loss can be helped *via* musical training versus, for example, more profound hearing loss where this tool would not be useful. More investigation is needed to clarify that important aspect.

Furthermore, none of these studies investigated if their participants had abnormal vestibular function. Results may have been different for all studies, not only for gait and posture, if that variable had been taken into account. There is a need to correlate the hearing and vestibular functions of participants with the results to provide a clear recommendation on whether or not to use musical therapy in audiological rehabilitation. Also, Maheu et al. (2019) suggested that auditory cues influence postural sway in individuals with vestibular dysfunction. Hearing aids can help to maintain balance for these individuals. No studies have yet asked their participants if they were using hearing aids or have reported to do so. More investigation is needed to correlate if there is a need to combine hearing aids used with musical training to obtain successful results with older individuals.

The measures of outcomes for each category of this review were disparate, which makes them difficult to compare. In the auditory abilities section, the authors were mostly focusing on speech-in-noise perception, but their methods of measuring this ability were different (word vs. phrases, different SNRs, etc.), which leads to a limited comparison. Also, whether musical training could improve other auditory analysis skills remains unknown. Both articles on gait and posture used really different outcomes too (balance on one foot, walk speed and functional reach vs. walking speed). For social aspects, the questionnaires were all different, but most of them were leading to similar conclusions about improvement of quality of life and social participation. There is still a main limitation for all social studies: they did not compare their participants to another group of older adults doing another leisure activity (not musical). Before concluding that musical training can help social participation, there is a need to investigate if just doing a leisure activity in general helps or if it is specific to music. Also, none of these studies reported if participants were doing more social activities in general after the musical training, which is an important question that needs to be addressed in further studies. Finally, the section on cognition was sparser in terms of evaluation. The batteries of evaluation were totally different from one study to another. Only one conclusion was comparable between two studies: both Thompson et al. (2005) and Mammarella et al. (2007) found improvements for verbal fluency. All other conclusions were supported by only one of the four studies described in the section. There were also huge differences between musical training among all studies. Some were focused on instrument learning or singing, which is more of a multisensory training, versus others that were more of a passive training (listening to music, exercise online, etc.). Further studies should investigate all aspects (auditory, vestibular, social, and cognition) described in this review in one study *via* a multidisciplinary evaluation. Also, there is a need to investigate which type of training is more efficient (music-based programs or learning an instrument, etc.). The duration of the training could also influence the outcome. There is also a need to investigate if improvement for each ability measured is maintained over time. Furthermore, investigation with electrophysiological measurement or imagery should be done to measure the extent of plasticity in older adults following the musical training in order to better understand neural correlates linked with their behavioral improvements. These will inform us about the extent of possibilities in terms of cortical reorganization by using musical training as a rehabilitation tool.

Similarly to this literature review, but in another area field of research, Sihvonen et al. (2017) have assessed the potential rehabilitative neurological effects of music-based intervention. In their case, they summarized results from musical therapy focusing on patients with stroke, dementia, Parkinson, epilepsy, multiple sclerosis, etc. They reached a similar conclusion to ours but for neurological rehabilitation: results tend to suggest a positive impact on patients follow a musical training program, but further controlled studies are needed to establish the efficacy of music in neurological recovery.

## CONCLUSION

Globally, music-based rehabilitation interventions are emerging as promising strategies. Further investigations are still needed, but all studies discussed above tend to suggest that musical training leads to improvements on all aspects targeted by this literature review (communication, posture and balance, cognitive abilities, and social participation). For future studies, researchers should measure hearing thresholds and vestibular function of their participants to evaluate if the improvements would be greater for elderly have hearing loss and/or vestibular impairments. Screening tools are available to do so, and clinicians working in rehabilitation should be encouraged to use them in order to fully understand the portrait of their patients. Also, a multidisciplinary point of view should be the focus of further research to take into account all the facets impacted by normal aging and promote aging well. The studies discussed, across all fields, used disparate musical therapies and outcome measurements. Further research is needed to identify the best musical training for each outcome because if simply listening to music can be helpful for elderly, it would be much easier to integrate into the retirement centers. However, this type of training may not be ideal for maximizing improvements considering that it does not necessarily include motor aspects compared to learning a musical instrument. Finally, neural correlates *via* EEG or fMRI should be integrated in future research to measure the possible extent of cortical plasticity as to objectify improvements.

Finally, a first step to improve the quality of life of elderly people with hearing loss would be simply to start integrating music in audiological rehabilitation programs. Also, giving access to music to elderly people in retirement homes could be a first step (e.g., to give them access to music lessons, to make musical instruments available to them, to develop choirs in centers for the elderly, or simply to give them access to portable music players to allow them to listen to the music they enjoy). In conclusion, music seems to be a powerful tool, and hearing healthcare professionals should start using it in audiological rehabilitation programs. Still, more investigations are needed to conclude the extent of benefits that this tool could bring to older adults.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This research was funded by the Cervo Brain Research Center startup grant of Sharp Laboratory.

## REFERENCES

Abrams, H. B., and Kihm, J. (2015). An introduction to MarkeTrak IX: a new baseline for the hearing aid market. *Hear. Rev.* 22:16.

Ahessy, B. (2016). The use of a music therapy choir to reduce depression and improve quality of life in older adults–a randomized control trial. *Music Med.* 8, 17–28. doi: 10.47513/mmd.v8i1.451

Alain, C., Zendel, B. R., Hutka, S., and Bidelman, G. M. (2014). Turning down the noise: the benefit of musical training on the aging auditory brain. *Hear. Res.* 308, 162–173. doi: 10.1016/j.heares.2013.06.008

Baloh, R. W., Enrietto, J., Jacobson, K. M., and Lin, A. (2001). Age—related changes in vestibular function: a longitudinal study. *Ann. N. Y. Acad. Sci.* 942, 210–219. doi: 10.1111/j.1749-6632.2001.tb03747.x

Bugos, J. A., Perlstein, W. M., McCrae, C. S., Brophy, T. S., and Bedenbaugh, P. H. (2007). Individualized piano instruction enhances executive functioning and working memory in older adults. *Aging Ment. Health* 11, 464–471. doi: 10.1080/13607860601086504

Chan, M. F., Wong, Z. Y., Onishi, H., and Thayala, N. V. (2012). Effects of music on depression in older people: a randomised controlled trial. *J. Clin. Nurs.* 21, 776-783.

Chasin, M., and Russo, F. A. (2004). Hearing aids and music. *Trends Amplif.* 8, 35–47.

Coffman, D. D., and Adamek, M. S. (1999). The contributions of wind band participation to quality of life of senior adults. *Music Ther. Perspect.* 17, 27–31. doi: 10.1093/mtp/17.1.27

Cohen, A., Bailey, B., and Nilsson, T. (2002). The importance of music to seniors. *Psychomusicology* 18:89. doi: 10.1037/h0094049

Cushing, S. L., Gordon, K. A., Rutka, J. A., James, A. L., and Papsin, B. C. (2013). Vestibular end-organ dysfunction in children with sensorineural hearing loss and cochlear implants: an expanded cohort and etiologic assessment. *Otol. Neurotology* 34, 422–428. doi: 10.1097/mao.0b013e31827b4ba0

Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verbal Learn. Verbal Behav.* 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6

Davis, A. (1995). *Hearing in Adults: The Prevalence and Distribution of Hearing Impairment and Reported Hearing Disability in the MRC Institute of Hearing Research's National Study of Hearing*. London: Whurr Publishers, 43–321.

Davis, L. J. (1995). *Enforcing Normalcy: Disability, Deafness, and the Body*. Brooklyn, NY: Verso.

Dubinsky, E., Wood, E. A., Nespoli, G., and Russo, F. A. (2019). Short-term choir singing supports speech-in-noise perception and neural pitch strength in older adults with age-related hearing loss. *Front. Neurosci.* 13:1153. doi: 10.3389/fnins.2019.01153

Feldmann, H., and Kumpf, W. (1988). Listening to music in hearing loss with and without a hearing aid. *Laryngol. Rhinol. Otol.* 67, 489–497.

Fleming, D., Belleville, S., Peretz, I., West, G., and Zendel, B. R. (2019). The effects of short-term musical training on the neural processing of speech-in-noise in older adults. *Brain Cogn.* 136:103592. doi: 10.1016/j.bandc.2019.103592

Gates, G. A., and Mills, J. H. (2005). Presbycusis. *Lancet* 366, 1111–1120.

Hamburg, J., and Clair, A. A. (2003). The effects of a movement with music program on measures of balance and gait speed in healthy older adults. *J. Music Ther.* 40, 212–226. doi: 10.1093/jmt/40.3.212

Hays, T., and Minichiello, V. (2005). The meaning of music in the lives of older people: a qualitative study. *Psychol. Music* 33, 437–451. doi: 10.1177/0305735605056160

Herholz, S. C., and Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron* 76, 486–502. doi: 10.1016/j.neuron.2012.10.011

Hillman, S. (2002). Participatory singing for older people: a perception of benefit. *Health Educ.* 102, 163–171. doi: 10.1108/09654280210434237

Hirokawa, E. (2004). Effects of music listening and relaxation instructions on arousal changes and the working memory task in older adults. *J. Music Ther.* 41, 107–127. doi: 10.1093/jmt/41.2.107

Houde, M. S., Landry, S. P., Pagé, S., Maheu, M., and Champoux, F. (2016). Body perception and action following deafness. *Neural Plast.* 2016:5260671.

Hülse, R., Biesdorf, A., Hörmann, K., Stuck, B., Erhart, M., Hülse, M., et al. (2019). Peripheral vestibular disorders: an epidemiologic survey in 70 million individuals. *Otol. Neurotol.* 40, 88–95. doi: 10.1097/mao.0000000000002013

Jahn, K. (2019). The aging vestibular system: dizziness and imbalance in the elderly. *Vestib. Disord.* 82, 143–149. doi: 10.1159/000490283

James, C. E., Altenmüller, E., Kliegel, M., Krüger, T. H., Van De Ville, D., Worschech, F., et al. (2020). Train the brain with music (TBM): brain plasticity and cognitive benefits induced by musical training in elderly people in Germany and Switzerland, a study protocol for an RCT comparing musical instrumental practice to sensitization to music. *BMC Geriatr.* 20, 1–19. doi: 10.1186/s12877-020-01761-y

Jansen, S., Luts, H., Wagener, K. C., Kollmeier, B., Del Rio, M., Dauman, R., et al. (2012). Comparison of three types of French speech-in-noise tests: a multi-center study. *Int. J. Audiol.* 51, 164–173. doi: 10.3109/14992027.2011.633568

Kaga, K., Shinjo, Y., Jin, Y., and Takegoshi, H. (2008). Vestibular failure in children with congenital deafness. *Int. J. Audiol.* 47, 590–599. doi: 10.1080/14992020802331222

Kishon-Rabin, L., Amir, O., Vexler, Y., and Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *J. Basic Clin. Physiol. Pharmacol.* 12(Suppl. 2), 125–144.

Krampe, R. T., and Ericsson, K. A. (1996). Maintaining excellence: deliberate practice and elite performance in young and older pianists. *J. Exp. Psychol. Gen.* 125, 331–359. doi: 10.1037/0096-3445.125.4.331

Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605. doi: 10.1038/nrn2882

Lin, F. R. (2011). Hearing loss and cognition among older adults in the United States. *J. Gerontol. A Biomed. Sci. Med. Sci.* 66, 1131–1136. doi: 10.1093/gerona/glr115

Lin, F. R., Yaffe, K., Xia, J., Xue, Q. L., Harris, T. B., Purchase-Helzner, E., et al. (2013). Hearing loss and cognitive decline in older adults. *JAMA Intern. Med.* 173, 293–299.

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., et al. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* 396, 413–446. doi: 10.1016/s0140-6736(20)30367-6

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., et al. (2017). Dementia prevention, intervention, and care. *Lancet* 390, 2673–2734.

Maclean, L. M., Brown, L. J., and Astell, A. J. (2014). The effect of rhythmic musical training on healthy older adults' gait and cognitive function. *Gerontologist* 54, 624–633. doi: 10.1093/geront/gnt050

Maclean, N., Pound, P., Wolfe, C., and Rudd, A. (2000). A critical review of the concept of patient motivation in the literature on physical rehabilitation. *Soc. Sci. Med.* 50, 495–506.

Maheu, M., Behtani, L., Nooristani, M., Houde, M. S., Delcenserie, A., Leroux, T., et al. (2019). Vestibular function modulates the benefit of hearing aids in people with hearing loss during static postural control. *Ear Hear.* 40, 1418–1424. doi: 10.1097/aud.0000000000000720

Maheu, M., Sharp, A., Landry, S. P., and Champoux, F. (2017a). Sensory reweighting after loss of auditory cues in healthy adults. *Gait Posture* 53, 151–154. doi: 10.1016/j.gaitpost.2017.01.015

Maheu, M., Sharp, A., Pagé, S., and Champoux, F. (2017b). Congenital deafness alters sensory weighting for postural control. *Ear Hear.* 38, 767–770. doi: 10.1097/aud.0000000000000449

Mammarella, N., Fairfield, B., and Cornoldi, C. (2007). Does music enhance cognitive performance in healthy older adults? The Vivaldi effect. *Aging Clin. Exp. Res.* 19, 394–399. doi: 10.1007/bf03324720

Marrone, N., Mason, C. R., and Kidd, G. Jr. (2008). Evaluating the benefit of hearing aids in solving the cocktail party problem. *Trends Amplifi.* 12, 300–315. doi: 10.1177/1084713808325880

McDermott, J. H. (2009). The cocktail party problem. *Curr. Biol.* 19, R1024–R1027.

Münte, T. F., Altenmüller, E., and Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nat. Rev. Neurosci.* 3, 473–478.

Nadhimi, Y., and Llano, D. A. (2020). Does hearing loss lead to dementia? A review of the literature. *Hear. Res.* 402:108038. doi: 10.1016/j.heares.2020.108038

Nondahl, D. M., Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., and Dalton, D. S. (2013). Sixteen-year change in acoustic-admittance measures among older adults: data from a population-based study. *J. Speech Lang. Hear. Res.* 56, 1745–1750. doi: 10.1044/1092-4388(2013/12-0381)

Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., and Kraus, N. (2011). Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PloS One* 6:e18082. doi: 10.1371/journal.pone.0018082

Pauley, S., Kopecky, B., Beisel, K., Soukup, G., and Fritzsch, B. (2008). Stem cells and molecular strategies to restore hearing. *Panminerva Med.* 50, 41–53.

Quach, J., and Lee, J. A. (2017). Do music therapies reduce depressive symptoms and improve QOL in older adults with chronic disease? *Nursing* 47, 58–63. doi: 10.1097/01.nurse.0000513604.41152.0c

Reitan, R. M., and Wolfson, D. (1985). *The Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation*, Vol. 4. Tucson, AZ: Neuropsychology Press.

Schlaug, G., Jancke, L., Huang, Y., and Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science* 267, 699–701. doi: 10.1126/science.7839149

Schneider, P., Andermann, M., Wengenroth, M., Goebel, R., Flor, H., Rupp, A., et al. (2009). Reduced volume of Heschl's gyrus in tinnitus. *Neuroimage* 45, 927–939. doi: 10.1016/j.neuroimage.2008.12.045

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* 5, 688–694. doi: 10.1038/nn871

Shinn-Cunningham, B. G., and Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplifi.* 12, 283–299. doi: 10.1177/1084713808325306

Sihvonen, A. J., Särkämö, T., Leo, V., Tervaniemi, M., Altenmüller, E., and Soinila, S. (2017). Music-based interventions in neurological rehabilitation. *Lancet Neurol.* 16, 648–660. doi: 10.1016/s1474-4422(17)30168-0

Slade, K., Plack, C. J., and Nuttall, H. E. (2020). The effects of age-related hearing loss on the brain and cognitive function. *Trends Neurosci.* 43, 810–821. doi: 10.1016/j.tins.2020.07.005

Solé, C., Mercadal-Brotons, M., Gallego, S., and Riera, M. (2010). Contributions of music to aging adults' quality of life. *J. Music Ther.* 47, 264–281. doi: 10.1093/jmt/47.3.264

Spiegel, M. F., and Watson, C. S. (1984). Performance on frequency−discrimination tasks by musicians and nonmusicians. *J. Acoust. Soc. Am.* 76, 1690–1695. doi: 10.1121/1.391605

Sweetow, R., and Palmer, C. V. (2005). Efficacy of individual auditory training in adults: a systematic review of the evidence. *J. Am. Acad. Audiol.* 16, 494–504. doi: 10.3766/jaaa.16.7.9

Thayer, R. E. (1978). Factor analytic and reliability studies on the activation-deactivation adjective check list. *Psychol. Rep.* 42, 747–756. doi: 10.2466/pr0.1978.42.3.747

Thompson, R. G., Moulin, C. J. A., Hayre, S., and Jones, R. W. (2005). Music enhances category fluency in healthy older adults and Alzheimer's disease patients. *Exp. Aging Res.* 31, 91–99. doi: 10.1080/03610730590882819

Tien, H. C., and Linthicum, F. H. Jr. (2002). Histopathologic changes in the vestibule after cochlear implantation. *Otolaryngol. Head Neck Surg.* 127, 260–264. doi: 10.1067/mhn.2002.128555

Vanderark, S., Newman, I., and Bell, S. (1983). The effects of music participation on quality of life of the elderly. *Music Ther.* 3, 71–81. doi: 10.1093/mt/3.1.71

White-Schwoch, T., Carr, K. W., Anderson, S., Strait, D. L., and Kraus, N. (2013). Older adults benefit from music training early in life: biological evidence for long-term training-driven plasticity. *J. Neurosci.* 33, 17667–17674. doi: 10.1523/jneurosci.2560-13.2013

Wise, G. W., Hartmann, D. J., and Fisher, B. J. (1992). Exploration of the relationship between choral singing and successful aging. *Psychol. Rep.* 70(Pt 2), 1175–1183. doi: 10.2466/pr0.70.4.1175-1183

World Health Organization. (2008). *Global Report on Falls Prevention in Older Age*.Geneva: World Health Organization.

World Health Organization. (2020). *Deafness and Hearing Loss*. Geneva: World Health Organization.

Worschech, F., Marie, D., Jünemann, K., Sinke, C., Krüger, T. H., Großbach, M., et al. (2021). Improved speech in noise perception in the elderly after 6 months of musical instruction. *Front. Neurosci.* 15:840. doi: 10.3389/fnins.2021.696240

Xu, X. D., Zhang, X. T., Zhang, Q., Hu, J., Chen, Y. F., and Xu, M. (2015). Ocular and cervical vestibular-evoked myogenic potentials in children with cochlear implant. *Clin. Neurophysiol.* 126, 1624–1631. doi: 10.1016/j.clinph.2014.10.216

Zendel, B. R., and Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychol. Aging* 27, 410–417. doi: 10.1037/a0024816

Zendel, B. R., West, G. L., Belleville, S., and Peretz, I. (2019). Musical training improves the ability to understand speech-in-noise in older adults. *Neurobiol. Aging* 81, 102–115. doi: 10.1016/j.neurobiolaging.2019.05.015

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 17 November 2021
doi: 10.3389/fpsyg.2021.650777

Check for updates

# Presence of Three-Dimensional Sound Field Facilitates Listeners' Mood, Felt Emotion, and Respiration Rate When Listening to Music

*Yuuki Ooishi[1*†], Maori Kobayashi[2,3†], Makio Kashino[1] and Kanako Ueno[3,4]*

*[1] NTT Communication Science Laboratories, NTT Corporation, Atsugi, Japan, [2] Faculty of Human Sciences, School of Human Sciences, Waseda University, Tokorozawa, Japan, [3] Department of Architecture, School of Science and Technology, Meiji University, Kawasaki, Japan, [4] Core Research for Evolutional Science and Technology, Japan Science and Technology Agency (CREST, JST), Tokyo, Japan*

OPEN ACCESS

**Edited by:**
*Marta Olivetti Belardinelli,
Sapienza University of Rome, Italy*

**Reviewed by:**
*Roger B. Dannenberg,
Carnegie Mellon University,
United States*
*Claudia Del Gatto,
Università Europea di Roma, Italy*

***Correspondence:**
*Yuuki Ooishi
ooishi-y@umin.org*

*†These authors have contributed equally to this work*

**Specialty section:**
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology*

**Received:** *08 January 2021*
**Accepted:** *15 October 2021*
**Published:** *17 November 2021*

**Citation:**
*Ooishi Y, Kobayashi M, Kashino M and Ueno K (2021) Presence of Three-Dimensional Sound Field Facilitates Listeners' Mood, Felt Emotion, and Respiration Rate When Listening to Music. Front. Psychol. 12:650777. doi: 10.3389/fpsyg.2021.650777*

Many studies have investigated the effects of music listening from the viewpoint of music features such as tempo or key by measuring psychological or psychophysiological responses. In addition, technologies for three-dimensional sound field (3D-SF) reproduction and binaural recording have been developed to induce a realistic sensation of sound. However, it is still unclear whether music listened to in the presence of 3D-SF is more impressive than in the absence of it. We hypothesized that the presence of a 3D-SF when listening to music facilitates listeners' moods, emotions for music, and physiological activities such as respiration rate. Here, we examined this hypothesis by evaluating differences between a reproduction condition with headphones (HD condition) and one with a 3D-SF reproduction system (3D-SF condition). We used a 3D-SF reproduction system based on the boundary surface control principle (BoSC system) to reproduce a sound field of music in the 3D-SF condition. Music in the 3D-SF condition was binaurally recorded through a dummy head in the BoSC reproduction room and reproduced with headphones in the HD condition. Therefore, music in the HD condition was auditorily as rich in information as that in the 3D-SF condition, but the 3D-sound field surrounding listeners was absent. We measured the respiration rate and heart rate of participants listening to acousmonium and pipe organ music. The participants rated their felt moods before and after they listened to music, and after they listened, they also rated their felt emotion. We found that the increase in respiration rate, the degree of decrease in well-being, and unpleasantness for both pieces in the 3D-SF condition were greater than in the HD condition. These results suggest that the presence of 3D-SF enhances changes in mood, felt emotion for music, and respiration rate when listening to music.

**Keywords: music, autonomic nerve, sound field, three dimension (3D), respiration, mood, emotion**

**Abbreviations:** BoSC, boundary surface control; 3D, three-dimensional; HRV, heart rate variability; HF, high frequency; LF, low frequency; AC, acousmonium; PM, pipe organ music; RVLM, rostral ventrolateral medulla; NA, nucleus ambiguus; DCMX, dorsal motor nucleus of the vagus.

## INTRODUCTION

Listening to music can elicit many kinds of emotional responses in humans (Chanda and Levitin, 2013; Koelsch, 2014). Psychological studies have revealed that music listening induces emotions and changes in moods (Blood et al., 1999; Nayak et al., 2000; Murrock and Higgins, 2009; Koelsch and Jancke, 2015). Psychophysiological studies have revealed that music-induced emotions are strongly associated with the modulation of the physiological system, which typically includes changes in heart rate, heart rate variability (HRV), and respiration rate. Listening to self-selected music that can cause chills induces increases in heart rate and respiration rate (Salimpoor et al., 2009, 2011). Listening to relaxing music suppresses stress-induced increases in heart rate (Knight and Rickard, 2001) and increases the amplitude of the high-frequency (HF) component of HRV, which is an index of parasympathetic nerve activity (Zhou et al., 2010). A similar symmetry property of music in physiological activities has been observed on other emotional axes such as happiness and sadness. An earlier study demonstrated that happy and sad excerpts can be distinguished with respiration rate (Etzel et al., 2006).

In investigations of the mechanisms that underlie music-induced emotion or its related physiological activities, musical features have been considered factors determining the effects of music listening on humans. For example, tempo has been considered a factor determining whether the effect of listening to music is exciting or relaxing. Listening to music with a fast tempo causes an increase in sympathetic nerve activity (Bernardi et al., 2006), while music with slow tempos reduces heart rate and is felt to be relaxing (Standley, 1986). The happy-sad music distinction depends on the combination of tempo and key (Hevner, 1935; Balkwill and Thompson, 1999). Gomez and Danuser (2007) have summarized the effects of several music features, such as intensity, tempo, rhythm, and accentuation, on the levels of arousal and valence or physiological activities.

However, the effect of being surrounded by three-dimensional (3D) sounds when listening to music has not been clarified. As represented by multichannel reproduction systems such as 22.2-channel sound systems, based on two-channel stereo (Hamasaki et al., 2008) or wave field synthesis techniques (Berkhout et al., 1993), technology for reproducing a 3D-sound field (3D-SF) has been developed to achieve a realistic sensation of sound, namely a sense of an acoustic field, which means that we feel as if we are in a real sound field, such as that of a live music performance. Developers of 3D-SF reproduction techniques have thought that being surrounded by 3D sounds is an important factor in music listening (Chen, 2001; Fukada, 2001). Actually, the presence of a 3D-SF enhances self-reported impressions of sound (e.g., presence, naturalness, brightness, reality, or beauty) compared with normal stereophonic sound (Hamasaki et al., 2008). In the sense of realizing a feeling of a sound field, however, binaurally recorded sounds also enable listeners to feel as if they were in the original sound field (Paul, 2009). Self-reported impressions of sound are higher when binaurally sounds recorded with a dummy head are reproduced through headphones than they are for normal stereophonic sounds through headphones (Paul, 2009). Both 3D-SF reproduction and binaural recording techniques have aimed at inducing a sense of sound realness, and the spatial audio information is almost the same between them. However, we believe that there are significant differences in psychological perception or physiological responses between a reproduced 3D-SF and binaurally recorded sound. Since our body is exposed to vibrations of air when we listen to sounds in a physically existing 3D-SF, we hypothesized that when we are exposed to music in the presence of a physical 3D-SF, the physical 3D-SF conveys information not only through auditory perception but also through other channels, such as tactile ones evoked by the effect of music-induced air vibration on the body of a listener, which might affect the psychological and physiological states of listeners. A couple of studies have indicated interactions between auditory and tactile information. One showed that musical meter recognition by musicians for auditory stimulation is disturbed with the presence of incongruent tactile stimulations (Huang et al., 2012). Another found that musicians have a higher ability of frequency discrimination in not only auditory but also tactile stimulation than non-musicians (Sharp et al., 2019). Since these two studies presented tactile stimulations on the fingertip or hand of participants independent of auditory stimulations, they did not examine the tactile sensation evoked by the air vibration that sound stimulations induce. Earlier studies showed the impact of listening to live musical performances. For example, live music was found to relieve tension-anxiety and enhance the vigor of patients with cancer compared with tape-recorded music (Bailey, 1983). More recent studies have discussed the effect of live music from the viewpoint of audiovisual interaction. The combination of audition and vision boosts the reward functions of an audience with higher electrodermal activities than when they are exposed to an audition-alone or vision-alone condition when listening to music (Chapados and Levitin, 2008). Other studies have shown the effects of the presence of other people at live music performances, including audiences and performers. The spontaneous presence of others has an influence on the reduction in cardiovascular reactivity to mental stress (Phillips et al., 2006). Sharing musical moments with lively performing musicians enhances the parasympathetic nerve activity more efficiently than when listening to recorded music, suggesting that not only verbal but also non-verbal communication occurs between the audience and performer (Shoda et al., 2016). However, these studies examining the effects of listening to live music have never assessed the potential effects of a physically existing 3D-SF on listeners of live music.

In this study, we examined whether the presence of a physical 3D-SF of music accelerates psychological and physiological responses when listening to music. To do this, we used a novel 3D-sound reproduction system (Omoto et al., 2015) and compared the effect of music reproduced through this system with that through headphones. Previous studies have measured mood (Nayak et al., 2000; Murrock and Higgins, 2009) and pleasantness or unpleasantness (Blood et al., 1999; Koelsch and Jancke, 2015) as psychological indices, and respiration rate and cardiac autonomic responses (Bernardi et al., 2006; Salimpoor et al., 2011) as physiological indices to examine the effects of music listening. Accordingly, this study also measured these indices.

## MATERIALS AND METHODS

### Participants

Thirty-two healthy students of Meiji University (age 22.5 ± 0.4 (SEM) years old, 3 females) participated in the experiment. None of them were musicians, and they did not have any musical training other than basic music courses in primary and secondary education. All procedures were conducted in accordance with the Declaration of Helsinki and approved by the Ethics and Safety Committees of Meiji University, and the participants were paid for their participation.

### Music Stimuli and Listening Environment

We selected two music genres for the experimental music stimuli: pipe organ music (PM) and acousmonium (AC). They were chosen because they both have rich echo, so listeners can effectively receive music-induced air vibration from all directions and feel as if they were surrounded by music. For the preparation of PM, a Japanese professional musician played Toccata and Fugue in D minor BWV 565 by Bach with a pipe organ in a church, which was recorded live using a fullerene-like microphone array[1] (**Figure 1A**). A piece of AC was composed by a Japanese musician, and he/she played it him/herself in a multipurpose hall with multiple speakers, which was also recorded live[2]. The lengths were 543 s for PM and 563 s for AC. The participants were very familiar with the introductory segment of Toccata and Fugue, but none of them had ever listened to it all the way through. They did not know the AC piece.

To elucidate the effect of the sound field of music, two types of sound reproduction conditions were prepared, namely a 3D-SF condition and an HD condition. We used the different sound presentation systems for each condition as follows.

### Three-Dimensional Sound Field Condition

In the 3D-SF condition, we used a 3D-SF reproduction system based on the boundary surface control (BoSC) principle (Ise, 1999), the details of which have been previously reported (Enomoto et al., 2011; Omoto et al., 2015). A 3D-SF reproduction system based on this principle is called a BoSC system (Enomoto et al., 2011). This system comprises a microphone array and a reproduction room. We used a fullerene-like microphone array consisting of 80 omnidirectional microphones (4060; DPA Microphones, Lillerød, Denmark), each of which was located in the position of the carbon molecule in C80 fullerene (**Figure 1A**). Music was recorded with the microphone array at a sampling rate of 48 kHz, and we used 24-bit quantization. Recorded music pieces were reproduced in the reproduction room, which contained seven-layer loudspeaker arrays based on height, forming what is called the Sound Cask (**Figure 1B**). Six loudspeakers (FE83E; Fostex, Tokyo, Japan) were installed on the ceiling plane. The other speakers were installed on surface of the wall at six different heights. Nine loudspeakers were installed at the highest and lowest points, and 18 were allotted to each of the remaining heights. The system accommodated 96 loudspeakers in total (**Figure 1B**). The BoSC principle supports the following idea (see **Figure 2**). An original sound is recorded with a microphone array in an enclosed area. This enclosed area is defined as recorded area $V$, the boundary surface enveloping recorded area $V$ is defined as $S$, and the sound field within recorded area $V$ is defined as $Sf$ (**Figure 2**, left). If the recorded original sound is reproduced in a reproduction room, and the sound field within the reproduced area $V'$, which has completely the same shape and size as recorded area $V$, is defined as $Sf'$, then $Sf'$ is mathematically equal to $Sf$ under the condition that the sound pressure at the boundary surface $S$ is perfectly reproduced at the boundary surface $S'$, which envelopes the reproduced area $V'$ (**Figure 2**, right) (Ise, 1999; Omoto et al., 2015). To best approximate this with our speakers, we set the fullerene-like microphone array in the listening position in the Sound Cask and recorded time stretched pulse (TSP) signals reproduced from the speakers. From the recorded sound data, we calculated the transfer function $T_{ij}$ between each speaker $SP_i$ (i = 1, 2, 3,. . .96) and each microphone $MC_j$ (j = 1, 2, 3,. . .80), and we prepared inverse filters (type-1 inverse filters) that consisted of the inverse transfer function $invT_{ij}$, by which the transfer function $T_{ij}$ between speakers and microphones can be cancelled (Enomoto et al., 2011) (**Supplementary Figure 1A**). We convolved the type-1 inverse filters to recorded music stimuli. By these processes, we were able to present the spatial information of the original sound field to listeners with state-of-the-art approximation, and the listeners were to experience the original sound field as if they existed in it (**Figure 3A**).

### HD Condition

In the HD condition, the same convolved music signals as in the 3D-SF condition were recorded in the Sound Cask once using a dummy head that has microphones in the position of the ear drums (HATS 4128C; Bruel & Kjaer, Nærum, Denmark; **Figure 1C**) to perform binaural recording. The sampling rate was 48 kHz, and we used 24-bit quantization as we did in the 3D-SF condition. The participants listened to the music signals through headphones (HDA 300; Sennheiser, Wedemark, Germany) in the listening area of the Sound Cask. To correct the music signals according to the frequency response of the headphones, like we did in 3D-SF condition, we prepared type 2 inverse filters that removed the frequency characteristics of the headphones (see **Supplementary Figure 1B**). With such a procedure, the participants in the HD condition were able to experience the spatial-audio information of music as rich as that in the 3D-SF condition (**Figure 3B**). The difference between the two conditions was the presence of the physical 3D-SF as the

[1]The pipe organ player kindly participated in this study on condition that we do not mention his/her name, and that the recording would be performed for only this tune.

[2]The piece is commercially available; it was not composed for the purpose of our research. The composer gave us permission to use the piece on condition that we do not mention its title or the name of the composer, and that only this piece would be used. The composer played this piece live so we could record it with a multichannel microphone array.

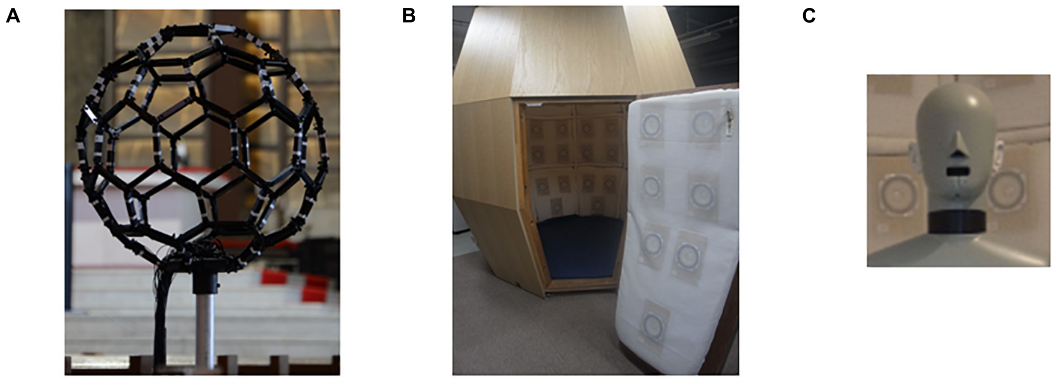



**FIGURE 1 |** Experimental equipment. **(A)** Photograph of fullerene-like microphone array with 80 channels of the BoSC system. **(B)** Photograph of reproduction system with 96 loudspeakers of the BoSC system. **(C)** Photograph of dummy head that had microphones in the position of the eardrums.

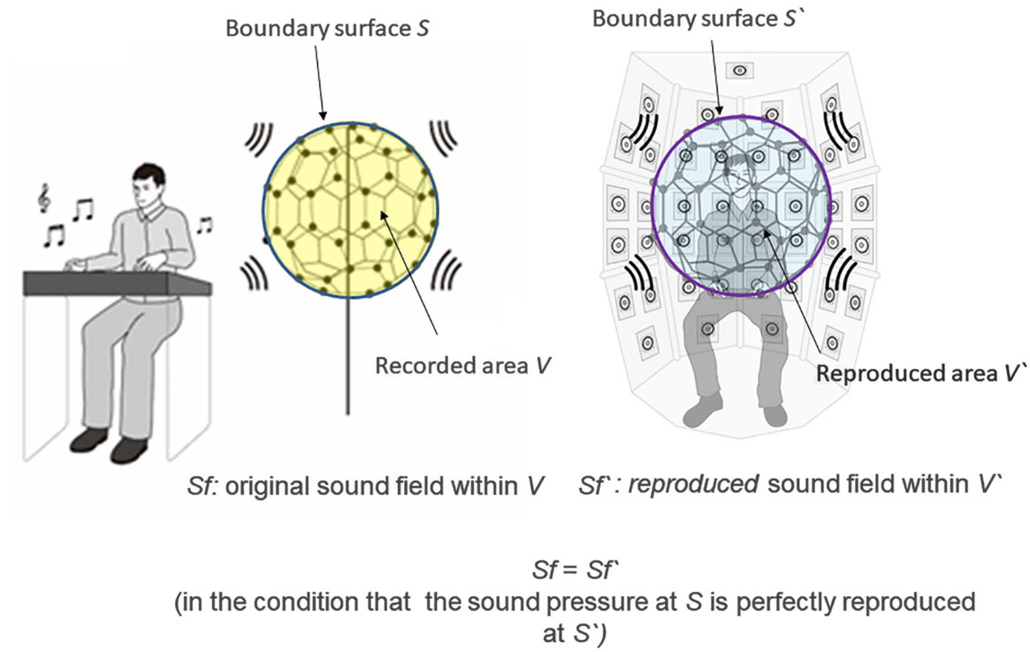



**FIGURE 2 |** Scheme of the boundary surface control (BoSC) principle. Left: Original sound was recorded with a microphone array. *Sf* stands for the original sound field within the enclosed space of the microphone array (defined as recorded area *V*), and *S* stands for the boundary surface of the recorded area *V*. Right: Recorded original sound was reproduced in a reproduction room. We imagine that there was the same microphone array as in the left panel at the listening position, and let *Sf′* stand for the reproduced sound field within the enclosed space of the virtual microphone array (defined as reproduced area *V′*) and let *S′* stand for the boundary surface of the reproduced area *V′*. If the original sound pressure at *S* is perfectly reproduced at *S′*, *Sf′* is mathematically completely the same as *Sf*.

vibration of air around the listeners or not. If the participants were affected by the presence of the physical 3D-SF through any channels other than auditory perception, such as tactile ones, we expected to observe psychological or physiological differences between the 3D-SF and HD conditions.

## Self-Reported Moods and Emotions

The method of multiple mood scaling (Terasaki et al., 1992) was used to examine the influence of music on the mood of the participants. We evaluated the strength of the following classified moods: Boredom, Liveliness, Well-being, Friendliness, Concentration, and Startle, each of which consisted of 10 mood-related adjectives such as tranquil or cautious, 60 adjectives in total. All the mood-related adjectives were written in Japanese. Before and after they had listened to music, the participants were asked to report the intensity in which they felt each mood-related adjective on a 4-point scale: 1. "don't feel at all"; 2. "don't feel so much"; 3. "feel to some extent"; and 4. "feel very much." The order of the adjectives was completely randomized for each participant. After the second rating of mood-related adjectives, the participants were asked to judge the emotions they felt from the music with "pleasant" and "unpleasant." After the judgment, they rated the degree of felt emotion for the music they had listened to on a 10-point scale: pleasant (1 = not pleasant at all, 10 = very pleasant); unpleasant (1 = not unpleasant at all, 10 = very unpleasant). When a participant selected one of "pleasant" or "unpleasant" and rated the score from 1 to 10, the score of the other was automatically 1.

## Physiological Measurements

Electrocardiograms and an elastic chest band (Polyam-RESP, Nihon Santeku, Osaka, Japan) were used to measure interbeat intervals (R-R intervals) of the heart and respiration, respectively, throughout the experiments. Analogue data were amplified and digitized with BIOPAC MP150 (BIOPAC Systems, Goleta, United States). The sampling rate was 1,250 Hz for both the ECGs and respiration measurements.

To calculate respiration rate, raw respiration signals were filtered with a.05–0.5-Hz band pass filter. The peak detection was performed with AcqKnowledge (analysis software of the BIOPAC MP150, United States) from the filtered respiration signals, and peak-to-peak intervals were obtained. The collected peak-to-peak interval data were resampled to 10 Hz by linear interpolation. For the respiration rate analysis, the interpolated data were converted to second-by-second values and expressed in cycles per minute (cpm) by dividing each value by 60.

To calculate the R-R intervals in the ECG measurement, R-wave detection was performed again with AcqKnowledge, and the result was visually screened to eliminate any inappropriate R-wave detection related to artifacts such as movement. The appropriately collected R-R interval data were resampled to 10 Hz by cubic spline interpolation. For the heart rate analysis, the interpolated R-R interval data were converted to second-by-second values and expressed in beats per minute (bpm) by dividing each value by 60.

For the HRV analysis, a fast Fourier transform (FFT) was applied to the interpolated R-R interval data after removing the linear trend to calculate the HRV power spectra using a Hanning window. Low-frequency (LF) and high-frequency (HF) components were obtained by integrating the power spectra over their respective ranges of 0.04–0.15 and 0.15–0.4 Hz. The magnitude of the HF and the ratio of LF to HF (LF/HF) correspond to the strength of the vagal activity (Porges, 1995) and the sympathovagal balance (Task Force, 1996), respectively. The magnitude of the LF involves both vagal and sympathetic nerve activities (Pomeranz et al., 1985). FFT was applied to each 3-min window of the interpolated data series of R-R intervals. The magnitude of each spectral component was evaluated using the natural logarithms of the power (lnLF and lnHF). The ratio of the LF component to the HF component (LF/HF ratio) was evaluated by dividing lnLF by lnHF (lnLF/lnHF).

## Experimental Procedure

The 32 participants were divided into two groups: the 3D-SF condition group (16 members, one female) and the HD condition group (16 members, two females). They came to the laboratory for 2 days, and on each day, they listened to one of the music pieces (AC or PM). On the first day of the experiment, they were given general information about the experiment upon arrival, and their written consent was obtained. The experimental procedure consisted of four periods: rest period → self-reported-mood period→ music period → self-reported-mood and emotion period. In the Sound Cask, the participants were asked to keep their eyes open during the experiment, and to steadily sit on a chair and avoid moving any part of their body, such as the head, as much as they could. The Sound Cask was kept dark during the experiment. The height of the chair was adjusted for each participant so that the position of the head was completely enclosed within the fullerene-like microphone array, which was used for the preparation of type-1 inverse filters. The participants were outfitted with ECG transducer electrodes and an elastic chest band for 10 min to familiarize them with the experimental environment. The participants in the HD condition group wore headphones. This time period is referred to as the "rest period." The last 3 min of the ECG and respiration recording was regarded as the baseline. Prior to the listening session, the participants were asked to rate the intensity of the 60 mood-related adjectives on a 4-point scale. When the listening session started (music period), the participants were presented with one of the music pieces with a maximum level of 84 dB SPL (A) for 540 s for PM and 563 s for AC. Then, there was a second period during which the participants were asked to rate the same 60 mood-related adjectives as in self-reported-mood period 1 and the felt emotion of this music on a 10-point scale (self-reported mood and emotion period). The order of the music pieces was completely randomized. The experimental procedure is summarized in **Figure 3C**.

## Data Analysis

Data are presented as means ± SEM, and a probability value $p < 0.05$ was considered to be statistically significant.

Time $t = 0$ was set as the onset of music. Physiological measures (respiration rate, heart rate, and HRV) were evaluated by averaging the data with a 3-min window ($t = 0$ to 3, $t = 3$ to 6, and $t = 6$ to 9 min for each) in the music period. The corresponding 3-min time regions were defined as T1, T2, and T3. To exclude the effect of the variation on the baseline value, the averaged data were normalized by dividing them by the basal value for each measurement in the baseline recording. The changes in these physiological measures normalized by the basal value were analyzed by three-factor repeated measures ANOVA with "sound reproduction condition" (3D-SF and HD) as a between-subjects factor and "music" (AC and PM) and "time" (baseline, T1, T2, and T3) as within-subjects factors.

The self-reported scores for the 60 mood-related adjectives were averaged among the corresponding 10 adjectives to determine the strength of each classified mood. The difference between the change in each mood (Boredom, Liveliness, Well-being, Friendliness, Concentration, and Startle) for different conditions was analyzed by two-factor repeated measures ANOVA with "sound reproduction condition" (3D-SF and HD) as a between-subjects factor and "music" (AC and PM) as a within-subjects factor.

Since all the participants rated their felt emotion as unpleasant for both AC and PM in both the 3D-SF and HD conditions, unpleasantness alone was analyzed by two-factor repeated

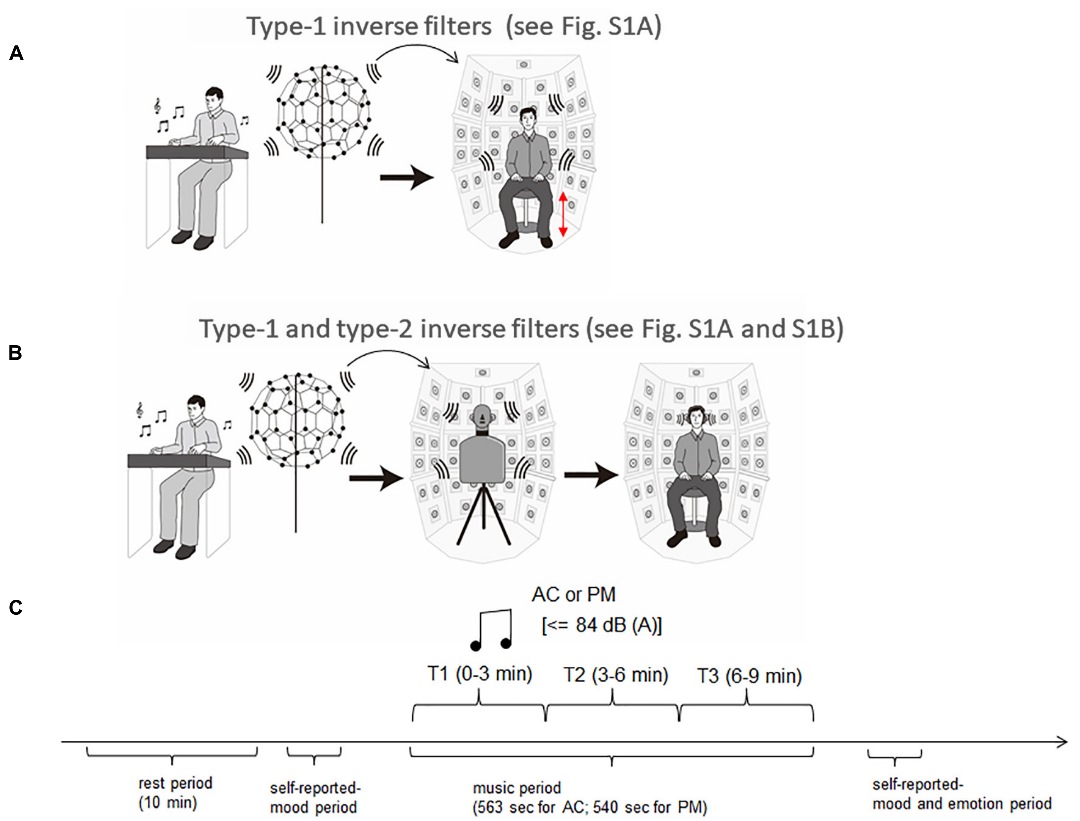



**FIGURE 3 |** Experimental setup and procedure. **(A)** Music presentation in the three-dimensional sound field (3D-SF) condition. Music was recorded with the fullerene-like microphone array with 80 channels, and was reproduced in the BoSC reproduction room with 96 channels. To reproduce the accurate sound field in the listening area, type 1 inverse filters were convolved with recorded music signals. The height of the chair was adjusted for each participant so that his/her head position was completely enclosed within the fullerene-like microphone array, which was used for the preparation of type 1 inverse filters. **(B)** Music presentation in the HD condition. Music in the 3D-SF condition was recorded in the Sound Cask through a dummy head and reproduced with headphones. To remove the frequency characteristics of the headphones, type 2 inverse filters were convolved with the music signals used in the 3D-SF condition. **(C)** Experimental procedure. Participants experienced one 10-min period of silence as a rest period. After that, they reported their mood and entered the session of music listening. When they finished the listening session, they performed a second self-report of mood and assessed the felt emotion of the music they had listened to.

measures ANOVA with “sound reproduction condition” (3D-SF and HD) as a between-subjects factor and ”music” (AC and PM) as a within-subjects factor.

Multiple comparisons between the three time regions of the music period (T1, T2, and T3) and baseline were analyzed with Ryan’s method. For the repeated measures ANOVA, Huynh-Feldt corrections were applied when appropriate.

## RESULTS

### Effect of Difference in Sound Reproduction Method on Self-Reported Moods and Emotions

Investigating the effect of the presence of physical 3D-SF on moods when listening to music, we found that “Well-Being” was more strongly reduced after listening to both PM and AC in the 3D-SF condition than in the HD condition (**Figure 4B**).

The ANOVA revealed the following significant effects: a significant interaction of “sound reproduction condition” × “music” [$F(1, 30) = 7.19$, $p = 0.012$, partial $\eta^2 = 0.19$] and a significant main effect of “music” [$F(1, 30) = 9.57$, $p = 0.0043$, partial $\eta^2 = 0.24$] for Boredom (**Figure 4A**); a significant main effect of “sound reproduction condition” [$F(1, 30) = 6.26$, $p = 0.018$, partial $\eta^2 = 0.17$] for Well-Being (**Figure 4B**); no significant effects for Concentration (**Figure 4C**); a significant main effect of “music” [$F(1,30) = 18.7$, $p = 0.0002$, partial $\eta^2 = 0.38$] for Liveliness (**Figure 4D**); a significant main effect of “music” [$F(1,30) = 10.29$, $p = 0.0032$, partial $\eta^2 = 0.26$] for Friendliness (**Figure 4E**); a significant main effect of “music” [$F(1, 30) = 20.87$, $p = 0.0001$, partial $\eta^2 = 0.41$] for Startle (**Figure 4F**). A simple effects test demonstrated that Boredom for PM in the HD condition was significantly greater than that for AC in the HD condition, while no significant difference was observed in the 3D-SF condition (**Figure 4A**).

The effect of the presence of the physical 3D-SF was also found on emotions felt for the music. The unpleasantness felt for both PM and AC was larger in the 3D-SF condition than in the

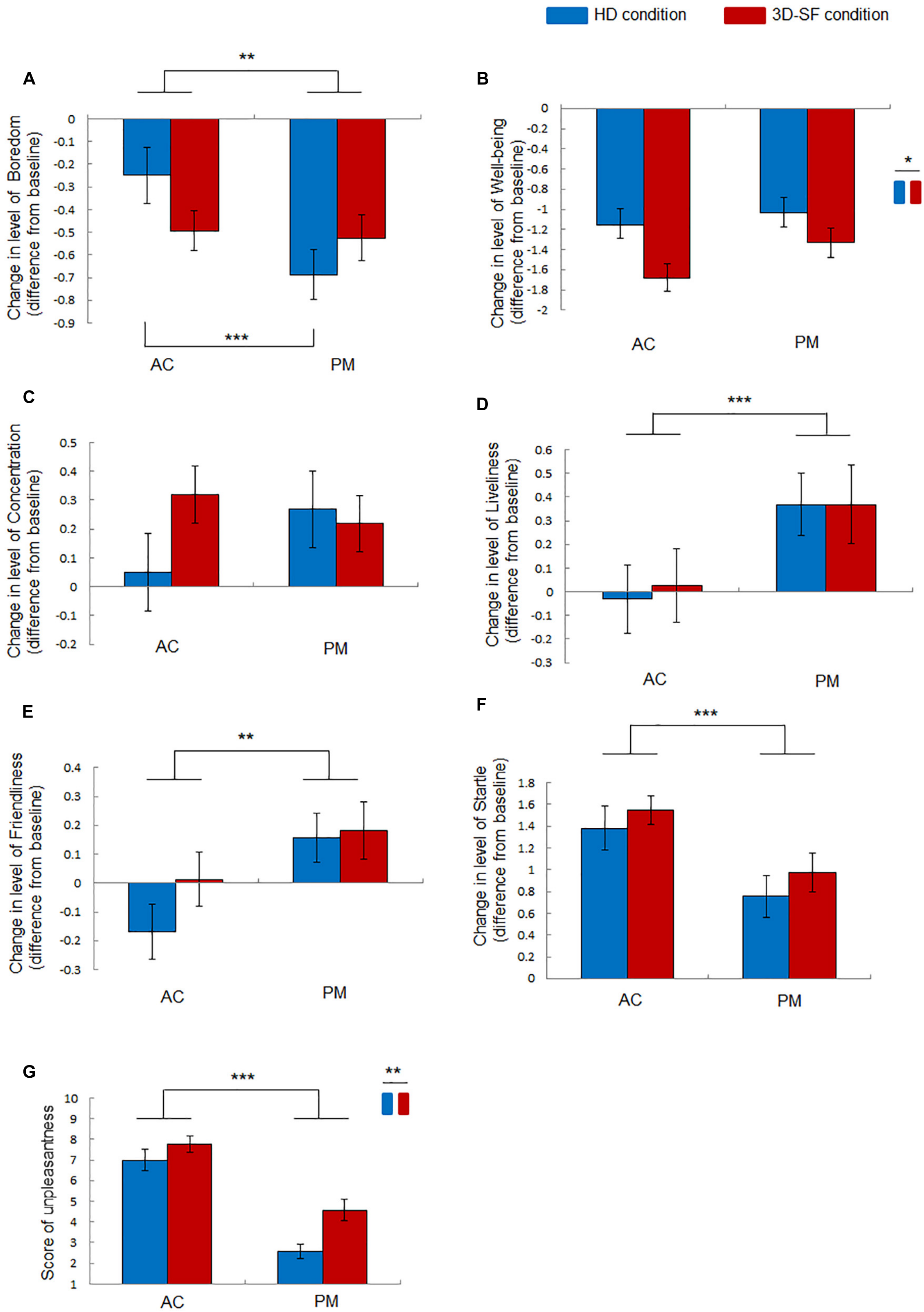



**FIGURE 4 |** Effect of music listening on moods and unpleasantness of music. **(A–F)** Change in the level of **(A)** Boredom, **(B)** Well-Being, **(C)** Concentration, **(D)** Liveliness, **(E)** Friendliness, and **(F)** Startle after music listening (AC or PM) in the different conditions (3D-SF or HD) compared with the level before it. **(G)** Felt unpleasantness of music rated after listening to music. Data are presented as means ± SEM; ****p* < 0.001, ***p* < 0.01, **p* < 0.05.

HD condition (**Figure 4G**). The felt pleasantness was considered to be not affected, because all the participants judged both AC and PM as "unpleasant" in both the 3D-SF and HD conditions. They only rated their felt unpleasantness, and their rate of pleasantness was automatically "1." Although four participants in the HD condition experiments rated PM as 1 (= not unpleasant at all), their felt emotion was more unpleasant than pleasant for listening to PM in the HD condition. For the four participants, the degree of unpleasantness was quite small, namely close to 1. ANOVA was applied to the data of unpleasantness alone. The ANOVA demonstrated a significant main effect of "music" [$F(1, 30) = 82.87$, $p < 0.0001$, partial $\eta^2 = 0.73$] and "sound reproduction condition" [$F(1, 30) = 8.35$, $p = 0.0071$, partial $\eta^2 = 0.22$] (**Figure 4G**).

As a reference, the significance of the change in mood from the baseline in each condition is summarized in **Table 1**, for which a paired *t*-test between the value before and after listening to music was performed.

## Effect of Difference in Sound Reproduction Method on Change in Respiration Rate

We found that the respiration rate of the participants was higher when they listened to both PM and AC in the 3D-SF condition than before they had listened (**Figure 5B**), whereas such a change was not observed in the HD condition (**Figure 5C**), suggesting that the presence of a physical 3D-SF can increase the respiration rate of listeners. The ANOVA revealed significant interactions of "sound reproduction condition" × "time" [$F(3, 90) = 5.59$, $p = 0.0015$, partial $\eta^2 = 0.16$] and "music" × "time" [$F(3, 90) = 3.68$, $p = 0.015$, partial $\eta^2 = 0.11$], and significant main effects of "sound reproduction condition" [$F(1, 30) = 6.83$, $p = 0.014$, partial $\eta^2 = 0.19$] and "time" [$F(3, 90) = 6.82$, $p = 0.0003$, partial $\eta^2 = 0.19$]. Concerning the difference between the reproduction conditions, a simple effects test demonstrated that the normalized respiration rate was larger for the 3D-SF condition than for the HD condition at the T1 [$F(1, 120) = 8.73$, $p = 0.0038$], T2 [$F(1, 120) = 12.41$, $p = 0.0006$], and T3 [$F(1, 120) = 4.6$, $p = 0.034$] regions (**Figure 5A**). Concerning the difference along the time region, a simple effects test demonstrated that there was a significant difference in the normalized respiration rate among the four time regions in the 3D-SF condition independent of the type of music [$F(3, 90) = 11.44$, $p < 0.0001$], in which Ryan's method of adjusting the *P* value showed that the normalized respiration rate in every music region was larger than the baseline ($t = 4.16$, $p = 0.0001$ for baseline vs. T1; $t = 5.36$, $p < 0.0001$ for baseline vs. T2; $t = 4.5$, $p < 0.0001$ for baseline vs. T3) (**Figure 5B**), but that no change in the normalized respiration rate was observed in the HD condition (**Figure 5C**).

## Effect of Difference in Sound Reproduction Method on Change in Heart Rate and Heart Rate Variability

We showed a positive effect of the presence of the physical 3D-SF, but we also observed a negative side. Heart rate was reduced during listening to AC in the HD condition, whereas the reduction was diminished in the 3D-SF condition (**Figure 6A**). The ANOVA revealed significant interactions of "sound reproduction condition" × "music" × "time" [$F(3, 90) = 3.16$, $p = 0.029$, partial $\eta^2 = 0.095$] and "music" × "time" [$F(3, 90) = 7.86$, $p = 0.0001$, partial $\eta^2 = 0.21$] and significant main effects of "music" [$F(1, 30) = 11.28$, $p = 0.0021$, partial $\eta^2 = 0.27$], and "time" [$F(3, 90) = 5.37$, $p = 0.0019$, partial $\eta^2 = 0.15$]. Concerning the difference between the reproduction conditions, a simple effects test demonstrated that the normalized heart rate in AC listening was significantly lower for the HD condition than for the 3D-SF condition in the T2 region [$F(1, 240) = 4.611$, $p = 0.033$] (**Figure 6A**), while no significant difference between the 3D-SF and HD conditions was observed for PM (**Figure 6B**). Concerning the difference along the time region, a simple effects test demonstrated that there was a significant difference in the normalized heart rate among the four time regions for AC [$F(3, 180) = 9.78$, $p < 0.0001$], in which Ryan's method of adjusting the *P* value showed that the normalized heart rate in T1 was lower than in the other regions [$t(180) = 4.98$, $p < 0.0001$ for T1 vs. baseline; $t(180) = 2.55$, $p = 0.012$ for T1 vs. T2; $t(180) = 4.25$, $p < 0.0001$ for T1 vs. T3] (**Figure 6A**).

In contrast to heart rate, neither a significant interaction nor a main effect was observed in HRV. The HRV data are summarized in **Table 2**.

# DISCUSSION

This study examined differences in psychological and physiological responses between two reproduction conditions, namely the 3D-SF and HD conditions, when participants listened to music. The participants could experience spatial-audio information in the HD condition as rich as that in the 3D-SF condition, and the difference between the 3D-SF and HD conditions was the presence or absence of the physical 3D-SF. This suggests that our results can be explained by the effects of non-auditory perception, such as tactile perception. In the psychological responses, we analyzed "unpleasantness" as a parameter of the emotional response to music stimuli and "Well-Being," "Boredom," "Liveliness," "Friendliness," "Concentration," and "Startle" as parameters of mood. We found a significant increase in unpleasantness. Although "Well-Being" showed a significantly larger decrease in the 3D-SF condition than in the HD-condition, none of the other moods showed significantly larger effects in the 3D-SF condition than in the HD-condition. These results suggest that a non-auditory cue provided by the presence of a 3D-SF facilitates emotions felt in response to music as well as changes in moods if listeners have low arousal and valence opposite to the felt emotion. In the physiological responses, we analyzed respiration rate, heart rate, and heart rate variability (HRV). We found that when the participants were in the 3D-SF condition, the respiration rate increased significantly during listening sessions regardless of the type of music, whereas no changes in HRV indices were observed in either the 3D-SF or HD conditions. Since an increase in the respiration rate is typically accompanied by an increase in sympathetic nerve

**TABLE 1 |** Summary of comparison of each mood between, before, and after listening to music.

| | Mood | Boredom | | Well-Being | | Concentration | | Liveliness | | Friendliness | | Startle | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Music** | **AC** | **PM** | **AC** | **PM** | **AC** | **PM** | **AC** | **PM** | **AC** | **PM** | **AC** | **PM** |
| HD | *t* | 2.04 | 6.28 | 7.15 | 7.01 | 0.37 | 2.03 | 0.22 | 2.82 | 1.77 | 1.84 | 6.91 | 3.88 |
| | *p* | 0.060 | <0.0001 | <0.0001 | <0.0001 | 0.72 | 0.061 | 0.83 | 0.013 | 0.097 | 0.086 | <0.0001 | 0.0015 |
| | Cohen's *d* | 0.51 | 1.57 | 1.79 | 1.78 | 0.093 | 0.51 | 0.054 | 0.70 | 0.46 | 0.46 | 1.70 | 0.95 |
| 3D-SF | *t* | 5.56 | 5.18 | 12.33 | 9.23 | 3.24 | 2.26 | 0.16 | 2.21 | 0.13 | 1.84 | 11.77 | 5.44 |
| | *p* | <0.0001 | 0.0001 | <0.0001 | <0.0001 | 0.0055 | 0.039 | 0.87 | 0.043 | 0.9 | 0.085 | <0.0001 | <0.0001 |
| | Cohen's *d* | 1.42 | 1.30 | 3.11 | 2.33 | 0.81 | 0.57 | 0.041 | 0.55 | 0.036 | 0.46 | 2.95 | 1.40 |

*The statistical significance was evaluated by a paired t-test.*
*t, t value of paired t-test; p, p value of paired t-test; Cohen's d, effect size of paired t-test.*

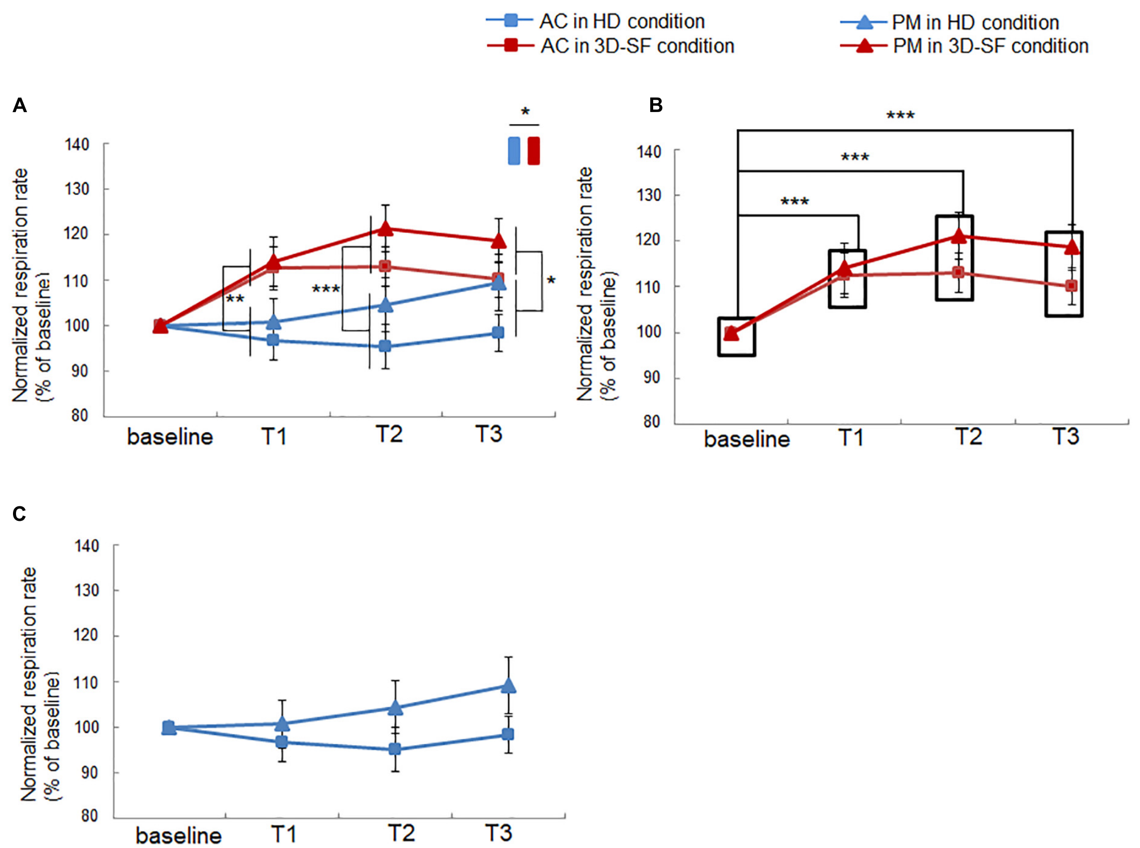



**FIGURE 5 |** Effect of music listening on normalized respiration rate. **(A)** Comparison between the HD and 3D-SF conditions. **(B)** Comparison among the four time regions (baseline, T1, T2, and T3) in the 3D-SF condition. **(C)** Comparison among the four time regions (baseline, T1, T2, and T3) in the HD condition. Data are presented as means ± SEM; ****p* < 0.001, ***p* < 0.01, **p* < 0.05.

activity (Bernardi et al., 2006; Salimpoor et al., 2011; Koelsch and Jancke, 2015), our finding of an increase in the respiration rate of the participants without any change in lnLF/lnHF when they listened to music is a novel one. Heart rate was reduced when they listened to AC in the HD condition, which could be regarded as an orienting response (Porges, 1995) induced by aversive stimuli (Bradley et al., 2001; Ooishi and Kashino, 2012). The reduction in heart rate was suppressed when they listened to AC in the 3D-SF condition. These results suggest that a non-auditory perception evoked by the presence of a 3D-SF activates the respiratory system without any effect on the sympathetic nervous system and, through the activation of the respiration system, suppresses the activity of neurons that evoke the aversive stimulus-induced orienting response.

The reproduction of a real sound field for listeners has been achieved in mainly two methodological ways. One is based on the wave field synthesis technique (Berkhout et al., 1993; Ise, 1999). The original sound is recorded with multichannel microphone arrays, and reproduced with multichannel speaker arrays. To enable the speaker arrays to reproduce the sound

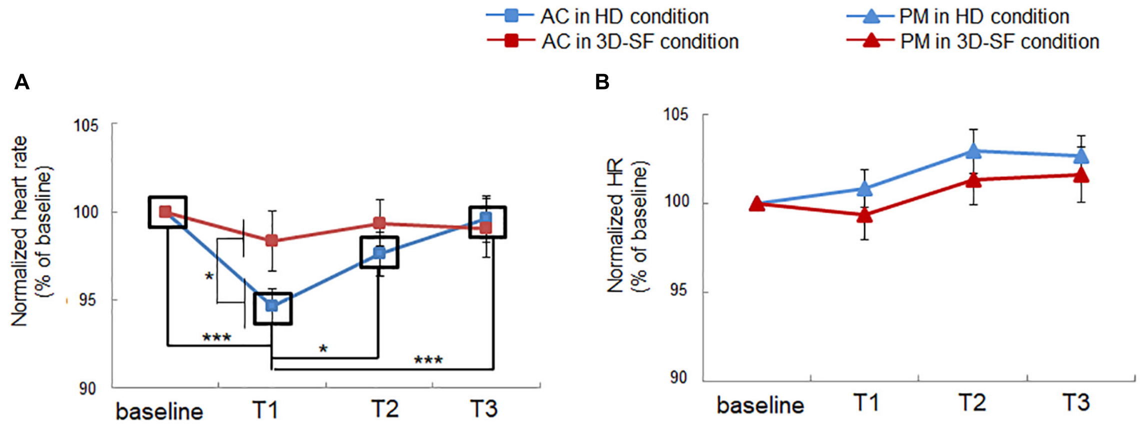



**FIGURE 6 |** Effect of music listening on normalized heart rate. **(A)** Comparison between the HD and 3D-SF conditions for AC. **(B)** Comparison between the HD and 3D-SF conditions when listening to PM.

field with the best approximation, inverse filters for the multi-input/multi-output system are convolved with sound signals. With this technique, listeners can experience as if they are surrounded with the original sound field. The other is based on the binaural reproduction technique (Paul, 2009). Contrary to listening through headphones to sound that is stereophonically recorded with two-channel microphones place at a chosen position in the sound field, binaurally recorded sound through headphones enables listeners to experience three-dimensional auditory perception. Binaural recording is usually performed with a dummy head that has two microphones at the position of the ear drums.

Listening to sound within a sound field reproduced with the wave field synthesis technique and listening to binaurally recorded sound through headphones provide almost the same spatial-audio information as that in the original sound field. However, sound produced by the wave field synthesis technique has a physical sound field, whereas binaurally recorded sound has none. As reported in earlier studies (Boone et al., 1995; Theile and Wittek, 2004), sound field synthesis enables listeners to feel as if they are surrounded by sound. We assume that this feeling is due to not only auditory perception but also other channels of perception, such as tactile ones. Another earlier study showed the effect of the "Music Bath," which can relieve pain or anxiety by presenting sounds from loudspeakers as a source of vibration (Skille, 1987a,b). This is consistent with our assumption in the sense that the effects of sound other than its auditory perception affect listeners. For example, we found that the felt unpleasantness was larger when the participants listened to music in the 3D-SF condition than in the HD condition, suggesting that non-auditory perception induced by the presence of a physical 3D-SF enhances the impression of music for listeners.

Terasaki et al. (1992) indicated that "Boredom," "Well-Being," "Concentration," "Liveliness," "Friendliness," and "Startle," the parameters used in this study, correspond to fatigue, well-being, concentration or anxiety, energy, social affection, and anxiety, respectively, as defined by Lebo and Nesselroade (1978). We found that the decrease in "Well-Being" was larger in the 3D-SF condition than in the HD condition after the participants had listened to music (**Figure 4B**). Considering this together with the results of unpleasantness, and considering that "Well-Being" (Levo's well-being) is similar to relaxation, it is suggested that unpleasant music reduced the relaxation level of the participants. Interestingly, the degree of unpleasantness does not determine the size of the reduction of relaxation, because the unpleasantness for AC was higher than that for PM (**Figure 4G**), while the size of the reduction of Well-Being was not significantly different between AC and PM (**Figure 4B**). Our results suggest that not the degree of unpleasantness itself but the increase in unpleasantness caused by the presence of physical 3D-SF boosts the reduction of relaxation level. In contrast to "Well-Being," "Liveliness" was not affected by the difference in the reproduction condition (**Figure 4D**). Considering that "Liveliness" (Lebo's energy) is similar to excitement, the difference between "Well-Being" and "Liveliness" is the degree of arousal; arousal is low for "Well-Being," while it is high for "Liveliness" (Russell et al., 1989). Therefore, it is suggested that moods with low arousal could be affected by the presence of a physical 3D-SF. Actually, "Friendliness" (Lebo's social affection, similar to love) and "Startle" (Lebo's anxiety) are categorized as high arousal (Russell et al., 1989). Although Lebo's concentration was not categorized by Russel, the "Concentration" we used corresponds to Lebo's anxiety as well as to Lebo's concentration, meaning that "Concentration" could also be categorized as high arousal (Russell et al., 1989). However, our assumption does not consider "Boredom" (Lebo's fatigue), because it is categorized as low arousal (Russell et al., 1989). The difference between "Well-Being" and "Boredom" is the degree of valence. "Well-Being" (similar to relaxation) is categorized as high valence, while "Boredom" is categorized as low valence (Russell et al., 1989). Since the music stimuli in this study were rated as unpleasant by all the participants, the non-auditory perception induced by the presence of the physical 3D-SF could have facilitated their moods if the moods were categorized as having low arousal and

**TABLE 2 |** Summary of heart rate variability (HRV).

| | | lnLF | | | lnHF | | | lnLF/lnHF | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | T1 | T2 | T3 | T1 | T2 | T3 | T1 | T2 | T3 |
| AC | HD | 1.00 ± 0.04 | 1.01 ± 0.04 | 1.01 ± 0.04 | 1.04 ± 0.06 | 1.02 ± 0.05 | 0.99 ± 0.05 | 0.98 ± 0.06 | 1.00 ± 0.04 | 1.04 ± 0.04 |
| | 3D-SF | 0.98 ± 0.03 | 0.98 ± 0.03 | 1.05 ± 0.04 | 1.01 ± 0.03 | 0.99 ± 0.03 | 1.00 ± 0.03 | 0.98 ± 0.03 | 1.00 ± 0.03 | 1.05 ± 0.04 |
| PM | HD | 1.00 ± 0.03 | 1.00 ± 0.03 | 1.04 ± 0.05 | 0.98 ± 0.03 | 0.97 ± 0.03 | 0.98 ± 0.04 | 1.03 ± 0.04 | 1.05 ± 0.04 | 1.07 ± 0.05 |
| | 3D-SF | 1.00 ± 0.03 | 0.97 ± 0.03 | 1.00 ± 0.03 | 1.03 ± 0.04 | 1.03 ± 0.04 | 1.00 ± 0.04 | 0.98 ± 0.03 | 0.95 ± 0.04 | 1.01 ± 0.03 |

*All the data are normalized by dividing them by the corresponding baseline value. Data are presented as means ± SEM.*

valence opposite to the emotion felt in response to music stimuli. We assume that the presence of a physical 3D-SF could facilitate Boredom when listening to music rated as pleasant.

The impact of the presence of the physical sound field on physiological activity when listening to music was also detected in this study. We observed a significant increase in respiration rate in the 3D-SF condition, regardless of the type of music (**Figure 5**). Previous studies have demonstrated the contribution of music features, such as tempo, to the modulation of respiration. In particular, the entrainment of respiration rate to music tempo has been frequently discussed (Haas et al., 1986; Bernardi et al., 2006; Etzel et al., 2006; Khalfa et al., 2008). Our results could be considered as novel findings in the sense that the response of respiration to music stimuli in our study was inconsistent with earlier studies. The increase in respiration rate during music listening is typically induced together with an increase in the sympathetic nerve activity (Bernardi et al., 2006; Salimpoor et al., 2011; Koelsch and Jancke, 2015), whereas our results showed no significant change in lnLF/lnHF with listening to both AC and PM in both the 3D-SF and HD conditions (**Table 2**). The LF/HF ratio of HRV is an index of the activities of neurons in the rostral ventrolateral medulla (RVLM) (Thomas and Spyer, 1999; Dun et al., 2000; Jiang et al., 2011; Shahid et al., 2012; Chen et al., 2013), one of the primary regulators of the sympathetic nervous system that controls cardiac activity (Shields, 1993; Kumagai et al., 2012), and is often used to represent music listening-induced excitation (Bernardi et al., 2006; Ooishi et al., 2017). Taking our results together with those of previous studies, we believe that the presence of the physical 3D-SF enhanced respiration rate during music listening through channels other than auditory perception without any effect on sympathetic cardiac control.

Interestingly, heart rate was reduced with listening to AC in the HD condition, which suggests a contribution of parasympathetic nerve activity. The nucleus ambiguus (NA) is one of the primary regulators of the parasympathetic nervous system that controls cardiac activity (Porges, 1995). The HF component of HRV can be an index of parasympathetic nerve activity originating from the NA and is often used to represent music-induced relaxation (Zhou et al., 2010; Ooishi et al., 2017). Therefore, no change in the HF component means no change in the cardiac parasympathetic modulation originating from the NA. In the literature, Porges proposed in his earlier study that there is a polyvagal mechanism that contains two types of nuclei inducing heart rate deceleration (Porges, 1995). One of the nuclei is the NA noted above, and the other is the dorsal motor nucleus of the vagus (DMNX). We assume that the heart rate deceleration that occurred while the participants were listening to AC in the HD condition originated from activation of the DMNX, which is categorized as an orienting response (Porges, 1995). Heart rate deceleration as an orienting response is frequently induced by aversive sensory stimuli (Bradley et al., 2001; Ooishi and Kashino, 2012). However, heart rate did not change with listening to AC in the 3D-SF condition, in which the participants rated it more unpleasant than in the HD condition (**Figure 4G**). We propose that the enhancer of respiration rate might suppress the activity of neurons in the DMNX. Actually, there is a connection between respiratory systems and the DMNX. For example, the DMNX receives input from the nucleus tractus solitarius (Kalia, 1981).

In conclusion, we found that the presence of a physical sound field when listening to music facilitates psychological and physiological responses. Music listening in the 3D-SF condition enhanced the decrease in Well-Being and the felt unpleasantness for both AC and PM compared with in the HD condition. Respiration rate was increased compared with the baseline only with music listening in the 3D-SF condition. Since reproduced music in both the 3D-SF and HD conditions has spatial-audio information as rich as that of the original one, we suggest that the presence of the physical sound field affects listeners through channels other than auditory perception, such as tactile ones, when listening to music.

## Limitations

Before starting this study, we prepared AC as the representative of unpleasant (e.g., feeling of fear or a thrill) but interesting music and PM as that of pleasant or impressive music. We hypothesized that the felt emotion for AC and PM would be larger when listening to them with a high-quality3D-SF reproduction system compared with listening to the same pieces binaurally recorded with a dummy head and reproduced with headphones. Our hypothesis was based on the possibility that the physical 3D-SF enhances the psychological and physiological responses of the listeners through channels other than auditory perception, such as tactile ones. Actually, we (the four authors of this study) regarded AC as extremely unpleasant music that could raise goose bumps due to aversion and PM as very pleasant and impressive, and we felt that the degree of such emotions was larger when listening to music in the 3D-SF condition than in the HD condition. However, all the students (six men, age 22∼24) from Meiji University recruited as participants in the preliminary experiment assessed PM as unpleasant. In particular, the unpleasantness of PM increased when they listened to it in the 3D-SF condition. In the experiments of this study, as well as in the preliminary experiment, all the participants judged both AC and PM in both the 3D-SF and HD conditions as unpleasant. We, therefore, asked the participants to rate the degree of felt unpleasantness without pleasantness when they listened to the music. Their rate of pleasantness was automatically "1." Therefore, we performed ANOVA only on the data of unpleasantness. Our results demonstrated that the unpleasantness of both AC and PM in the 3D-SF condition was larger than in the HD condition. A future direction will include examining the effect of the presence of the 3D-SF when participants listen to music that they appreciate as pleasant.

Among the 32 participants recruited for this study, there were only 3 females, so there may have been biases in how they responded to the music. An earlier study demonstrated that emotional and psychophysiological responses to music are larger for females than for males (Gupta and Gupta, 2016). In addition, all the participants were non-musicians. Differences in responses to music have also been observed between musicians and non-musicians. For example, musicians have greater respiratory sensitivity to tempo than non-musicians (Bernardi et al., 2006). A future direction will include considering the balance of genders (male and female) or music-training histories (musicians and non-musicians) to examine the differences or similarities between them (male vs. female, musicians vs. non-musicians).

We used a dummy head to perform binaural recording of music because it enables listeners to perceive spatial-audio information as rich as that in the original sounds due to its inclusion the effect of the head-related transfer function (HRTF) (Paul, 2009). Strictly speaking, the HRTF is different between individuals because of differences in the shape of the head, ear pinna, body, and features of clothes they wear (Riederer, 2003; Wersényi and Illényi, 2005). However, the effect of binaurally recorded sounds with a dummy head reproduced through headphones is sufficiently larger than it is for normal two-channel stereophonic sounds through headphones (Paul, 2009). Currently it is too difficult to calculate individual HRTFs in each experiment, but we hope for a technology that will make it possible in the future.

In the HD condition, we re-recorded the reproduced music stimuli in the 3D-SF condition with the dummy head, although it would have been better to record the music pieces played live with the dummy head. The possible loss of sound information by re-recording could have affected the results of this study. The reason we performed the re-recording was to ensure that there would be no difference in the direction of the sound source between conditions. If the music pieces had been recorded with the fullerene-type microphone array and the dummy head simultaneously, the direction of the sound source would have differed between the respectively recorded sounds, because the position of the microphone array and the dummy head was different. To reduce the effect of re-recording, we set the sampling rate at 48 kHz and performed 24-bit quantization, which can sufficiently cover the human audible range, when recording music pieces with both the fullerene-type microphone array and the dummy head.

In this study, we proposed that the presence of a physical 3D-SF affects listeners through non-auditory perception. Since all the experiments were performed in a dark reproduction room, the contribution of vision could be ignored when considering the effect of the 3D-SF. Therefore, we predicted that tactile perception would be a possible source of the effect of the3D-SF; however, none of the data we obtained could prove this prediction. Furthermore, we asked the participants to keep their eyes open and to steadily sit on a chair and avoid bodily movement, including the head, as much as they could, but, of course, they might not have been able to keep their eyes open during the entire listening session (excluding eye blinks) or achieve complete stillness. Closing the eyes while listening to negative music has been shown to produce greater ratings of emotionality than opening them (Lerner et al., 2009), and head movement affects spatial-audio perception (Kondo et al., 2012). Therefore, failure to keep the eyes open the whole time and the lack of head fixation might weaken the possibility of the contribution of non-auditory perception, such as tactile to the effect of music listening in the presence of a physical 3D-SF. Future studies should include an attempt to directly reveal that tactile perception contributes to the effect of music listening in the presence of a physical 3D-SF on listeners.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics and Safety Committees of Meiji University. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

YO, MKo, and KU designed the psychological experiments. MKo collected the autonomic response data and self-reported ratings, and prepared the experimental speech stimuli. YO analyzed the autonomic responses data and self-reported ratings, and evaluated the results. KU provided the three-dimensional sound reproduction system. YO, MKo, and MKa wrote the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This study was supported by a grant from Core Research for Evolutional Science and Technology, Japan Science and Technology Agency (CREST, JST) (JPMJCR10T2).

## ACKNOWLEDGMENTS

We thank Ise of Tokyo Denki University for the opportunity to use the BoSC reproduction system, including the BoSC reproduction room and multichannel microphone array.


## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.650777/full#supplementary-material

## REFERENCES


Bailey, L. M. (1983). The effects of live music versus tape-recorded music on hospitalized cancer patients. *Music Ther.* 3, 17–28. doi: 10.1093/mt/3.1.17

Balkwill, L.-L., and Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: psychophysical and cultural cues. *Music Percept.* 17, 43–64. doi: 10.1371/journal.pone.0241196

Berkhout, A. J., De Vries, D., and Vogel, P. (1993). Acoustic control by wave field synthesis. *J. Acoust. Soc. Am.* 93, 2764–2778. doi: 10.1121/1.405852

Bernardi, L., Porta, C., and Sleight, P. (2006). Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: the importance of silence. *Heart* 92, 445–452. doi: 10.1136/hrt.2005.064600

Blood, A. J., Zatorre, R. J., Bermudez, P., and Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nat. Neurosci.* 2, 382–387. doi: 10.1038/7299

Boone, M. M., Verheijen, E. N., and Van Tol, P. F. (1995). Spatial sound-field reproduction by wave-field synthesis. *J. Audio Eng. Soc.* 43, 1003–1012.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., and Lang, P. J. (2001). Emotion and motivation I: defensive and appetitive reactions in picture processing. *Emotion* 1:276. doi: 10.1037/1528-3542.1.3.276

Chanda, M. L., and Levitin, D. J. (2013). The neurochemistry of music. *Trends Cogn. Sci.* 17, 179–193. doi: 10.1016/j.tics.2013.02.007

Chapados, C., and Levitin, D. J. (2008). Cross-modal interactions in the experience of musical performances: physiological correlates. *Cognition* 108, 639–651. doi: 10.1016/j.cognition.2008.05.008

Chen, J., Xia, C., Wang, J., Jiang, M., Zhang, H., Zhang, C., et al. (2013). The effect of orexin-A on cardiac dysfunction mediated by NADPH oxidase-derived superoxide anion in ventrolateral medulla. *PLoS One* 8:e69840. doi: 10.1371/journal.pone.0069840

Chen, T. (2001). "Surround sound beyond 5.1," in *Proceedings of the 19th Audio Engineering Society International Conference: Surround Sound-Techniques, Technology, and Perception*, (New York, NY: Audio Engineering Society).

Dun, N. J., Le Dun, S., Chen, C. T., Hwang, L. L., Kwok, E. H., and Chang, J. K. (2000). Orexins: a role in medullary sympathetic outflow. *Regul. Pept.* 96, 65–70. doi: 10.1016/s0167-0115(00)00202-0

Enomoto, S., Ikeda, Y., Ise, S., and Nakamura, S. (2011). "3-D sound reproduction system for immersive environments based on the boundary surface control principle," in *International Conference on Virtual and Mixed Reality*, ed. R. Shumaker (Berlin: Springer), 174–184. doi: 10.1007/978-3-642-22021-0_20

Etzel, J. A., Johnsen, E. L., Dickerson, J., Tranel, D., and Adolphs, R. (2006). Cardiovascular and respiratory responses during musical mood induction. *Int. J. Psychophysiol.* 61, 57–69. doi: 10.1016/j.ijpsycho.2005.10.025

Fukada, A. (2001). "A challenge in multichannel music recording," in *Proceedings of the 19th Audio Engineering Society International Conference: Surround Sound-Techniques, Technology, and Perception*, (New York, NY: Audio Engineering Society). doi: 10.1121/1.4835815

Gomez, P., and Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion* 7, 377–387. doi: 10.1037/1528-3542.7.2.377

Gupta, U., and Gupta, B. (2016). Gender differences in psychophysiological responses to music listening. *Music Med.* 8, 53–64. doi: 10.47513/mmd.v8i1.471

Haas, F., Distenfeld, S., and Axen, K. (1986). Effects of perceived musical rhythm on respiratory pattern. *J. Appl. Physiol.* 61, 1185–1191. doi: 10.1152/jappl.1986.61.3.1185

Hamasaki, K., Nishiguchi, T., Okumura, R., Nakayama, Y., and Ando, A. (2008). A 22.2 multichannel sound system for ultrahigh-definition TV (UHDTV). *SMPTE Motion Imaging J.* 117, 40–49. doi: 10.5594/j15119

Hevner, K. (1935). The affective character of the major and minor modes in music. *Am. J. Psychol.* 47, 103–118. doi: 10.3389/fpsyg.2013.00464

Huang, J., Gamble, D., Sarnlertsophon, K., Wang, X., and Hsiao, S. (2012). Feeling music: integration of auditory and tactile inputs in musical meter perception. *PLoS One* 7:e48496. doi: 10.1007/978-1-4614-1590-9_50

Ise, S. (1999). A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems. *Acta Acust. United Acust.* 85, 78–87.

Jiang, M. Y., Chen, J., Wang, J., Xiao, F., Zhang, H. H., Zhang, C. R., et al. (2011). Nitric oxide modulates cardiovascular function in the rat by activating adenosine A2A receptors and inhibiting acetylcholine release in the rostral ventrolateral medulla. *Clin. Exp. Pharmacol. Physiol.* 38, 380–386. doi: 10.1111/j.1440-1681.2011.05521.x

Kalia, M. P. (1981). Anatomical organization of central respiratory neurons. *Annu. Rev. Physiol.* 43, 105–120. doi: 10.1146/annurev.ph.43.030181.000541

Khalfa, S., Roy, M., Rainville, P., Dalla Bella, S., and Peretz, I. (2008). Role of tempo entrainment in psychophysiological differentiation of happy and sad music? *Int. J. Psychophysiol.* 68, 17–26. doi: 10.1016/j.ijpsycho.2007.12.001

Knight, W. E., and Rickard, D. N. (2001). Relaxing music prevents stress-induced increases in subjective anxiety, systolic blood pressure, and heart rate in healthy males and females. *J. Music Ther.* 38, 254–272. doi: 10.1093/jmt/38.4.254

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666

Koelsch, S., and Jancke, L. (2015). Music and the heart. *Eur. Heart J.* 36, 3043–3049.

Kondo, H. M., Pressnitzer, D., Toshima, I., and Kashino, M. (2012). Effects of self-motion on auditory scene analysis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6775–6780.

Kumagai, H., Oshima, N., Matsuura, T., Iigaya, K., Imai, M., Onimaru, H., et al. (2012). Importance of rostral ventrolateral medulla neurons in determining efferent sympathetic nerve activity and blood pressure. *Hypertens Res.* 35, 132–141.

Lebo, M. A., and Nesselroade, J. R. (1978). Intraindividual differences dimensions of mood change during pregnancy identified in five P-technique factor analyses. *J. Res. Pers.* 12, 205–224. doi: 10.1016/0092-6566(78)90098-3

Lerner, Y., Papo, D., Zhdanov, A., Belozersky, L., and Hendler, T. (2009). Eyes wide shut: amygdala mediates eyes-closed effect on emotional experience with music. *PLoS One* 4:e6230. doi: 10.1371/journal.pone.0006230

Murrock, C. J., and Higgins, P. A. (2009). The theory of music, mood and movement to improve health outcomes. *J. Adv. Nurs.* 65, 2249–2257. doi: 10.1111/j.1365-2648.2009.05108.x

Nayak, S., Wheeler, B. L., Shiflett, S. C., and Agostinelli, S. (2000). Effect of music therapy on mood and social interaction among individuals with acute traumatic brain injury and stroke. *Rehabil. Psychol.* 45:274.

Omoto, A., Ise, S., Ikeda, Y., Ueno, K., Enomoto, S., and Kobayashi, M. (2015). Sound field reproduction and sharing system based on the boundary surface control principle. *Acoust. Sci. Technol.* 36, 1–11. doi: 10.1250/ast.36.1

Ooishi, Y., and Kashino, M. (2012). Habituation of rapid sympathetic response to aversive timbre eliminated by change in basal sympathovagal balance. *Psychophysiology* 49, 1059–1071. doi: 10.1111/j.1469-8986.2012.01393.x

Ooishi, Y., Mukai, H., Watanabe, K., Kawato, S., and Kashino, M. (2017). Increase in salivary oxytocin and decrease in salivary cortisol after listening to relaxing slow-tempo and exciting fast-tempo music. *PLoS One* 12:e0189075. doi: 10.1371/journal.pone.0189075

Paul, S. (2009). Binaural recording technology: A historical review and possible future developments. *Acta Acust. United Acust.* 95, 767–788. doi: 10.3813/aaa.918208

Phillips, A. C., Carroll, D., Hunt, K., and Der, G. (2006). The effects of the spontaneous presence of a spouse/partner and others on cardiovascular reactions to an acute psychological challenge. *Psychophysiology* 43, 633–640. doi: 10.1111/j.1469-8986.2006.00462.x

Pomeranz, B., Macaulay, R. J., Caudill, M. A., Kutz, I., Adam, D., Gordon, D., et al. (1985). Assessment of autonomic function in humans by heart rate spectral analysis. *Am. J. Physiol.* 248, H151–H153.

Porges, S. W. (1995). Orienting in a defensive world: mammalian modifications of our evolutionary heritage. A polyvagal theory. *Psychophysiology* 32, 301–318. doi: 10.1111/j.1469-8986.1995.tb01213.x

Riederer, K. A. (2003). Effects of eye-glasses, hair, headgear, and clothing on measured head-related transfer functions Part Ib. *J. Acoust. Soc. Am.* 114, 2388–2388. doi: 10.1121/1.4777835

Russell, J. A., Weiss, A., and Mendelsohn, G. A. (1989). Affect grid: a single-item scale of pleasure and arousal. *J. Pers. Soc. Psychol.* 57:493. doi: 10.1037/0022-3514.57.3.493

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat. Neurosci.* 14, 257–262. doi: 10.1038/nn.2726

Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., and Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PLoS One* 4:e7487. doi: 10.1371/journal.pone.0007487

Shahid, I. Z., Rahman, A. A., and Pilowsky, P. M. (2012). Orexin A in rat rostral ventrolateral medulla is pressor, sympatho-excitatory, increases barosensitivity and attenuates the somato-sympathetic reflex. *Br. J. Pharmacol.* 165, 2292–2303. doi: 10.1111/j.1476-5381.2011.01694.x

Sharp, A., Houde, M. S., Maheu, M., Ibrahim, I., and Champoux, F. (2019). Improved tactile frequency discrimination in musicians. *Exp. Brain Res.* 237, 1575–1580.

Shields, R. W. Jr. (1993). Functional anatomy of the autonomic nervous system. *J. Clin. Neurophysiol.* 10, 2–13. doi: 10.1097/00004691-199301000-00002

Shoda, H., Adachi, M., and Umeda, T. (2016). How live performance moves the human heart. *PLoS One* 11:e0154322. doi: 10.1371/journal.pone.0154322

Skille, O. (1987a). "Low frequency sound massage—the music bath—a follow-up-report," in *Musik in Der Medizin/Music in Medicine*, eds R. Spintge and R. Droh (Berlin: Springer), 257–259. doi: 10.1007/978-3-642-71697-3_25

Skille, O. (1987b). "The music bath—possible use as an anxiolytic," in *Musik in Der Medizin/Music in Medicine*, eds R. Spintge and R. Droh (Berlin: Springer), 253–255. doi: 10.1007/978-3-642-71697-3_24

Standley, J. M. (1986). Music research in medical/dental treatment: meta-analysis and clinical applications. *J. Music Ther.* 23, 56–122. doi: 10.1093/jmt/23.2.56

Task Force (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. task force of the european society of cardiology and the North American society of pacing and electrophysiology. *Eur. Heart J.* 17, 354–381.

Terasaki, M., Kishimoto, Y., and Koga, A. (1992). Construction of a multiple mood scale. *Shinrigaku kenkyu* 62, 350–356. doi: 10.4992/jjpsy.62.350

Theile, G., and Wittek, H. (2004). Wave field synthesis: a promising spatial audio rendering concept. *Acoust. Sci. Technol.* 25, 393–399.

Thomas, T., and Spyer, K. M. (1999). A novel influence of adenosine on ongoing activity in rat rostral ventrolateral medulla. *Neuroscience* 88, 1213–1223. doi: 10.1016/s0306-4522(98)00296-6

Wersényi, G., and Illényi, A. (2005). Differences in dummy-head HRTFs caused by the acoustical environment near the head. *Electron. J. Tech. Acoust.* 1, 1–15.

Zhou, P., Sui, F., Zhang, A., Wang, F., and Li, G. (2010). "Music therapy on heart rate variability," in *Proceedings of the 2010 3rd International Conference on Biomedical Engineering and Informatics*, (Yantai: Institute of Electrical and Electronics Engineers), 965–968.

**Conflict of Interest:** YO and MKa are employees of NTT Communication Science Laboratories, which is a basic-science research section of Nippon Telegraph and Telephone (NTT) Corporation. There are no pending patents involving the reported research. There are no products in development or marketed products to declare. Thus, the authors adhere to the policies of Frontiers.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

# Advantages of publishing in Frontiers

**OPEN ACCESS**
Articles are free to read for greatest visibility and readership

**FAST PUBLICATION**
Around 90 days from submission to decision

**HIGH QUALITY PEER-REVIEW**
Rigorous, collaborative, and constructive peer-review

**TRANSPARENT PEER-REVIEW**
Editors and reviewers acknowledged by name on published articles

**Frontiers**
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

**Visit us:** www.frontiersin.org
**Contact us:** frontiersin.org/about/contact

**REPRODUCIBILITY OF RESEARCH**
Support open data and methods to enhance research reproducibility

**DIGITAL PUBLISHING**
Articles designed for optimal readership across devices

**FOLLOW US**
@frontiersin

**IMPACT METRICS**
Advanced article metrics track visibility across digital media

**EXTENSIVE PROMOTION**
Marketing and promotion of impactful research

**LOOP RESEARCH NETWORK**
Our network increases your article's readership