NEUROSCIENCE AND NEW MUSIC: ASSESSING BEHAVIORAL AND CEREBRAL ASPECTS OF ITS PERCEPTION, COGNITION, INDUCTION, AND ENTRAINMENT

EDITED BY: Thomas James Lundy, Clara Eline James and Mark Reybrouck
PUBLISHED IN: Frontiers in Psychology and Frontiers in Neuroscience

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-291-4
DOI 10.3389/978-2-83250-291-4

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

NEUROSCIENCE AND NEW MUSIC: ASSESSING BEHAVIORAL AND CEREBRAL ASPECTS OF ITS PERCEPTION, COGNITION, INDUCTION, AND ENTRAINMENT

Topic Editors:
Thomas James Lundy, Cuttlefish Arts, United States
Clara Eline James, University of Applied Sciences and Arts of Western Switzerland, Switzerland
Mark Reybrouck, KU Leuven, Belgium

Citation: Lundy, T. J., James, C. E., Reybrouck, M., eds. (2022). Neuroscience and New Music: Assessing Behavioral and Cerebral Aspects of its Perception, Cognition, Induction, and Entrainment. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-291-4

Table of Contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 27 July 2022
DOI 10.3389/fpsyg.2022.979570

OPEN ACCESS

EDITED AND REVIEWED BY
William Polk Cheshire,
Mayo Clinic Florida, United States

*CORRESPONDENCE
Thomas Lundy
lundy.thomas.james@gmail.com

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology

RECEIVED 27 June 2022
ACCEPTED 12 July 2022
PUBLISHED 27 July 2022

CITATION
Lundy T, James C and Reybrouck M (2022) Editorial: Neuroscience and new music: Assessing behavioral and cerebral aspects of its perception, cognition, induction, and entrainment. *Front. Psychol.* 13:979570.
doi: 10.3389/fpsyg.2022.979570

Editorial: Neuroscience and new music: Assessing behavioral and cerebral aspects of its perception, cognition, induction, and entrainment

Thomas Lundy[1,2]*, Clara James[3,4] and Mark Reybrouck[5,6]

[1]Cuttlefish Arts, Wheat Ridge, CO, United States, [2]Department of Mathematics, University of Colorado Denver, Denver, CO, United States, [3]Geneva Musical Minds Lab (GEMMI lab), Geneva School of Health Sciences, University of Applied Sciences and Arts Western Switzerland HES-SO, Geneva, Switzerland, [4]Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland, [5]Musicology Research Group, Faculty of Arts, KU Leuven-University of Leuven, Leuven, Belgium, [6]IPEM, Department of Art History, Musicology and Theatre Studies, Ghent, Belgium

KEYWORDS

New Music, neuroscience, editorial, cognition, music perception

Editorial on the Research Topic
Neuroscience and new music: Assessing behavioral and cerebral aspects of its perception, cognition, induction, and entrainment

Defining "New Music" is challenging both temporally (there is no clear starting point) and musically. It comprises any musical idiom diverging from the common practice period, unfamiliar to many listeners. Unfamiliar because of the partial or total abandon of tonality, compositional conventions, and the utilization of unusual sounds and rhythms. All of these aspects require learning new ways of arranging musical notes according to the idiom of each composer or even each piece, soliciting effortful new cognitive and cerebral strategies from the listener. New Music, therefore, is an umbrella term that embraces diverse musical genres: Musique Concrète, electronic synthesis, Elektronische Musik, the controversial second Vienna School, free and serial atonal and pointillistic compositions, spectral music, microtonal music, and many others. One may argue that popular genres such as electronics, dance music, Ambient Music, Sound Art, and Postclassical Minimal Music can join the list. Nowadays, there is a dissipation of the barriers between so-called serious music and more popular music—they all rely on musical sounds and rhythms and their possible transformations as essential constituents.

We launched the Special Research Topic on "*Neuroscience and New Music: Assessing Behavioral and Cerebral Aspects of Its Perception, Cognition, Induction, and Entrainment*" to address the growing need for theoretical and empirical grounding regarding New Music's perception, valuation, and production. Two overarching themes were specified initially: (i) the actual processing of New Music, with a distinction between perception, cognition, induction, and entrainment, and (ii) the neural correlates of these levels of processing and the neuroimaging techniques for their assessment.

The answer to the call for papers was not abundant, which was somewhat unexpected, given the current explosion of music and brain studies. Even if the COVID-19 crisis curtailed and even derailed some exciting proposals, narrowing the scope to New Music reduced the pool of relevant studies, clearly illustrating the salient character of this Research Topic that is underrepresented in the literature.

We finally selected ten papers revolving mainly around topics as divergent as embodied cognition, creative composing, behavioral responses, perceptual issues, aesthetic judgment, predictive processing, the role of technology, and electrophysiological and fMRI measurements.

Foster et al. compared the tempo judgment of disk jockeys (DJ), percussionists, and melodic instrumentalists (all professionals) with untrained controls. The authors show that there is no difference in performance between DJs and other professional musicians in any tempo range, i.e., DJs are on par with professional musicians regarding tempo judgment. The book review by Besada sets Mariusz Kozak's 2020 volume within the broader context of rhythm and musical time. In particular, Besada draws attention to Kozak's emphasis on embodiment and enaction to target the listener's experience while noting that more high-level and abstract representations of time are valuable tools for composers. Besada also commends Kozak for conducting original empirical research to support his reasoning. The contribution by Arkhipova et al. examines the Different Hearing Program (DHP)—which involves creative group composing in the classroom. The authors applied functional brain imaging and behavioral techniques to determine if DHP induces changes in subjective appreciation of different classes of music. The results imply that DHP training altered the activation of functional brain networks, with default mode network activation increasing and executive network activation decreasing in relation to creative thinking. The study by Dauer et al. explored whether inter-subject correlation (ISC) of electroencephalographic responses (EEG) can capture engagement with minimalist music. The authors collected EEG and continuous behavioral (CB) data while 30 adults listened to an excerpt from Steve Reich's *Piano Phase*, three controlled manipulations of Piano Phase, and a popular-music remix of the work. Results show that EEG responses and CB ISC levels were highest for the remix stimuli and lowest for the most recurrent manipulations. Godøy's contribution discusses the concept of the sound object, as embodied in the *Musique Concrète* of Pierre Schaffer. Godøy argues that the sound object is relevant to many musical genres and styles, not just Musique Concrète. Furthermore, he posits that sound objects can better encapsulate time-dependent music features, i.e., various dynamic, textural, timbral, and expressive envelopes, than traditional Western music theory frameworks. Touizrar et al. examines the relationship between the perception of repetition, comprehensibility, and aesthetic judgment in the domain of post-tonal music. Sixty participants identified repetition in 14 three-minute excerpts from ensemble and orchestra pieces selected from three categories: Tonal, Modernist, and Post-1970. Statistical analysis demonstrates a degree of independence between repetition and aesthetic preference. The study by Færøvik et al. notes that the temporal, repetitive nature of musical rhythms is ideal for investigating the phenomena of auditory repetition suppression and omission activation. The study successfully replicated previous findings that left dominant superior temporal brain activation decreases during repetition of an unaltered rhythmic stimulus and right lateralized middle temporal brain activation increases in response to omissions. Generalized activation in error detection areas likely evidenced working memory involvement. These results show that tailored musical stimuli in an fMRI setting permit robust investigation of such neural phenomena. Washburn et al. is a detailed examination of the interaction between acoustic transmission latency (ATL) and asynchrony between performers. The authors found that **n**etworked music performances create longer ATL between performers than traditional in-person settings and, conversely, longer transmission latencies and more significant temporal asynchrony between performers. The perspective article by Mencke et al. notes that the tonal or metrical hierarchy of atonal music exhibits low predictability, whereas Western tonal music is more predictable owing to its hierarchical structure. Thus, listening to atonal music requires generating predictive models. This behavior is characterized in behavioral neuroscience as fulfilling an innate drive to reduce uncertainty but has received little attention in empirical music research. They also suggest new research avenues for deepening our understanding of the aesthetic experience of atonal music in particular and revealing core qualities of the aesthetic experience in general. Finally, Phillips et al. examines listeners' segmentation decisions in a piece of contemporary music, Ligeti's *Fanfares*, *via* the Practice & Research in Science & Music (PRiSM) Perception smartphone App. Listeners tapped when they felt that a section had ended. Listeners showed high levels of agreement. Moreover, familiarity with contemporary repertoire did not seem to influence the responses, and the differential effect of musical training was small.

In order to grasp New Music, we must adopt alternate tonal and rhythmic frameworks. These include embodiment, attention on recurring or missing elements, innovative thinking, perceiving music as sound objects, and more. Using functional brain imaging (fMRI, EEG) may allow investigating how the brain can process these new musical texts into coherent representations. Clearly, this field requires more comprehensive and well-coordinated research. We hope this Research Topic contributes to that.

10.3389/fpsyg.2022.979570

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

Author TL was employed by Cuttlefish Arts.

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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 05 October 2021
doi: 10.3389/fpsyg.2021.709979

Check for updates

Accuracy of Tempo Judgments in Disk Jockeys Compared to Musicians and Untrained Individuals

Nicholas E. V. Foster[1,2,3]*, *Lauriane Beffa*[2,3] *and Alexandre Lehmann*[1,2,3]

[1]*Department of Otolaryngology Head and Neck Surgery, McGill University, Montreal, QC, Canada,* [2]*International Laboratory for Brain, Music and Sound Research (BRAMS), Montreal, QC, Canada,* [3]*Center for Research on Brain, Language and Music (CRBLM), Montreal, QC, Canada*

Professional disk jockeys (DJs) are an under-studied population whose performance involves creating new musical experiences by combining existing musical materials with a high level of temporal precision. In contemporary electronic dance music, these materials have a stable tempo and are composed with the expectation for further transformation during performance by a DJ for the audience of dancers. Thus, a fundamental aspect of DJ performance is synchronizing the tempo and phase of multiple pieces of music, so that over seconds or even minutes, they may be layered and transitioned without disrupting the rhythmic pulse. This has been accomplished traditionally by manipulating the speed of individual music pieces "by ear," without additional technological synchronization aids. However, the cumulative effect of this repeated practice on auditory tempo perception has not yet been evaluated. Well-known phenomena of experience-dependent plasticity in other populations, such as musicians, prompts the question of whether such effects exist in DJs in their domain of expertise. This pilot study examined auditory judgments of tempo in 10 professional DJs with experience mixing by ear, compared to 7 percussionists, 12 melodic instrumental musicians, and 11 untrained controls. Participants heard metronome sequences between 80 and 160 beats per minute (BPM) and estimated the tempo. In their most-trained tempo range, 120–139 BPM, DJs were more accurate (lower absolute percent error) than untrained participants. Within the DJ group, 120–139 BPM exhibited greater accuracy than slower tempos of 80–99 or 100–119 BPM. DJs did not differ in accuracy compared to percussionists or melodic musicians on any BPM range. Percussionists were more accurate than controls for 100–119 and 120–139 BPM. The results affirm the experience-dependent skill of professional DJs in temporal perception, with comparable performance to conventionally trained percussionists and instrumental musicians. Additionally, the pattern of results suggests a tempo-specific aspect to this training effect that may be more pronounced in DJs than percussionists and musicians. As one of the first demonstrations of enhanced auditory perception in this unorthodox music expert population, this work opens the way to testing whether DJs also have enhanced rhythmic production abilities, and investigating the neural substrates of this skill compared to conventional musicians.

Keywords: DJ, training, practice, auditory, tempo, rhythm, beat, absolute perception

OPEN ACCESS

Edited by:
Clara Eline James, University of Applied Sciences and Arts of Western Switzerland, Switzerland

Reviewed by:
Blake Edward Butler, Western University, Canada
Paul J Colombo, Tulane University, United States

****Correspondence:***
Nicholas E. V. Foster
nicholas.foster@umontreal.ca

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

Received: *14 May 2021*
Accepted: *10 September 2021*
Published: *05 October 2021*

Citation:
Foster NEV, Beffa L and Lehmann A (2021) Accuracy of Tempo Judgments in Disk Jockeys Compared to Musicians and Untrained Individuals. Front. Psychol. 12:709979. doi: 10.3389/fpsyg.2021.709979

INTRODUCTION

An essential element of musical rhythm is the "beat" or regular rhythmic pulse to which we often dance or clap. While listening to music, the beat—a periodic, isochronous pulse—is readily perceived from the music even in musically untrained individuals, and even when the rhythmic patterns present in the music do not always coincide with the pulse (Cannon and Patel, 2021). Together, the beat and the rate of the beat (the musical tempo, commonly expressed in beats per minute, "BPM") represent fundamental temporal features in music perception (Geiser et al., 2014). In turn, tempo plays an important role in musical expression and appreciation (Palmer, 1997; Quinn and Watt, 2006). Although a melody's identity can be defined in terms of relative structure, such as pitch and time intervals, the absolute "surface features" of tempo, key, and timbre are important attributes by which we can identify and remember particular musical performances (Levitin and Cook, 1996; Schellenberg et al., 2014; Jakubowski et al., 2018) and musical works (Clynes and Walker, 1986; Bailes and Barwick, 2011).

Precise perception and synchronization with the beat are important for performing musicians, as well as for other expert populations whose performance involves manipulating and combining music, such as disk jockeys (DJs). However, although musicians have been highly studied as a model of experience-dependent plasticity in various aspects of auditory perception, DJs have only recently gained attention as an expert population in this domain (Butler and Trainor, 2015). Consequently, much remains to be discovered about how the particular experience of DJs may lead to similar or unique patterns of ability compared to conventionally trained musicians.

Musical training is associated with greater accuracy of temporal perception and sensorimotor synchronization. Compared to untrained individuals, musicians have improved detection of timing differences for individual sound durations as well as for temporal manipulations embedded within sound sequences (Yee et al., 1994; Rammsayer and Altenmüller, 2006). Musicians are also sensitive to smaller changes of tempo than nonmusicians (Ellis, 1991; Sheldon, 1994). Percussionists, whose training emphasizes rhythmic timing, exhibit particular advantages over non-percussionists in performing fine temporal discrimination (Ehrlé and Samson, 2005; Manning and Schutz, 2016) as well as in sensorimotor timing synchronization and reproduction tasks (Krause et al., 2010; Cicchini et al., 2012; Cameron and Grahn, 2014).

Given the importance of musical rhythm and timing in DJ performance, and a comparable degree of long-term practice between DJs and many musicians (e.g., Butler and Trainor, 2015), we may predict that DJs also possess experience-dependent expertise on rhythmic tasks. A central skill in DJ performance is combining and transitioning between different musical works in a rhythmically seamless way. In its most long-standing form, this involves modulating the playback of a sound recording by directly interacting with a vinyl record and a variable-speed turntable in order to bring two songs into precise temporal alignment at a common tempo, and further adjusting playback as necessary to maintain synchrony (Broughton and Brewster, 2003, p. 50). This approach therefore relies on a combination of precise temporal perception to detect asynchrony and sensorimotor control to apply appropriate corrections to the record or turntable. In the context of an entire DJ performance, this process is aided by memory for the musical structure of each particular song, including its rhythm and beat tempo (Brewster and Broughton, 2000, p. 8). Similar to musicians, for whom tempo is often indicated explicitly on musical scores and controlled during practice by the use of a metronome, DJs are aided in choosing and synchronizing music by knowledge of the explicit tempo value of each song (Broughton and Brewster, 2003, p. 83). Whereas enhanced rhythmic perception and sensorimotor precision have been measured in DJs, as reviewed below, the accuracy of tempo judgments in DJs remains unexamined and is the focus of the present study.

Current evidence for training-related expertise in rhythmic perception in DJs comes from research by Butler and Trainor (2015). In that study, professional club DJs, percussionists, and untrained controls were tested on their ability to detect deviations in auditory rhythmic patterns. Compared to the untrained group, both the DJ and percussionist groups showed enhanced detection of deviations from "on beat" timing for a probe stimulus following an isochronous pattern and a brief silent period. The DJ group's detection sensitivity was comparable to the percussion group across all levels of deviation. The authors interpret the enhanced performance on this task to reflect both greater sensitivity to small timing discrepancies and more accurate memory for rhythm through the brief silent period.

Although memory for musical tempo has not yet been studied in DJs, several studies have demonstrated that absolute tempo is remembered for highly familiar songs in untrained individuals, and this tempo memory also exists for newly heard songs when tested at shorter intervals (e.g., minutes or hours). An early demonstration of this phenomenon was made by Levitin and Cook (1996), who found that when asked to sing, hum, or whistle from memory a popular song they knew well, most participants' renditions had a tempo within 8% of the recorded version. Subsequent studies have found similar accuracy in the range of 3–18% for highly familiar music in participants unselected for musical experience (Jakubowski et al., 2015, 2016, 2018). Additional support for the stability of absolute tempo representation comes from the observation that songs sung by mothers to their infants deviate in tempo by only 3% across sessions (Bergeson and Trehub, 2002), and evidence for memory of absolute tempo for unfamiliar birdsong (Roeske et al., 2020) and periodic environmental sounds like footsteps (Boltz, 2010) show that this ability generalizes beyond music. Evidence for a link between absolute tempo ability and musical training currently remains unclear and limited by methodological approach. A study of musicians and nonmusicians by Gratton et al. (2016) showed that untrained individuals can learn a novel absolute tempo scale but did not show clearly higher performance in musicians compared to untrained controls. In that study, the spacing between unique items was quite large (12%) and may not have had sufficient resolution to find differences between the training groups. Another study that attempted to relate individual differences in musical training to tempo accuracy of remembered familiar music did not find a significant relation (Jakubowski et al., 2018), but the number of participants was small. Nonetheless, these studies validate the idea that at least at a coarse level, tempo is remembered regardless of musical experience.

While a basic fidelity of musical tempo memory has been established in previous research, DJs present an ideal population to test whether this ability can be enhanced by sustained practice. In turn, this approach offers an opportunity to broaden the empirical description of an under-studied model of experience-dependent plasticity in professional DJs. Hence, in consideration of the importance of musical tempo to DJ performance, the goal of the present pilot study was to characterize tempo perception in highly trained DJs and compare its accuracy with conventionally trained experts in rhythm and music, i.e., percussionists and melodic musicians. The inclusion of three expert groups, with similar training duration but different training characteristics, allowed us to assess the specificity of tempo estimation ability to different training profiles. Tempo perception was tested by presenting a range of metronome tempos and measuring the accuracy of tempo estimates in groups of DJs, percussionists, melodic musicians, and untrained individuals. The results were examined to determine the effects of training and tempo upon the accuracy of this judgment. We predicted that DJs would show an enhanced ability to discern the tempo of metronome sounds compared to untrained individuals, and similar accuracy when compared to percussionists and melodic musicians.

MATERIALS AND METHODS

Participants

Forty participants (23 male, mean age 29.3 years, SD 8.4) participated in the study, comprising 10 professional DJs with experience mixing by ear, 7 percussionists, 12 melodic instrumental musicians, and 11 untrained controls. **Table 1** summarizes the characteristics of the study groups. DJs were required to have at least 10 years of experience DJing for at least 6 h per week and to mix music without the aid of synchronizing software (e.g., using turntables). Melodic musicians and percussionists were required to have at least 10 years of musical practice for at least 6 h per week. Individuals in the untrained group were required to have less than 3 years of musical training or DJ experience. Participants completed a questionnaire to verify their history of musical or DJ experience. Group matching was not achieved for gender or age. The DJ group consisted of only male participants, whereas the other groups had between 27 and 72% male participants. Mean age was greater in the DJ group and did not differ among the other groups [$F(3,36) = 13.9$, $p < 0.001$; Tukey-corrected pairwise comparisons with DJ group $p \leq 0.014$]. These potential group confounds were addressed in additional analyses as described later. The experimental protocol was approved by the Comité d'éthique de la recherche en arts et en sciences at the University of Montreal. Participants gave informed consent and were compensated for their time.

Stimuli

Stimuli in the tempo estimation task consisted of isochronous patterns of a metronome sound. The full set of 81 stimulus rates included every integer BPM rate from 80 to 160 (i.e., inter-onset intervals of 750 ms at 80 BPM to 375 ms at 160 BPM). The metronome sound was 25 ms in duration, with a strong initial transient, and having spectral energy concentrated between 500 Hz and 3 kHz. Each stimulus was created from the metronome sound repeated at the BPM rate with silence in the intervening samples. The total duration of each stimulus varied between 5 and 10 s and was chosen randomly on each trial.

Procedure

At the beginning of the testing session, participants completed a questionnaire to assess their history of DJ and musical experience. DJs were also asked to indicate the range of musical tempos they played most often.

The tempo estimation task then took place in a soundproof audio booth. The task was implemented using the Psychophysics Toolbox extensions (Psychtoolbox; RRID:SCR_002881) in Matlab (RRID:SCR_001622), running on a MacBook Pro computer connected to an RME Fireface audio interface, and Behringer HA8000 headphone amplifier. Participants heard audio stimuli *via* Beyerdynamic DT 770 PRO 250 Ohm headphones and received visual prompts and feedback on a computer display.

Before testing commenced, the tempo estimation task was explained to the participant. Music recordings were played to

TABLE 1 | Participant characteristics.

Group	*n*	n Male	% Male	Age*		Years DJ experience		Years music experience†	
				Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)
DJ	10	10	100.0	30–52	39.2(7.7)	9–40	19.6(11.1)	0–12	3.3(5.1)
Percussionist	7	5	71.4	22–45	29.9(7.6)	0–5.5	0.8(2.1)	11–27	16.9(5.1)
Musician	12	5	41.7	20–33	26.2(4.8)	0–0.5	0.1(0.2)	10–28.5	16.7(5.2)
Untrained	11	3	27.3	20–28	23.7(3.3)	0–0	0.0(0.0)	0–3	1.0(1.1)

*Mean age is greater in the DJ group and does not differ among the other groups [$F(3,36) = 13.9$, $p < 0.001$; Tukey-corrected pairwise comparisons with DJ group $p \leq 0.014$].
†Music experience is based on years of formal practice on a musical instrument.

illustrate specific tempo values, covering the range used in the task: "Halo" by Beyonce (80 BPM), "Stayin' Alive" by the Bee Gees (104 BPM), "Raise Your Glass" by Pink (122 BPM), "Viva la Vida" by Coldplay (138 BPM), and "Happy" by Pharrell Williams (160 BPM). These song tempos were determined by automatic analyzer (Traktor software, Native Instruments, Berlin, Germany) and confirmed *via* online databases[1,2] and manually by an experienced DJ.

In the tempo estimation task, each trial started with a 500 Hz warning beep for 1 s, followed by 500 ms of silence, followed by the metronome stimulus. Upon completion of the metronome stimulus, the participant was prompted to estimate the stimulus tempo in BPM using the computer keyboard. After the participant's response, the actual stimulus tempo was displayed on the screen for 1 s. The participant was then presented with a random number between 20 and 100 and asked to count backward from that number, out loud, until a subsequent stop message 5 s later. In combination, the feedback and subsequent counting phase were intended to help orient the participants (especially the untrained group) to the absolute tempo scale, while interfering with relative comparison of successive trial stimuli by increasing irrelevant load on auditory working memory. Finally, a 3 s delay followed each trial.

Participants completed four practice trials at randomly selected tempos and then had the opportunity to ask questions before the testing phase began. In the testing phase, participants completed 36 trials of the tempo estimation task at tempos randomly selected from the full set of 81 tempos between 80 and 160 BPM, with no tempos repeated. To reduce fatigue, at an interval of every nine trials, the participant was allowed to take a pause before resuming the task. The entire task was about 30–45 min in duration, including instructions and practice trials.

Analyses

Given the stimulus tempo range of 80–160 BPM, tempo responses of <20 BPM were considered to be at risk of typing errors and were discarded prior to analysis (N = 8 trials out of 1,440 total). The remaining responses had a range of 63–172 BPM.

These BPM responses were then converted to absolute percent error, i.e., 100*abs[(response BPM—actual BPM)/(actual BPM)]. When stimulus BPM was used as an independent variable in analysis, it was binned into four tempo ranges: 80–99 BPM, 100–119 BPM, 120–139 BPM, and 140–160 BPM.

Statistical analysis was performed using R (R Core Team, 2019, RRID:SCR_001905). In consideration of the relatively small sample size and lack of balance between groups (McNeish and Harring, 2017), a mixed-effects model was used to test the effects of group and tempo range on absolute percent error scores on the BPM judgment task, using lmerTest (Kuznetsova et al., 2017; RRID:SCR_015656) with lme4 (Bates et al., 2015; RRID:SCR_015654). Group and tempo range were parameterized using sum-to-zero contrasts, and a random effect of participant was added to account for the hierarchical structure of the data. Omnibus tests of the main effects and their interaction were performed on this model *via* the car package (Fox and Weisberg, 2011), yielding F and *p* values calculated from Wald tests using Kenward-Roger approximated degrees of freedom (Halekoh and Højsgaard, 2014) and type 3 sum of squares. Following a significant interaction effect, the emmeans R package (Lenth, 2020; RRID:SCR_018734) was used to test post-hoc comparisons between groups within a tempo range (24 pairwise comparisons) and differences between tempo ranges within a group (24 pairwise comparisons). The post-hoc tests were corrected for 48 multiple comparisons using the Sidak method (Abdi, 2007).

In order to assess potential confounds between group, age, and gender, additional mixed-effects models were run in specific groups to test age and gender as predictors of absolute percent error scores. The first model included age, tempo range, and their interaction as predictors, with participant as a random effect. The second model included gender, tempo range, and their interaction, with participant as a random effect. Both of these models were tested in the untrained group alone, as well as with a pooled sample composed of the untrained, percussion, and musician groups.

RESULTS

Participants in the DJ group were asked to report the typical range tempos of the music they play. The distribution of those ranges is shown in **Figure 1**. A total of 72.9% of this distribution falls into the range 120–130 BPM, with a peak at 125 BPM, and the remaining 27.1% of the distribution falls into the range 110–119 BPM.

The mixed-effects analysis of absolute percent error on BPM judgments found main effects of group [$F(3, 36.02) = 5.67$, $p = 0.003$] and tempo range [$F(3, 1389.56) = 19.86$, $p < 0.001$], and interaction between group and tempo range [$F(9, 1389.55) = 2.70$, $p < 0.001$]. Mean error by group and tempo range is shown in **Figure 2**. Post-hoc comparisons showed that in their most-trained tempo range of 120–139 BPM, DJs were more accurate than untrained participants (DJ error 3.10%, untrained error 7.91%, $p < 0.001$). Additionally, within the DJ group, this 120–139 BPM range exhibited greater accuracy than for slower tempos of 80–99 (error 7.54%, $p < 0.001$) or 100–119 BPM (error 7.59%, $p < 0.001$). When compared with the groups having conventional musical training, DJs did not differ in accuracy compared to percussionists or melodic instrumental musicians on any BPM range ($p \geq 0.84$).

Similar to DJs, percussionists were more accurate than untrained individuals on both the 100–119 (percussionist error 5.14%, untrained error 9.39%, $p = 0.017$) and 120–139 BPM ranges (percussionist error 3.84%, untrained error 7.91%, $p = 0.018$). Melodic musicians did not differ in accuracy with any other group, although within 120–139 BPM, the difference in accuracy between musicians and controls was near significance (120–139 BPM musicians vs. controls, $p = 0.065$; other comparisons $p \geq 0.255$).

[1]http://beatport.com
[2]http://songbpm.com

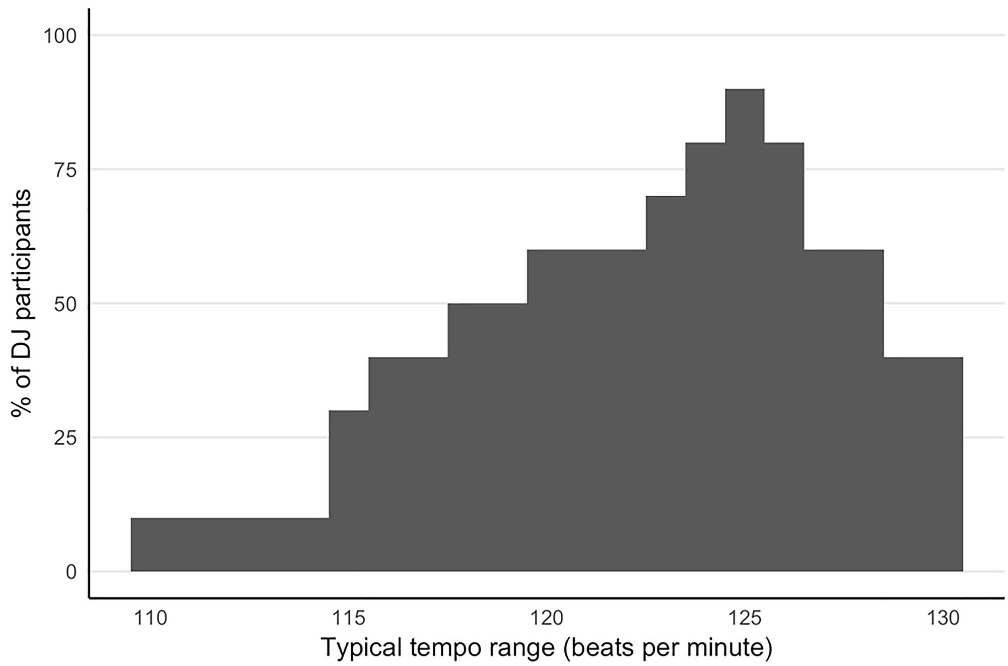

FIGURE 1 | Distribution of tempos reported as typically played by participants in the DJ group. Each participant reported a range of tempos (minimum and maximum), and the distribution shows at each BPM what proportion of DJ participants included this tempo in their range.

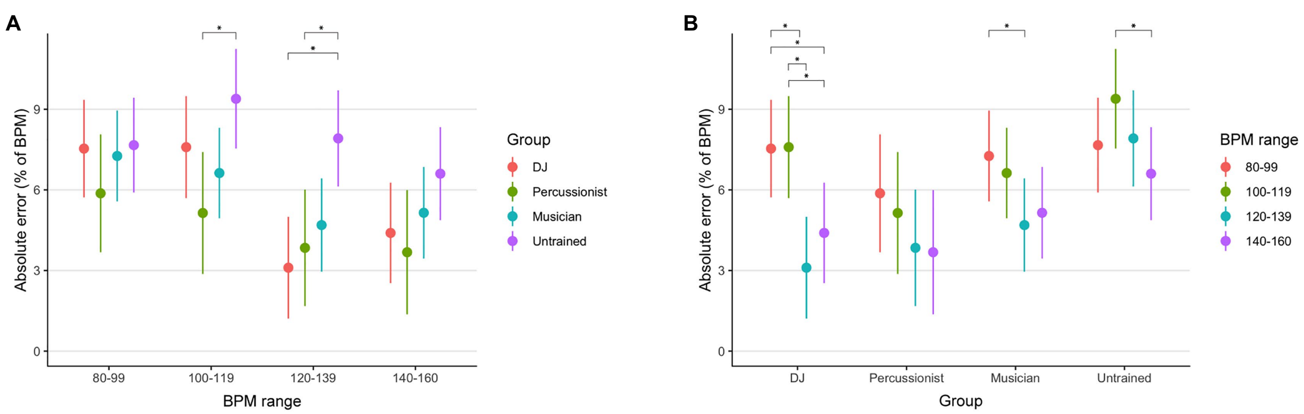

FIGURE 2 | Mean accuracy of tempo judgments by participant group and tempo range. In panel **(A)**, the values are organized by group, and Sidak-corrected significance of pairwise differences between tempo ranges within groups is indicated. In panel **(B)**, values are organized by tempo range and Sidak-corrected significance of pairwise differences between groups is indicated. Error bars indicate 95% confidence intervals of the estimated marginal means.

Because of the greater mean age and exclusively male composition of the DJ group, it was important to test whether age or gender could have imparted a confounding effect on performance differences found in the DJ group. For both age and gender, additional analyses on error scores were performed in the untrained group alone, as well as across a pooled sample of the untrained, percussion, and musician groups. No significant effects of age or gender were found. In untrained individuals, there was no main effect of gender [$F(19.07)=0.19$, $p=0.671$] nor an interaction between gender and tempo range [$F(3,379.56)=1.4750$, $p=0.221$]. Similarly, across all the non-DJ groups, there was no main effect of gender [$F(128.03)=2.45$, $p=0.13$] nor an interaction between gender and tempo range [$F(3,1044.16)=0.59$, $p=0.62$]. In the untrained group, in which ages ranged from 20 to 28 years, there was no main effect of age [$F(19.03)=0.12$, $p=0.73$] nor an interaction between age and tempo range [$F(3,378.75) = 0.67$, $p = 0.57$]. Across all the non-DJ groups, in which ages ranged from 20 to 45 years, there was a marginal effect of age [$F(128.31) = 4.10$, $p = 0.052$] and no interaction between age and tempo range [$F(3,1046.09) = 0.18$, $p = 0.912$]. Here, even given the possibility of a main effect of age, there is no evidence for a specificity of effect to any particular tempo range, in contrast to the performance differences noted for the DJ and percussion groups earlier.

DISCUSSION

These pilot results provide evidence that DJs and percussionists are more accurate at estimating auditory metronome tempo than untrained individuals. In DJs, this increased accuracy showed a specificity to the musical tempo ranges that participants reported as typical for their practice (120 to 139 BPM) and approached the previously reported perceptual discrimination threshold for tempo comparisons (Drake and Botte, 1993). The results add to the evidence that DJs are a highly trained expert population who exhibit heightened musical perception, comparable to percussionists, even though the nature of their craft and training is different from conventionally trained musicians (Butler and Trainor, 2015). As both the tempo estimation task and the study population are relatively novel, in the following discussion, we will examine several aspects of the task with consideration to both the baseline of tempo perception in the general untrained population and the tempo-specific enhancement observed in DJs.

The task in the present study asked participants to estimate the tempo of metronome stimuli using numeric values of beats per minute. To orient participants for whom this scale is unfamiliar, several pieces of popular music were played before the experiment, which together covered the range of tested tempos. Additionally, the true tempo was displayed following each trial, after the participant recorded their response. These elements of the procedure made it feasible to test tempo estimation accuracy using a continuous, concrete response scale in both untrained and trained individuals. An interfering step was added after the feedback to guard against performing the task simply using auditory working memory. This was done because the display of the true tempo of each trial could also make it feasible to perform the task in sequential and relative manner, simply comparing the tempo of the present item with the label and perceptual memory of the previous item; this relative aspect of absolute perceptual tasks exists even in the absence of feedback (Lockhead, 2004). To interfere with this effect, participants were required to count backward, out loud, for a 5 s period between each trial. Nonetheless, part of the increased accuracy in tempo estimation by the trained groups that we observe could reflect a greater resilience of memory from trial to trial and decreased interference from the backward counting.

The perceptual and associative operations involved in the task may be useful to consider separately (Kim and Knösche, 2017), especially because much of the existing research on absolute tempo memory does not involve explicit labeling of tempo values. The tempo judgment task in our study asks participants to respond to each stimulus by estimating its tempo in beats per minute and hence involves a combination of perceptual extraction of the absolute tempo and an associative process of labeling this tempo with a numeric value. We may imagine that the cumulative experience of practicing as a DJ or musician improves both the perceptual and associative abilities involved in the absolute tempo memory phenomenon examined here, but potentially to different extents or with different training trajectories.

Some studies of absolute tempo perception have used tasks without any explicit tempo labeling, and their results may help elucidate the perceptual component. Halpern and Müllensiefen (2008) exposed participants having various degrees of musical experience to unfamiliar melodic excerpts, then in a subsequent testing phase, presented tempo-modified and original versions of these melodies along with novel melodies. The tempo-modified versions, which had been speeded or slowed by 15–20%, were rated as less familiar than the original versions, and this effect was not modulated by musical experience. A more recent study by Schellenberg et al. (2014) has replicated this result using a greater tempo change. Furthermore, the basic ability to remember the tempo of musical pieces appears to emerge quite early in development, as shown in infants using tempo alterations of 25% (Trainor et al., 2004).

In quantitative terms, the untrained participants in the present study had a mean absolute error of about 7.5% over the full range of tempos; DJs had the best error rate in the 120–139 BPM range, at 3%. Although no previous studies have tested absolute tempo labeling on an explicit BPM scale, the magnitude of error is comparable to that found in other types of tasks of absolute tempo memory. Studies that involve reproducing a familiar tune's tempo, whether by singing, humming, or tapping, generally report error values between 3 and 18% in participants unselected for musical experience. For example, Levitin and Cook (1996) found that 72% of the productions on two consecutive trials came within 8% of the actual tempo. Two studies of tapping to involuntary musical imagery, or "earworms," found mean differences of 11% (Jakubowski et al., 2015) and 8% (Jakubowski et al., 2018) as compared to the song's canonical sound recording. When tunes are generally well known but not specifically selected by each participant, Jakubowski et al. (2016) found somewhat higher error levels of 18% when tapping to the imagined song, but 8% when listening and adjusting the playback rate of the actual song recording. The lowest value of 3% was found for songs sung by mothers to infants, measured between different renditions by the same mother (Bergeson and Trehub, 2002).

One common aspect of the latter literature is that participants are evaluated on tempo memory for auditory material that is highly familiar, either as popular music canon (e.g., Levitin and Cook, 1996; Jakubowski et al., 2016) or personal familiarity (e.g., Bergeson and Trehub, 2002; Jakubowski et al., 2015). As mentioned previously, another experimental approach familiarizes participants with novel musical material (Halpern and Müllensiefen, 2008; Schellenberg et al., 2014; Schellenberg and Habashi, 2015) or non-musical periodic environmental sounds (Boltz, 2010). These studies tested whether tempo manipulations reduce subsequent familiarity ratings, with the inference that lower ratings for tempo-modified items provide evidence of absolute tempo memory for the original stimulus. The tempo shifts in these tasks were 15–60% and therefore help establish an upper bound on absolute tempo resolution for unfamiliar auditory material. This may be particularly relevant for the untrained participants in the present study, because while the examples of several specific popular song tempos may have been familiar to them, both the absolute BPM tempo scale and the specific metronome stimuli were novel to them prior to the present experiment.

A study of absolute tempo labeling and production by Gratton et al. (2016) has several important points of comparison with the present results. In that study, seven metronome stimuli at systematically spaced tempos between 71 and 143 BPM were familiarized in two groups of musicians and nonmusicians, and then tested with tasks of absolute identification and finger tapping production. The steps between the stimulus levels were about 12% by tempo, and in the perceptual identification task, nonmusicians selected the correct tempo item 45% of the time, whereas musicians had an identification accuracy rate of 53% that was not significantly different from nonmusicians. Those results validate the idea that untrained individuals can learn an unfamiliar absolute tempo labeling scale, as they were similarly asked to do in the present study. The lack of a clear musical training effect in the study by Gratton and colleagues likely reflects the wide granularity of their tempo intervals, which were 9–16 BPM apart, whereas the spacing of 1 BPM in the present study provides a finer resolution to estimate absolute tempo perception accuracy. An additional important aspect of the Gratton study is that their task involved explicit labeling of the test stimuli on an ordinal tempo scale. The participants were able to learn this novel tempo scale and then label items on the scale, which requires both the perceptual component of extracting the tempo on an absolute scale and the associative component of selecting a discrete name for the value (Kim and Knösche, 2017), as in the present task.

Another notable finding in the absolute tempo perception study by Gratton and colleagues is a U-shaped distribution of accuracy. In studies of absolute perception, this is often referred to as the "bow" or edge effect, where identification is more accurate toward the minimum and maximum of the tested range (Brown et al., 2008). The present results appear to be consistent with this effect in the untrained group, where the intermediate 100–119 BPM range has greater error than the fastest 140–160 BPM range (**Figure 2B**), whereas the pattern is not seen for the trained groups. Indeed, participants having DJ or musical training may have prior, stable BPM referents in the intermediate tempo ranges that would counteract this bow effect.

Finally, the error level achieved by DJs in the 120–139 BPM range approaches the just noticeable difference (JND) of 2–3% previously reported for tempo discrimination tasks using isochronous auditory sequences (Drake and Botte, 1993). This performance is notable both for its high accuracy and its tempo specificity. The 120–139 BPM tempo range coincides with the tempos reported by the same participants as most frequently played during their DJ experience (**Figure 1**) and is also in concordance BPM values previously reported as typical for dance music played by DJs (Moelants, 2008). Several models of absolute perception emphasize the importance of stable memory for specific referents on the absolute scale. In the ANCHOR model by Petrov and Anderson (2005), these referent anchors serve the mapping from the perceptual continuum to the response scale, are competitively selected among when a response is needed, and provide a stable basis to estimate the actual response value. Subsequent research by others has reinforced the importance of longer-term absolute referents over short-term relative comparisons (SAMBA model; Brown et al., 2008, 2009) and determined that bow effect (decreased performance in the middle of the scale range compared to the extremes) is diminished when intermediate values are frequently encountered (Kent and Lamberts, 2016). In DJs, high familiarity with the tempos of music in their repertoire, especially those songs that are most frequently played, likely provides stable anchors referents to support highly accurate tempo estimation in this task. Additionally, as DJs are often aware of the numeric BPM value of their most familiar music, their experience could serve to provide both stable anchors that refine an internal continuum of perceptual tempo, and also the mapping to BPM values when an explicit value is to be estimated.

Limitations

This is a pilot study with 40 participants, and consequently, there is likely a diminished sensitivity to detect group differences and a greater influence of individual participants on the analytic results. A mixed-effects model was used in order to minimize bias in model estimates and obtain statistical power with a relatively small sample with unbalanced groups (McNeish and Harring, 2017). For a more detailed profile of absolute tempo accuracy in DJs and percussionists, a larger study will be necessary.

As this is a cross-sectional study, it cannot be excluded that a preexisting disposition to precise tempo estimation influenced people to become DJs or percussionists, rather than (or in addition to) the training of these individuals honing their tempo perception abilities. There is suggestive evidence that even a single week of DJ training can improve rhythm perception (Butler and Trainor, 2015). However, a longer longitudinal study is needed to confirm if tempo or rhythm perception accuracy is driven primarily by practice effects rather than predisposition in DJs.

While this study demonstrates the phenomenon of increased absolute tempo ability in DJs, it is not able to dissociate several presumed components of this ability. The accuracy of tempo memory that is measured presumably depends in turn upon the perceptual resolution for tempo, the extent and quality of one's internal representation of tempo scale, and associations between anchors on this scale and explicit tempo value labels. The degree to which greater accuracy observed in DJs and percussionists may result from enhancement of one of these abilities (e.g., knowledge of labels for scale anchors) or a coordination of multiple abilities is unknown. To tease apart these abilities and better define the key mechanisms involved, future research on this phenomenon would benefit from comparing these groups with untrained individuals on additional tasks that do not involve explicit tempo labeling. At a low level, a tempo JND task (e.g., Drake and Botte, 1993) could test whether greater sensitivity to fine tempo differences accounts for some variation in tempo memory fidelity. A more intermediate task, like reproduction of the tempo of familiar songs (as in Jakubowski et al., 2016), could better define the stability of anchors on the absolute tempo scale, without the requirement that these anchors be associated with explicit tempo values.

Directions for Future Research

With very little existing empirical research on DJs, many interesting questions remain about the skill profile and underlying experience-dependent plasticity in this expert population. In addition to dissociating the components of absolute tempo estimation and confirming their refinement by DJ training, as described in the previous section, future research may examine the potential role of different technical approaches in DJing and seek a more detailed profile of tempo and rhythm abilities in DJs.

Through much of the history of DJing, proficiency has involved the development of sufficient perceptual and motor skills to synchronize different musical recordings, "by hand" and "by ear" (Brewster and Broughton, 2000). As we are interested in the perceptual expertise of DJs, the present study required DJ participants to have experience with playing vinyl records and mixing them by ear. In recent years, technological aids are more readily available to DJs, and many of these are aimed at easing the task of synchronizing different pieces of music. At its most basic, this support includes tempo counters in mixer hardware and extends to automated synchronization between songs provided by DJ hardware or software (Broughton and Brewster, 2003, p. 97). The extent to which tempo judgment accuracy in our DJ group may generalize to other DJs whose training embraces this synchronization assistance is unknown. It may be that tempo labeling ability remains comparably accurate due to the experience of consistent software-provided tempo displays while practicing and performing, whereas detection of beat asynchrony may be less accurate for DJs who rely upon automated synchronization of the music they play.

Given the importance of rhythm to music in general, and to the musical practice of DJs, there is a constellation of related rhythmic perception and production skills that may also be improved in DJs. The previous study by Butler and Trainor (2015) has already demonstrated that temporal changes in rhythmic sequences are better detected in highly trained DJs (at a comparable sensitivity to percussionists) and are improved in untrained individuals who undergo 1 week of DJ practice. Detecting fine temporal structure is important for DJs when synchronizing music, as the tempo must be matched and the phase kept in alignment. One beat-based perceptual task that would be well suited to evaluating DJs' expertise is found in the Beat Alignment Test battery of Iversen and Patel (2008), which assesses sensitivity to detect a mismatch of tempo or phase in a metronome overlaid on music. Butler and Trainor (2015) also highlight the relevance of DJs' sensorimotor expertise, finding that finger tapping to the beat of their auditory stimuli aided the perceptual detection of rhythmic deviations. Indeed, neural systems involved in motor control are believed to contribute to the precision of rhythm perception when there is a regular beat-based structure (Grube et al., 2010; Teki et al., 2011). Fine temporal motor control is also essential for DJs when they manually manipulate music playback to bring two songs into temporal alignment and apply corrections to maintain synchrony. Tasks measuring temporal sensorimotor production, such as those found in the BAASTA (Bégel et al., 2018) and H-BAT (Fujii and Schlaug, 2013) rhythmic assessment batteries, offer a means for future research to test whether these abilities are enhanced in DJs.

CONCLUSION

In sum, the present work highlights that DJs are a highly trained population having expertise in musical tempo perception that is comparable to percussionists, but with a unique performance profile shaped by their most frequently practiced musical repertoire. Parallels between DJs and musicians, both in their extensive experience and enhanced perception, point to many potential avenues for future research in this new model of experience-dependent plasticity, with the opportunity to more broadly understand the effects and neural bases of sustained practice in the brain.

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 Comité d'éthique de la recherche en arts et en sciences at the University of Montreal. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

Abdi, H. (2007). "Bonferroni and Sidák corrections for multiple comparisons," in *Encyclopedia of Measurement and Statistics*, ed. N. J. Salkind (Thousand Oaks, CA: Sage).

Bailes, F., and Barwick, L. (2011). Absolute tempo in multiple performances of aboriginal songs: analyzing recordings of Djanba 12 and Djanba 14. *Music. Percept.* 28, 473–490. doi: 10.1525/mp.2011.28.5.473

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.*, 67, 1–48. arXiv. [Preprint]

Bégel, V., Verga, L., Benoit, C.-E., Kotz, S. A., and Dalla Bella, S. (2018). Test-retest reliability of the battery for the assessment of auditory sensorimotor and timing abilities (BAASTA). *Ann. Phy. Rehabil. Med.* 61, 395–400. doi: 10.1016/j.rehab.2018.04.001

Bergeson, T. R., and Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychol. Sci.* 13, 72–75. doi: 10.1111/1467-9280.00413

Boltz, M. G. (2010). Rate and duration memory of naturalistic sounds. *Acta Psychol.* 135, 168–181. doi: 10.1016/j.actpsy.2010.06.004

Brewster, B., and Broughton, F. (2000). *Last Night a DJ Saved my Life: The History of the Disc Jockey*. New york: Grove Press.

Broughton, F., and Brewster, B. (2003). *How to DJ (Properly): The Art and Science of Playing Records*. New York: Grove Press, Random House.

Brown, S. D., Marley, A. A. J., Dodds, P., and Heathcote, A. (2009). Purely relative models cannot provide a general account of absolute identification. *Psychon. Bull. Rev.* 16, 583–593. doi: 10.3758/PBR.16.3.583

Brown, S. D., Marley, A. A. J., Donkin, C., and Heathcote, A. (2008). An integrated model of choices and response times in absolute identification. *Psychol. Rev.* 115, 396–425. doi: 10.1037/0033-295X.115.2.396

Butler, B. E., and Trainor, L. J. (2015). The musician redefined: A behavioral assessment of rhythm perception in professional Club DJs. *Timing Time Percept.* 3, 116–132. doi: 10.1163/22134468-03002041

Cameron, D. J., and Grahn, J. A. (2014). Enhanced timing abilities in percussionists generalize to rhythms without a musical beat. *Front. Human Neurosci.* 8:1003. doi: 10.3389/fnhum.2014.01003

Cannon, J. J., and Patel, A. D. (2021). How beat perception co-opts motor neurophysiology. *Trends Cogn. Sci.* 25, 137–150. doi: 10.1016/j.tics.2020.11.002

Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., and Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *J. Neurosci.* 32, 1056–1060. doi: 10.1523/JNEUROSCI.3411-11.2012

Clynes, M., and Walker, J. (1986). Music as Time's measure. *Music. Percept.* 4, 85–119. doi: 10.2307/40285353

Drake, C., and Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: evidence for a multiple-look model. *Percept. Psychophys.* 54, 277–286. doi: 10.3758/BF03205262

Ehrlé, N., and Samson, S. (2005). Auditory discrimination of anisochrony: influence of the tempo and musical backgrounds of listeners. *Brain Cogn.* 58, 133–147. doi: 10.1016/j.bandc.2004.09.014

Ellis, M. C. (1991). Research note. Thresholds for detecting tempo change. *Psychol. Music* 19, 164–169. doi: 10.1177/0305735691192007

Fox, J., and Weisberg, S. (2011). *An R Companion to Applied Regression (Second)*. California, United States: Sage.

Fujii, S., and Schlaug, G. (2013). The Harvard beat assessment test (H-BAT): A battery for assessing beat perception and production and their dissociation. *Front. Hum. Neurosci.* 7:771. doi: 10.3389/fnhum.2013.00771

Geiser, E., Walker, K. M. M., and Bendor, D. (2014). Global timing: A conceptual framework to investigate the neural basis of rhythm perception in humans and non-human species. *Front. Psychol.* 5:159. doi: 10.3389/fpsyg.2014.00159

Gratton, I., Brandimonte, M. A., and Bruno, N. (2016). Absolute memory for tempo in musicians and non-musicians. *PLoS One* 11:e0163558. doi: 10.1371/journal.pone.0163558

Grube, M., Cooper, F. E., Chinnery, P. F., and Griffiths, T. D. (2010). Dissociation of duration-based and beat-based auditory timing in cerebellar degeneration. *Proc. Natl. Acad. Sci. U. S. A.* 107, 11597–11601. doi: 10.1073/pnas.0910473107

Halekoh, U., and Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models – The R package pbkrtest. *J. Stat. Software* 59, 1–30. doi: 10.18637/jss.v059.i09

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

Iversen, J. R., and Patel, A. D. (2008). The Beat Alignment Test (BAT): Surveying beat processing abilities in the general population. *Proceedings of the 10th International Conference on Music Perception and Cognition*, 465–468.

Jakubowski, K., Bashir, Z., Farrugia, N., and Stewart, L. (2018). Involuntary and voluntary recall of musical memories: A comparison of temporal accuracy and emotional responses. *Mem. Cogn.* 46, 741–756. doi: 10.3758/s13421-018-0792-x

Jakubowski, K., Farrugia, N., Halpern, A. R., Sankarpandi, S. K., and Stewart, L. (2015). The speed of our mental soundtracks: tracking the tempo of involuntary musical imagery in everyday life. *Mem. Cogn.* 43, 1229–1242. doi: 10.3758/s13421-015-0531-5

Jakubowski, K., Farrugia, N., and Stewart, L. (2016). Probing imagined tempo for music: effects of motor engagement and musical experience. *Psychol. Music* 44, 1274–1288. doi: 10.1177/0305735615625791

Kent, C., and Lamberts, K. (2016). Stimulus probability effects in absolute identification. *J. Exp. Psychol. Learn. Mem. Cogn.* 42, 740–748. doi: 10.1037/xlm0000194

Kim, S.-G., and Knösche, T. R. (2017). On the perceptual subprocess of absolute pitch. *Front. Neurosci.* 11:557. doi: 10.3389/fnins.2017.00557

Krause, V., Pollok, B., and Schnitzler, A. (2010). Perception in action: The impact of sensory information on sensorimotor synchronization in musicians and non-musicians. *Acta Psychol.* 133, 28–37. doi: 10.1016/j.actpsy.2009.08.003

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Statistical Software* 82, 1–26.

Lenth, R. (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. https://CRAN.R-project.org/package=emmeans

Levitin, D. J., and Cook, P. R. (1996). Memory for musical tempo: additional evidence that auditory memory is absolute. *Percept. Psychophys.* 58, 927–935. doi: 10.3758/BF03205494

Lockhead, G. R. (2004). Absolute judgments are relative: A reinterpretation of Some psychophysical ideas. *Rev. Gen. Psychol.* 8, 265–272. doi: 10.1037/1089-2680.8.4.265

Manning, F. C., and Schutz, M. (2016). Trained to keep a beat: movement-related enhancements to timing perception in percussionists and non-percussionists. *Psychol. Res.* 80, 532–542. doi: 10.1007/s00426-015-0678-5

McNeish, D. M., and Harring, J. R. (2017). Clustered data with small sample sizes: comparing the performance of model-based and design-based approaches. *Commun. Stat. Simulation Computation* 46, 855–869. doi: 10.1080/03610918.2014.983648

Moelants, D. (2008). Hype vs. Natural tempo: A long-term study of dance music tempi. *In The 10th International Conference on Music Perception and Cognition.*

Palmer, C. (1997). Music performance. *Ann. Rev. Psychol.* 48, 115–138. doi: 10.1146/annurev.psych.48.1.115

Petrov, A. A., and Anderson, J. R. (2005). The dynamics of scaling: A memory-based Anchor model of category rating and absolute identification. *Psychol. Rev.* 112, 383–416. doi: 10.1037/0033-295X.112.2.383

Quinn, S., and Watt, R. (2006). The perception of tempo in music. *Perception* 35, 267–280. doi: 10.1068/p5353

R Core Team. (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available at: https://www.R-project.org/

Rammsayer, T., and Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music. Percept.* 24, 37–48. doi: 10.1525/mp.2006.24.1.37

Roeske, T., Larrouy-Maestri, P., Sakamoto, Y., and Poeppel, D. (2020). Listening to birdsong reveals basic features of rate perception and aesthetic judgements. *Proc. R. Soc. B Biol. Sci.* 287:20193010. doi: 10.1098/rspb.2019.3010

Schellenberg, E. G., and Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Mem. Cogn.* 43, 1021–1031. doi: 10.3758/s13421-015-0519-1

Schellenberg, E. G., Stalinski, S. M., and Marks, B. M. (2014). Memory for surface features of unfamiliar melodies: independent effects of changes in pitch and tempo. *Psychol. Res.* 78, 84–95. doi: 10.1007/s00426-013-0483-y

Sheldon, D. A. (1994). Effects of tempo, musical experience, and listening modes on tempo modulation perception. *J. Res. Music. Educ.* 42, 190–202. doi: 10.2307/3345699

Teki, S., Grube, M., Kumar, S., and Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *J. Neurosci.* 31, 3805–3812. doi: 10.1523/JNEUROSCI.5561-10.2011

Trainor, L. J., Wu, L., and Tsang, C. D. (2004). Long-term memory for music: infants remember tempo and timbre. *Dev. Sci.* 7, 289–296. doi: 10.1111/j.1467-7687.2004.00348.x

Yee, W., Holleran, S., and Jones, M. R. (1994). Sensitivity to event timing in regular and irregular sequences: influences of musical skill. *Percept. Psychophys.* 56, 461–471. doi: 10.3758/BF03206737

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

BOOK REVIEW
published: 09 August 2021
doi: 10.3389/fpsyg.2021.727768

Book Review: Enacting Musical Time: The Bodily Experience of New Music

*José L. Besada**

Department of Musicology, Universidad Complutense de Madrid, Madrid, Spain

Keywords: time, enaction, embodied cognition, affordances, listening, contemporary music

A Book Review on

Enacting Musical Time: The Bodily Experience of New Music

Mariusz Kozak (New York: Oxford University Press), 2020, 310 pages, ISBN: 970-0-19-008020-4

The Oxford Studies in Music Theory are recently strengthening the focus on issues around rhythm and musical time (Yust, 2018; Ohriner, 2019). Kozak's volume in this collection pursues this path by scrutinizing contemporary music through the lenses of phenomenology and cognitive science. His principal goal is to reconsider the widespread approach to time by music theorists as external to the listener's situated experience. As stated in the introduction, the author regards musical time as "constituted by the moving bodies of participants engaged in musical activities," which leads to his main thesis: "musical time emerges when the listener enacts his or her implicit *kinesthetic knowledge* about 'how music goes"' (p. 4–5).

Chapter 1 criticizes the impact of Newtonian and Cartesian conceptions in music theory for an objective time which spread from the eighteenth century. Kozak embraces in turn the idea of a lived time "as part of the unfolding dynamical system that emerges between an embodied consciousness and the world" (p. 34), thus endorsing the phenomenological tradition by Edmund Husserl and Maurice Merleau-Ponty. Informed by James J. Gibson's and Shaun Gallagher's approaches, Kozak considers the significance of music "in and through the dynamical system that forms when acoustical phenomena elicit responses from enculturated listeners that make these phenomena musical"; consequently, time "is not a condition of music, but something that emerges from it" (p. 53). Affordances are the topic of Chapter 2. Instead of limiting himself to preexistent musicological uses of this term in performance studies, he widens the ecological context by borrowing the notion of social affordances and proposing the temporal ones which jointly specify aesthetic behavior and frame musical affordances. After reviewing Anthony Chemero's approach to radical embodied cognitive science and dynamical systems theory, Kozak specifically defines temporal affordances as "information specified in perceived events in the dynamical relationship between two or more physical affordances," and highlights their relevance as "temporal alignment [...] critical to the successful realization of an intended action" (p. 90). Chapter 3 is mainly devoted to bodily matters. The fundamental contribution in this section is the definition of kinesthetic knowledge as "a contextual enactment of the dynamics, affectivity, and intercorporeality of our bodily involvement with the world," wherein the body "enact[s] its agency in response to both physical and cultural constraints" (p. 129). I highlight in this chapter his ecological distinction between synchronization and coordination, which is chiefly illustrated with a musical example by Brian Ferneyhough. This distinction leads to privilege the contextual joint action instead of any underlying metrical beat or rhythmic patterns. Merleau-Ponty emerges again in Chapter 4 through the notion of flesh–*la chair* –for depicting the body secreting time via enaction. Particularly, the proposal offered by Kozak moves beyond Merleau-Ponty's predilection of haptic and visual examples toward a phenomenological framework that highlights the experience of listening. Some reasonings in this section are aimed at revealing that, beyond highly rationalized conceptions of musical analysis, "the

OPEN ACCESS

Edited by:
Mark Reybrouck, KU Leuven, Belgium

Reviewed by:
Luc Nijs, Ghent University, Belgium
Dylan van der Schyff, University of Melbourne, Australia

Correspondence:
José L. Besada
besadajl@gmail.com

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

Received: *19 June 2021*
Accepted: *16 July 2021*
Published: *09 August 2021*

Citation:
Besada JL (2021) Book Review: Enacting Musical Time: The Bodily Experience of New Music. Front. Psychol. 12:727768. doi: 10.3389/fpsyg.2021.727768

body is already doing analytical work on its own terms" (p. 183), which speaks to the central role it plays during the enactive listening experience. Chapters 5 and 6 are finally governed by two main case studies–from Louis Andriessen's and Toshio Hosokawa's oeuvre–for further developments of the previous framework. This choice makes these two last chapters quite more meaningful for scholars in the field of music theory than those dealing first and foremost with the psychology of music.

Kozak's argumentative style is clearly rooted into the theoretical production of Zbikowski (2002, 2017), who was his Ph.D. advisor. This is particularly noticeable in the choice of apparently simple musical examples which raise important questions that are addressed through a rigorous methodological framework and with very subtle terminological precision. In addition, in Chapter 3 incorporates some empirical evidence–from research carried out by himself–for supporting some of his reasonings. This direct participation in empirical research is relatively unusual from the side of music theorists. Finally, I consider that a short conclusion by the end would have enhanced the global scope of the whole essay.

Kozak's focus on embodiment and enaction targets the listener's experience of time, which is a very appealing approach for both the music theory and the psychology of music communities. His insistence on embodiment and enaction pushes forward new directions beyond some canonic perspectives (London, 2004; Toussaint, 2013) which have conceptualized time more abstractly. By taking this path, some high-level visual representations of time, like ubiquitous timelines, have been underestimated or overtly neglected, in my opinion, through several pages of his book. However, these kinds of representations are often significant from the composers' side, in a quest of anchoring their particular struggle with temporal conceptions. This topic currently elicits scholarly discussion from cognitive perspectives which acknowledge the importance of embodiment and enaction–from a less radical viewpoint, though–as substantial features of compositional practices (Besada and Pagán Cánovas, 2020; Besada et al., 2021). Rather than a critique of Kozak's position, my last remark is a challenge for future collaboration around overlapping research questions.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This review has been funded by an Attraction of Research Talent Grant (ref. 2019-T2/HUM-14477) from the Comunidad de Madrid.

REFERENCES

Besada, J. L., Barthel-Calvet A.-S., and Pagán Cánovas, C. (2021). Gearing time toward musical creativity: conceptual integration and material anchoring in Xenakis' *Psappha*. *Front. Psychol.* 11:611316. doi: 10.3389/fpsyg.2020.611316

Besada, J. L., and Pagán Cánovas, C. (2020). Timelines in spectral composition: a cognitive approach to musical creativity. *Organ. Sound* 25, 142–155. doi: 10.1017/S1355771820000059

London, J. (2004). *Hearing in Time: Psychological Aspects of Musical Meter*. New York, NY: Oxford University Press.

Ohriner, M. (2019). *Flow: The Rhythmic Voice in Rap Music.* New York, NY: Oxford University Press.

Toussaint, G. (2013). *The Geometry of Musical Rhythm: What Makes a "Good" Rhythm Good?* Boca Raton, FL: CRC Press.

Yust, J. (2018). *Organized Time: Rhythm, Tonality, and Form.* New York, NY: Oxford University Press.

Zbikowski, L. M. (2002). *Conceptualizing Music: Cognitive Structure, Theory, and Analysis*. New York, NY: Oxford University Press.

Zbikowski, L. M. (2017). *Foundations of Musical Grammar*. New York, NY: Oxford 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 Neuroscience

ORIGINAL RESEARCH
published: 01 October 2021
doi: 10.3389/fnins.2021.703620

Check for updates

Changes in Brain Responses to Music and Non-music Sounds Following Creativity Training Within the "Different Hearing" Program

*Anna Arkhipova[1], Pavel Hok[1], Jan Valošek[1,2], Markéta Trnečková[1,3], Gabriela Všetičková[4], Gabriela Coufalová[4], Jaromír Synek[4], Vít Zouhar[4] and Petr Hluštík[1]**

[1] Department of Neurology, Faculty of Medicine and Dentistry and University Hospital Olomouc, Olomouc, Czechia, [2] Department of Biomedical Engineering, University Hospital Olomouc, Olomouc, Czechia, [3] Department of Computer Science, Faculty of Science, Palacký University Olomouc, Olomouc, Czechia, [4] Department of Music Education, Faculty of Education, Palacký University Olomouc, Olomouc, Czechia

The "Different Hearing" program (DHP) is an educational activity aimed at stimulating musical creativity of children and adults by group composing in the classroom, alternative to the mainstream model of music education in Czechia. Composing in the classroom in the DHP context does not use traditional musical instruments or notation, instead, the participants use their bodies, sounds originating from common objects as well as environmental sounds as the "elements" for music composition by the participants' team, with the teacher initiating and then participating and coordinating the creative process, which ends with writing down a graphical score and then performing the composition in front of an audience. The DHP methodology works with a wide definition of musical composition. We hypothesized that the DHP short-term (2 days) intense workshop would induce changes in subjective appreciation of different classes of music and sound (including typical samples of music composed in the DHP course), as well as plastic changes of the brain systems engaged in creative thinking and music perception, in their response to diverse auditory stimuli. In our study, 22 healthy university students participated in the workshop over 2 days and underwent fMRI examinations before and after the workshop, meanwhile 24 students were also scanned twice as a control group. During fMRI, each subject was listening to musical and non-musical sound samples, indicating their esthetic impression with a button press after each sample. As a result, participants' favorable feelings toward non-musical sound samples were significantly increased only in the active group. fMRI data analyzed using ANOVA with *post hoc* ROI analysis showed significant group-by-time interaction (opposing trends in the two groups) in the bilateral posterior cingulate cortex/precuneus, which are functional hubs of the default mode network (DMN) and in parts of the executive, motor, and auditory networks. The findings suggest that DHP training modified the behavioral and brain response to diverse sound samples, differentially changing the engagement of functional networks known to be related to creative thinking, namely, increasing DMN activation and decreasing activation of the executive network.

Keywords: creativity, brain plasticity, music training, task-related fMRI, auditory perception, music composition, music education

OPEN ACCESS

Edited by:
Thomas James Lundy,
Cuttlefish Arts, United States

Reviewed by:
Tatsuya Daikoku,
The University of Tokyo, Japan
Meng Zhang,
Xinxiang Medical University, China
Xin Wu,
Xinxiang Medical University, China
Radwa Khalil,
Jacobs University Bremen, Germany

Correspondence:
Petr Hluštík
phlustik@upol.cz

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

Received: *30 April 2021*
Accepted: *06 September 2021*
Published: *01 October 2021*

Citation:
Arkhipova A, Hok P, Valošek J, Trnečková M, Všetičková G, Coufalová G, Synek J, Zouhar V and Hluštík P (2021) Changes in Brain Responses to Music and Non-music Sounds Following Creativity Training Within the "Different Hearing" Program. Front. Neurosci. 15:703620. doi: 10.3389/fnins.2021.703620

INTRODUCTION

Creativity is one of the essential human-specific constructs and it has been consensually defined as an ability to produce novel and useful/appropriate/valuable ideas/works, both in the general public and among researchers (Runco and Jaeger, 2012). Creativity applies not only to specific domains such as music, visual arts, sciences, and industry, but also to many details of daily work and life, thus, high creativity has a big impact on the society and quality of many scenes in human life. Domain-general creativity can be enhanced by training of a certain modality, such as musical creativity, embedded in the process of composing. Many research efforts focus on specialized music teaching and further development of musical abilities (Hickey, 2002; Lapidaki, 2007; Running, 2008; Webster, 2016). In our research, we focus on the development of musical creativity through group compositional techniques in children and students. Against the background of the many proposed definitions of musical creativity (see, e.g., Cook, 2011; Burnard, 2012; Hargreaves, 2012), our work defines creativity operationally as the ability to include non-musical sounds and silence to make music, to engage in group composing, yielding a concrete result (composition), subsequently presented in public and providing satisfaction to participants, teachers/instructors as well as to audience (which was not part of the composing process).

The "Different Hearing" program (DHP) is a musical cognitive training aimed at stimulating creativity by means of group music composing in the classroom, using objects from everyday life rather than traditional musical instruments, employing non-musical sounds (Zouhar and Medek, 2010; Coufalová and Synek, 2014). The program was established in 2001 at the Department of Music Education of Palacký University Olomouc, Czechia, as an alternative method to the mainstream model of music education. The DHP methodology is based on Cage's (1973) wide meaning of musical composition: "The material of music is sound and silence. Integrating these is composing." (p. 62). Participants in the DHP workshop gain both knowledge that all kinds of sounds can be used and put together to create music and the practical skills how to do so. Group composing in the DHP is similar to the approach of composers and educators Paynter and Aston (1970), Schafer (1986), Schneider et al. (2000), or Laycock (2005). DHP has primarily focused on children and young people age five to eighteen, although young adults have been repeatedly studied as well (Zouhar and Medek, 2010; Coufalová and Synek, 2014). Detailed evaluations of a series of DHP workshops since 2002 are performed through structured questionnaires and clearly indicate a potential to enhance self-perceived musical creativity and change subjective appreciation of music and sounds (Synek, 2008; Medek et al., 2014). Changes of non-musical sound perception could be illustrated by a participant's statement: "Začala jsem více vnímat zvuky (v hlučném městě zpěv ptáků)./I started paying more attention to sounds (birds singing in the noisy town)" (Synek, 2008, p. 119).

Although we have not found published evidence for neuroanatomical correlates of behavioral changes after musical creativity training, functional neuroimaging studies of domain-general creativity suggest association/relationship between the creative performance and brain function.

For example, functional neuroimaging studies in creativity suggest that there is a correlation between domain-general creativity and functional connectivity within well-defined resting state networks, namely, in the default mode network (DMN) and executive-control network (EN) as well as the salience network (Beaty et al., 2017, 2018). Observational studies of creative performance, however, identify only the brain structures associated with static creative traits. Further insight may be gained by evaluating dynamic processes, such as increasing one's capacity for creative performance. For example, in a fMRI study by Fink et al. (2015), activation changes after 3-week domain-general creativity training (20 min/day) were observed in the left inferior parietal lobule (IPL) and the left middle temporal gyrus (MTG). Also, Sun et al. (2016) observed change in brain function in the dorsal anterior cingulate cortex (dACC) and dorsolateral prefrontal cortex (DLPFC) as well as increased gray matter volume in dACC after 20 sessions (28 min/day) of creativity training. Successful training in musical creativity would be similarly expected to induce functional plasticity of the participating brain systems. Training-induced changes may subsequently be observed not only during creative activity, but also during perception of different sound classes. This approach (using a different task to probe the brain networks changed by training or stimulation) has been used in other plasticity-inducing protocols, including but not limited to music training (Schlegel et al., 2015; Tierney et al., 2015; Herholz et al., 2016; Sachs et al., 2017).

Since DHP has a potential to develop and/or enhance musical creativity and change subjective appreciation of music and sounds (Synek, 2008; Medek et al., 2014), we hypothesized that a 2-day intense workshop would induce changes in subjective appreciation of different classes of music and sound (including typical DHP samples), as well as plastic changes of the brain systems engaged in creativity and music perception, in response to diverse auditory stimuli.

To address these hypotheses, we designed a randomized behavioral and fMRI study, where an active group would be scanned twice, before and after participating in the DHP workshop, and a control group would also be scanned twice within the same time interval. FMRI has been repeatedly used to describe neuroplastic changes related to behavioral training and learning. The subsequent analysis was designed to evaluate the effect of the intervention either (1) with respect to the perception of different sample classes or (2) regardless of the sample class.

MATERIALS AND METHODS

Subjects

Forty-six healthy volunteers with normal hearing and with no history of neurological disorders participated in this study (40 females and 6 males, mean age 21.6 ± 1.4). The subjects were Teacher Training for Primary Schools students from the

Faculty of Education at Palacký University Olomouc, who had 10.3 ± 5.0 years of music education. Four participants were left-handed, two were ambidextrous, and 41 were right-handed as assessed by the Edinburgh handedness inventory (Oldfield, 1971). The study was carried out in accordance with World Medical Association Declaration of Helsinki. Written informed consent was obtained from all participants prior to their inclusion in the study and the study was approved by the Ethics Committee of the Faculty of Education at Palacký University in Olomouc, approval number 02/2017. Twenty-two participants were randomized to participate in the DHP workshop with the length of 10.5 h over two consecutive days (active group) and 24 students were randomized to the control group to receive no special training and continued normal daily activities and student life. Both groups underwent two fMRI examinations 8 days apart - for the active group, one was before and one after the workshop (**Figure 1**). Both groups were matched in terms of music education, according to computed results from self-reported questionnaires (unpaired t-test on 7 items scoring music education, all p-values > 0.05).

Description of the "Different Hearing" Program Workshop

DHP aims at stimulating and enhancing children's own creativity by means of learning group music composing in the classroom. As already mentioned in the Introduction, participants in the DHP workshop gain both knowledge that all kinds of sounds can be used and put together to create music and the practical skills how to do so (see below). The program was established in 2001 at the Department of Music Education of Palacký University Olomouc, Czechia, as an alternative method to the mainstream model of music education. The DHP methodology is based on Cage's (1973) wide meaning of musical composition: "The material of music is sound and silence. Integrating these is composing." (p. 62). Group composing activities are oriented similar to composers and educators Paynter and Aston (1970), Schafer (1986), Schneider et al. (2000), or Laycock (2005). It has primarily focused on children and young people age five to eighteen, although young adults have been repeatedly studied as well (Zouhar and Medek, 2010; Coufalová and Synek, 2014).

In our study, the length of the DHP workshop was 10.5 h, divided over two consecutive days, and the instructors functioned as partners/co-performers initiating and then participating and coordinating the creative process, rather than acting as teachers. On the first day, participants were trained to discover a new sonic world, namely by creating sounds using their own bodies/voices, and by creating original musical instruments made of materials from everyday life such as plastic bottles, wooden chairs, recorded ambient sounds from the streets, etc., but without using traditional musical instruments in the traditional way. Next, they learned basic principles of improvisation and composition, as well as creation of their own graphic notations using original symbols. On the second day, the participants were divided into three groups and composed music using their original sounds and prepared the graphic scores, then performed the compositions in front of the other groups and instructors, followed by a discussion for evaluations and feedback.

fMRI Task

Each fMRI examination included two functional imaging acquisitions during the task of listening to musical and non-musical sound samples from five different classes, i.e., Classic music, New music, DHP samples (composed and performed by previous participants during the workshops in the past), Nature sounds, and Industrial noises. Classical music and New music samples were extracted from musical recordings (CD tracks). The Classical music samples cover the classes from Middle Ages (Leonin, Machaut) and Baroque music (Monteverdi, Handel, Vivaldi), through Classical Period (Gluck, Stamitz, Beethoven) to Romantic music (Chopin, Verdi, Bruckner). The selection of New music spans a wide range of 20th century esthetics from Schoenberg, Hába, Cowell, Varèse to Boulez, Stockhausen, Xenakis, Ligeti, Lutosławski and Štědroň. The samples were chosen according to various parameters – in the classical music group, compositions from different stylistic periods (Middle Ages – 20th century) were chosen in such a way that vocal, instrumental and vocal-instrumental, chamber and large-scale compositions were represented. Samples from the New music section represent different styles and various sound qualities related to contemporary music with the aim to provide a colorful selection of sounds. DHP sound compositions samples were cut from tracks recorded during different former workshops. Nature sounds and industrial sounds are field recordings – in the city, nature, etc., by members of the DHP team.

For each of these sound classes, 12 unique samples 15 s long were prepared. During each imaging run, fifteen different samples were played through MR-compatible headphones in a counterbalanced order across subjects, such that three different samples from each sound class were presented in each run. Each session consisted of two such imaging runs with different sound samples, thus, no sample was repeated twice for any subject. The volume level of sounds was adjusted in the audio editing software Audacity[1] to be equal across all the samples. Furthermore, participants were asked to keep their eyes open and watch a fixation cross during the listening phase. They were instructed to press one of two buttons (like/not like) after listening to each sample as soon as the question "Did you like the sample?" (in Czech) appeared on screen (displayed for 4 s following the sound sample offset).

MRI Data Acquisition

MRI examinations were performed twice with a 8-day interval for all subjects (the second examination of the active group was scheduled within 2 days after the workshop), using a 3T scanner (Siemens Prisma, Erlangen, Germany) with a standard 64-channel head and neck coil in the Multimodal and Functional Imaging Laboratory (MAFIL), Central European Institute of Technology (CEITEC) in Brno. The subject's head was immobilized with cushions to assure maximum comfort

[1] www.audacityteam.org

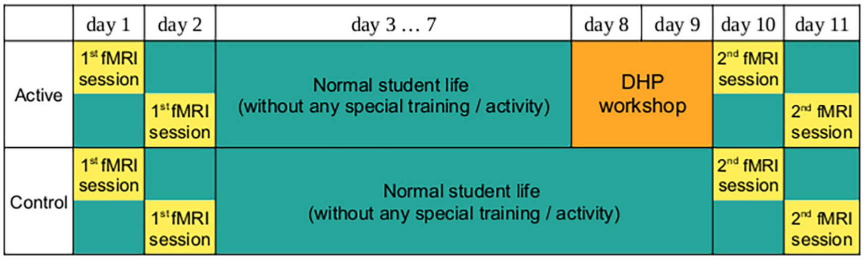

FIGURE 1 | Flow Chart of the Study. All subjects had a same-length interval (8 days) between the first and second MRI measurements. The second examination of the active group was performed within 2 days after the DHP workshop.

and minimize head motion. The MRI protocol included task-related blood oxygenation level-dependent (BOLD) fMRI data acquisition (T2*-weighted echo-planner imaging; 48 slices, 3mm slice thickness; repetition time/echo time = 780/35ms; flip angle 45°; field of view = 192mm; matrix 64 × 64; 465 volumes; repeated twice) during listening to the sound samples through MR-compatible headphones. Gradient-echo phase and magnitude fieldmap images were acquired to allow correction of the echo planar imaging distortions. A high resolution T1-weighted structural image was acquired using Magnetization-Prepared rapid Gradient-Echo (MPRAGE) sequence for anatomical reference. In addition, resting-state fMRI with BOLD EPI imaging data were obtained before the task-related acquisition, but the data is not reported here. Heart rate (pulse oximetry) and respiration (respiratory belt) were monitored during BOLD scanning.

BOLD MRI Data Pre-processing

The fMRI data were processed using FEAT Version 6.00, part of FSL (FMRIB's Software Library), version 5.0.9 (Jenkinson et al., 2012). The fMRI data were initially checked for susceptibility artifacts and none of the subjects were excluded. The pre-processing consisted of: correction of B0 distortions, motion correction, non-brain tissue removal, and spatial smoothing using a Gaussian kernel with 8.0 mm full width at half maximum (FWHM). During pre-processing, an affine registration matrix between the functional images and the respective structural image was obtained using FLIRT (Jenkinson and Smith, 2001; Jenkinson et al., 2002) and a non-linear transformation between the structural space and the MNI 152 standard space was calculated using FNIRT (Grabner et al., 2006). Next, using partially pre-processed data, motion-related artifacts were regressed out from functional time-series by ICA-AROMA automatic noise classifier (Pruim et al., 2015), followed by high-pass temporal filtering with sigma = 60.0 s. Finally, pre-processing included estimation of nuisance signal regressors based on the RETROICOR method (Glover et al., 2000).

Statistical Analysis of BOLD Imaging Data

Voxelwise general linear model (GLM) analysis was carried out using FILM (Woolrich et al., 2001). At the single-subject level, the GLM consisted of 5 regressors to separately model activation evoked by each sound type, 2 regressors to model positive and negative feedback responses, and a single regressor to model activation due to visual presentation of the instruction. Temporal derivatives of each regressor were added to account for non-uniform slice timing and haemodynamic response function (HRF) delay. Eight more nuisance regressors obtained from RETROICOR were added to account for physiological noise. Single-subject contrasts were set to provide mean activation for each sound type.

Next, average effects per subject were computed using a fixed-effects analysis. The resulting beta parameters and residuals were then carried over to the group-level mixed effects analysis. At the group level, an analysis of variance (ANOVA) with 3 factors (time, group, and sound type) accounting for repeated measures was employed to test the main hypotheses. Following F-tests were evaluated: (1) the goodness-of-fit of the ANOVA model (global F-test), (2) mean activation/deactivation, (3) group-by-time-by-sound type interaction, and (4) group-by-time interaction. The mixed-effects analysis was performed using FLAME (FMRIB's Local Analysis of Mixed Effects) stage 1 (Beckmann et al., 2003; Woolrich et al., 2004). The Z (Gaussianised T) statistic images were thresholded using clusters determined by $Z > 3$ and family-wise error (FWE) corrected cluster significance threshold was $p < 0.05$ (Worsley, 2001). To assess the directionality of significant changes, a *post hoc* t-test was performed for the interaction where significant clusters were detected. Only voxels falling within significant clusters in F-test 1, 2, and 3 (for group-by-time-by-sound interaction) and F-tests 1, 2, and 4 (for group-by-time interaction) were considered. The *post hoc* analysis images were thresholded voxel wise at the FWE corrected $p < 0.05$. Significant clusters bigger than 100 voxels were anatomically classified according to an overlap with the Harvard-Oxford Cortical and Subcortical Structural Atlases (Desikan et al., 2006), and the Probabilistic Cerebellar Atlas labels (Diedrichsen et al., 2009). The resulting statistical images were rendered in Mango v4.0 (Research Imaging Institute, UT Health Science Center at San Antonio, TX, United States).

Analysis of In-Scanner Behavioral Data

The effect of time/session on subjective like/not like response to individual stimulus classes was tested within subject using Wilcoxon signed rank tests in the active and control groups.

RESULTS

Behavioral Data

Comparison of (behavioral) responses to sound samples showed that favorable feelings toward DHP, New music and non-musical sound samples (Nature and Industrial) significantly increased only in the active group ($p < 0.05$, Wilcoxon signed rank tests). The change for the DHP sound samples was most robust and survived optional Bonferroni correction for multiple testing. Descriptive statistics for the individual sound classes across groups and sessions is provided in **Table 1**.

Imaging Results

The main hypotheses of our study were tested within the framework of a comprehensive statistical model, including all the main effects (group, time, and sample class), interactions and specific *post hoc* contrasts. Firstly, the global F-test for the goodness-of-fit of the ANOVA model yielded significant clusters in a number of areas, including but not limited to the temporal and frontal cortices and cerebellum. Next, the highest level interaction (group-by-time-by-sound type, F-test 3) yielded an empty map (data not shown), meaning that there were no statistically significant clusters showing this particular effect. In other words, there was no significant difference among the interactions for each individual sample class.

Our next hypothesis addressed the differential effect of time (training) between the active and control groups, which was captured by the group-by-time interaction (F-test 4). As described in the Methods, only voxels significant on the global F-test (F-test 1) and manifesting significant mean activation/deactivation effect (F-test 2) were considered further. **Figure 2** shows the statistical parametric maps for (A) the overall effect (F-test 1), (B) mean activation (F-test 2), (C) group-by-time interaction (F-test 4) and (D) the region of interest (ROI) mask created by conjunction analysis of F-tests 1, 2, and 4.

The group-by-time interaction and subsequent *post hoc* analysis of fMRI data showed that, regardless of auditory sample class, activation in the active group apparently increased after DHP training (loss of deactivation) in the bilateral posterior cingulate cortex (PCC) and precuneus, which are functional hubs of the DMN. Moreover, significantly decreased activation in the active group was observed in the left inferior frontal gyrus (IFG), bilateral frontal orbital cortex (OFC) and right anterior cingulate cortex (ACC), which are parts of the EN, as well as in the regions in the motor network (bilateral supplementary motor cortex [SMA], right pre-supplementary motor area [pre-SMA] and left cerebellum) and auditory network (left MTG, right superior temporal gyrus [STG], planum temporale [PT] and temporal pole [TP]), whereas activation in all of these networks (EN, motor network, auditory network) in the control group was significantly increased (**Figures 3**, **4**). For detailed description of the *post hoc* clusters, see **Table 2**.

No *post hoc* analysis was performed for the three-way interaction as there were no significant clusters detected, as already mentioned.

DISCUSSION

"Different Hearing" Program and Creativity

Before we proceed to the discussion of the study results, we will discuss in more detail the concept and design of the DHP and its relationship to creativity, as defined by our group.

The DHP was developed as an alternative music education approach. Similarly to other contemporary approaches (Hickey, 2002; Lapidaki, 2007; Running, 2008; Webster, 2016), DHP uses group composition in the classroom, encouraging the participants to move beyond traditional use of traditional music instruments to make music; instead, to include any non-musical sounds as well as silence in the course of group composing, organizing their work with the help of graphical scores and finishing with a concrete result (composition). From the cognitive neuroscience perspective, the workshop becomes a special case of musical cognitive training.

The new skills acquired during the DHP workshop represent a particular/specific type of musical creativity. In fact, Burnard (2012) pointed out that in current cultural, social and activity systems it is not possible to define a single definition of musical creativity. On the contrary, she argues for "the broadening of the concept of "musical creativity" to include a plurality of equally valid creativities through which musicians may fluidly move or situate within realms of creating and receiving musical artworks and cultural products." Among which she mentions, e.g., individual creativity, collaborative (group) creativity, communal creativity, emphatic creativity, performance creativity and others (p. 15-16). Cook (2011) used a similar argument by pointing out that the term refers to "an indefinite number of related concepts or behaviors." Cook suggests that musical creativity "revolves round social interaction, and is embedded and embodied in the practices of everyday life" (p. 451). According to Hargreaves (2012), musical creativity is "only one facet of a much broader phenomenon, the central core of which is imagination, which incorporates creative perception as well as production."

At conception of our study, we have not found a musical creativity scale or assessment that would, in our view, adequately capture the musical behavioral changes induced by the DHP. The lack of objective creativity measurement has been clearly admitted in the Limitations section of the Discussion.

The behavioral outcomes of DHP workshops have thus been evaluated with detailed structured questionnaires (18 questions), including one question about perceived change in musical creativity and another about change in broadening the perception of sounds from the environment ("opening the ears") after taking the course.

The behavioral observations from the previous DHP workshops (Synek, 2008; Medek et al., 2014) have inspired the hypotheses of the present study, both the expected behavioral change and the neuroimaging correlates.

Behavioral Results

Behavioral results showed changes in esthetic/emotional perception, predominantly of New music and non-musical

TABLE 1 | Behavioral responses to individual sound classes by group and session.

		Session 1			Session 2			
Group	**Sound class**	**Median**	**25%**	**75%**	**Median**	**25%**	**75%**	***P*-value**
Active	Classical	100.0%	100.0%	100.0%	100.0%	87.5%	100.0%	0.8359
	New	16.7%	0.0%	45.8%	50.0%	33.3%	66.7%	*0.0389*
	DHP	16.7%	16.7%	33.3%	50.0%	20.8%	66.7%	**0.0043**
	Nature	83.3%	66.7%	100.0%	100.0%	100.0%	100.0%	*0.0244*
	Industrial	33.3%	16.7%	55.4%	66.7%	33.3%	83.3%	*0.0397*
Control	Classical	100.0%	87.5%	100.0%	100.0%	100.0%	100.0%	*0.0479*
	New	50.0%	33.3%	66.7%	25.0%	16.7%	50.0%	0.1878
	DHP	50.0%	20.8%	66.7%	25.0%	16.7%	33.3%	0.1195
	Nature	100.0%	100.0%	100.0%	83.3%	72.9%	100.0%	0.0902
	Industrial	66.7%	33.3%	83.3%	33.3%	19.2%	50.0%	0.2651

P-value: Wilcoxon signed-rank test (session 1 – session 2). Bonferroni-corrected threshold is 0.05/10 = 0.005. Statistically significant changes are displayed in ***bold*** *(corrected) and italics (uncorrected). DHP, "Different Hearing" program.*

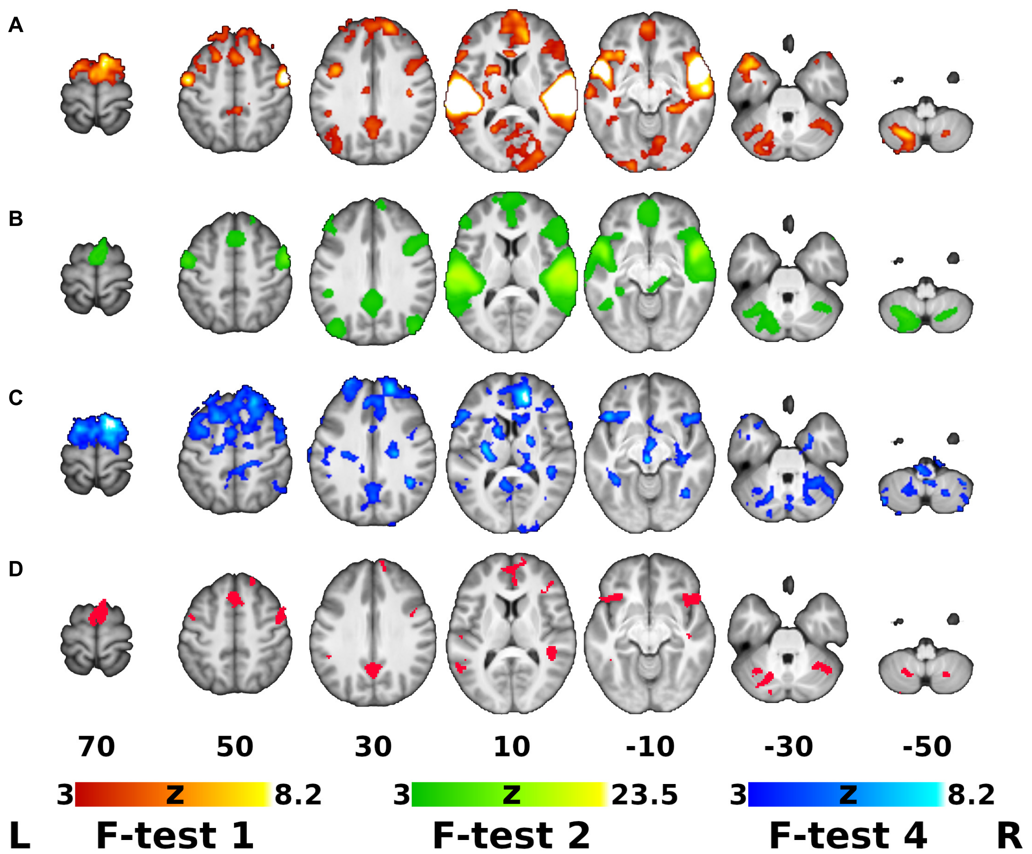

FIGURE 2 | Global F-test of the ANOVA Model, Overall Activation/Deactivation, Group-by-Time Interaction and the ROI Mask for *Post hoc* Analysis. **(A–C)** Thresholded Z-score (normalized F-statistics) maps on top of an average T1-weighted structural image. **(A)** Global F-test of the analysis of variance (ANOVA) model (F-test 1, red-yellow overlay). **(B)** Mean activation/deactivation map across all sound classes (F-test 2, green overlay). **(C)** Areas showing differential effect of training/time between the groups (Group-by-time interaction, F-test 4, blue overlay). **(D)** Binary region of interest (ROI) mask resulting from conjunction of maps **A** ∩ **B** ∩ **C** (red). Maps **(A–C)** were cluster-wise thresholded using corrected cluster significance of $p < 0.05$ and cluster-forming threshold of $Z > 3.0$. Right is right according to neurological convention.

sounds, possibly induced by the DHP training. Increased music liking has been described after repeated exposure (McDermott, 2012), however, this cannot be the sole explanation of our results. First, sound samples were randomly selected for each sound class each time, so the subjects would not hear exactly the same selections in sessions 1 and 2, even though some repetitions may

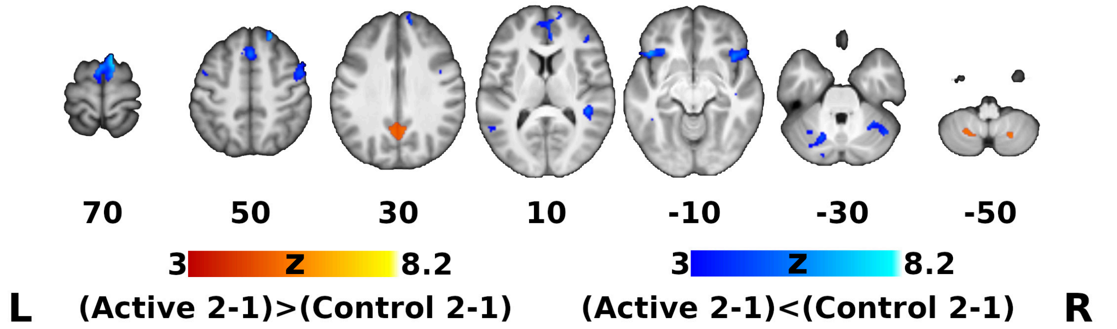

FIGURE 3 | Thresholded Statistical Parametric Maps for *Post hoc* Contrasts within the ROI Mask (**Figure 2D**). Orange: Clusters manifesting significantly greater activation difference over time for the Active group compared to the Control group (posterior cingulate/precuneus, inferior cerebellum). Blue: Clusters manifesting significantly smaller activation difference over time for the Active group compared to the Control group (see **Table 1** for the list of clusters). Other conventions same as **Figure 2**.

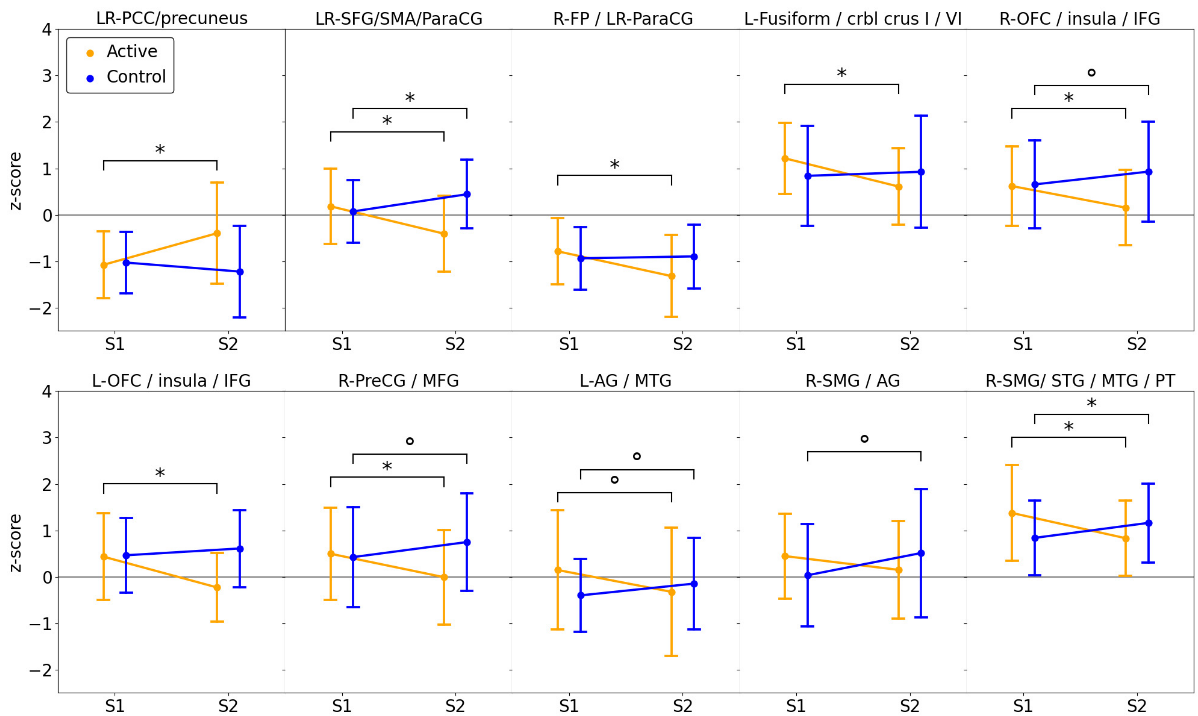

FIGURE 4 | Line Graphs Demonstrating Average Activation for Each Group and Session within Each Cluster Displayed in **Figure 3**. Statistically significant differences between sessions (paired t-test) are shown by brackets and an asterisk (*), trends ($0.05 < P < 0.1$) are indicated by a circle (°).

have occurred. Second, and more importantly, repetition effects would manifest in both groups, whereas most increases in the "like" scores happened in the active (DHP training) group. The only statistically increase in the control group (uncorrected) appeared for Classical music, which, interestingly, was far away from achieving significance in the Active group. However, this effect and the inter-group discrepancy may be due to a ceiling effect, since both groups' likings were close to 100% (see **Table 1**). The DHP sounds may present a special case for the active group. The DHP course by definition exposed the participants to many different sounds from the same DHP sound class, even if not identical to those presented during the fMRI examination. Perhaps they were similar enough to increase the liking by repetition, a marked change, which was statistically most robust and significant, even after Bonferroni correction. The composing process itself is important as well, when the active group had an experience with producing/creating sounds and creating sound compositions related to DHP and New Music samples. The DHP group while composing uses pre-recorded environmental sounds, whether nature or industrial, as well. These, however, represent only a portion of the DHP "elements," which would explain why the change in the active group was smaller and

TABLE 2 | Significant group-by-time interaction – *Post hoc* analysis.

Contrast	Cluster index	Volume [cm^3]	Atlas label	Z_{max}	Z_{max} Coordinates [x, y, z (mm)]
(Active S2-S1) > (Control S2-S1)	1	2.79	31.2% L cingulate gyrus, posterior division 30.7% L precuneous cortex 25.8% R precuneous cortex 12.3% R cingulate gyrus, posterior division	7.06	−4, −54, 18
	1	10.40	35.5% R superior frontal gyrus 17.5% R SMA 16.2% L superior frontal gyrus 14.0% L SMA 11.1% R paracingulate gyrus 5.8% L paracingulate gyrus	9.01	10, 22, 66
	2	3.86	55.3% R frontal pole 21.3% R paracingulate gyrus 14.5% L paracingulate gyrus 6.8% R cingulate gyrus, anterior division	7.82	14, 58, 20
	3	3.20	51.0% L CRBL crus I 38.3% L occipital fusiform Gyrus 34.0% L CRBL VI 25.0% L temporal occipital fusiform cortex	7.01	−36, −80, −24
(Active S2-S1) < (Control S2-S1)	4	2.49	47.9% R frontal orbital cortex 23.2% R insular cortex 12.2% R inferior frontal gyrus, pars triangularis 9.6% R temporal pole	7.15	42, 16, −6
	5	2.23	65.6% L frontal orbital cortex 11.5% L insular cortex 11.5% L inferior frontal gyrus, pars opercularis 5.4% L frontal operculum cortex	8.37	−44, 20, −6
	6	1.51	79.4% R precentral gyrus 19.0% R middle frontal gyrus	5.66	50, 2, 48
	7	1.46	45.6% L angular gyrus 41.8% L middle temporal gyrus, temporooccipital part 7.7% L lateral occipital cortex, superior division	5.98	−48, −52, 2
	8	1.27	49.7% R supramarginal gyrus, posterior division 48.4% R angular gyrus	4.70	54, −48, 18
	9	1.08	40.0% R supramarginal gyrus, posterior division 23.7% R superior temporal gyrus, posterior division 17.8% R middle temporal gyrus, temporooccipital part 13.3% R planum temporale	6.39	44, −36, 8

CRBL, cerebellum; L, left; R, right; S1, session 1; S2, session 2; SMA, supplementary motor cortex (Juxtapositional Lobule Cortex); Z_{max}, maximum Z score.

less robust (only significant without Bonferroni correction). One might also argue that environmental sounds would be very familiar to the adult participants and thus not likely to experience liking change by repetition during the DHP course. Here, we also consider that during the DHP course, they are used with motivation and purpose (not just passive listening), which may support the observed increase in liking. Finally, the New music samples have certain similarity to all three mentioned sound classes (DHP, Nature, Industrial), so again, taking the DHP course (composing, improvising, listening and performing activities) may influence New music perception in a similar way as actual repeated hearing New music samples. The increase in New music liking, like that for Nature and Industrial, is smaller and less robust than for the DHP samples, as would be expected.

As for the necessity of Bonferroni correction, we suggest that it depends on the perspective of the data. If we assumed that listening to each sound class was an independent process, generating a separate dependent liking variable, then we believe Bonferroni correction would not be required, as we were not performing repeated testing on the same data. If, however, we take liking as one common dependent variable and the 5 particular sound classes and 2 sessions as mere instances, then a correction would be necessary.

Imaging Results

Our hypothesis was that DHP training with composing music would change the brain activation as a functional correlate of enhancement of general creativity, i.e., widening of participants' possibilities/flexibility to incorporate new sounds as elements of composition of music. Our imaging findings suggest that DHP training modified the response to diverse sound samples, differentially changing the engagement of functional networks known to be related to domain-general creative thinking, namely, decreasing DMN deactivation and decreasing activation of EN (see, e.g., Beaty et al., 2017).

In studies of neural correlates underlying the musical creativity, researchers have been investigating brain activity during novel music creation, as Bengtsson et al. (2007) mentioned "improvisation arguably satisfies the demands of a prototypical creative behavior." For example, Limb and Braun (2008) suggested in their fMRI study that dissociated pattern of activity in medial and lateral prefrontal cortices is associated with the neural substrate of improvisation and spontaneous musical creativity. Furthermore, in a meta-analysis (Boccia et al., 2015) of the fMRI studies in three domains of creativity, i.e., musical, verbal and visuo-spatial, revealed that musical creativity is associated with activations in bilateral medial frontal gyrus, in the left cingulate gyrus, middle frontal gyrus (MFG), and IPL and in the right postcentral and fusiform gyri, while the general meta-analysis in all the three domains showed the activated clusters in the bilateral occipital, parietal, frontal, and temporal lobes.

However, the comparison of our results with the studies involving active creative process is not straightforward as we investigated effects of the creativity training during passive listening task with subsequent sample evaluation. Hence, in the following sections, we primarily discuss our results according to distinct functional systems that were modulated by the training intervention, followed by interpretation of the results in the context of the neural correlates of creativity.

Auditory and Motor Networks

Primarily, regions in the auditory network and motor networks were expected to be activated during our fMRI measurements, as functional neuroimaging studies document these two networks interacting during auditory perception as well as music production (Chen et al., 2007; Zatorre et al., 2007). There are many recent studies that investigated brain activation during music/sound/speech sample perception (Hanke et al., 2014; Hong and Santosa, 2016; Casey, 2017; for review in music perception, see Angulo-Perkins and Concha, 2014; Janata, 2015). The motor cortical areas are considered to play an important role in perception of temporal patterns during music listening, as a meta-analysis of the studies of perception rhythmic patterns suggested common activations of premotor areas (Janata and Grafton, 2003). Especially, activation in pre-SMA and SMA have been observed during beat perception (together with premotor cortex, basal ganglia and cerebellum) (Grahn and Brett, 2007), temporal perception (together with in ACC) (Pastor et al., 2006), and during distinguishing changes of rhythmical features (together with the premotor area) (Popescu et al., 2004). Activation in several areas in motor network can also be observed during music improvisation, possibly functioning as sequence processing centers: pre-SMA and dorsal premotor cortex (dPMC) (de Manzano and Ullén, 2012; Donnay et al., 2014); pre-SMA, dPMC as well as DLPFC, and the left posterior part of the STG (improvise > reproduce) (Bengtsson et al., 2009); and the dPMC, ventral PMC, together with areas in EN (IFG and ACC) in a different study (improvisation > playing patterns) (Berkowitz and Ansari, 2008).

In our model, activation changes in regions of the auditory network (i.e., decreased activation in right STG, PT, TP, right MTG, and enhanced deactivation in left MTG) and the motor network [i.e., deactivation in bilateral superior frontal gyrus [SFG] (pre-SMA and SMA), and decreased activation in the left cerebellum, right precentral gyrus [preCG] (dPMC)] are observed in the DHP trained group, whereas activation in these networks was increased in the control group. The strengthening with repetition in the control group can be regarded as a priming effect, engaging both the "listening" network and the "music-making" network, a learning effect which may be expected. On the other hand, the weakening of the response in the active group may come as a surprise, especially since the group has been engaged in 2-day group composition and performance, employing both the auditory and motor networks. Perhaps the weaker response in the "modality-specific" networks actually reflects a change of balance toward "modality-independent" networks engaged in music perception and creation, as a result of the DHP training. Previously, activation decrease with repetition was reported, so-called repetition suppression (e.g., Brown et al., 2013), however, the effect was observed in short-term within-session studies, not across sessions. Also, our study design did not involve exact repetition, as our sound samples were randomly selected and ordered. Our longitudinal design including a control group permits separation of general time and repetition effects from training-induced plasticity (Olszewska et al., 2021).

Default Mode and Executive Networks

Next, as a result of the *post hoc* analysis, we observed significant loss of task-related deactivation in PCC and precuneus in the second session in the active (DHP-trained) group. These areas are the main hubs of the DMN (Raichle et al., 2001), which was originally known to be deactivated during focused cognitive task performance and oppositely activated during the rest/mind-wondering. Whereas cognitive tasks requiring focused attention deactivate the DMN, DMN apparently becomes active during internally focused tasks, spontaneous cognition and/or broad awareness of the environment (Raichle et al., 2001; Buckner et al., 2008; Andrews-Hanna et al., 2010). For our data, it is tempting to consider the observed "loss of task-related deactivation" in the DMN as "broadening" of the attention during listening to the sound samples after training, in parallel to the above mentioned behavioral "widening" of the inner concept of (pleasant) music.

In previous studies in creativity, several researchers suggested relationship between creativity and DMN (Jung et al., 2010; Takeuchi et al., 2011, 2012; Wei et al., 2014; Bashwiner et al., 2016; Fink et al., 2018). For example, Takeuchi et al. (2011) observed that reduced task-related deactivation in the precuneus was associated with higher individual domain-general creativity, where the relationship between creativity and brain activity during working memory task was investigated. In these studies, creativity was assessed using divergent thinking (DT) tests (Guilford, 1967), which have been recognized as indicators of creative ability and often been used in neuroscientific studies of creativity due to their relationship to flexibility, fluency and originality, since psychologists have demonstrated these characteristics associated with highly creative people (Guilford, 1950; Drevdahl, 1956). The neurobiological correlates of improvement in those individual DT components (flexibility, fluency, originality) were evaluated in three neuroimaging studies of domain-general creativity training (Wei et al., 2014; Fink et al., 2015; Sun et al., 2016). Wei et al. (2014) especially observed positive correlation of originality with the strength of resting-state functional connectivity between medial prefrontal cortex (mPFC, a hub node of DMN) and MTG, whereas other two studies showed improvement of fluency and originality associated either with increased activation in the left IPL and decreased activation in the left MTG (Fink et al., 2015), or with increased activation in the bilateral DLPFC, dACC, right precuneus and left IPL (Sun et al., 2016). Since the DHP method of group composing using non-traditional instruments and environmental "non-musical" sounds is closer to DT training than domain-specific musical training such as musical improvisation, the functional changes observed in the DMN in this study may, similarly, reflect improvement in the originality and fluency components of domain-general creativity. However, such interpretation of the loss of deactivation in PCC and precuneus would be speculative since the previous changes in DMN were located elsewhere (mPFC and IPL). Furthermore, no direct assessment of DT components was performed in our study since the DHP had been originally developed independent of the concept of DT.

Structural MRI studies have also demonstrated association between creativity and gray matter change in the DMN. Jung et al. (2010) observed a positive correlation between high domain-general creativity and regional cortical volume and thickness of the right PCC, right angular gyrus (AG) and lower left lateral OFC, whereas a study by Bashwiner et al. (2016) showed that high musical creativity correlated with increased cortical surface area or volume not only in the regions associated with motor activity and sound processing, but also in three out of four nodes of the DMN (i.e., dorsomedial PFC, lateral temporal cortex, and TP).

In our study, greater deactivation in the right ACC, left OFC/IFG, AG and decreased activation in the right posterior MFG and supramarginal gyrus (SMG)/AG (IPL) was also observed in the second session in the active group. These areas constitute the EN (Seeley et al., 2007), which has been associated with working memory, problem solving, and decision making. Interestingly, a fMRI study revealed negative association with total hours of improvisation experience and activation in the EN during musical improvisation by professional musicians compared to rest (Pinho et al., 2014), while another study showed greater deactivation of EN during the generation phase of new poetry in experts compared to that in novices (Liu et al., 2015). With respect to the salience network, the observed training-related decreases in the right-sided ACC and bilateral insula (see **Table 2**) may, in fact, reflect decreased activation in the salience network.

Existing studies suggest a relationship between creativity and activity patterns engaging both DMN and EN in a specific manner. For example, Beaty et al. (2015) observed functional networks consisting of regions within the DMN (PCC, precuneus, and IPL) and EN (DLPFC) changing their coupling according to the current stage of the DT task performance in temporal (time-resolved) connectivity analysis. Similarly, during poetry composition, in which generation and revision phases were separated, Liu et al. (2015) observed that a hub region of DMN (mPFC) was activated during both phases, whereas activation in EN (DLPFC and IPS) was increased only during the revision phase. Also, in a study by Marron et al. (2018), the subjects with higher-creative potential (assessed by scores in the DT task) showed greater activation in the DMN as well as reduced activation in the EN during the task, compared to the lower-creative potential groups.

Even though the above studies investigated brain networks engaged in creative tasks, whereas our active group passively listened to sound samples, apparently, there was a common observation of greater DMN activation (or less DMN deactivation) together with decreased EN activation (or greater EN deactivation). We may therefore speculate that the DHP training modulates these brain networks in the same way as domain-general creativity training.

Taken together, our results from the imaging analysis may suggest that the participants of DHP were "differently hearing" sound samples in the second session. Potentially, this could mean that they were not only passively hearing but also actively perceiving sounds as materials/elements of music, as in a pre-stage (thinking phase) of composing music. In such case, DHP training could possibly widen/enhance the flexibility of subjects' music perception toward non-musical sounds, in the similar way of the concept of divergent thinking. Of course, this conclusion would be much stronger if it were supported by specific behavioral data, as well as by sound-class specific training effects in the active group (i.e., group by time by sound class interaction), which was not found in the present data.

Limitations

Neither cognitive tests of creativity nor in-scanner task of creativity were performed. Thus, our study was mainly focused on the "wondering phase" but not "producing process." The results of such testing, if compared with the task performed here focusing on the "wondering phase," could provide further validation of our conclusions and shed more light on the "producing process."

Choice of musical/non-musical sound samples did not systematically cover all music genres (e.g., There were no samples from Pop, Rock or electric music representative as more popular musical sounds among young people, instead, some sound compositions were quite similar between the samples of New music class and DHP).

Furthermore, since the control group only experienced normal student activities without any control/sham training and only the active group underwent an intervention, the possibility of "placebo" effect (e.g., due to subconscious expectations) in the active group could not be completely ruled out.

Finally, although we recognize greater potential for plasticity of the children's brains, we opted for recruitment of university students due to methodological and ethical limitations of neuroimaging in under-age subjects. Still, previous data suggested that behavioral data of DHP training are similar in the young adults as in children.

Future Directions

Because of previous studies suggesting association between creativity and functional connectivity within/between brain networks such as DMN/EN, we plan to analyze functional connectivity changes induced by DHP training.

We also plan to further explore the brain responses of the active group to Natural sounds. Individual responses in the active group showed lower brain activation for the class of Natural sounds (stream, birds, and rain, etc.) compared to all other sound classes. If this variable phenomenon were to be confirmed as a statistically significant group effect, this type of sound could be used in educational, artistic and therapeutic activities.

CONCLUSION

We suggest that DHP seems effective to broaden sound and music preferences and perception, as reflected in both behavior and brain function. Neuroanatomical location and character of the training-induced changes suggest their possible relationship to domain-general creative 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

The studies involving human participants were reviewed and approved by Ethics Committee of the Faculty of Education at Palacký University Olomouc. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PHl and VZ conceived the study. GV, PHl, VZ, and AA performed the literature search. VZ, GV, GC, and JS recruited the participants and trained them for the experiments, PHl, VZ, GV, GC, JS, and PHo designed the fMRI experiment, prepared the stimuli, and supervised data acquisition. PHo and PHl designed the fMRI analysis. JV and MT designed and performed the statistical analysis of behavioral data. AA, PHo, and PHl drafted the manuscript. All authors have participated in interpretation of the results and critical revisions of the manuscript.

FUNDING

This study was supported by OpenAccess grant of Czech-Bioimaging (LM2015062 and LM2018129) to the MAFIL core facility of CEITEC, MUNI, Brno. The Faculty of Education, Palacký University Olomouc, supported student travel grants and scholarships.

ACKNOWLEDGMENTS

The authors specifically thank Ing. Radek Mareček of the MAFIL staff for expert and friendly support.

REFERENCES

Andrews-Hanna, J. R., Reidler, J. S., Huang, C., and Buckner, R. L. (2010). Evidence for the default network's role in spontaneous cognition. *J. Neurophysiol.* 104, 322–335. doi: 10.1152/jn.00830.2009

Angulo-Perkins, A., and Concha, L. (2014). Music perception: information flow within the human auditory cortices. *Adv. Exp. Med. Biol.* 829, 293–303. doi: 10.1007/978-1-4939-1782-2_15

Bashwiner, D. M., Wertz, C. J., Flores, R. A., and Jung, R. E. (2016). Musical creativity "revealed" in brain structure: interplay between motor, default mode, and limbic networks. *Sci. Rep.* 6:20482. doi: 10.1038/srep20482

Beaty, R. E., Benedek, M., Barry Kaufman, S., and Silvia, P. J. (2015). Default and executive network coupling supports creative idea production. *Sci. Rep.* 5:10964. doi: 10.1038/srep10964

Beaty, R. E., Christensen, A. P., Benedek, M., Silvia, P. J., and Schacter, D. L. (2017). Creative constraints: brain activity and network dynamics underlying semantic interference during idea production. *NeuroImage* 148, 189–196. doi: 10.1016/j.neuroimage.2017.01.012

Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q., et al. (2018). Robust prediction of individual creative ability from brain functional connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 115, 1087–1092. doi: 10.1073/pnas.1713532115

Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *NeuroImage* 20, 1052–1063. doi: 10.1016/S1053-8119(03)00435-X

Bengtsson, S. L., Csíkszentmihályi, M., and Ullén, F. (2007). Cortical regions involved in the generation of musical structures during improvisation in pianists. *J. Cogn. Neurosci.* 19, 830–842. doi: 10.1162/jocn.2007.19.5.830

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

Berkowitz, A. L., and Ansari, D. (2008). Generation of novel motor sequences: the neural correlates of musical improvisation. *NeuroImage* 41, 535–543. doi: 10.1016/j.neuroimage.2008.02.028

Boccia, M., Piccardi, L., Palermo, L., Nori, R., and Palmiero, M. (2015). Where do bright ideas occur in our brain? Meta-analytic evidence from neuroimaging studies of domain-specific creativity. *Front. Psychol.* 6:1195. doi: 10.3389/fpsyg.2015.01195

Brown, R. M., Chen, J. L., Hollinger, A., Penhune, V. B., Palmer, C., and Zatorre, R. J. (2013). Repetition suppression in auditory-motor regions to pitch and temporal structure in music. *J. Cogn. Neurosci.* 25, 313–328. doi: 10.1162/jocn_a_00322

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011

Burnard, P. (2012). "Rethinking 'musical creativity' and the notion of multiple creativities in music," in *Musical Creativity: Insights from Music Education Research*, ed. O. Odena (Farnham: Ashgate Publishing, Ltd), 5–27.

Cage, J. (1973). *Silence: Lectures and Writings*. Middletown, CT: Wesleyan University Press.

Casey, M. A. (2017). Music of the 7Ts: predicting and decoding multivoxel fMRI responses with acoustic, schematic, and categorical music features. *Front. Psychol.* 8:1179. doi: 10.3389/fpsyg.2017.01179

Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2007). 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

Cook, N. (2011). "Afterword: beyond creativity?," in *Musical Imaginations: Multidisciplinary Perspectives on Creativity, Performance and Perception*, eds D. Hargreaves, D. Miell, and R. MacDonald (Oxford: Oxford University Press), 451–459.

Coufalová, G., and Synek, J. (2014). "Composing in the classroom "Different Hearing" program: experiences in czech music education," in *Proceedings of the SGEM2014 Conference Proceedings*, (Sofia), 169–176. doi: 10.5593/sgemsocial2014/B13/S3.023

de Manzano, Ö, and Ullén, F. (2012). Activation and connectivity patterns of the presupplementary and dorsal premotor areas during free improvisation of melodies and rhythms. *NeuroImage* 63, 272–280. doi: 10.1016/j.neuroimage.2012.06.024

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021

Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., and Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *NeuroImage* 46, 39–46. doi: 10.1016/j.neuroimage.2009.01.045

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

Drevdahl, J. E. (1956). Factors of importance for creativity. *J. Clin. Psychol.* 12, 21–26. doi: 10.1002/1097-4679(195601)12:1<21::AID-JCLP2270120104<3.0.CO;2-S

Fink, A., Benedek, M., Koschutnig, K., Papousek, I., Weiss, E. M., Bagga, D., et al. (2018). Modulation of resting-state network connectivity by verbal divergent thinking training. *Brain Cogn.* 128, 1–6. doi: 10.1016/j.bandc.2018.10.008

Fink, A., Benedek, M., Koschutnig, K., Pirker, E., Berger, E., Meister, S., et al. (2015). Training of verbal creativity modulates brain activity in regions associated with language- and memory-related demands. *Hum. Brain Mapp.* 36, 4104–4115. doi: 10.1002/hbm.22901

Glover, G. H., Li, T. Q., and Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: retroicor. *Magn. Reson. Med.* 44, 162–167.

Grabner, G., Janke, A. L., Budge, M. M., Smith, D., Pruessner, J., and Collins, D. L. (2006). Symmetric atlasing and model based segmentation: an application to the hippocampus in older adults. *Med. Image Comput. Comput.Assist. Interv.* 9, 58–66.

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

Guilford, J. P. (1950). Creativity. *Am. Psychol.* 5, 444–454. doi: 10.1037/h0063487

Guilford, J. P. (1967). *The Nature of Human Intelligence*. New York, NY: McGraw-Hill.

Hanke, M., Baumgartner, F. J., Ibe, P., Kaule, F. R., Pollmann, S., Speck, O., et al. (2014). A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *Sci. Data* 1:140003. doi: 10.1038/sdata.2014.3

Hargreaves, D. J. (2012). Musical imagination: perception and production, beauty and creativity. *Psychol. Music* 40, 539–557. doi: 10.1177/0305735612444893

Herholz, S. C., Coffey, E. B. J., Pantev, C., and Zatorre, R. J. (2016). Dissociation of neural networks for predisposition and for training-related plasticity in auditory-motor learning. *Cereb. Cortex* 26, 3125–3134. doi: 10.1093/cercor/bhv138

Hickey, M. (2002). "Creativity research in music, visual art, theater, and dance," in *The New Handbook of Research on Music Teaching and Learning*, eds R. Colwell and C. Richardson (Oxford: Oxford University Press), 398–415.

Hong, K.-S., and Santosa, H. (2016). Decoding four different sound-categories in the auditory cortex using functional near-infrared spectroscopy. *Hear. Res.* 333, 157–166. doi: 10.1016/j.heares.2016.01.009

Janata, P. (2015). Neural basis of music perception. *Handb. Clin. Neurol.* 129, 187–205. doi: 10.1016/B978-0-444-62630-1.00011-1

Janata, P., and Grafton, S. T. (2003). Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nat. Neurosci.* 6, 682–687. doi: 10.1038/nn1081

Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17, 825–841. doi: 10.1006/nimg.2002.1132

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

Jung, R. E., Segall, J. M., Bockholt, H. J., Flores, R. A., Smith, S. M., Chavez, R. S., et al. (2010). Neuroanatomy of creativity. *Hum. Brain Mapp.* 31, 398–409. doi: 10.1002/hbm.20874

Lapidaki, E. (2007). Learning from masters of music creativity: shaping compositional experiences in music education. *Philos. Music Educ. Rev.* 15, 93–117.

Laycock, J. (2005). *A Changing Role for the Composer in Society*. Bern: Peter Lang.

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

Liu, S., Erkkinen, M. G., Healey, M. L., Xu, Y., Swett, K. E., Chow, H. M., et al. (2015). Brain activity and connectivity during poetry composition: toward a multidimensional model of the creative process. *Hum. Brain Mapp.* 36, 3351–3372. doi: 10.1002/hbm.22849

Marron, T. R., Lerner, Y., Berant, E., Kinreich, S., Shapira-Lichter, I., Hendler, T., et al. (2018). Chain free association, creativity, and the default mode network. *Neuropsychologia* 118, 40–58. doi: 10.1016/j.neuropsychologia.2018.03.018

McDermott, J. H. (2012). "Chapter 10 - auditory preferences and aesthetics: music, voices, and everyday sounds," in *Neuroscience of Preference and Choice*, eds R. Dolan and T. Sharot (San Diego, CA: Academic Press), 227–256.

Medek, I., Synek, J., and Zouhar, V. (2014). *Composing in the Classroom. Different Hearing: Experiences in Czech Music Education*. Brno: Janáček Academy of Music and Performing Arts in Brno.

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

Olszewska, A. M., Gaca, M., Herman, A. M., Jednoróg, K., and Marchewka, A. (2021). How musical training shapes the adult brain: predispositions and neuroplasticity. *Front. Neurosci.* 15:630829. doi: 10.3389/fnins.2021.630829

Pastor, M. A., Macaluso, E., Day, B. L., and Frackowiak, R. S. J. (2006). The neural basis of temporal auditory discrimination. *NeuroImage* 30, 512–520. doi: 10.1016/j.neuroimage.2005.09.053

Paynter, J., and Aston, P. (1970). *Sound and Silence: Classroom Projects in Creative Music*. London: Cambridge University Press.

Pinho, A. L., de Manzano, O., Fransson, P., Eriksson, H., and Ullen, F. (2014). Connecting to create: expertise in musical improvisation is associated with increased functional connectivity between premotor and prefrontal areas. *J. Neurosci.* 34, 6156–6163. doi: 10.1523/JNEUROSCI.4769-13.2014

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

Pruim, R. H. R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., and Beckmann, C. F. (2015). ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage* 112, 267–277. doi: 10.1016/j.neuroimage.2015.02.064

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci.* 98, 676–682. doi: 10.1073/pnas.98.2.676

Runco, M. A., and Jaeger, G. J. (2012). The standard definition of creativity. *Creat. Res. J.* 24, 92–96. doi: 10.1080/10400419.2012.650092

Running, D. J. (2008). Creativity research in music education: a review (1980–2005). *Update Appl. Res. Music Educ.* 27, 41–48. doi: 10.1177/8755123308322280

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

Schafer, R. M. (1986). *The Thinking Ear*. Tronto: Arcana Editions.

Schlegel, A., Alexander, P., Fogelson, S. V., Li, X., Lu, Z., Kohler, P. J., et al. (2015). The artist emerges: visual art learning alters neural structure and function. *NeuroImage* 105, 440–451. doi: 10.1016/j.neuroimage.2014.11.014

Schneider, H., Bösze, C., and Stangl, B. (2000). *Klangnetze: Ein Versuch, die Wirklichkeit mit den Ohren zu erfi nden*. Saarbrücken: PFAU Verlag.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *J. Neurosci.* 27, 2349–2356. doi: 10.1523/JNEUROSCI.5587-06.2007

Sun, J., Chen, Q., Zhang, Q., Li, Y., Li, H., Wei, D., et al. (2016). Training your brain to be more creative: brain functional and structural changes induced by divergent thinking training: the neural plasticity of creativity. *Hum. Brain Mapp.* 37, 3375–3387. doi: 10.1002/hbm.23246

Synek, J. (2008). *Elementární Komponování a jeho Význam v Edukaci [Elementary Composing and Its Significance in Education]*. PhD Thesis. Olomouc: Palacký University.

Takeuchi, H., Taki, Y., Hashizume, H., Sassa, Y., Nagase, T., Nouchi, R., et al. (2011). Failing to deactivate: the association between brain activity during a working memory task and creativity. *NeuroImage* 55, 681–687. doi: 10.1016/j.neuroimage.2010.11.052

Takeuchi, H., Taki, Y., Hashizume, H., Sassa, Y., Nagase, T., Nouchi, R., et al. (2012). The association between resting functional connectivity and creativity. *Cereb. Cortex* 22, 2921–2929. doi: 10.1093/cercor/bhr371

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

Webster, P. R. (2016). Creative thinking in music, twenty-five years on. *Music Educ. J.* 102, 26–32. doi: 10.1177/0027432115623841

Wei, D., Yang, J., Li, W., Wang, K., Zhang, Q., and Qiu, J. (2014). Increased resting functional connectivity of the medial prefrontal cortex in creativity by means of cognitive stimulation. *Cortex* 51, 92–102. doi: 10.1016/j.cortex.2013.09.004

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage* 21, 1732–1747. doi: 10.1016/j.neuroimage.2003.12.023

Woolrich, M. W., Ripley, B. D., Brady, M., and Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage* 14, 1370–1386. doi: 10.1006/nimg.2001.0931

Worsley, K. J. (2001). "Statistical analysis of activation images," in *Functional MRI: An Introduction to Methods*, eds P. Jezzard, P. M. Matthews, and S. M. Smith (Oxford: Oxford University Press).

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

Zouhar, V., and Medek, I. (2010). "Music making in the classroom: possibilities for Czech music education," in *Proceeding of the 2nd World Conference on Arts Education, 25–28 May 2010* (Seoul), 1–3, (Paris: UNESCO).

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 Neuroscience

ORIGINAL RESEARCH
published: 09 December 2021
doi: 10.3389/fnins.2021.702067

Check for updates

Inter-subject Correlation While Listening to Minimalist Music: A Study of Electrophysiological and Behavioral Responses to Steve Reich's *Piano Phase*

Tysen Dauer[1,2]*, *Duc T. Nguyen*[1,3], *Nick Gang*[1], *Jacek P. Dmochowski*[4], *Jonathan Berger*[1] *and Blair Kaneshiro*[1,3]

[1] Center for Computer Research in Music and Acoustics, Stanford University, Stanford, CA, United States, [2] Rotman Research Institute, Baycrest Centre, Toronto, ON, Canada, [3] Center for the Study of Language and Information, Stanford University, Stanford, CA, United States, [4] Department of Biomedical Engineering, City College of New York, New York, NY, United States

OPEN ACCESS

Edited by:
Mark Reybrouck, KU Leuven, Belgium

Reviewed by:
Zachary Wallmark, University of Oregon, United States
Anja Thiede, University of Helsinki, Finland
Anna Czepiel, Max Planck Society, Germany

***Correspondence:**
Tysen Dauer
tysend@ccrma.stanford.edu

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

Received: *29 April 2021*
Accepted: *18 November 2021*
Published: *09 December 2021*

Citation:
Dauer T, Nguyen DT, Gang N, Dmochowski JP, Berger J and Kaneshiro B (2021) Inter-subject Correlation While Listening to Minimalist Music: A Study of Electrophysiological and Behavioral Responses to Steve Reich's Piano Phase. Front. Neurosci. 15:702067. doi: 10.3389/fnins.2021.702067

Musical minimalism utilizes the temporal manipulation of restricted collections of rhythmic, melodic, and/or harmonic materials. One example, Steve Reich's *Piano Phase*, offers listeners readily audible formal structure with unpredictable events at the local level. For example, pattern recurrences may generate strong expectations which are violated by small temporal and pitch deviations. A hyper-detailed listening strategy prompted by these minute deviations stands in contrast to the type of listening engagement typically cultivated around functional tonal Western music. Recent research has suggested that the inter-subject correlation (ISC) of electroencephalographic (EEG) responses to natural audio-visual stimuli objectively indexes a state of "engagement," demonstrating the potential of this approach for analyzing music listening. But can ISCs capture engagement with minimalist music, which features less obvious expectation formation and has historically received a wide range of reactions? To approach this question, we collected EEG and continuous behavioral (CB) data while 30 adults listened to an excerpt from Steve Reich's *Piano Phase*, as well as three controlled manipulations and a popular-music remix of the work. Our analyses reveal that EEG and CB ISC are highest for the remix stimulus and lowest for our most repetitive manipulation, no statistical differences in overall EEG ISC between our most musically meaningful manipulations and Reich's original piece, and evidence that compositional features drove engagement in time-resolved ISC analyses. We also found that aesthetic evaluations corresponded well with overall EEG ISC. Finally we highlight co-occurrences between stimulus events and time-resolved EEG and CB ISC. We offer the CB paradigm as a useful analysis measure and note the value of minimalist compositions as a limit case for the neuroscientific study of music listening. Overall, our participants' neural, continuous behavioral, and question responses showed strong similarities that may help refine our understanding of the type of engagement indexed by ISC for musical stimuli.

Keywords: inter-subject correlation (ISC), engagement, continuous behavioral measure, minimalism (music), EEG, music cognition

1. INTRODUCTION

The genre of musical minimalism is (in)famously characterized by highly recurrent, starkly restricted pitch and rhythmic collections. From the early days of scholarship on minimalist, or "repetitive music" as it was often called, commentators described the music's timbral and rhythmic staticity and its limited pitch patterns (Mertens, 1983, p. 12). While many advocates reported what we might call blissing out to this "meditative music" (to use yet another early term for this repertoire), some composers went on record to state their intention that the music should be listened to carefully (Strongin, 1969; Henahan, 1970). For example, the composer Steve Reich wrote in 1968 that he wanted to write works with musical processes that any listener could perceive: works where the process unfolded very gradually in order to "facilitate closely detailed listening" (Reich, 2009, p. 34). Numerous professional musicians and critics have asserted that listeners do not engage in such detailed listening—in part, they argue, because the music is overly simple and has insufficient substance to be cognitively engaging (see summaries of such negative appraisals in Fink, 2005, p. 19; Dauer, 2020, p. 24). Some music scholars have argued that minimalism's simplicity contains complexities upon analysis (Epstein, 1986; Cohn, 1992; Quinn, 2006). Beyond professionally trained listeners, do listeners tend to find the music engaging? If yes, do specific compositional details and techniques drive patterns of engagement?

Reich's *Piano Phase* (1967) offers a case study of how engagement and detailed listening might unfold. The piece, written for two pianos or marimbas, alternates between two distinct and highly repetitive states resulting from a single process. During in-phase sections, the two performers play a short musical unit in rhythmic unison, though varying in pitch alignment (**Figure 1**). In between these in-phase sections, one performer gradually accelerates, resulting in unpredictable note onsets (i.e., phasing sections). Over time these phasing sections lead to a new pitch alignment in the subsequent in-phase section.[1] The driving phasing process offers the listener an outline of how the piece unfolds at a macro-level while leaving many details unpredictable—for example, rhythms during the phasing sections and accent patterns during in-phase sections. For a listener interested in detailed minutia and slight variation, the work may fascinate; in other moods or with other listening priorities, the piece can bore, confuse, and even anger (Rockwell, 1973). With such a plethora of responses (Dauer, 2020), we aimed this initial study at better understanding engagement in general, operationalized for participants as "being compelled, drawn in, connected to what is happening, and interested in what will happen next" (Schubert et al., 2013).

Recent research using the high temporal resolution of electroencephalography (EEG) has suggested that the correlation of neural responses among participants (inter-subject correlation, or ISC) in response to natural audio-visual stimuli objectively indexes a state of "engagement." Foundational ISC work using fMRI has highlighted across-participant synchronization of neural responses to natural stimuli such as film excerpts (Hasson et al., 2004) and spoken narratives (Simony et al., 2016), and has uncovered relationships between neural synchronization and stimulus characteristics such as emotional arousal (Hasson et al., 2004) and narrative coherence (Lerner et al., 2011). fMRI ISC has also been used to study music processing: Abrams et al. (2013) reported greater synchronization when hearing intact music compared to temporally or spectrally manipulated controls, while Farbood et al. (2015) related hierarchical structural coherence of music to hierarchical neural processing. ISC for EEG was introduced in a film-viewing study by Dmochowski et al. (2012), who found that neural correlation was higher in response to film excerpts containing intact (vs. temporally scrambled) narratives, and peaked during periods of high tension and suspense—leading the authors to frame EEG-ISC as a measure of *engagement*, which they defined as "emotionally laden attention." Dmochowski et al. (2012) note that the brain state of engagement "lacks a rigorous definition" yet can be "readily describe[d] subjectively," and that it implies not only a state of attention, but an attentive state that "entails emotional involvement." The engagement interpretation of EEG ISC in the context of audiovisual processing was further investigated by Dmochowski et al. (2014), who found ISC of an experimental sample to reflect "engagement or interest of a large population" in television viewing. EEG ISC has subsequently been shown to index attentional state (Ki et al., 2016) and to predict memory retention (Cohen and Parra, 2016) and test scores (Cohen et al., 2018).

Ensuing studies have demonstrated how EEG ISC may be a powerful tool for analyzing music listening (Madsen et al., 2019; Kaneshiro et al., 2020, 2021). Madsen et al. (2019) drew on instrumental compositions (19 Western classical musical works in a variety of styles, and one Chinese folk song) to establish that ISCs decrease over repeated exposures to familiar music (though ISCs were sustained for participants with musical training). Kaneshiro et al. (2020) presented popular, Hindi-language songs from "Bollywood" films to participants and reported higher behavioral ratings and ISCs for their original versions when compared with phase-scrambled manipulations. Most recently, Kaneshiro et al. (2021) investigated participants' time-resolved

[1]The piece begins with one pianist (Pianist 1) playing a twelve-note pattern consisting entirely of 16th notes and containing five unique pitches in the treble register. The pattern can be divided into two groups of six 16th notes, and Reich gave a metronome marking of 72 beats per minute to the dotted quarter note (one group of six 16th notes). The score consists of numbered modules that are repeated an indeterminate number of times: Reich noted approximate ranges for the number of repetitions above each module. After the pattern is established in the first module, the second pianist (Pianist 2) fades in, playing the identical pattern in unison with Pianist 1. After repeating the pattern in unison for some time, Pianist 2 accelerates very slightly while Pianist 1 holds the opening tempo, causing the sound from the two pianos to "wobble" out of sync to varying degrees as the pattern is repeated at different tempos (we call these portions *phasing* sections). Various and unpredictable rhythm and pitch events emerge and disappear in these phasing sections. Eventually Pianist 2's acceleration process culminates in another unison module where each pianist's 16th notes are once again realigned (which we label *in-phase* sections). Although the pianists' rhythms are realigned, the pitch content of the pattern will have shifted: In this example, Pianist 2 aligns the second pitch of the opening pattern with the first pitch of the pattern (played by Pianist 1). *Piano Phase* proceeds by alternating between phasing and in-phase sections, where each successive in-phase section presents the next shifted alignment of the opening, twelve-note pattern (note three aligns with the first note of the pattern, a phasing section occurs, then note four aligns with the first note of the pattern, etc.).

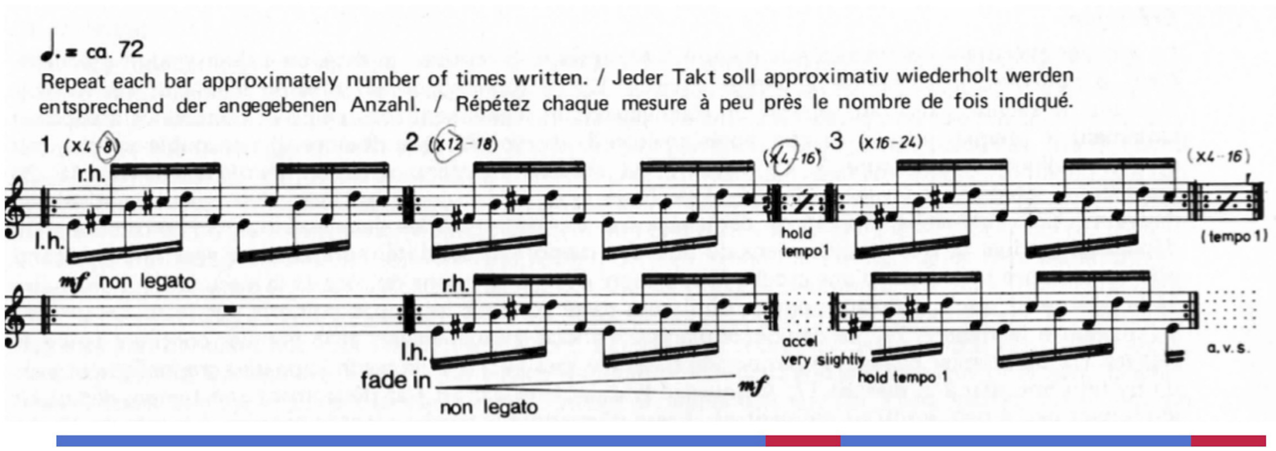

FIGURE 1 | The opening modules from Steve Reich's *Piano Phase*. Lines under the staff indicate sections: blue lines are in-phase sections and red lines are phasing sections.

ISCs in response to the first movement of Edward Elgar's Cello Concerto in E minor, Op. 85. In contrast to the stimuli used in these previous studies, and true to minimalism's stereotypical characteristics, Reich's *Piano Phase* features a high level of repetition, unchanging timbre, and narrow pitch content.[2]

Other researchers have used minimalist compositions as experimental stimuli, similarly taking advantage of the works' unusual musical properties. Musicologist Keith Potter and computer science colleagues used two early works by Philip Glass to compare information dynamics and musical structure (Potter et al., 2007). Psychologist Michael Schutz worked with percussionist Russell Hartenberger to examine desynchronization among performers of Reich's *Drumming* (Hartenberger, 2016),[3] and Daniel Cameron and colleagues have studied experiences of groove and neural entrainment using Reich's *Clapping Music* (Cameron et al., 2017, 2019). Dauer et al. (2020) examined preattentive cortical responses to various types of formal repetition using synthesized melodies based on early minimalist compositional techniques. The current study takes minimalism as an edge case in the applicability of neural correlation, uniting the repertoire's extreme musical techniques (and unique reception history) with multivariate techniques for analyzing brain data. While we focus on phasing as one important type of musical repetition, we anticipate that some aspects of the results may meaningfully generalize to other repetitive repertoires such as music used to accompany trancing (Becker, 2004). Future work could interrogate such generalizations.

Our primary research question was to uncover whether participants shared engagement patterns (as measured by ISC) while listening to *Piano Phase*. In particular, we hypothesized that phasing sections (sections where one pianist is changing tempo) would be more collectively engaging (i.e., elicit more correlated responses) than in-phase sections, due to phasing sections' rhythmic variety, rhythmic unpredictability, and a wider variety of pitch interactions (see above and **Figure 1** for musical details about *Piano Phase*). If listeners deployed the hyper-detailed listening strategy described above, phasing sections would offer rich content with which to engage. On the other hand, detailed listening during phasing sections could lead to divergent engagement between listeners as they lock on to different aspects of the music during these more eventful sections. Since ISC depends on time-locked similarities in neural data, these divergent but equally engaged listening experiences may result in lower correlations than in-phase sections. Using ISC as a way to index collective engagement, we explored whether phasing sections contribute to ISC by introducing a manipulation of *Piano Phase* without phasing sections (which we called Abrupt Change). We anticipated that ISC would be lower for this manipulation if phasing sections contributed to ISC in the original version. We also examined whether the gradual nature of the phasing process in *Piano Phase* might be critical for engagement. To this end, we included a manipulation of *Piano Phase* with frequent and random changes in the content (Segment Shuffle). By reshuffling 5-s segments of the original excerpt, we rendered unrecognizable the gradual phasing process and the alternations between in-phase and phasing sections. If the phasing process meaningfully contributes to engagement, we expected lower ISC values for the shuffled version as it lacked gradual phasing. To examine the possibility of listeners being bored or disengaged by the original work, we also introduced a third control stimulus with more extreme repetition that should be less engaging than the original work (Tremolo). Finally, we included a commercial remix of Reich's original work in a popular style (Remix), which we conjectured would reliably engage listeners and elicit EEG ISC comparably to previous experiments with popular music stimuli (Kaneshiro, 2016; Kaneshiro et al., 2020). Remix also provided a stylistic contrast with *Piano Phase*: we expected that the remix would engage listeners more than *Piano Phase* because the remix has more attention-catching musical events. In sum, if the core musical features of *Piano*

[2]We note that Madsen et al. (2019) did include Philip Glass's *String Quartet No. 5* (1991): a more popular or "post-minimalist" work by comparison.
[3]https://maplelab.net/reich/

Phase drive engagement, we hypothesized that the manipulated versions (Abrupt Change, Segment Shuffle, and Tremolo) would elicit lower ISC. We expected ISC in response to Remix to be comparable with values found in previous popular-music pieces (Kaneshiro et al., 2020).

In line with recent work, we computed ISCs over entire excerpts and in shorter, overlapping time windows, giving us a sense of overall engagement as well as moment-to-moment patterns shared between audience members (Dmochowski et al., 2012; Kaneshiro et al., 2021). To provide complementary measures of what ISC is reliably indexing, participants rated the stimuli and additionally completed a second experimental block where they continuously reported their level of engagement with the stimuli. This allowed us to compare relationships for both overall and time-resolved neural and behavioral measures. These continuous EEG and behavioral measures allowed us to examine our expectations at a more granular level: We expected significant ISC during phasing sections of the original version and at the onset of new phasing sections in Abrupt Change, scarce ISC for Segment Shuffle, less for Tremolo, and frequent ISC in response to the dramatic musical events in Remix.

2. METHODS

2.1. Stimuli

All five stimuli in the experiment are related to Steve Reich's *Piano Phase*, a much-anthologized example of American minimalism for two pianos or marimbas (**Figure 1**). In the experiment we used pianists Nurit Tilles and Edmund Neimann's 1987 recording on the album *Reich "Early Works"* released by Double Edge (Reich, 1987). The performers take an appropriate tempo (see footnote 1), use detached articulation, and create an overall energetic feel. We used the first 5 min and 5 s (5:05) of the track's 20:26 duration. We refer to this excerpt of *Piano Phase* used in the experiment as the *Original* condition (**Figure 2A**).[4]

Piano Phase offers contrasting sections (phasing and in-phase) with slightly varying musical content for comparison while holding many other musical parameters constant: timbre, dynamics (largely), instrumentation, pitch content, and absence of lyrics or vocal content. These features make it uncommonly amenable to the creation of the stimulus manipulations used in this study.

Using MATLAB software, we created three additional stimulus conditions of equal duration, each based on the content of the excerpt used in the Original condition. First, in the *Abrupt Change* condition, (**Figure 2B**) all phasing sections from the Original excerpt were replaced with exact repetitions of the preceding in-phase material. The stimulus thus presents repetitions of an in-phase motif through the section where the phasing would have occurred, and then shifts abruptly to the next in-phase section as closely as possible to its occurrence in the original recording. For example, the stimulus begins with the in-phase section where Pianist 1 and Pianist 2 align the first notes of the twelve-note pattern. This continues without phasing

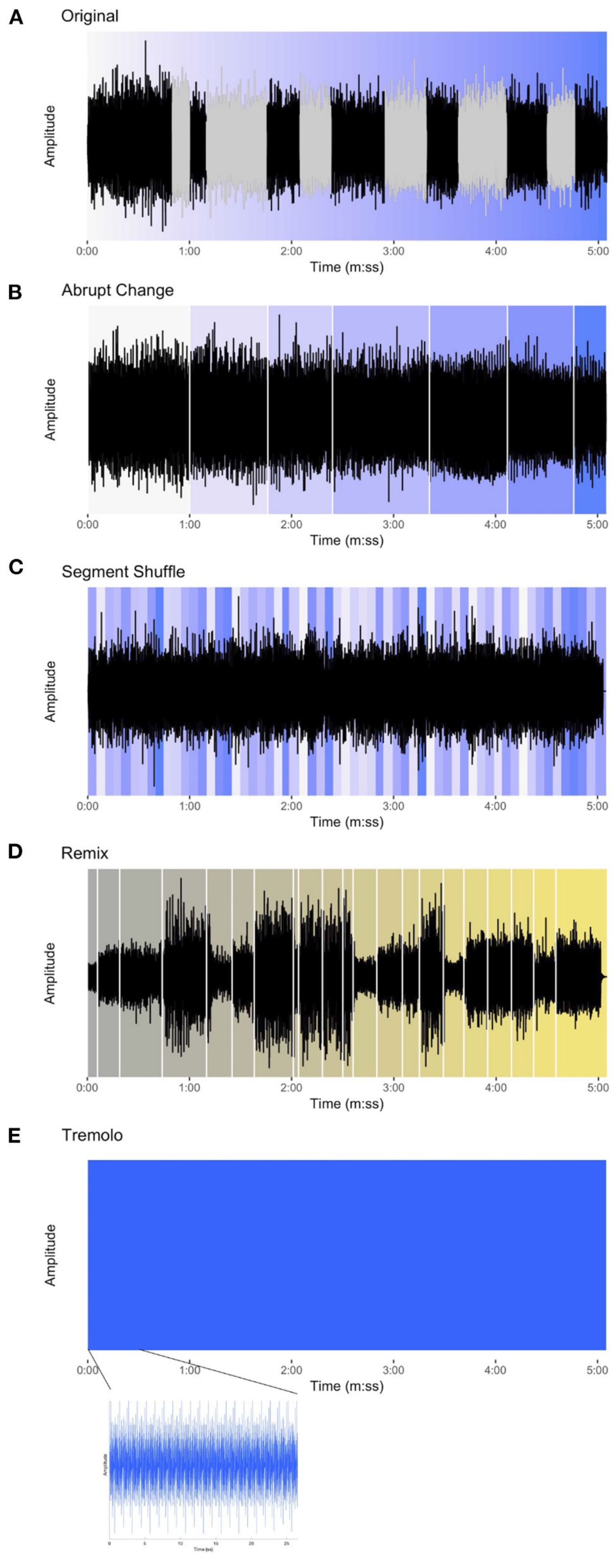

FIGURE 2 | The waveforms for each of the stimuli in the experiment. **(A)** Original, with phasing sections colored gray and the progression of events *(Continued)*

[4] A meter shift and accompanying pattern change occur later in the piece, but after the excerpt used in the experiment.

FIGURE 2 | represented by the gradual change of color from white to blue. **(B)** Abrupt Change, white lines denoting sudden shift from one in-phase section to the next and background color showing approximate location of in-phase material in the Original condition. **(C)** Segment Shuffle, random re-ordering of 5-s units shown using original color in Original. **(D)** Remix [Winn's *Piano Phase (D*Note's Phased & Konfused Mix)*], gradual progression of events represented with color change from gray to yellow and key musical events beginning with white lines. **(E)** Tremolo, appearing as an unchanging block when zoomed out, but in the lower plot, zoomed in to show the reiterated pitch material.

until suddenly the next in-phase section emerges, where Pianist 2 aligns the second note of the pattern with the first note of the pattern played by Pianist 1. Thus, the Abrupt Change condition is, in essence, form without function: where regular markers of formal sections (i.e., points of arrival at the alignments of in-phase sections) are situated without the functional transitions (i.e., the phasing sections).

As a contrast to the sudden changes embodied by the Abrupt Change condition, we created the *Segment Shuffle* condition (**Figure 2C**). Here we divided the Original audio into 5-s segments and randomly reordered them (i.e., "shuffled" them). In order to avoid abrupt disjunct shifts, the transitions between segments were smoothed by applying a linear crossfade. The 5-s segments included both phasing and in-phase material, meaning that upcoming content was unpredictable for listeners. In contrast with the Abrupt Change condition, Segment Shuffle featured function without form: constant, potentially surprising changes with no overarching formal scheme.

Finally, we synthesized a stimulus with neither form nor function, taking the repetition aspect of minimalist music to an extreme. Our *Tremolo* condition (**Figure 2E**) consisted solely of the aggregated pitch content of *Piano Phase* presented as a block chord, reiterated at Reich's opening tempo marking and lasting the duration of the Original excerpt.

For comparison with the more popular genres of audio materials used in previous ISC studies, we also included Matt Winn's Piano Phase *(D*Note's Phased & Konfused Mix)*, an homage to Reich's piece released on the 1999 *Reich Remixed* album (Reich, 1999); we refer to this condition as *Remix* for short (**Figure 2D**). Winn's dance music group, D*Note, draws on sounds from electronica and jazz, and these influences show up in Remix alongside samples from Reich's piece.[5] The entire track was used in the experiment and its duration (5:05) informed the length of the other stimuli. Listening to Remix, we identified moments (musical events) that we predicted would engage listeners (for a full list, see **Supplementary Table S1**). These events guided our interpretation of time-resolved EEG and continuous behavioral (CB) results.

All stimuli were presented to participants as mono .wav files; the second audio channel was embedded with intermittent square-wave pulses which were used as precise timing triggers (see § 2.3 and Kaneshiro et al., 2020).

[5]https://www.mattwinn.co.uk/about

2.2. Participants

We were interested in listeners' initial experiences of Reich's piece and sought participants who were unlikely to have heard the composition before. Participants had to be 18–35 years old, have normal hearing, be right-handed, have no cognitive or decisional impairments, be fluent in English, and have had no individual musical instrument or vocal training, nor musical education after high school (or equivalent).

The participant sample ($N = 30$; 19 female, 11 male) had a mean age of 23.8 years (ranging from 18 to 35 years). Twelve participants reported some formal musical training ranging from 2 to 16 years (average of 4.5 years) including activities such as elementary school band and orchestra and piano lessons in middle school. Only two participants reported ongoing musical activities (one was an amateur ukulele player and another noted participating in occasional jam sessions). All participants reported listening to music regularly, from 0.2 to 8 h a day (average of 2.4 h per day).

2.3. Experimental Paradigm and Data Acquisition

The Stanford University Institutional Review Board approved this research, and all participants gave written informed consent before completing the experiment. After discussing and signing the consent form, each participant completed questionnaires about demographic information and musical experience. Each participant then completed two blocks: one EEG (Block 1) and one behavioral (Block 2), both conducted in an acoustically and electrically shielded ETS-Lindgren booth (**Figure 3**). The participant completed a brief training session to acquaint them with the interface and task before the experimenter donned the EEG net. The participant was told to sit comfortably in front of the monitor and view a fixation image while EEG was recorded. Participants listened to each of the five stimuli once in random order with their eyes open. Participants did not perform any task during the presentation of the stimuli and were told to refrain from moving their body in response to the music: they were told not to tap their feet or hands, or bob their heads. After each stimulus in Block 1, the participant rated how pleasant, well ordered, musical, and interesting the preceding stimulus was on a scale of 1 (not at all) to 9 (very) via key press using a computer keyboard. Participants were permitted to move and take short breaks in between stimuli (during which time a "break" screen appeared). When ready, the participant initiated the next stimulus by pressing the space bar on the keyboard.

The EEG net was removed after Block 1, and the participant returned to the sound booth to complete Block 2. Here the participant heard the same five stimuli (in random order) and this time completed a continuous behavioral task while listening. Their task was to continuously report their level of engagement—which was defined as "being compelled, drawn in, connected to what is happening, and interested in what will happen next" (Schubert et al., 2013)—over the duration of each stimulus. We consider this definition to be aligned with the original definition in EEG-ISC research of "emotionally laden attention" (Dmochowski et al., 2012), while also providing participants a clearer, more elaborated way of understanding engagement in order to perform the task. This more detailed definition has also been used in previous studies involving continuous reporting of engagement in response to dance (Schubert et al., 2013) and music (Olsen et al., 2014). The definition of engagement was provided only in the second, behavioral block and not in the EEG block as participants in prior EEG-ISC studies of engagement were not informed that their neural responses would be related to this state (Dmochowski et al., 2012, 2014; Madsen et al., 2019; Kaneshiro et al., 2020, 2021).

To perform this task, the participant used a computer mouse to control a slider shown on the computer monitor. The screen displaying the slider contained the prompt "Rate your level of engagement as the excerpt plays," and the endpoints of the slider were labeled "Not at all" and "Very engaged," corresponding to continuous scale values of 0 and 100, respectively. The slider was positioned at the bottom of its range (0 value) at the start of each trial. After each stimulus, the participant rated how engaging they found the preceding stimulus to be overall, using the same 1–9 key press scale used in Block 1. The ordering of blocks was not randomized (i.e., the EEG block always preceded the CB block) because we wanted to ensure that during recording of EEG data in Block 1, participants would not be biased with the definition of engagement and the continuous reporting task that came in Block 2.

The experiment was programmed in MATLAB using the Psychophysics Toolbox (Brainard, 1997). Stimuli were played through two Genelec 1030A speakers located 120 cm from the participant. The stimuli were scaled to a common loudness level based on perceptual assessments from three researchers, all of whom were trained musicians. Stimulus onsets were precisely timed by sending square-wave pulses to the EEG amplifier from a second audio channel (not heard by the participant). We used the Electrical Geodesics, Inc., (EGI) GES 300 platform (Tucker, 1993), a Net Amps 300 amplifier, and 128-channel electrode nets to acquire EEG data with a 1 kHz sampling rate and Cz vertex reference. Before beginning the EEG block, we verified that electrode impedances were below 60 kΩ (Ferree et al., 2001). In the CB block, data were acquired at a sampling rate of 20 Hz.

2.4. Data Preprocessing

Continuous EEG recordings were preprocessed offline in MATLAB after export using Net Station software. The data preprocessing procedure used here is described in detail in Kaneshiro et al. (2021), which itself was adapted from the preprocessing procedure of Kaneshiro et al. (2020). Briefly, data were preprocessed on a per-recording basis: Each recording was highpass (above 0.3 Hz), notch (between 59 and 61 Hz) and lowpass (below 50 Hz) zero-phase filtered before being downsampled from 1 kHz to 125 Hz. Epochs for each stimulus were 5 min (5:00; 37501 time samples) in length; we used a slightly shorter analysis epoch than the length of the stimuli (excluding the last 5 s) because the Remix stimulus included 5 s of silence at the end. Stimulus onsets were precisely timed from the audio pulses. Ocular and EKG artifacts were removed using ICA (Jung et al., 1998): Components whose activity reflected ocular activity (identified according to the procedure described in Kaneshiro et al., 2020) or EKG artifacts (identified via visual inspection of projected activity in the first 30 components) were zeroed out before projecting data back to electrode space. Finally, data were converted to average reference, and data from bad electrodes or noisy transients were replaced with a spatial average of data from neighboring electrodes. After preprocessing, each trial of data was a 2D electrode-by-time matrix ($125 \times 37,501$). The matrices contained data from 125 electrodes as we excluded the four sensors over the face (electrodes 125–128) and reconstituted the reference sensor during preprocessing (Kaneshiro, 2016; Losorelli et al., 2017; Kaneshiro et al., 2020, 2021). During preprocessing, participant S08's response to the Tremolo stimulus was flagged as containing excessive noise artifacts; therefore we excluded this trial from further analysis, but retained other trials from this participant.

After preprocessing, we aggregated trials into 3D electrode-by-time-by-participant data matrices for each stimulus. As a result, responses to Original, Abrupt Change, Segment Shuffle, and Remix stimuli were stored in $125 \times 37,501 \times 30$ matrices, while responses to Tremolo were stored in a $125 \times 37,501 \times 29$ matrix.

CB data were similarly segmented into 5-min (5:00) epochs and aggregated into a single time-by-participant-by-stimulus matrix. As the data remained at the acquisition sampling rate of 20 Hz for analysis, the matrix was of size $6,000 \times 30 \times 5$. Behavioral ratings from the EEG and CB blocks were aggregated into a single .csv file for statistical analyses.

2.5. Data Analysis

Figure 3 summarizes our analysis pipeline for the EEG and CB data. EEG was recorded from participants in Block 1, and participants provided CB reports of engagement in Block 2. Participants also rated the stimuli in both blocks. We computed ISC of both the EEG and CB measures, and also computed mean CB across participants. Finally, we analyzed the ratings to determine whether they differed significantly according to stimulus.

2.5.1. Spatial Filtering of EEG Data

Previous EEG ISC studies have prepended a spatial filtering operation before computing correlations in order to maximize signal-to-noise ratio of the data while also reducing the dimensionality of each EEG trial from a space-by-time matrix to time vectors from one or a few components (Dmochowski et al., 2012). Therefore, we filtered the EEG data using Reliable Components Analysis (RCA) prior to computing ISC (Dmochowski et al., 2012, 2015). RCA maximizes across-trials covariance of EEG responses to a shared stimulus relative to within-trials covariance, and therefore maximizes correlated activity across trials (i.e., ISC). It is similar to PCA, but maximizes correlation across trials as opposed to variance explained in a single response matrix. Like PCA, RCA involves an eigenvalue decomposition of the data, returning multiple spatial filters as eigenvectors and corresponding coefficients as eigenvalues (Dmochowski et al., 2012). The components are returned in descending order of reliability explained; in other words, the first component RC1 is that in which ISC of component-space data

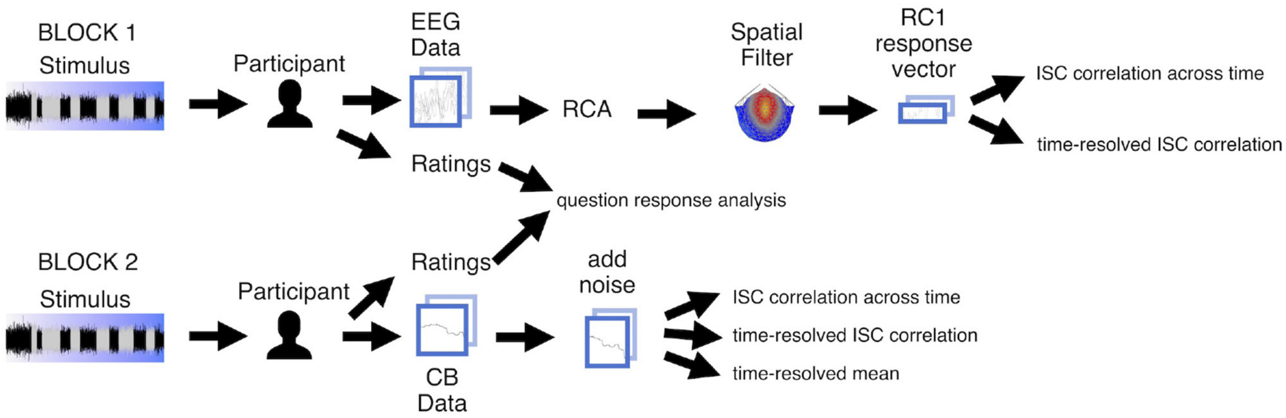

FIGURE 3 | Analysis pipeline for experiment data. Participants heard each of the five stimuli twice, once in each block. During Block 1 we recorded EEG, and during Block 2 participants completed the continuous behavioral (CB) task. Participants answered questions about each stimulus after hearing it. For the EEG data we computed spatial components maximizing temporal correlation and projected electrode-by-time response matrices to component-by-time vectors. For vectorized EEG as well as CB vectors, we then computed inter-subject correlation (ISC) of the vectors on a per-stimulus basis, across time and in a time-resolved fashion. We additionally computed the time-resolved mean values between participants. We aggregated and analyzed ratings.

is maximized, followed by RC2, RC3, etc. We use the RC1 of Tremolo in what follows but note two important limitations: no RCs for Tremolo are statistically significant, and the topography of RC1 is qualitatively different from the first RCs for the other four stimuli (**Figure 5A** and **Supplementary Figure S1**).

We used a publicly available MATLAB implementation (Dmochowski et al., 2015), computing RCA separately for each stimulus. Following Kaneshiro et al. (2020), we computed the top five reliable components (RCs). We observed a sharp drop in RC coefficients after the first, most-correlated component (RC1); given that past research has reported negligible ISC in subsequent RCs in this scenario (Kaneshiro et al., 2021), we proceeded with ISC analyses using RC1 data only, as was done by Kaneshiro et al. (2020). In presenting the forward-model projections of component weights as scalp topographies (Parra et al., 2005), each weight vector was first multiplied by ± 1 such that frontal electrodes were associated with positive weightings; this was for visualization only, and polarity of the projected data does not impact computed correlations.

2.5.2. Inter-subject Correlation Analyses

In computing the EEG ISC of RC1 response vectors, we first computed ISC across the entire duration of each stimulus (Kaneshiro et al., 2020, 2021). Following this, we computed ISC in a time-resolved fashion. Following past research (Dmochowski et al., 2012; Poulsen et al., 2017; Kaneshiro et al., 2021), we used a 5-s window with a 1-s shift between windows. These parameters provide an adequate number of data points (625 EEG samples, 100 CB samples) to produce sufficient signal to noise to measure correlation, while this window length in conjunction with the 1-s hop size leads to an 80% overlap between windows, which smooths the resulting time series and facilitates interpretation. These window length and window shift parameters produced a total of 296 time-resolved ISC points across each stimulus with a temporal resolution of 1 s. ISC for each participant was computed in a one-against-all fashion (the correlation of each participant's RC1 response vector with every other participant's response vector for a given stimulus). We report the mean ISC across participants and additionally visualize single-participant correlations for all-time ISC and standard error of the mean for time-resolved ISC.

For the CB responses, we computed mean CB at each time sample, as well as CB ISC both across entire excerpts and in the short time windows described above. CB responses were already in vector form for each participant, so we did not perform any operation akin to EEG spatial filtering before computing means and ISC. At times, individual participants did not move the slider in a given 5-s window, which produced missing values when computing correlations. To address this issue, for the CB ISC analyses *only* we added a small amount of noise, uniformly distributed over the interval ± 0.001, independently to each CB response vector prior to computing ISC. As with the EEG data, we report means and single-participant values for analyses across entire stimuli, and means with standard error of the mean for time-resolved measures.

2.5.3. Statistical Analyses

Significance of each EEG result was computed using permutation testing. As described in detail in previous studies (Kaneshiro et al., 2020, 2021), we conducted each EEG analysis 1,000 times; in each iteration, the phase spectrum of each EEG trial input to RCA had been randomized (Prichard and Theiler, 1994). The distribution of 1,000 outcomes for each analysis then served as the null distribution for assessing significance of the observed result. We performed a similar procedure to create null distributions for CB ISC, independently phase scrambling each CB response vector prior to computing ISC—also over 1,000 iterations. We compare each observed ISC result to the corresponding null distribution in order to compute p-values, and as effect size d we report the number of standard deviations between the observed result and the expected result under the null distribution (Nakagawa and Cuthill, 2007).

Behavioral ratings, EEG ISC computed over entire stimuli, and CB ISC computed over entire stimuli were each analyzed using R (Ihaka and Gentleman, 1996; R Core Team, 2019) and the lme4 package (Bates et al., 2012). We performed a linear mixed-effects analysis of the relationship between response values and stimulus conditions, with fixed effect of condition (Original, Abrupt Change, Segment Shuffle, Remix, and Tremolo) and random effect of participant in each model. We then tested each model against a null model without the fixed effect of condition using the anova function in lme4. This produced a chi-squared statistic and associated p-value (Winter, 2013). As in Kaneshiro et al. (2020), ordinal behavioral ratings were treated as approximately continuous (Norman, 2010). Following this we conducted two-tailed pairwise t-tests to assess differences between pairs of stimulus conditions. Effect size (Cohen's D) was also calculated and reported.

Results for analyses involving multiple comparisons were corrected using False Discovery Rate (FDR, Benjamini and Yekutieli, 2001). For discrete results, we corrected for multiple comparisons on a per-stimulus basis (EEG ISC and CB ISC data: 10 paired comparisons over five stimulus conditions; behavioral ratings: 10 paired comparisons per stimulus; RC coefficients: five unpaired comparisons per stimulus). We performed no temporal cluster correction on the time-resolved ISC: as noted by Kaneshiro et al. (2021), temporal dependence was accounted for in the phase-scrambling procedure underlying the permutation testing, which preserves autocorrelation characteristics of the original response data (Prichard and Theiler, 1994; Lancaster et al., 2018).

3. RESULTS

In order to examine engagement with an example of musical minimalism, we used inter-subject correlation (ISC) to analyze EEG and continuous behavioral (CB) responses from 30 adult participants who heard an intact excerpt of Steve Reich's *Piano Phase*, three manipulated control stimuli, and a professional remix of Reich's piece. We analyzed EEG and CB ISC in two ways: an aggregate ISC value for each stimulus (overall EEG ISC, overall CB ISC) and time-resolved ISCs for both EEG and CB data. Each participant also gave ordinal ratings of each stimulus (behavioral ratings).

3.1. Remix Stimulus Garnered Highest Behavioral Ratings

After hearing each stimulus in Block 1, participants used a 1–9 scale to rate how pleasant, musical, well ordered, and interesting they found each excerpt. Later, in Block 2, they used the same scale to report their overall level of engagement with each stimulus. Ratings for all five questions were found to differ significantly by condition (**Figure 4**): pleasant [$\chi^2_{(4)} = 126.03$, $p < 0.001$], musical [$\chi^2_{(4)} = 139.78$, $p < 0.001$], well ordered [$\chi^2_{(4)} = 37.996$, $p < 0.001$], interesting [$\chi^2_{(4)} = 104.29$, $p < 0.001$], and engaging [$\chi^2_{(4)} = 127.92$, $p < 0.001$].

Follow-up pairwise t-tests comparing responses between conditions showed a similar pattern for four of the five questions (see **Supplementary Tables S2–S6** for all p-values and d-values). For pleasant, musical, interesting, and engaging ratings, responses to Remix were significantly higher than to the other four conditions ($p_{FDR} < 0.01$, 10 comparisons) and responses to Tremolo were significantly lower than the other four conditions ($p_{FDR} < 0.01$). However, these ratings did not differ significantly between Original, Abrupt Change, and Segment Shuffle conditions ($p_{FDR} > 0.05$).

Ratings for how "well ordered" the stimuli were followed a slightly different pattern. While Remix was rated significantly higher than all other conditions (see **Supplementary Table S4**), Tremolo was rated significantly lower than all other conditions except Segment Shuffle ($p_{FDR} = 0.719$, $d = 0.065$). In addition, Segment Shuffle was rated significantly lower than Abrupt Change ($p_{FDR} = 0.036$, $d = 0.543$).

3.2. Overall EEG ISC Is Highest for Remix, Lowest for Tremolo

In computing the EEG ISCs, we first spatially filtered the responses for each stimulus in order to reduce their dimensionality from 125 electrodes to a single, maximally correlated spatial component (RC1) for each stimulus. These components are shown in **Figure 5A**. For all but the Tremolo, RC1 was maximally weighted over the fronto-central region. While our spatial filtering technique returned multiple components, we focus only on the first component because it is the only component with statistically significant coefficients for the majority of stimuli: **Figure 5B** demonstrates that RC1 was the only significant component for most stimuli (permutation testing; Original, Abrupt Change, Segment Shuffle, Remix $p_{FDR} < 0.001$; Tremolo $p_{FDR} = 0.379$; see **Supplementary Table S7** for all p-values). Remix also had a significant RC4 and Tremolo had no significant RCs. The topographies and coefficient significance for RC1 are in line with those computed in previous music EEG ISC studies (Kaneshiro et al., 2020, 2021); given that subsequent RCs did not correspond to significant ISC in a closely related study with similar distributions of coefficients (Kaneshiro et al., 2021), here we compute ISC only for RC1.

When computed over the entire duration of a stimulus, EEG ISC was statistically significant in response to Original (permutation test, $p < 0.001$, $d = 2.4$), Abrupt Change ($p < 0.001$, $d = 3.1$), Segment Shuffle ($p < 0.001$, $d = 2.7$), and Remix ($p < 0.001$, $d = 5.1$), but not Tremolo ($p = 0.41$, $d = 0.7$). ISC also differed significantly by condition [$\chi^2_{(4)} = 96.002$, $p < 0.001$]; follow-up pairwise comparisons indicated that Original, Abrupt Change, Segment Shuffle, and Tremolo all significantly differed from Remix ($p_{FDR} < 0.001$), and Original, Abrupt Change, Segment Shuffle, and Remix all differed from Tremolo ($p_{FDR} < 0.001$). **Figure 5C** shows the direction of these significant differences: Remix garnered higher overall EEG ISC values than the other conditions, while Tremolo received the lowest overall values. Despite their structural differences, ISC among Original, Abrupt Change, and Segment Shuffle did not

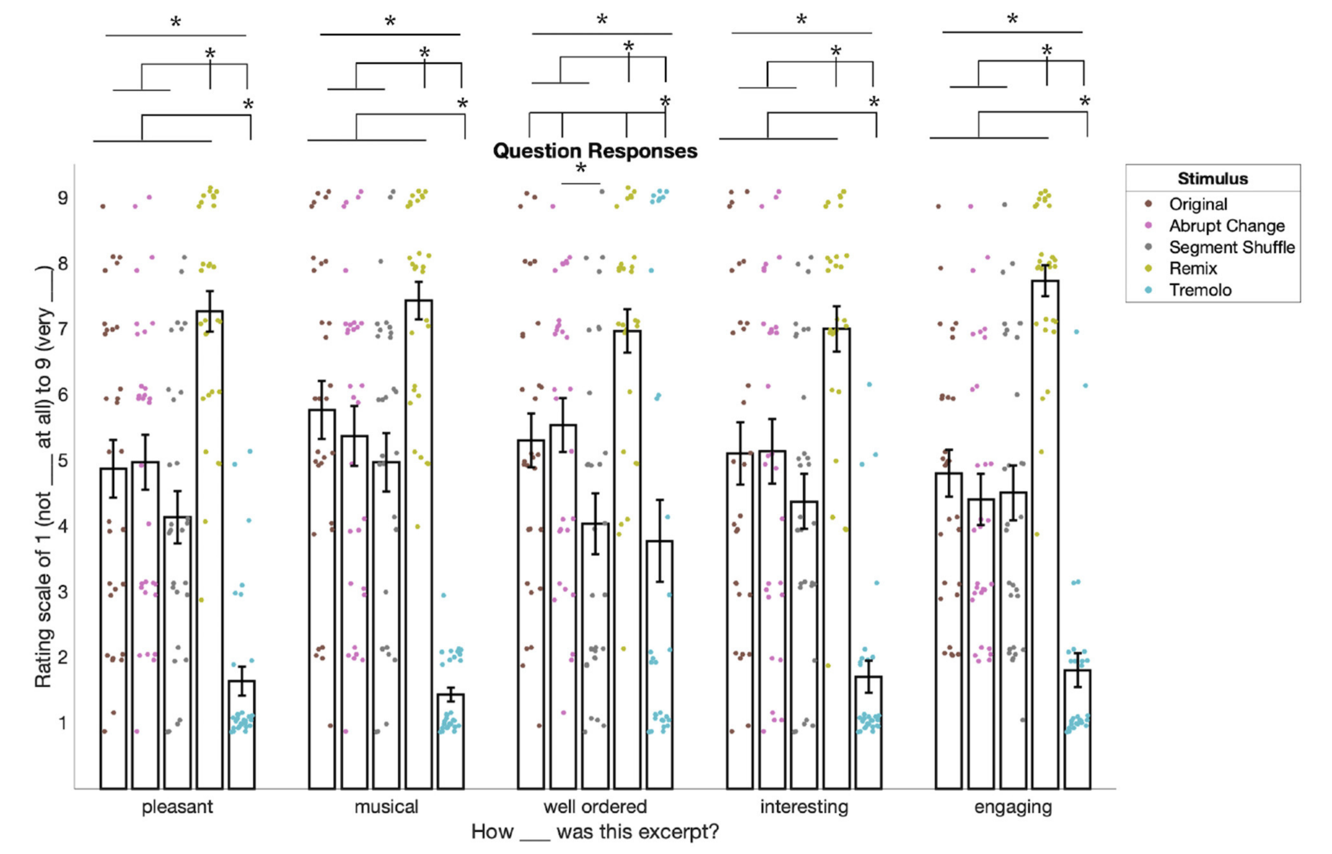

FIGURE 4 | Behavioral ratings for all questions in the experiment (responses were ordinal and are slightly jittered for visualization only). Ratings for "pleasant," "musical," "well ordered," and "interesting" come from Block 1 and ratings for "engaging" come from Block 2. For pleasant, musical, interesting, and engaging, responses for Remix were significantly higher than for the other conditions. For these same questions, responses were also significantly lower for Tremolo compared to all other conditions. For ratings of well ordered, we saw a similar pattern except that Abrupt Change was significantly higher than Segment Shuffle. Asterisks denote significance of $p < 0.05$. Please refer to the online version of the paper for the full-color figure.

differ significantly from one another when computed over entire excerpts (see **Supplementary Table S8** for a full list of p-values and d-values).

3.3. Overall CB ISC Aligns Broadly With EEG ISC

To analyze the CB ISC values (**Figure 6**), we followed the same procedures used for comparing EEG ISC computed over entire stimuli. Statistically significant CB ISC was observed in responses to Original (permutation test, $p < 0.001$, $d = 6.5$), Abrupt Change ($p < 0.001$, $d = 6.6$), Segment Shuffle ($p < 0.001$, $d = 12.7$), and Remix ($p < 0.001$, $d = 34.7$) stimuli, but not Tremolo ($p = 0.22$, $d = 0.6$). CB ISC significantly differed by condition [$\chi^2_{(4)} = 180.2$, $p < 0.001$]. Pairwise comparisons revealed that Remix had higher ISC than all other conditions, Tremolo had lower ISC than all other conditions, and Segment Shuffle had higher ISC than all conditions except Remix. All condition comparisons were significant except for Original vs. Abrupt Change ($p_{FDR} = 0.87$, $d = 0.06$; all other comparisons, $p_{FDR} < 0.05$, see **Supplementary Table S9** for a full list). Cross-correlations between time-resolved EEG ISC and CB ISC showed a similar pattern (maximum normalized r-values with a maximum lag of 10 s were: Original: 0.29, Abrupt Change: 0.29, Segment Shuffle: 0.37, Remix: 0.70, and Tremolo: 0.18).

3.4. Time-Resolved Measures Coincide With Musical Events

In addition to calculating the overall ISC for EEG and CB data, we were also interested in observing changes in ISC over the course of the stimuli (see **Supplementary Figure S2** for individual, time-resolved CB responses underlying CB ISC). After computing ISC over short, shifting time windows, we visualized the ISC trajectory over time. Permutation testing provided a time-varying statistical significance threshold, allowing us to see when participants, as a group, had significantly correlated responses. Below we give a qualitative assessment of these results (**Figure 7**). Note that although EEG and CB ISC data had different sampling rates, we used identical time window lengths (5 s) and shifts (1 s) to facilitate comparison. We plot time-resolved ISC at the center of each temporal window. This means significant ISC implicates activity from ±2.5 s around each time point. Because all participants experienced the EEG block first and the CB block

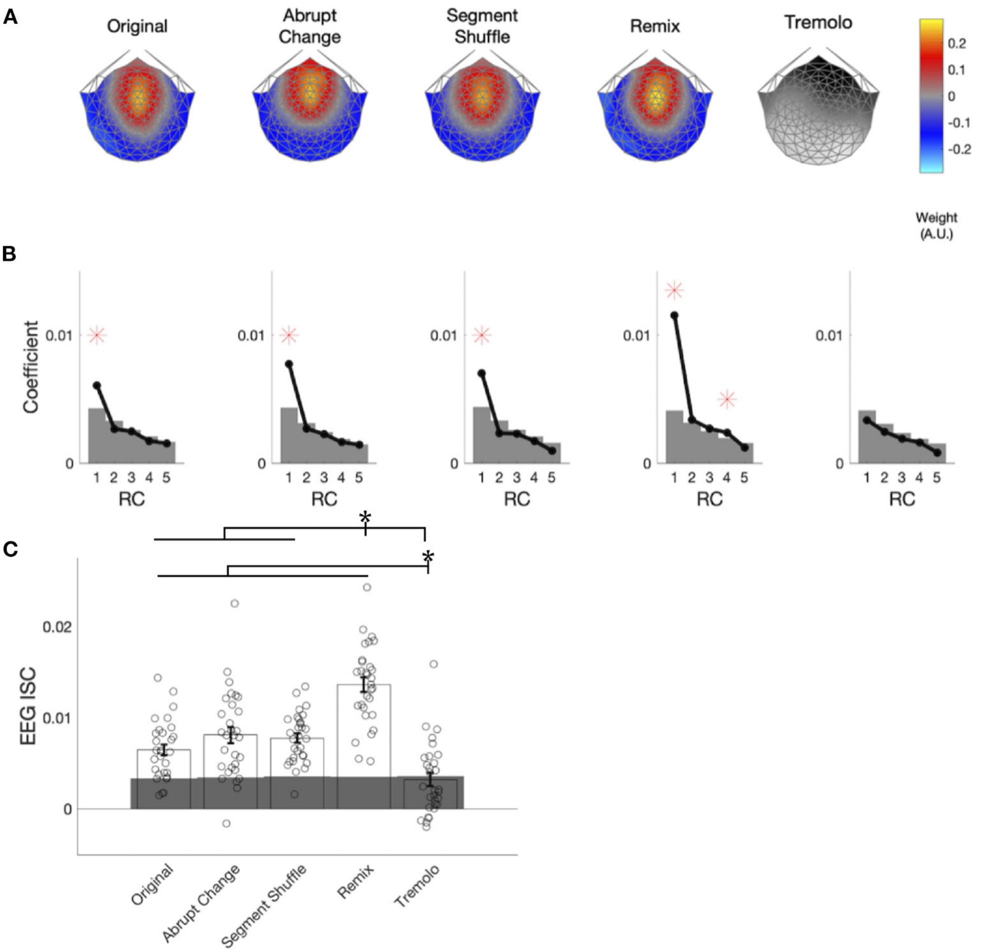

FIGURE 5 | EEG components, coefficients, and aggregate ISC. **(A)** Spatial filter weights are visualized on a scalp model using forward-model projections. Maximally reliable components (RC1) exhibit consistent auditory topographies for all stimulus conditions except Tremolo. **(B)** Spatial filter eigenvalues serve as component coefficients. Significant coefficients are marked with red asterisks and significance thresholds; gray areas denote the 95th percentile of the null distribution. RC1 is statistically significant for all conditions except Tremolo. **(C)** ISC was computed over the entire duration of each stimulus. Remix elicited significantly higher ISC than all the other conditions, and Tremolo elicited significantly lower ISC than all other conditions. Individual participants' EEG ISC values are denoted with dots. Gray areas denote the 95th percentile of the null distribution. Asterisks denote significance of $p < 0.05$.

second, differences between the two could be due to repeated exposure (Madsen et al., 2019). In addition, although there is precedent in fMRI ISC research to remove the initial seconds of participants' responses to avoid including an onset response (Wilson et al., 2008), we decided to include the responses for the entire stimulus duration because even these early responses varied by stimulus condition.

Responses to the Original stimulus show small but significant ISC peaks in the EEG data (permutation test $p < 0.05$, uncorrected, see Methods; time-varying effect sizes are given in **Supplementary Figures S3**, **S4**), with statistically significant ISC in 16.9% of the time windows (**Table 1**). The largest ISC peaks appear around the approximate start times of phasing sections, or shortly thereafter. Each of the phasing section onsets (marked in **Figure 7A** with dotted lines) is accompanied by a significant peak with the exception of the third phasing section (which may have a perceptually smoother transition than the other phasing sections). While phasing elicits ISC peaks relatively consistently, in-phase sections fail to correspond to any significant ISC peaks. Both EEG and CB ISC also contain a significant peak at the start of the excerpt. In the time-resolved CB ISC data, only a handful of small peaks occur above the significance threshold after the initial drop; they seem unrelated to phasing and in-phase musical events (peaks one and four fall in-phase sections, peaks two, three, and five fall in phasing sections), and only 4.7% of the ISC values are significant (**Table 1**). In contrast with phasing sections eliciting consistent peaks in the EEG ISC data, the CB mean data shows an increase in mean engagement rating after the start of each in-phase section. There also appears to be a slight decrease across the length of the stimulus.

EEG ISC data for the Abrupt Change condition shows significant peaks within seconds of the in-phase shifts (shifts

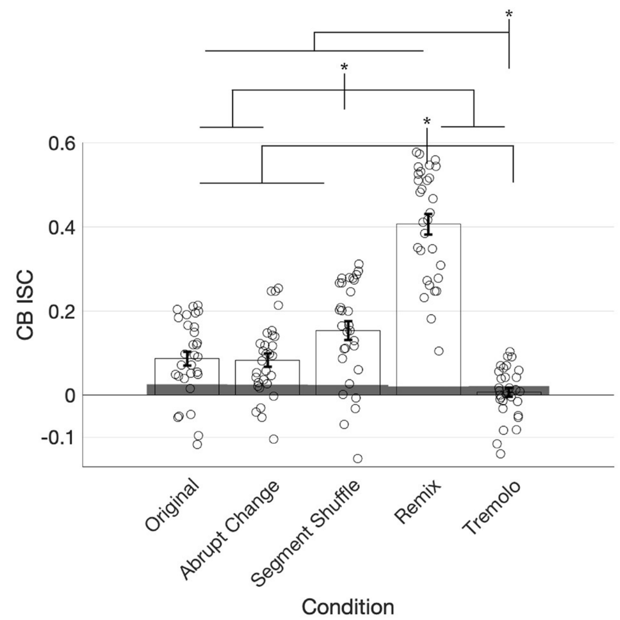

FIGURE 6 | ISC of continuous behavioral (CB) reports of engagement for each condition with individual participant data and standard error of the mean plotted. Shaded gray regions denote the 95th percentile of the null distribution. Remix elicited significantly higher ISC than all the other conditions and Tremolo elicited significantly lower ISC than all the other conditions. Segment Shuffle also differs significantly from all other conditions. Asterisks denote significance of $p < 0.05$.

number two, three, five and six as marked in solid lines in **Figure 7B**; (18.6% of ISC values are significant; see **Table 1**).[6] In contrast with the Original condition, in the Abrupt Change condition, where in-phase sections begin suddenly, they seem to elicit ISC peaks in the EEG data. The other small significance peaks in the EEG data come between in-phase changes, perhaps as participants anticipate stimulus alterations during the long stretches of unchanging material (perhaps something like the hazard function between warning and imperative stimuli (Tecce, 1972; Nobre et al., 2007). After an initial descent, the CB ISC data shows significant peaks around the first two and final two in-phase changes (percentage of significant time-resolved CB ISCs = 7.0%; see **Table 1**). The other two significant peaks appear between in-phase changes, perhaps related to the effect noted above. As in the Original condition, time-resolved CB mean data shows slight increases in engagement ratings after all six abrupt changes and an overall decline in engagement.

The perennially unpredictable changes in Segment Shuffle were met with frequent, small bursts of significant ISC correlations in the EEG data (**Figure 7C**; 15.9% significant ISC values; see **Table 1**). Comparing EEG and CB ISC time courses reveals unreliable alignment: After the initial drop in CB data, eight significant peak bursts unfold; about half of them align with EEG peaks (see peaks around time 1:30 and 3:05) while the other half do not (see peaks around time 0:15 and 2:30). CB means show small bumps in engagement ratings in the midst of a long-term downward trend (percentage of significant time-resolved CB ISCs = 10.3%; see **Table 1**).

Time-resolved ISCs for the Remix condition give ample opportunity to correlate peaks with musical events, with statistically significant EEG ISC in 45.6% time windows and significant CB ISC in 25.9% of time windows (**Table 1**). We selected the coded events in **Figure 7D** based on moments in the work that we deemed most musically salient (see **Supplementary Table S1** for the timings and descriptions of all 19 events). Note that not all of these events aligned with ISC peaks, but here we discuss some that did. After a sample from *Piano Phase* is presented for the first few seconds of Remix, a dramatic drum machine attack builds into simultaneous entrances for a synth countermelody and marimba riff (0:06). This build up and entrance align with the first and largest peak in the EEG data. The second peak in the EEG data comes at what might be the most dramatic moment in the piece, a beat drop anticipated with a drum machine lick (0:44). Note the potentially related peak in the CB ISC data following this event. But ISC peaks are not always elicited in both EEG and CB data. For example, the neighboring musical moments around minute 2:00 arise from a sudden dropping out of the percussion for a few seconds (2:01), leaving only a low, meandering synth line and a *Piano Phase* sample until the percussion reenters (2:04). This double event seems associated with an EEG ISC peak but no significant CB activity. A similar compositional technique plays out before minute 3:00. Two coded lines before that time (2:36), all instruments drop out except for the *Piano Phase* sample. It goes on, unchanging, until lush pitched percussion (a marimba) and additional synth lines enter at 2:50 (the line just before minute 3:00 in **Figure 7D**). The ISC peaks in both the EEG and CB data anticipate the reentry of additional instrumental lines, possibly in line with the previously mentioned hazard function: an anticipation that something must be coming given the static situation.

We did not expect any significant EEG ISC peaks for Tremolo, with its static, stark content. We see only occasional, small peaks above significance in the EEG ISC, an initial pair of significant points in the CB ISC, and a low and relatively unchanging CB mean (**Figure 7E**; percentage of significant time-resolved EEG ISCs = 7.4%; percentage of significant time-resolved CB ISCs = 1.0%; see **Table 1**). We also note that in contrast to the other stimulus conditions, the time-resolved EEG ISC for this condition does not include a significant peak at the beginning of the excerpt. However, similar to the control condition in Kaneshiro et al. (2020), this RC1 differs in topography from the other conditions (**Figure 5A**) and is not statistically significant (**Figure 5B**), making for uneven comparison between the Tremolo EEG ISC time course and the EEG responses to the other stimuli.

Comparing the present percentages of significant time-resolved ISCs for EEG data in RC1 with those reported by Kaneshiro et al. (2021) shows that our highest EEG ISC (for Remix) eclipses their finding of 37% (in response to Elgar's cello concerto); our Original, Abrupt Change, and Segment Shuffle stimuli elicit higher percentages of significant ISC than their control condition (an envelope-scaled but otherwise temporally unstructured manipulation, 8%); and our Tremolo condition

[6] Smaller peaks are easiest to see in the online version of the paper.

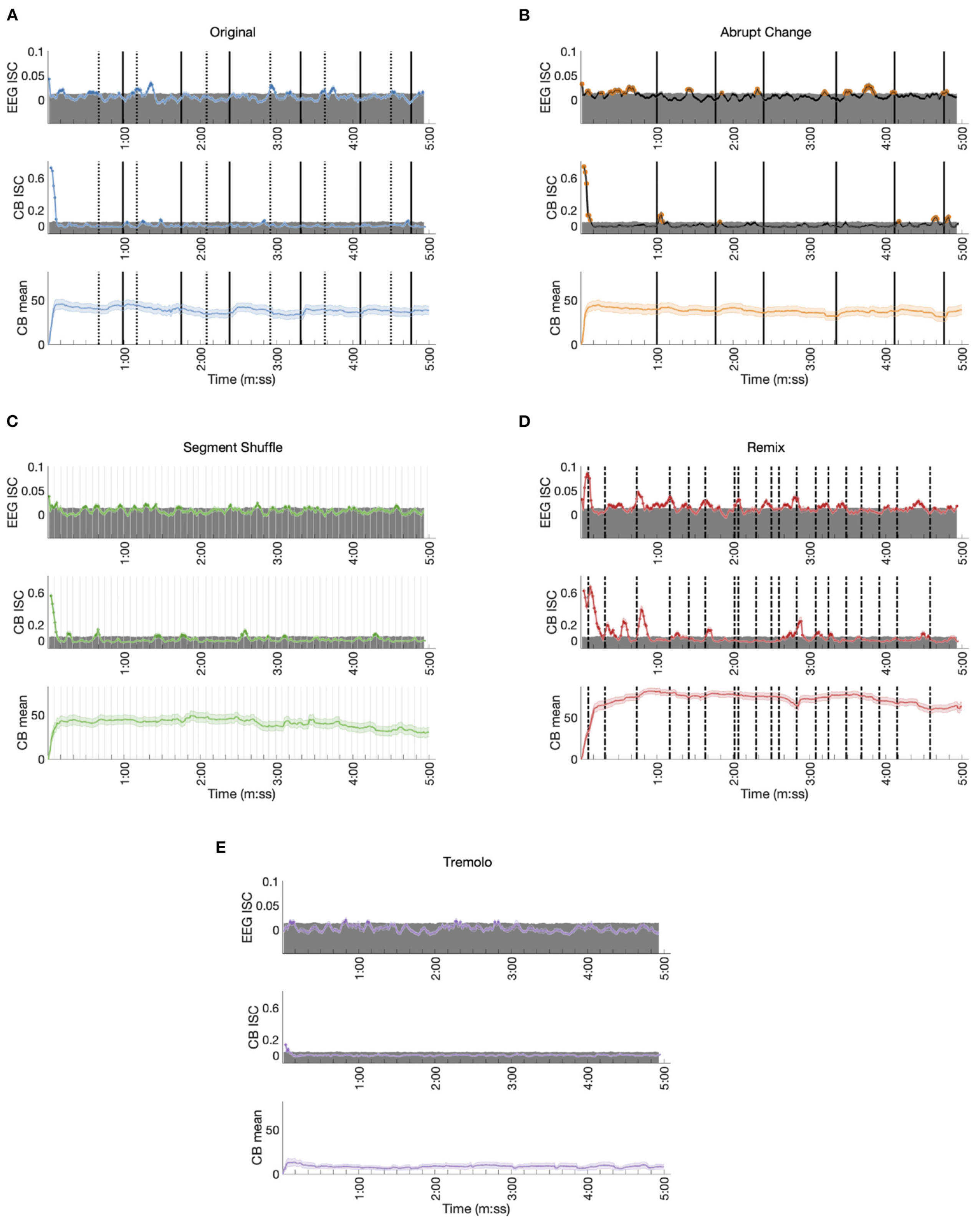

FIGURE 7 | Time-resolved EEG ISC, CB ISC, and CB means for each condition. The top of each shaded gray region represents the 95th percentile of the corresponding null distribution. **(A)** Original: Dotted lines mark the start of phasing sections, solid lines mark the start of in-phase sections. **(B)** Abrupt Change: Solid lines mark the start of each new in-phase section. **(C)** Segment Shuffle: Light gray lines mark the start of each new segment. **(D)** Remix: Dashed lines mark musical events expected to be significant to listeners. **(E)** Tremolo.

TABLE 1 | For each stimulus: the percentage of statistically significant EEG and CB time-resolved ISC windows, description of statistically significant EEG and CB peaks and CB mean changes, and qualitative assessment of alignment between time-resolved EEG and CB ISC and time-resolved EEG ISC and CB mean.

Stimulus	% sig. EEG ISC	% sig. CB ISC	Desc. of sig. EEG ISC peaks	Desc. of sig. CB ISC peaks	Desc. of CB mean	Alignment: EEG and CB ISC	Alignment: EEG ISC and CB mean
Original	16.9	4.7	Most phasing sections	Unrelated to phasing	increase after each in-phase section, general decrease over time	Weak	Weak
Abrupt change	18.6	7.0	Most in-phase shifts	Most in-phase shifts	Increase after each in-phase section, general decrease over time	Moderate	Moderate
Segment shuffle	15.9	10.3	After many shifts	After many shifts	Infrequent increases after shifts, general decrease over time	Moderate	Moderate
Remix	45.6	26.0	After most musical events	After many musical events	After many musical events	Strong	Strong
Tremolo	7.4	1.0	Infrequent and small	Only at opening	Infrequent and small	Weak	Weak

approximates the percentage found for their control condition. Even the present Remix stimulus elicits a lower percentage of significant ISC windows, however, than RC1 ISC reported by Dmochowski et al. (2012) during film viewing, where over 50% of time windows contained significant ISC.

4. DISCUSSION

We used inter-subject correlation (ISC) as a measure of engagement with Steve Reich's *Piano Phase* and manipulated and remixed versions of the work. We expected the phasing process at the heart of *Piano Phase* to drive electroencephalographic (EEG) and continuous behavioral (CB) ISC. At the overall-level, we found no statistically significant differences between the EEG ISC for the original work and our phasing-related manipulations (Abrupt Change and Segment Shuffle). At the time-resolved level, however, we noted the impact of phasing and in-phase sections in the confluence of the start of phasing sections in the Original with significant EEG ISC and CB mean activity, in-phase sections in Abrupt Change with significant CB ISC and EEG ISC, and Segment Shuffle shifts with significant EEG ISC and CB mean activity. At the overall level, we found that Original, Abrupt Change, and Segment Shuffle had significantly higher EEG ISC levels than Tremolo (the extremely repetitive manipulation). The remixed version, more related to popular music, resulted in the highest ISC. Overall CB ISC results were similar, but Segment Shuffle had significantly higher ISC than Original, Abrupt Change, and Tremolo, and significantly lower than Remix. From this overall stance, EEG and CB ISC values for Original, Abrupt Change, and Segment Shuffle generally align with participants' behavioral ratings (with the single exception of ratings for "well ordered"). In general, we found alignment between behavioral and neural measures of engagement.[7]

[7]See **Supplementary Figure S5** for the correlations between mean CB and engagement rating values.

In addition to the overall alignment, we also noticed differences between EEG ISC and CB measures at the time-resolved level (We note that because EEG data collection always preceded CB data collection, it is possible that order effects play a role: perhaps participants focus more on lower-level features in the initial hearing when compared with subsequent hearings). Phasing sections in the Original, with their many and unpredictable onsets, elicited neural ISC but failed to generate significant CB ISC. Participants had higher CB mean ratings at the start of in-phase sections, perhaps returning attention to the stimulus when it emerged from complex phasing sections back toward unison clarity (in-phase sections)—a phenomenon not seen in the time-resolved EEG ISC data. We also noted the mix of alignment and independence between neural and behavioral measures in Abrupt Change, Segment Shuffle, and Remix, again with some significant EEG ISC unaccompanied by behavioral ISC. One way to understand the differences between EEG and CB measures is to connect them with the previous ISC finding that frequently and unexpectedly changing stimuli seem capable of driving correlated neural responses, perhaps pointing to a relationship between ISC and something like the orienting response (voluntary and automatic neural and behavioral responses to novel information, Sokolov, 1990; Sokolov et al., 2002). Dmochowski and colleagues reported relationships between EEG ISC and population ratings of Super Bowl commercials and found that an audio-visual stimulus with "repeated and jarring scene cuts" associated with "relatively strong neural reliability" drove ISC measures above population ratings (this stimulus was ultimately excluded in order to maintain stronger predictive performance of population ratings; Dmochowski et al., 2014, **Supplementary Note 3**). Ki et al. (2016) found that narratives in a foreign language elicited higher ISC than a narrative in the participants' native language. Using two films as stimuli, Poulsen et al. (2017) reported a significant correlation between ISC and average luminance difference, suggesting that ISC for their primary component of interest "may indeed be driven by low-level visual evoked responses" (p. 5). Finally, Kaneshiro et al. (2020) noted that a stimulus manipulation in which measures of music were randomly re-ordered (and thus musically less meaningful but more surprising) resulted in higher EEG ISC than intact music. If EEG ISC is heavily influenced by such contrastive changes in acoustic features, perhaps some of them "break through" to the behavioral level and others do not. While this could explain the differences in time-resolved EEG ISC and CB ISC findings, we argue that it does not point to a break between ISC and engagement. Rather, the strong *overall* similarities between our EEG, CB, and question response data suggests that such contrastive changes may heavily contribute to participants' feeling of engagement and narrow the type of engagement that ISC indexes. Future studies could also work to remove the possible influence of order effects by alternating or randomizing the sequence of blocks in which EEG and CB data are collected (but without biasing participants about the nature of the experiment in the instructions for the CB block when it comes first).

A previous study reported decreased ISC (i.e., lower correlations of brain data between participants) when familiar music stimuli are repeated (Madsen et al., 2019). The authors argued that because EEG ISC tracks rapid responses to stimuli, it likely indexes more stimulus-driven responses, as opposed to cognitive elaborations that likely occur at longer temporal durations. One explanation of our findings is that highly repetitive music (such as minimalism and Reich's phasing process) will elicit lower engagement, and thus, lower ISC values. Our Tremolo condition offers an extreme test and seeming confirmation of this hypothesis. More varied stimuli still featuring high levels of repetition—i.e., Original, Abrupt Change, and Segment Shuffle—yielded higher EEG and CB ISC than Tremolo. Remix's frequently changing musical parameters resulted in rather high ISC. One could argue that the more repetitive the stimulus was, the less interesting it may have been, and thus, less engaging.

Yet, as some have pointed out (Madsen et al., 2019; Kaneshiro et al., 2021), ISC measures *shared* engagement. Put another way, ISC can only pick up on forms of engagement that unfold similarly between multiple participants. Other types of engagement, be they idiosyncratic, or only shared by a few participants, would not show up. The strongest empirical evidence for such a view of our current data comes from individual CB responses (**Supplementary Figure S2**). In said data, at least two participants (the highest two lines of raw data) show patterns of high and dynamic engagement in the Tremolo condition, a condition where we predicted and found very low EEG and CB ISC. Further supporting the notion of idiosyncratic engagement patterns is the fact that these two participants did not have unusually high EEG ISC responses,[8] nor is their behavior explainable via musical background: one had a musical background and one did not. Previous theoretical and empirical work bolsters the idea of multiple styles of engagement. The transportation and cognitive elaboration framework for engagement (Green and Brock, 2000) posits two strands of engagement: transportation, where audience members are locked into the content of the art object, tracking details; and cognitive elaboration, where an observer or listener is prompted by the stimulus to reflect on the artwork, drawing connections with other experiences and other knowledge. David Huron's (2002) listening styles offer even more potential types or modes of engagement, ranging from mentally singing along to mentally reminiscing about musically associated memories. ISC would be unlikely to pick up on these listening styles equally, and it would be odd if a single measure could.

Some cognitive science of music scholars have argued that repetition could augment individualized, internally focused experiences by gradually demanding less processing power and attention over time. Such a process may open up reflective space for listeners (Margulis, 2012, 2014).[9] (This is in contrast with the type of engagement that might occur during dramatic moments like the beat drop in the first minute of Remix.) In *Piano Phase*, such a trajectory could be cyclical, with listeners drifting off into individual experience and tugged back into the details of the ongoing external stimulus events by changes in the music. If enough participants were drawn back to the stimulus details at the same time, neural responses could become sufficiently correlated to produce an ISC peak (perhaps something like the peak around minute 3:00 in the Original EEG ISC time-resolved data). In this line of thought, musicologists and music theorists have noted the long trajectories of expectation formation in minimalist music such as Reich's. Cadences in tonal music (i.e., the ends of phrases) often drive and ultimately resolve such expectations (what key are we in? where are we in the phrase? what harmonic and melodic activity is likely to come next?). Cadences and their accompanying harmonic trajectories are also present in minimalism but often in a stretched out form (Fink, 1996). Some listeners may lose interest along the way, while others may be drawn into granular detail and vary in what layer of granularity they are caught up in. Perhaps most move from state to state: For examples of the former situation, two participants in the present study noted that the Tremolo stimulus was difficult to listen to—"intense" in the words of one. Another participant stated that to them the stimuli were "all the same but with different layers."

One potential route forward for this line of research would be to use the current results to hypothesize quantifiable musical features that may be driving time-resolved EEG and CB ISC peaks (Alluri et al., 2012). This could lead to a fruitful exploration of how far specific compositional techniques such

[8]We note the caveat that the RC1 for Tremolo was not statistically significant and did not appear to be auditory in nature.

[9]Margulis's account could be framed as a type of cognitive elaboration, where listeners are able to chunk or process ever longer swaths of the highly repetitive stimulus: "...the horizon of involvement widens with additional exposures, so that the music doesn't seem to be coming at the listener in small bits, but rather laying out broader spans for consideration" (Margulis, 2014, p. 9). In the original Reich piece, this process could happen within each in-phase unit: the musical material is short enough that even within a few repetitions, some listeners may "zoom out." If a listener is engaged in detailed listening, phasing sections offer little exact repetition, but, if engaged in a temporally broader manner, the pattern of in-phase to phasing could similarly encourage gradual zooming out as the listener groks the structure of the composition.

as phasing generalize into other repetition-based techniques like Philip Glass's additive and subtractive modular technique (York, 1981) or electronic dance music (Solberg and Dibben, 2019). It also reveals new layers of detail for scholars who work on the repertoire—a testing ground for theories of how the music can function for individuals. On that front, this study suggests important follow up research. For instance, alpha activity is thought to reflect meditative states (Lee et al., 2018). Therefore, alternative approaches to analyzing the EEG data—e.g., by assessing alpha power, or correlation thereof—may prove more appropriate measures for indexing listener states while listening to minimalist music. We might hypothesize that when participants are diversely engaged with a stimulus, a similar psychological state may be shared—but one that is better indexed by other means than EEG ISC. As alpha activity has been shown to index multiple states in varying locations (Nunez et al., 2001; Başar, 2012; Lee et al., 2018), future research could also include interviews with music listeners to provide complementary insights into inter-individual differences in music listening. Such mixed-methods work could reveal patterns for calm vs. bored listeners or time periods of boredom, interest, and relaxation. While we limited ourselves to exploring a general type of engagement, future research could work to distinguish between types of engagement and even diverse forms on non-engagement (distinguishing boredom from confusion, for example).

Our hypothesis that the core compositional feature of Reich's *Piano Phase* would differentially drive engagement (measured using inter-subject correlation, ISC) between an excerpt of that work and conditions that manipulated the phasing process was consistent with time-resolved EEG and behavioral ISC data which showed that the timing of key musical moments often corresponded with these measures of engagement. Overall, our participants' neural, continuous behavioral, and question responses show that a popular-music style Remix of Reich's *Piano Phase* was more engaging than the original work and two conditions that manipulated its core compositional technique. In turn, these three stimuli were more engaging than an intensely repetitive condition that featured no compositional changes. Although research continues to unravel the specifics of what EEG ISC measures when participants are presented with musical stimuli, we found that participants EEG ISC, CB ISC, and question responses broadly align: evidence that *some* type of engagement is tracked by EEG ISC. We propose that the nature of the engagement indexed by EEG ISC with musical stimuli seems to be a mixture of attention, acoustic features, and the level of contrastive change in those acoustic features.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://purl.stanford.edu/kt396gb0630.

ETHICS STATEMENT

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

AUTHOR CONTRIBUTIONS

TD, DN, JB, and BK designed the experiment. TD, NG, JB, and BK created the stimuli. DN and BK created participant interfaces for the experiment. TD and DN collected the data. TD, DN, and BK curated the data. TD, JD, and BK specified formal and statistical analyses. TD and BK analyzed the data, created the visualizations, and drafted the original manuscript. DN, NG, JD, and JB reviewed and edited the manuscript. JB and BK supervised the research. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Wallenberg Network Initiative: Culture, Brain Learning (TD, DN, JB, and BK); the Patrick Suppes Gift Fund (DN and BK); the Roberta Bowman Denning Fund for Humanities and Technology (JB and BK); the Army Research Laboratory W911-NF-10-2-0022 (JD); the Stanford Humanities Center (TD); and a Ric Weiland Graduate Fellowship (TD). Open access publication fees were paid by the Stanford Center for Computer Research in Music and Acoustics.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.702067/full#supplementary-material

REFERENCES

Abrams, D. A., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D. J., et al. (2013). Inter-subject synchronization of brain responses during natural music listening. *Eur. J. Neurosci.* 37, 1458–1469. doi: 10.1111/ejn.12173

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., and Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *Neuroimage* 59, 3677–3689. doi: 10.1016/j.neuroimage.2011.11.019

Başar, E. (2012). A review of alpha activity in integrative brain function: fundamental physiology, sensory coding, cognition and pathology. *Int. J. Psychophysiol.* 86, 1–24. doi: 10.1016/j.ijpsycho.2012.07.002

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., et al. (2012). *Package 'lme4'*. Vienna: CRAN. R Foundation for Statistical Computing.

Becker, J. (2004). *Deep Listeners: Music, Emotion, and Trancing, Vol. 1*. Bloomington, IN: Indiana University Press.

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat*. 29, 1165–1188. doi: 10.1214/aos/1013699998

Brainard, D. (1997). The psychophysics toolbox. *Spat. Vis*. 10, 433–436. doi: 10.1163/156856897X00357

Cameron, D., Potter, K., Wiggins, G., and Pearce, M. (2017). Perception of rhythmic similarity is asymmetrical, and is influenced by musical training, expressive performance, and musical context. *Timing Time Percept*. 5, 211–227. doi: 10.1163/22134468-00002085

Cameron, D. J., Zioga, I., Lindsen, J. P., Pearce, M. T., Wiggins, G. A., Potter, K., et al. (2019). Neural entrainment is associated with subjective groove and complexity for performed but not mechanical musical rhythms. *Exp. Brain Res*. 237, 1981–1991. doi: 10.1007/s00221-019-05557-4

Cohen, S. S., Madsen, J., Touchan, G., Robles, D., Lima, S. F., Henin, S., et al. (2018). Neural engagement with online educational videos predicts learning performance for individual students. *Neurobiol. Learn. Mem*. 155, 60–64. doi: 10.1016/j.nlm.2018.06.011

Cohen, S. S., and Parra, L. C. (2016). Memorable audiovisual narratives synchronize sensory and supramodal neural responses. *eNeuro* 3:ENEURO.0203-16.2016. doi: 10.1523/ENEURO.0203-16.2016

Cohn, R. (1992). Transpositional combination of beat-class sets in steve reich's phase-shifting music. *Perspect. New Music* 30, 146–177. doi: 10.2307/3090631

Dauer, T. (2020). *The Varieties of Minimalist Experience: The Roles of Psychological States in the Reception of American Minimalism During the Long Sixties* (Ph.D. thesis). Stanford University.

Dauer, T., Nerness, B., and Fujioka, T. (2020). Predictability of higher-order temporal structure of musical stimuli is associated with auditory evoked response. *Int. J. Psychophysiol*. 153, 53–64. doi: 10.1016/j.ijpsycho.2020.04.002

Dmochowski, J. P., Bezdek, M. A., Abelson, B. P., Johnson, J. S., Schumacher, E. H., and Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nat. Commun*. 5, 1–9. doi: 10.1038/ncomms5567

Dmochowski, J. P., Greaves, A. S., and Norcia, A. M. (2015). Maximally reliable spatial filtering of steady state visual evoked potentials. *Neuroimage* 109, 63–72. doi: 10.1016/j.neuroimage.2014.12.078

Dmochowski, J. P., Sajda, P., Dias, J., and Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention–a possible marker of engagement? *Front. Hum. Neurosci*. 6:112. doi: 10.3389/fnhum.2012.00112

Epstein, P. (1986). Pattern structure and process in steve reich's" piano phase". *Musical Q*. 72, 494–502. doi: 10.1093/mq/LXXII.4.494

Farbood, M. M., Heeger, D. J., Marcus, G., Hasson, U., and Lerner, Y. (2015). The neural processing of hierarchical structure in music and speech at different timescales. *Front. Neurosci*. 9:157. doi: 10.3389/fnins.2015.00157

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

Fink, R. (2005). *Repeating Ourselves*. Berkeley: University of California Press.

Fink, R. W. (1996). "*Arrows of Desire": Long-Range Linear Structure and the Transformation of Musical Energy* (Ph.D. thesis). University of California, Berkeley.

Green, M. C., and Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *J. Pers. Soc. Psychol*. 79, 701. doi: 10.1037/0022-3514.79.5.701

Hartenberger, R. (2016). *Performance Practice in the Music of Steve Reich*. Cambridge: Cambridge University Press.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. doi: 10.1126/science.1089506

Henahan, D. (1970). Steve Reich presents a program of pulse music at Guggenheim. *The New York Times*.

Huron, D. (2002). "Listening styles and listening strategies," in *Society for Music Theory Annual Conference*. (Columbus, Oh).

Ihaka, R., and Gentleman, R. (1996). R: A language for data analysis and graphics. *J. Computat. Graph. Stat*. 5, 299–314. doi: 10.1080/10618600.1996.10474713

Jung, T.-P., Humphries, C., Lee, T.-W., Makeig, S., McKeown, M. J., Iragui, V., et al. (1998). "Extended ICA removes artifacts from electroencephalographic recordings," in *Advances in Neural Information Processing Systems* (Cambridge), 894–900.

Kaneshiro, B., Nguyen, D. T., Norcia, A. M., Dmochowski, J. P., and Berger, J. (2020). Natural music evokes correlated EEG responses reflecting temporal structure and beat. *Neuroimage* 214:116559. doi: 10.1016/j.neuroimage.2020.116559

Kaneshiro, B., Nguyen, D. T., Norcia, A. M., Dmochowski, J. P., and Berger, J. (2021). Inter-subject EEG correlation reflects time-varying engagement with natural music. *bioRxiv*. doi: 10.1101/2021.04.14.439913

Kaneshiro, B. B. (2016). *Toward an Objective Neurophysiological Measure of Musical Engagement* (Ph.D. thesis). Stanford University.

Ki, J. J., Kelly, S. P., and Parra, L. C. (2016). Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. *J. Neurosci*. 36, 3092–3101. doi: 10.1523/JNEUROSCI.2942-15.2016

Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V., and Stefanovska, A. (2018). Surrogate data for hypothesis testing of physical systems. *Phys. Rep*. 748, 1–60. doi: 10.1016/j.physrep.2018.06.001

Lee, D. J., Kulubya, E., Goldin, P., Goodarzi, A., and Girgis, F. (2018). Review of the neural oscillations underlying meditation. *Front. Neurosci*. 12:178. doi: 10.3389/fnins.2018.00178

Lerner, Y., Honey, C. J., Silbert, L. J., and Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci*. 31, 2906–2915. doi: 10.1523/JNEUROSCI.3684-10.2011

Losorelli, S., Nguyen, D. T., Dmochowski, J. P., and Kaneshiro, B. (2017). "NMED-T: a tempo-focused dataset of cortical and behavioral responses to naturalistic music," in *Proceedings of the 18th International Society for Music Information Retrieval Conference* (Suzhou), 339–346.

Madsen, J., Margulis, E. H., Simchy-Gross, R., and Parra, L. C. (2019). Music synchronizes brainwaves across listeners with strong effects of repetition, familiarity and training. *Sci. Rep*. 9, 1–8. doi: 10.1038/s41598-019-40254-w

Margulis, E. H. (2012). Musical repetition detection across multiple exposures. *Music Percept*. 29, 377–385. doi: 10.1525/mp.2012.29.4.377

Margulis, E. H. (2014). *On Repeat: How Music Plays the Mind*. Oxford: Oxford University Press.

Mertens, W. (1983). *American Minimal Music: La Monte Young, Terry Riley, Steve Reich, Philip Glass*. London: Kahn & Averill.

Nakagawa, S., and Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol. Rev*. 82, 591–605. doi: 10.1111/j.1469-185X.2007.00027.x

Nobre, A. C., Correa, A., and Coull, J. T. (2007). The hazards of time. *Curr. Opin. Neurobiol*. 17, 465–470. doi: 10.1016/j.conb.2007.07.006

Norman, G. (2010). Likert scales, levels of measurement and the "laws"9D of statistics. *Adv. Health Sci. Educ*. 15, 625–632. doi: 10.1007/s10459-010-9222-y

Nunez, P. L., Wingeier, B. M., and Silberstein, R. B. (2001). Spatial-temporal structures of human alpha rhythms: Theory, microcurrent sources, multiscale measurements, and global binding of local networks. *Hum. Brain Mapp*. 13, 125–164. doi: 10.1002/hbm.1030

Olsen, K. N., Dean, R. T., and Stevens, C. J. (2014). A continuous measure of musical engagement contributes to prediction of perceived arousal and valence. *Psychomusicology* 24, 147. doi: 10.1037/pmu0000044

Parra, L. C., Spence, C. D., Gerson, A. D., and Sajda, P. (2005). Recipes for the linear analysis of EEG. *Neuroimage* 28, 326–341. doi: 10.1016/j.neuroimage.2005.05.032

Potter, K., Wiggins, G. A., and Pearce, M. T. (2007). Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music. *Musicae Sci*. 11, 295–324. doi: 10.1177/102986490701100207

Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., and Hansen, L. K. (2017). EEG in the classroom: Synchronised neural recordings during video presentation. *Sci. Rep*. 7, 1–9. doi: 10.1038/srep43916

Prichard, D., and Theiler, J. (1994). Generating surrogate data for time series with several simultaneously measured variables. *Phys. Rev. Lett*. 73, 951–954. doi: 10.1103/PhysRevLett.73.951

Quinn, I. (2006). Minimal challenges: Process music and the uses of formalist analysis. *Contemporary Music Rev*. 25, 283–294. doi: 10.1080/07494460600726537

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reich, S. (1987). *Reich "Early Works"*. New York, NY: Double Edge.

Reich, S. (1999). *Reich Remixed*. New York, NY: Arthrob/Nonesuch.
Reich, S. (2009). *Writings on Music, 1965-2000*. Oxford: Oxford University Press.
Rockwell, J. (1973). Records: Roiling work: Reich's 'Four Organs,' which created a stir at concert, is on Angel Disk Hyman's Piano. *New York Times*.
Schubert, E., Vincs, K., and Stevens, C. J. (2013). Identifying regions of good agreement among responders in engagement with a piece of live dance. *Empir. Stud. Arts* 31, 1–20. doi: 10.2190/EM.31.1.a
Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., et al. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun*. 7, 1–13. doi: 10.1038/ncomms12141
Sokolov, E. (1990). The orienting response, and future directions of its development. *Pavlov. J. Biol. Sci*. 25, 142–150.
Sokolov, E. N., Spinks, J. A., Näätänen, R., and Lyytinen, H. (2002). *The Orienting Response in Information Processing*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
Solberg, R. T., and Dibben, N. (2019). Peak experiences with electronic dance music: Subjective experiences, physiological responses, and musical characteristics of the break routine. *Music Percept*. 36, 371–389. doi: 10.1525/mp.2019.36.4.371
Strongin, T. (1969). Is timelessness out of style? *New York Times* 21.
Tecce, J. J. (1972). Contingent negative variation (CNV) and psychological processes in man. *Psychol. Bull*. 77, 73. doi: 10.1037/h0032177
Tucker, D. M. (1993). Spatial sampling of head electrical fields: the geodesic sensor net. *Electroencephalogr. Clin. Neurophysiol*. 87, 154–163. doi: 10.1016/0013-4694(93)90121-B
Wilson, S. M., Molnar-Szakacs, I., and Iacoboni, M. (2008). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cereb. Cortex* 18, 230–242. doi: 10.1093/cercor/bhm049
Winter, B. (2013). Linear models and linear mixed effects models in r with linguistic applications. *arXiv preprint* arXiv:1308.5499.
York, W. (1981). Form and process in two pages of philip glass. *Sonus* 1, 28–50.

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

CONCEPTUAL ANALYSIS
published: 20 May 2021
doi: 10.3389/fpsyg.2021.672949

Perceiving Sound Objects in the *Musique Concrète*

Rolf Inge Godøy[1,2]*

[1] Department of Musicology, University of Oslo, Oslo, Norway, [2] RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway

In the late 1940s and early 1950s, there emerged a radically new kind of music based on recorded environmental sounds instead of sounds of traditional Western musical instruments. Centered in Paris around the composer, music theorist, engineer, and writer Pierre Schaeffer, this became known as *musique concrète* because of its use of concrete recorded sound fragments, manifesting a departure from the abstract concepts and representations of Western music notation. Furthermore, the term *sound object* was used to denote our perceptual images of such fragments. Sound objects and their features became the focus of an extensive research effort on the perception and cognition of music in general, remarkably anticipating topics of more recent music psychology research. This sound object theory makes extensive use of metaphors, often related to motion shapes, something that can provide holistic representations of perceptually salient, but temporally distributed, features in different kinds of music.

Keywords: ***musique concrète***, **perception, sound object, typology, morphology**

OPEN ACCESS

Edited by:
Thomas James Lundy, Cuttlefish Arts, United States

Reviewed by:
Tim Ziemer, University of Bremen, Germany
Christian Grueny, Max Planck Institute for Empirical Aesthetics, Max Planck Society (MPG), Germany

***Correspondence:**
Rolf Inge Godøy
r.i.godoy@imv.uio.no

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

Received: *26 February 2021*
Accepted: *19 April 2021*
Published: *20 May 2021*

Citation:
Godøy RI (2021) Perceiving Sound Objects in the Musique Concrète. Front. Psychol. 12:672949. doi: 10.3389/fpsyg.2021.672949

INTRODUCTION

In parallel with the emergence of the *musique concrète*, Pierre Schaeffer and coworkers also dedicated much attention to the perceptual issues of this new kind of music, in particular to the so-called *sound objects* and their various features. Although based on recorded fragments of sound, typically in the 0.5-5 s duration range, sound objects are actually perceptual images of these sound fragments, i.e., images in our minds of sound fragments we listen to. These perceptual images are largely influenced by our individual attitudes and schemas, including our attentional focus during listening. Yet there was also an attempt to find some common features among individual experiences of sound objects by the use of metaphor labels. These labels, largely related to motion shapes such as *impulsive*, *iterative*, *sustained*, *rough*, *smooth*, etc., grew out of practical composition work in the *musique concrète* and ensuing discussions at the *Groupe de recherches musicales* in Paris.

This focus on sound objects and their perceptually salient features was, and still is, a remarkable development in music theory. It is top-down in the sense of taking the subjective (or what we could call experiential) features as point of departure for systematic exploration, all the time guided by the seemingly naïve method of numerous listening to any sound fragment and trying to depict what we are hearing. In being founded on perceived sound, and not on Western notation or other abstract paradigms, this theory emerged ahead of its time, and may from our present day vantage point be regarded as just as much a project of music psychology as of music theory.

The aim of the present paper is first of all to highlight the extraordinary role of the sound object focus in music theory, not only relevant for the *musique concrète,* but relevant well beyond that to many different musical genres and styles, because sound objects are inherently holistic and thus capable of conceptualizing temporally distributed features of music such as the various dynamic, textural, timbral, and expressive envelopes that has not be possible within the more traditional Western music theory frameworks. A follow-up aim here is to demonstrate how the metaphors of Schaeffer and co-workers may be related to the acoustic substrate of sound objects, now that we have readily available methods and technologies to explore such relationships. The empirical material in this paper is then the collection of metaphor labels and their corresponding sound examples in Schaeffer's theory, as manifest in the text and sound work *Solfège de l'objet sonore* (Schaeffer et al., 1998, here called *Solfège*), specifically the labels of the *typology* (the overall dynamic and pitch-related shapes of sound objects), and of the *morphology* (the various detail contents of sound objects).

The challenge in this paper is to explore how to bridge the gaps between the subjective labels and the acoustic features of the sound objects in research, in particular as these labels are applied across different sounds e.g., across that of a drumroll, a deep bassoon tone, and of a processed sound, all with prominent so-called *iterative* feature (cf. *Solfège* CD3, tracks 25, 26, and 27, see **Figure 1** below). The central question is then: is it possible to document the acoustic features of sound objects that constitute the basis for the subjective labels in the *Solfège*? In other words: can we correlate the subjective feature labels of the sound objects with their acoustic features, as was indeed the long-term project of musical research proposed by Schaeffer?

The scheme of this paper is to first present the reasoning for the sound object focus, and then to illustrate the correlation of metaphors and acoustics by going backward from the subjective feature labels to the acoustic features of some sound examples in the *Solfège*, and using available sound analysis tools (such as *Praat* and the *MIRtoolbox*), evaluate how these subjective feature labels actually relate to acoustic features. The hoped for outcome is to point out some correlations between subjective labels and their acoustical substrates in the *Solfège*, and contribute to future research bridging the gaps between subjective labels and acoustic features in music perception studies (be that behavioral or neurocognitive), as well as in practical sound design work.

To better understand this sound object focus, the next two sections will present some background material on the sources and context of the *Musique Concrète*. This will be followed by a presentation of the main challenges of the *Musique Concrète*, its object focus, listening ontologies, and object features, before sections on acoustic correlates, typology, morphology, and multidimensional modeling. Lastly, there will be a discussion of the gains and possible future developments of the sound object theory, in particular in view of readily available technology for analysis and creation.

Sources of the *Musique Concrète*

Besides being a composer and an engineer, Schaeffer was also a writer with numerous publications (including novels and plays), and he wrote an extensive account of the development, experiments, and thoughts, leading up to the *musique concrète* in his 1952 book, *A la recherche d'une musique concrète* Schaeffer, 1952, available in English translation as (Schaeffer, 2012). Part diary, part protocol, and with extensive discussions of the perceptual issues involved, this book is a testimony to Schaeffer's fascination with the aesthetics of concrete sounds. This book also includes accounts of the composing of early concrete works such as *Etude aux chemins de fer,* using sounds of locomotive engines and horns put together in a kind of score (Schaeffer, 2012, pp. 26-27).

As for the label "musique concrète," Schaffer wrote: "When in 1948 I suggested the term *musique concrète*, I intended, by this adjective, to express a *reversal* of the way musical work is done. Instead of notating musical ideas using the symbols of music theory, and leaving it to known instruments to realize them, the aim was to gather concrete sound, wherever it came from, and to abstract the musical values it potentially contained." (Schaeffer, 2017, p. 7). And: "I have coined the term Musique Concrete for this commitment to compose with materials taken from "given" experimental sound in order to emphasize our dependence, no longer on preconceived sound abstractions, but on sound fragments that exist in reality and that are considered as discrete and complete sound objects, even if and above all when they do not fit in with the elementary definitions of music theory." (Schaeffer, 2012, p. 12).

The terms *concrete* and *abstract* are recurrent in Schaeffer's writings, and they denote not only the crucial difference between recorded sound and Western music notation symbols, but also the difference between more open-ended concepts of musical sound and the relatively strict categorical schemes of Western notation, i.e., of pitch and duration. In the words of Schaeffer: "The adjective "abstract" is applied to ordinary music because it is initially conceived in the mind, then notated theoretically, and finally executed in an instrumental performance. As for "concrete" music, it is made up of preexisting elements, taken from any sound material, noise, or musical sound, then composed experimentally by direct montage, the result of a series of approximations, which finally gives form to the will to compose contained in rough drafts, without the help of an ordinary musical notation, which becomes impossible." (ibid, p. 25). It is also interesting that the conceptual apparatus and associated terminology for the sound object focus, was largely developed already in 1952, as can be seen in the section called "Twenty five initial words for a vocabulary," developed in collaboration with Abraham Moles (Schaeffer, 2012, pp. 194-221).

Context of the *Musique Concrète*

The concrete-abstract antinomies should also be seen on the background of Western musical modernism in the post-1945 area. Several aesthetic directions existed side by side, but there were some prominent directions that we see allusions to in the writings of Schaeffer. There is the so-called *Darmstadt School* of integral serialism fronted by Boulez, Stockhausen, and Nono (among others), and in spite of its proclamations of radical aesthetic ideas, it was also practicing a conventional Western musical framework of tempered tuning and discrete notation.

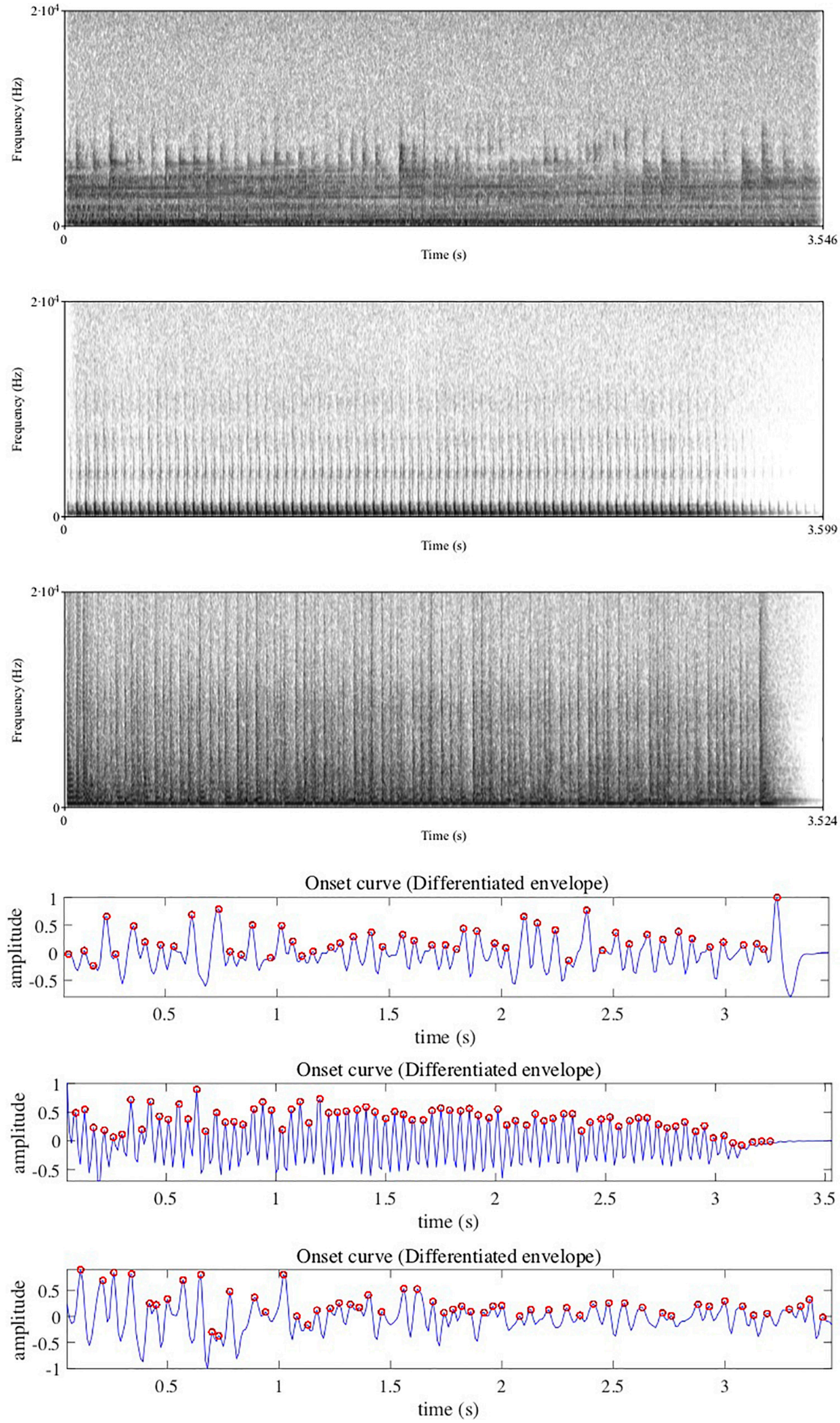

FIGURE 1 | Spectrograms (top part) and MIRtoolbox plotting (bottom part) of amplitude peaks (indicated with small red circles) in the sounds of a drumroll, a deep bassoon tone, and an electronic sound, from *Solfège* CD3, respectively, tracks 25, 26, and 27. These series of amplitude peaks, variably in the region between 15 Hz and 20 Hz, are all perceived as so-called *iterative* textures of the three sound objects in spite of their different origins, illustrating the principle of generic typological features across sounds with different origins. The MIRtoolbox plotting picking out the peaks from the continuous sound signal are based on a derivative function for peak finding in the MIRtoolbox, hence will be quite accurate in locating the peak points in the signal.

In the opinions of its critics (including Schaeffer), there was also a lack of focus on the perceptual outcomes of this music. But in a kind of middle position, we may find Xenakis with his schemes of formal and statistical distributions of musical sound, hence partially focused on the emergent sound object shapes and textures (Xenakis, 1992).

In parallel, there was a development of acoustics research, and Schaeffer was evidently quite well versed in this field, both by training as an engineer and by being familiar with mainstream publications in musical acoustics (Helmholtz, Stumpf, as well as Moles, with whom Schaeffer worked). What emerges from Schaeffer's discussion of acoustics topics is a recognition of the acoustic basis for musical sound combined with reservations about the limitations of acoustic theory in view of experienced sound features. Schaeffer basically claimed there was a non-linear relationship between acoustics and perception, manifest in what he called *anamorphosis*, sometimes rendered in English as "warping," implying a need for an empirical feature-by-feature mapping between signal and percept. This reservation in Schaeffer's writings toward relevant insights from acoustics, is something that Michel Chion ascribes to a kind of "scientific" idea in the 1950s and 1960s of the sound itself without taking into consideration the complexities and non-linearities of perception and the listeners intentionality (Chion, 1983, p. 30).

Between the polarities on the one hand, of inherited western notation concepts, and on the other hand, the more physicalist acoustical research, Schaeffer sought to establish a domain of research into perception of musical sound where the sound was considered concrete in the sense of not limited by the abstractions of Western music notation symbols. The core of this research was on sound objects, and as is the main focus of this paper, Schaeffer saw sound objects as an autonomous domain of research based on our subjective experiences, however with a long-term ambition to correlate these sound objects with acoustic features.

The principal source for Schaeffer's sound object research in this paper is his monumental *Traité des objets musicaux* (Schaeffer, 1966), now fortunately available in English (Schaeffer, 2017). A very useful introduction and overview of this work is in Michel Chion's *Guide des objects sonores* (Chion, 1983), also available in English (Chion, 2009). The mentioned pedagogical work, *Solfège de l'object sonore* (originally published in 1967, but a renewed version containing 3 CDs and a text booklet was made available in Schaeffer et al., 1998), containing sound examples and Schaeffer's narration, gives an excellent presentation of the main ideas of Schaeffer's research and will be much referred to here. Schaeffer's *Traité* (Schaeffer, 2017) may seem dauting in its extension and richness of ideas, and in comparison, the *Solfège* is a simplified yet highly reliable presentation of the main elements of Schaeffer's theories.

CHALLENGES OF THE *MUSIQUE CONCRÈTE*

Reading Schaeffer's accounts of the *musique concrète* developments in 1948-1952, we sense a profound fascination with the features of concrete sounds as well as various ideas of how to use these sounds in musical compositions. This included the idea of a kind of "noise piano" akin to Cage's prepared piano (Schaeffer, 2012, pp. 7-8), yet there was also a need to make some more general kind of structure to all this sound material:

> Faced with so many disparate objects, totally without grouping, without their conventions or their natural patrimony, a *classification*, even approximate, is essential, a sort of "grid" completely replacing instrumental tablature or the natural repertoire of noises. For how can we study an infinity of sounds that are not identified in any way? We will therefore use "sound identification criteria." They will give us the means to isolate sound objects from each other, since we refuse to do this through the usual sound or musical structures. In addition, they will lead us to a practical classification of sound objects, an obvious prerequisite for any further musical regrouping. (Schaeffer, 2017, p. 289).

In (Schaeffer, 2017) there are numerous passages about how to make sense of this disparate material, and with excursions into linguistics, gestalt theory, and phenomenology (to name the most important domains), yet at the same time also recognizing that there was a rather pragmatic beginning to the pervasive focus on sound objects, namely the experience of the so-called *closed groove* (*sillon fermé* in the French original).

The experience of closed groove stems from the use of disks with looped sounds in the *musique concrète* during the years before tape recorders became available. In composition work, this meant mixing sounds from several different phonographs to make the wished for sound texture, but it had the side effect of making people listen to innumerable repetitions of sound fragments. This in turn diverted the listeners' attention to various features of any sound fragment, first of all to the subjectively perceived overall dynamic shape, or envelope, of the sound fragment, and then, also to the more internal timbral-textural features of the fragment. In the words of Schaeffer:

> First, using the closed groove in the early stages of our work with the gramophone (without the closed groove our method would doubtless have never come into being), we made ourselves extract "something" out of the most disparate sound continuum. Thus this surrealist violation, so far removed from the earnestness of our colleagues in electronic music, obliged us to cut up sound and face up to what was most ill-assorted, most resistant to organization. (Schaeffer, 2017, p. 310).

Another and related experience was the so-called *cut bell* (*cloche coupée* in the French original), denoting the experience of a looped bell sound containing only the resonant part of the sound, i.e., that the attack transient of the mallet stroke had been removed, resulting in a sound that more resembled a sustained flute sound than a bell sound.

The experiences of "closed groove" and "cut bell," seemingly serendipitous in origin, became decisive for the development of research on sound objects for two main reasons:

(1) It forced the listeners' attention to the overall shape features of the sound fragment, i.e., to highly salient sound object timescale features that tend to be swallowed up in a longer context of continuous musical sound, e.g., as in that of the

timescale of Western musical works, or on the other hand, ignored by the more abstract focus on notated tones.

(2) It forced listeners to accept that sound features may be based on elements spread out in time, and that perception may work on a more holistic basis of taking some segment of musical sound into account as a whole, e.g., that the strike phase, and not only the quasi-stationary phase of the bell sound, is crucial for the perception of is characteristic sound.

Object Focus

There are other possible approaches to object-formation in auditory perception, and in the literature, we may encounter terms such as *chunking*, *grouping*, *parsing,* often based on qualitative discontinuities such as alternations between sound and silence, shifts in register, changes in spectral features, etc. Yet qualitative discontinuities alone may often be insufficient because of competing cues, missing information, or ambiguities requiring cues in accompanying modalities (See Godøy, 2008 for a summary). In the case of Schaeffer, the approach is first of all based on a top-down subjective perception of the overall shapes of the sound fragment, proceeding downward to more detail shapes, in a process where the closed groove gave the sound object a kind of solid appearance:

> The closed groove did, indeed, give an object in the sense of *a thing*, hidden away, as it were, by destroying another object. We have just observed that this involves not so much an objective discovery as putting the participant in a different situation. What does he see now that he had not seen by similarly, breaking up an elementary object, such as the sound of a bell, for example? Breaking it up informs him about the object, which he has—momentarily— destroyed only to hear it better. But if we bring the two experiments together, the closed groove and the cut bell, artificial, strange, anti-musical objects, and if we open our ears, we begin to hear whatever it is, sound or musical, differently, thanks to *reduced listening*, an experience that these two exercises in disruption taught us. (Schaeffer, 2017 p. 311).

The experience of the closed groove and cut bell would then serve to shift our attention as listeners to the overall shape as well as to the more internal features of the sound, away from the usual everyday significations, e.g., away from perceiving the squeaking of a door as a cue that someone is coming, toward the dynamic and spectral features of the squeak. This shift of focus in listening is what Schaeffer called *reduced listening*, with reference to intentionally disregarding everyday significations of sounds in favor of their more intrinsic sound features. But notably so, Schaeffer stated that this shift to focusing on the sound object timescale also was a strategic choice:

> We could say, in the most everyday language, that we could tackle the investigation of the musical from both ends—material and works—and *that we have exclusively chosen material*. But to put forward such a clear separation would be to forget the essential connectedness that articulates structures from the simple to the composite and that does not necessarily start with the simple: we enter into such relationships at any level, so we gain access as much to the higher as to the lower levels. In other words we perpetually keep in our minds and ears the part played in every work by *objects* (sound building blocks) that we can isolate and compare with each other independently of the context from which they come. (Schaeffer, 2017, p. 17).

Upon reflection, we may realize that the object focus is not altogether foreign to other music theory. For instance, in music theory texts we encounter the explanation of various clichés, e.g., in the treatment of dissonance (suspension, cambiata, etc.) and ornaments (mordent, trill, etc.) as well as instrumental idioms, as objects, and in this line of thinking, Schaeffer actually lists Messiaen's birdsongs as objects (Schaeffer, 2012, p. 171). The advantage of having the twin experiences of closed groove and cut bell is then to demonstrate that sound object perception is holistic, and that what occurs sequentially may be perceived holistically.

Listening Ontologies

Another crucial feature of the concrete music was the use of loudspeakers. With reference to the myth of Pythagoras that he was hiding behind a screen when teaching so that his students should not be distracted by seeing him, Schaeffer adopted the term *acousmatic* to denote the listening situation of concrete music as based on loudspeakers. Having no visual sound-producing source present except for loudspeakers, this meant that all sensations had to come by way of listening. With concrete sound objects in most cases having multiple significations and features, e.g., in terms of everyday significations as well as spectral, dynamic, textural, etc., features, the listening experience will seldom, if ever, be unambiguous, and will depend on the intentional focus of the listener at any time. For this reason, Schaeffer emphasized that listening would produce varying results, yet also suggested that we to a certain extent may control our attentional focus in listening:

> Every object perceived through sound is only so because of our listening intention. Nothing can prevent a listener from destabilizing this, going unconsciously from one system to another or else from *reduced* listening to one that *is not*. We should perhaps even congratulate ourselves on this. It is through such swirling intentions that links are established, information exchanged. (Schaeffer, 2017, p. 272).

The general point is that sound sensations can have multiple meanings, i.e., be what we could call *ontologically composite*. And on the way to a comprehensive theory of sound object perception, it could be useful to make an analysis of listening in view of focusing on the musical potential of concrete sounds. To this end, Schaeffer suggested that there are four components in listening (here with the French terms in parentheses for clarity): listening (écouter), perceiving (ouïr), hearing (entendre), and comprehend (comprendre), and that although they in most situations interact, we should distinguish them because they concern different aspects of what we perceive. To summarize the relationship between these four components, Michel Chion made this example sentence: "*I perceived (ouïr) what you said despite myself, although I did not listen (écouter) at the door, but I didn't comprehend (comprendre) what I heard (entendre).*" (Chion, 2009, p. 20).

Furthermore, Schaeffer suggested that listening may proceed by sketches, i.e., that sound object images may develop in our minds by repeated listening experiences: "…the process of *qualified listening*, the diversity of which arises from this fundamental law of perception, which is to proceed "by a series of sketches," without ever exhausting the object…" (Schaeffer, 2017, p. 77). With reference to Husserl's ideas on the constitution of objects in our minds by way of multiple but different sensations of the same object (Husserl, 1982), and where the cumulative perceptual image is called the *transcendence* of the object, Schaeffer comments:

> And then I notice that it is *in my experience* that this transcendence is *formed:* in other words, the *style* of perception itself, the fact that it never uses up its object, proceeds by rough sketches and always refers to other experiences that may contradict the previous ones and make them appear illusory, is not the sign of an accidental and regrettable imperfection that prevents me from knowing the external world "as it is." This style is, in fact, the mode in which the world is given to me as distinct from me. It is a particular style that allows me to distinguish the perceived object from the products of my mind or imagination that have other structures of consciousness. *So every domain of objects has its type of "intentionality." Each of their properties refers back to the activities of the consciousness that are "constitutive" of it: and the perceived object is no longer the cause of my perception. It is its "correlate."* (Schaeffer, 2017, p. 210).

The point of the correlate is crucial for understanding the relationship between the sound object as a perceptual entity and the acoustic contents of any sound fragment. Exploring this correlation was seen by Schaeffer as the long-term aim of research, whereas his more here-and-now project consisted in mapping out the subjective features of sound objects, notably so that there would be some kind of constancy in the object images in spite of the incessant variations in our images. Under the heading of *Objectivity of the object*, in the *Solfège* CD2, tracks 88 to 95, we can hear a series of examples combining constancy and variation, and also illustrating some of the major sound object feature categories.

Object Features

In (Schaeffer, 2012, 2017) Schaeffer recounts how he arrived at considering the sound object timescale as the most important in musical experience through some experimental work using (by our present standards) rather simple technologies. The mentioned closed groove and cut bell were the beginning, and this was followed by some other discoveries regarding the role of the attack transients and spectral non-linearities, summarized in the concept of *anamorphosis*, i.e., of warping.

The cut bell experience was considered a *temporal anamorphosis* in the sense of the instrument identity being dependent on sequentially occurring elements. There were also experiences of non-linearities in other domains, such as in the spectral composition of sounds across the range of any single instrument. For instance, a deep piano tone when shifted up a couple of octaves (by increasing the playback speed) sounds more like a harpsichord than a piano, hence, our perception of the unitary "piano-like" sound across the full range of the piano must be due to a more complicated set of factors than a linear shifting of the acoustic signal. This and similar experiences led Schaeffer to suggest that perceptual sound features should be seen as an independent domain, however, related to the acoustic substrates by the mentioned relationship of correlation. This correlation serves to show that the point of departure should be the subjective experience of a sound object and its perceptual features, in other words, that we should proceed in a top-down manner starting with seemingly simple questions as to what we are hearing. Notably so, this top-down feature differentiation may become quite complex, consisting of several main features, sub-features of these main features, and sometimes also sub-sub-features, as can be seen in the *Summary diagram of the theory of musical objects* (Schaeffer, 2017, pp. 464-467).

In view of the possible use of sound objects in musical compositions, Schaeffer introduced the idea of the *suitable object*. The suitable object fulfills some very general criteria, criteria that are flexible and context-dependent, yet interesting here in view of perceptual features. In brief, these criteria may be summarized as:

- The sound object should not be too long, nor too short.
- The sound object should not be too diverse, nor not to uniform.

The most interesting here is probably that or duration, which we can correlate with theories of attention spans found in various phenomenological and/or cognitive science contexts (e.g., in Michon, 1978; Husserl, 1991; Pöppel, 1997) and in recent research on motor control (see Godøy, 2018 for a summary).

As for non-suitable objects, they are either too short or too long, and/or they are either too varied or too unchanging, what Schaeffer denotes as *redundant*. One such sound object is the so-called *large note*, denoting a sound object that in spite of having clear gestalt coherence and closure, is just too long to be a suitable object in future musical compositions (examples of large note in the *Solfège* are up to 30 s). Another is the *ostinato* object that is redundant in its manifold repetitions, and at the other extreme, there are objects that are too short and/or too dense to be called suitable. CD3, tracks 43 to 65 offers examples of these various non-suitable objects.

It may be interesting to compare these durations with some other projects, e.g., that of (Gjerdingen and Perrott, 2008) suggesting that the overall timbral features may be perceived in fragments as short as 250 ms, whereas recognition of various rhythmic and melodic features of course would require longer durations. At the longer end of the timescale, we find object durations typically in the 3 – 15 s duration range, such as the UST (*Unités Sémiotiques Temporelles*) project, partially inspired by Schaeffer's research, but more focused on semiotic and affective features of sound objects (Delalande et al., 1996).

Acoustic Correlates of Sound Objects

That Schaeffer characterized the relationship between sound objects and acoustics as that of correlation and not of identity does not mean that there is no basis for sound objects in musical acoustics, but rather that he thought it necessary to make an analysis of factors involved in the perception of sound objects, and then explore the relationship between subjective perceptual sensations and acoustic features. This analysis may include features of what we presently refer to as *psychoacoustics* (as e.g., in Fastl and Zwicker, 2007), but Schaeffer's theory is broader in scope in that it also includes composition-oriented features, as well as features at the timescale of entire sound objects.

For Schaeffer, the point of departure for this correlation research was the initial subjective sensations of the sound object as a whole, as manifest in the *closed groove* experience, and then proceed to differentiate its various features. The approach is thus top-down, starting with an overall subjective image of any sound object, and progressing downward into the signal-based acoustic substrates.

The first consideration of acoustic correlates is then that of *timescales*, i.e., that acoustic features of sound objects are found at concurrent timescales ranging from the slow of the shape of the entire sound object, i.e., typically in the 0.5-5 s duration range, with overall dynamical, timbral, or pitch-related envelopes, down to the fast oscillations in the range of perceived pitch and/or spectral features, i.e., in the 20-20000 Hz range. A crucial point with the experiences of *closed groove* and *cut bell* was the need to take the entire sound object into consideration, in that all events within the entire sound object, e.g., the strike tone followed by the sustain tone and the ensuing overall dynamical, spectral, pitch-related, etc., envelopes of the entire sound object, contribute to its perceptual image. This requires that the entire sound object be kept in echoic memory (Snyder, 2000), so that sequentially unfolding elements may be present in our minds "all-at-once."

The second consideration of acoustic correlates is that sound objects may be situated in a *multidimensional model* consisting of main features, sub-features, sub-sub-features, as well as various values for these different feature dimensions, i.e., scales between minimum and maximum values, e.g., the feature of tremolo (amplitude modulation) may occur at a range between a maximum and a minimum speed. And with the possibility of intentional focus in listening, we may also zoom in and out of features at different timescales, i.e., from the overall shape of the sound object to its most minute transients, something that Schaeffer denoted as the "two infinities" of sound objects (Schaeffer, 2017, p. 220). A timescale analysis is then crucial for understanding the differentiation of acoustic features in Schaeffer's theory, as reflected in these three main categories:

- The *typology* timescale, with various sub-categories, can encompass features in the duration range from that of the entire sound object down to patterns within a sound object, typically concerning overall dynamic and spectral shapes of sound objects, providing a coarse sorting of sound objects.
- The *morphology* timescale, with various sub-categories, typically including more internal features of the sound objects such as its pitch, timbre, spectral shapes, as well as various rapid internal fluctuations and transients.
- The combination of all main features in the multidimensional model, rendered in English as *"Summary diagram of the theory of musical objects"* (Schaeffer, 2017, and pp. 464-467), which presents an overview of the various typological and morphological main dimensions and sub-dimensions, as well as relative values for these dimensions, e.g., the sound object of a rapid harp glissando combining an overall dynamic and pitch-related shape with detail shapes of individual tone onsets and timbres.

The idea of such multidimensional modeling of musical sound has later been developed in the works of e.g., John Gray, David Wessel, and others, and with implementations in software such as e.g., in the *Timbre Toolbox* (Peeters et al., 2011) as well as in some projects within so-called *music information retrieval*, e.g., the *MIRtoolbox* (Lartillot and Toiviainen, 2007). Besides being remarkably ahead of his time, Schaeffer's multidimensional model is also striking in its generality, making it applicable to any kind of music (e.g., *musique concrète*, various avantgarde or various non-Western music). Another advantage is the strong top-down direction of detecting and qualifying perceptually salient features, rather than a more "blind" and/or purely signal-based, bottom-up search for significant features.

Guided by subjective perceptual categories, this multidimensional model can relate to most (or all) traditional musical acoustic elements such as pitch, stationary spectra, various time domain and frequency domain envelopes, but with the additional advantage of naming salient features that are based on more composite acoustic features such as that of *gait* and *grain* in the morphology (see section "Morphology" below). This is in particular useful for capturing components of timbre and musical textures that rely on transients and fluctuations, hence are not limited to stationary spectra.

This multidimensional scheme can distinguish different values for salient components, e.g., the rate and amplitude of fluctuations within a sound objects. Changes in rates and amplitude may then help us distinguish different categorical thresholds for salient features, e.g., that of the rate and amplitude of a frequency modulation: if the rate is slow (say no more than 8 Hz) we may perceive a vibrato, but if it is significantly faster (say above 20 Hz) it will become a timbral feature. Similar value thresholds may be found in most (or all) other features and can also be explored by an *analysis-by-synthesis* approach (Risset, 1991), similar to what we may hear on CD2, track 89 and onward. Categorical threshold explorations may also be found at the object shape timescale, such as on CD3, track 60-63, with an incremental change from a protracted ostinato sound object to a series of singular impulse objects.

The key to exploring the acoustic correlates of sound object is then the two-step process of (a) distinguishing and naming some perceptually salient feature of the sound objects, e.g., its fluctuation in amplitude, and then (b) qualify its value, i.e., its rate, shape, regularity, etc., of fluctuation. We could also add a third step, (c) if we have the means to do so, to generate variants of this feature and evaluate the result, i.e., make incrementally different sound objects by varying the rate, amplitude, shape, regularity, etc., hence, engage in a process of analysis-by-synthesis to explore various categorical thresholds. In brief, Schaeffer's multidimensional view of acoustics is a remarkable project of making a large number of previously non-thematized but highly salient perceptual features accessible for intentional focus in research and in artistic creation.

Throughout these considerations of salient perceptual features of sound objects and their acoustic correlates, there is a pervasive use of shape-related metaphors, in particular for time domain features such as dynamic envelopes, as well as for spectral elements. This penchant for what we could call "shape cognition" was present already in Schaeffer's early work on the *musique concrète* in a number of graphical images (Schaeffer, 2012), in particular concerning the attack shapes of sound objects (see e.g., Schaeffer, 2012, p. 203) and is a further testimony to perceptual features as distributed and holistic, i.e., not reducible to more abstract symbolic representation as has been the dominant trend in Western music theory.

Typology

The typology has two basic dimensions, one concerned with the dynamic shapes, i.e., what we could call the energy envelopes of the sound objects, and what Schaeffer called *facture* types, and one concerned with the spectral features of the sound objects, what Schaeffer called *mass* types, be that as clear pitch sensations or as more ambiguous inharmonic and/or noise-dominated sensations. The three main facture types are the following:

- *Impulsive*, a fast and short sound, typically as by striking or plucking.
- *Sustained*, a prolonged sound with a steady level energy envelope.
- *Iterative*, fast back-and-forth or rotational motion, resulting in a stream of impulses as can be seen in **Figure 1** where the sounds of a drumroll, a deep bassoon tone, and a processed sound are compared in view of the common feature of a pronounced iterative feature.

The concept of facture refers to how things are made, so the mentioned three categories can also be related to similar categories of sound-producing motions. This means that an impulsive motion is fast and brief (sometimes also called *ballistic*), and a sustained motion is a protracted and smooth motion, whereas the iterative motion typically is a back-and-forth shaking or rotating motion. The link between the sound facture and motion categories is interesting also in that motion categories are mutually exclusive (e.g., motion can't be sustained and impulsive at the same time) and refer to basic biomechanical and motor control constraints. Furthermore, there is the link with what may be called the *motormimetic* element in perception and cognition (Godøy, 2003), suggesting that we may also make a multimodal view of these categories in Schaeffer's theory (Godøy, 2006).

In terms of pitch and/or spectrum, the typology has three main mass types:

- *Tonic*, meaning a perceivable and stable pitch.
- *Complex*, denoting inharmonic or noise-dominated sound.
- *Varied*, encompassing any sound that changes in pitch or frequency placement.

All these categories are well illustrated with examples in the *Solfège* CD3, tracks 28, 29, 30 with respectively, a tonic mass, complex mass, and varied mass.

The typology is a coarse but flexible and universally applicable categorical scheme in which the facture and mass categories can also be combined in a 3 × 3 matrix as follows, first with traditional musical instruments in CD3, tracks 31, 32, and 33:

- Track 31: tonic impulsive, tonic sustained, tonic iterative.
- Track 32: impulsive complex mass, complex sustained mass, iterative complex mass.
- Track 33: impulsive varied mass, sustained varied mass, iterative varied mass.

Then follows a similar scheme with prepared piano sounds repeated in tracks 34, 35, and 36, and with concrete sounds repeated in tracks 37, 38, and 39, as well as with electronic sounds repeated in track 40, 41, and 42. In **Figure 2**, we can see spectral representations in the 3 × 3 typological matrix based on the electronic sounds of tracks 40, 41, and 42. The overall facture shapes of these 9 different sounds should be quite clear, in spite of different mass content.

Morphology

The morphology is mainly concerned with the internal features of sound objects, and has several dimensions and sub-dimensions, and for some features, also relative values, e.g., indicating the speed and amplitude of various fluctuations. The morphology is a rather extensive domain, here illustrated with only two of its dimensions:

- *Grain* designates the fast and small fluctuations in the sound, and has attributes such as variations in amplitude, speed, and consistency.
- *Gait* denotes the slower fluctuations in the sound, similar in pace with walking or dancing, but may have variable speed, amplitude, and quality such as mechanical, living, natural, orderly, disorderly, etc.

Additionally, there are morphology dimensions such as *mass* (denoting the overall spectral content, cf. the definition above in the context of the typology), *dynamics* (for the overall loudness), *harmonic timbre* (spectral distribution), *melodic profile* (pitch-related shapes), and *profile of mass* (spectral shapes). This means that the morphology contains both features that are stationary (or quasi stationary) and changing, be that as rapid fluctuations of the grain or more slowly of the gait. Crucially, the morphology features are eminently top-down with sub-dimensions, sub-sub-dimensions, etc., where the guiding principle of exploration is the increasing distinguishing of detail features.

Lastly, there is the possibility of so-called *phase transitions* between various typological and morphological categories. Depending on the density and duration of events, there may be a shift from one category to another, for instance, if an iterative sound is speeded up, it may turn into a grain feature, and conversely, if it is slowed down, it may turn into a series of impulsive sounds. Or: if an impulsive sound is prolonged beyond a certain limit, it may turn into a sustained sound, and conversely, a sustained sound shortened beyond a certain limit will turn into an impulsive sound. The main point here is that we have categorical thresholds where features

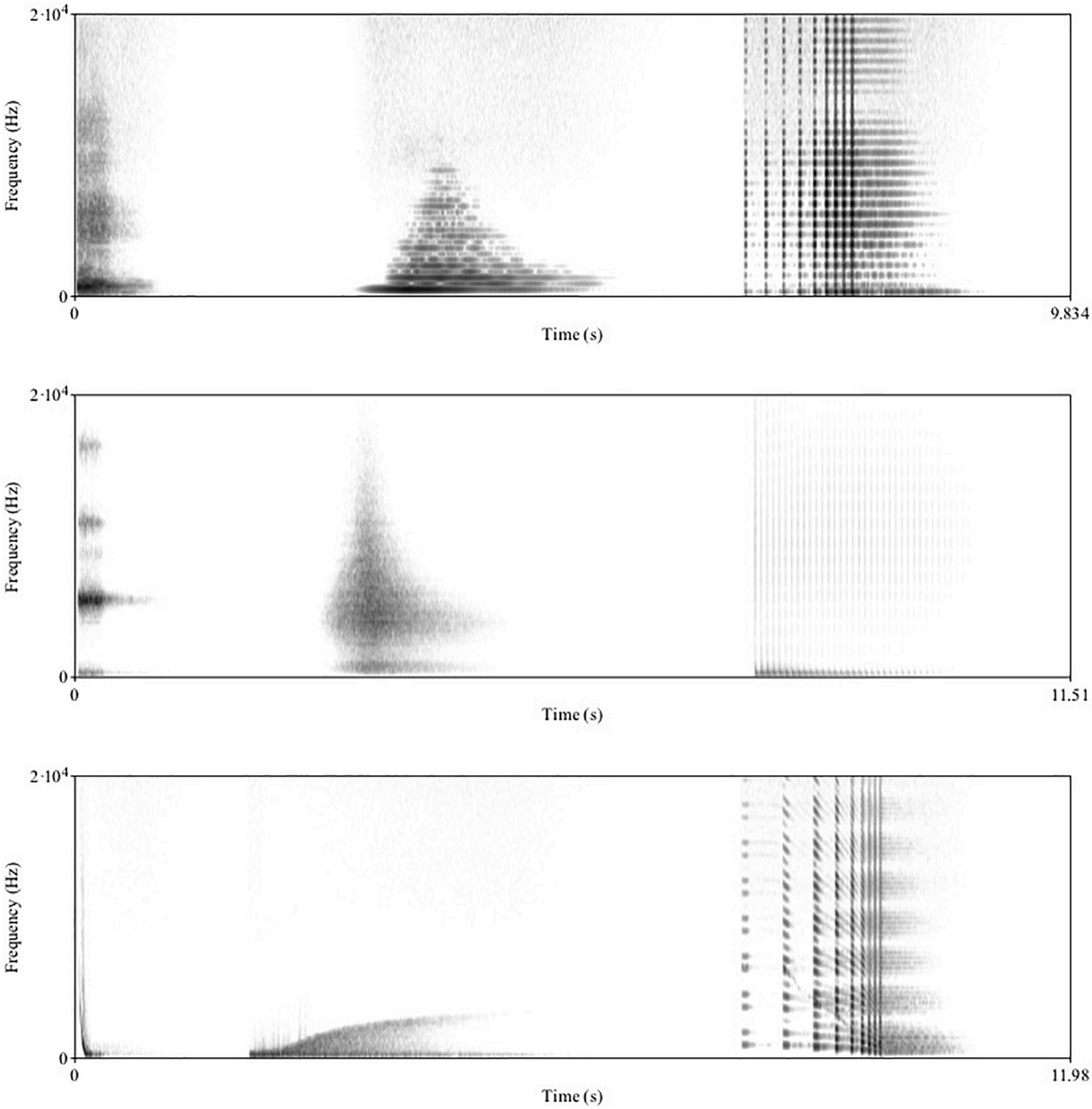

FIGURE 2 | A 3 × 3 typological matrix of spectrograms based on electronic sounds from *Solfège* CD3, tracks 40, 41, and 42. In the first row, spectrograms of track 40 with tonic mass, i.e., clearly perceivable pitches, in the second row, spectrograms of track 41 with complex mass, i.e. inharmonic and noise-dominated sounds, and in the third row, track 42 spectrograms with varied mass, i.e., glissando pitches. In each row, the envelopes have respectively, an impulsive, a sustained, and an iterative facture (see main text for explanation of "facture"). Notice the overall dynamic envelopes of these different factures, i.e., the abrupt of the impulsive, the prolonged of the sustained, and the punctual and accelerating of the iterative.

have more or less typical value ranges, value ranges that may very well be rooted in various perceptual-cognitive schemata of our organism.

As an illustration of some of these different typological and morphological features, and linked with what Schaeffer called *Objectivity of the object* in track 88 of the *Solfège* CD2, it is instructive to consider the perception of invariance across the series of variants of a sound object starting with track 89: in track 90, a variant of its overall shape, in track 91, a variant of its mass, in track 92, a variant of its grain, in track 93, an exaggerated harmonic timbre, in track 94, an exaggerated gait, and then all combined in the resultant exaggerated object in track 95. This was meant by Schaeffer as an illustration of what a series of listening intentions on the one and same object may bring about, but here manifest in the acoustic correlates of the sound object. In **Figure 3**, we can see the spectrum and spectral flux (changes in the spectral width) of track 89 into the exaggerated gait in track 94.

Summary Diagram of the Theory of Musical Objects

Toward the end of CD3 of the *Solfège*, from track 64 and onward, Schaeffer's conceptual apparatus is put into practice, with an anecdotic account of a composition factory where masses of sound arrive by truckloads, is processed further, and then put together. In these tracks of CD3, we can hear how electronic (both concrete and synthetic) and instrumental sounds, can all be handled with the same perceptual categorical apparatus.

We have an overview of this perceptual categorical apparatus in the mentioned *Summary diagram of the theory of musical objects* (Schaeffer, 2017, pp. 464-467). There we see the main dimensions, sub-dimensions, sub-sub-dimensions, and relative values for these dimensions, enabling a positioning of any sound object in a multidimensional feature model. And this is the take home message from Schaeffer's theory: we may finetune our perceptual images of sound objects, both in the *musique concrète* and other music, and also apply this conceptual scheme to sound design or composition as part of musical craftsmanship.

Michel Chion reminds us that this diagram is a tool for investigation, that "The general procedure in this music theory is to move forward in a series of approximations rather than in a straight line." (Chion, 2009, 100). This back-and-forth of sound object and its typological and morphological features, resembling a kind of hermeneutic circle, is one of the key features of exploring sound objects: detecting and naming features enhance our awareness, and this awareness makes us in turn detect more detail features, progressively building up richer and more many-facetted images of sound objects.

DISCUSSION: EXPLORING SOUND OBJECTS BEYOND THE *MUSIQUE CONCRÈTE*

Musique concrète was a remarkable project of aesthetic innovation combined with reflections on perception, and in this respect quite different from other contemporary music (cf. section "Listening Ontologies" above). This intrinsic focus on perceptual issues seems to have been driven by the more or less total lack of conceptual tools in mainstream Western music theory for new sounds and sound features, encouraging Schaeffer and co-workers to think outside the box on what were (and are) fundamental questions of sound features and aesthetic judgment (cf. the mentioned ideas about the *suitable object*).

The main elements of Schaeffer's and co-workers strategy for developing a new and more comprehensive music theory that also included the perception of concrete sound, can be summarized in the concept of the sound object and its typological and morphological features. As found in the *Summary diagram of the theory of musical objects* mentioned above, the most striking elements of this typology and morphology are:

- The multidimensional scheme of detecting, and further differentiating, what are considered salient features of any sound object.
- The all-pervasive element of what we can call *shape cognition* (Godøy, 2019). This concerns not only the verbal labels and graphic signs in the *Summary diagram of the theory of musical objects*, but also the numerous graphical figures in the course of Schaeffer's work, representing the various attack and spectral components of the sound objects (e.g., shape images for attacks in Schaeffer, 2017, p. 425, and spectral components in Schaeffer, 2012, p. 210).

As for the sound object focus, it is about having holistic perceptions of temporally unfolding fragments of sound. That there always is this holistic element at work for sound objects is reflected by Schaeffer in the example of cutting an object into smaller parts: each new part has a head, body, and tail, just like a magnet cut into two parts will immediately have polarities in each new part (Schaeffer et al., 1998). In this way, any arbitrary cut in a continuous sound recording will result in new sound objects, albeit in the case of a totally random cut, the resultant sound object may not be particularly useful in a musical context, cf. the abovementioned criteria for the so-called suitable object.

It is the overall energy envelope of the sound object that usually will be most prominent, cf. the mentioned typological facture categories, and as such, may be linked with various criteria for chunk-formation in the cognitive sciences, ranging from the classics of Lashley (1951); Miller (1956) to the more recent of Gobet et al. (2016), as well as some ecologically oriented schemes for auditory chunk formation (e.g., Bregman, 1990; Gaver, 1993; Bizley and Cohen, 2013, see Godøy, 2018 for overviews). Importantly, there are also links to chunking in sound-producing motion (Godøy, 2013, 2014), in turn related to chunk formation in body motion (Grafton and Hamilton, 2007; Klapp and Jagacinski, 2011; Loram et al., 2011).

Considering the crucial role of the shapes of sound objects and their features as depicted in the typology and morphology, implies that shape cognition goes right to the core of *musique concrète*. In having moved outside the Western note-symbol domain, we arrive at a more general and sound-centered domain where we are concerned with temporally (and usually also spectrally) distributed, non-abstract entities. This also means we open the door to many traditionally non-thematized features in Western music, in particular concerning timbre, but also various expressive fluctuations of intonation, dynamics, and tempo, with the common feature of not being reducible to singular symbols, but actually requiring some kind of shape representation.

Several projects within the cognitive sciences converge in regarding shape images as fundamental for human perception and cognition (see Godøy, 2019 for an overview). In particular, the so-called *morphodynamical* domain has contributed strongly to recognizing shape images as crucial in understanding and handling complex sensory streams (Thom, 1983; Petitot, 1990;

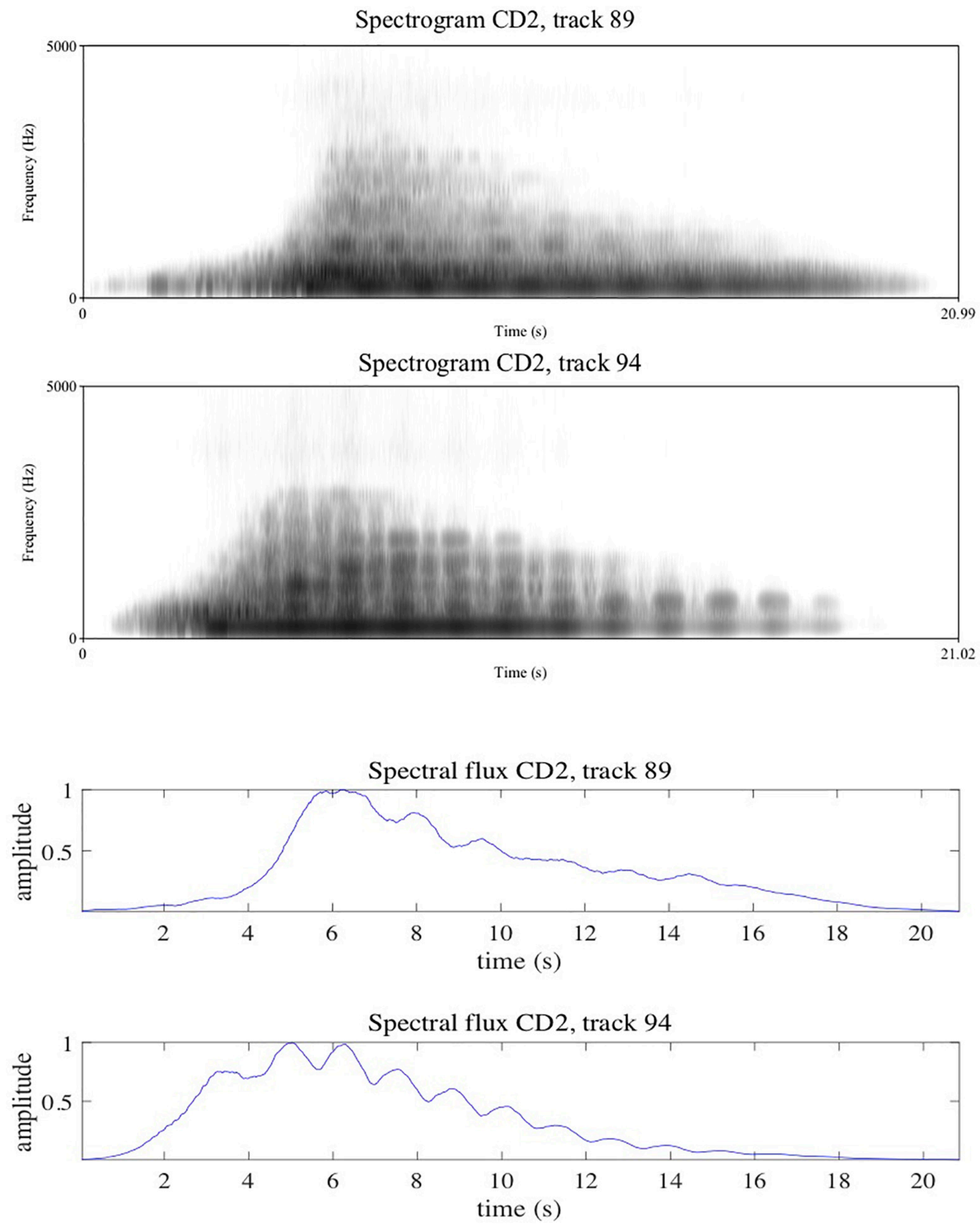

FIGURE 3 | The gait (*allure* in the French original) from *Solfège* CD2, track 89 in the first row and track 94 in the second row, with the gait exaggerated in this second row, both shown here as spectrograms and MIRtoolbox plotting of spectral flux, i.e., plotting of changes in spectral spread, reflecting the wavy motion of the gait. The sound in track 89 is of a tremolo crescendo on a tam-tam followed by a long decay tail with the wavy gait motion at around 2 Hz, and in track 94, the same sequence of events, however, here the amplitude of the wavy gait is strongly exaggerated.

Godøy, 1997). In the words of René Thom: "…the first objective is to characterize a phenomenon as shape, as a "spatial" shape. To understand means first of all to geometrise." (Thom, 1983, p. 6).

From the various sound examples in the *Solfège*, it is clear that there was an affinity of electroacoustic and ordinary acoustic music in the mind of Schaffer: the typology, morphology, and associated concepts are equally applicable to all kinds of music, i.e., just as well applicable to instrumental, vocal, orchestral, etc., music, as to electroacoustic music. In particular, the typological and morphological categories could be relevant in the analysis of orchestration (see Godøy, 2018 for some examples).

As for sources of shape representations, we have the signal-based, i.e., time-domain and frequency-domain, images. These may be subject to further levels of processing and schemes of representation, selectively representing a variety of perceptually salient features as suggested by the typology and morphology of sound objects. Additionally, there are also the connection to motor theory images of sound, with similar schemes of shape cognition (Godøy, 2003, 2006, 2010), providing multimodal links using motion data in synchrony with sound features data.

Lastly, these shape-oriented explorations of perceptual features can be linked with large-scale perceptual surveys/experiments using music information retrieval software such as the *MIRtoolbox* (Lartillot and Toiviainen, 2007) for exploring correlations between sound objects feature shapes and acoustic signal features, as was the long-term project of Schaeffer.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

The work on this manuscript has been partially supported by the Research Council of Norway through its Centre of Excellence scheme, project number 262762 and by the University of Oslo.

ACKNOWLEDGMENTS

Many thanks to colleagues at the RITMO Centre of Excellence for interesting discussions on topics of sound and motion shapes.

REFERENCES

Bizley, J. K., and Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.* 14, 693–707. doi: 10.1038/nrn3565

Bregman, A. (1990). *Auditory Scene Analysis*. Cambridge: The MIT Press.

Chion, M. (1983). *Guide Des Objets Sonores*. Paris: Eìditions Buchet/Chastel.

Chion, M. (2009). *Guide to Sound Objects (English Translation by John Dack and Christine North)*. Electro Acoustic Resource Site (EARS). http://ears.huma-num.fr*

Delalande, F., Formosa, M., Freìmiot, M., Gobin, P., Malbosc, P., Mandelbrojt, J., et al. (1996). *Les Uniteìs Seìmiotiques Temporelles: Eìleìments Nouveaux D'analyse Musicale*. France: EìditionsMIM-DocumentsMusurgia.

Fastl, H., and Zwicker, E. (2007). *PsychoacousticsFacts and Models*. Berlin: Springer.

Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecol. Psychol.* 5, 1–29. doi: 10.1207/s15326969eco0501_1

Gjerdingen, R. O., and Perrott, D. (2008). Scanning the Dial: the rapid recognition of music genres. *J. New Music Res.* 37, 93–100. doi: 10.1080/09298210802479268

Gobet, F., Lloyd-Kelly, M., and Lane, P. C. R. (2016). What's in a name? The multiple meanings of "Chunk" and "Chunking.". *Front. Psychol.* 7:102. doi: 10.3389/fpsyg.2016.00102

Godøy, R. I. (1997). *Formalization and Epistemology*. Oslo: Scandinavian University Press.

Godøy, R. I. (2003). Motor-mimetic music cognition. *Leonardo* 36, 317–319. doi: 10.1162/002409403322258781

Godøy, R. I. (2006). Gestural-sonorous objects: embodied extensions of schaeffer's conceptual apparatus. *Organised Sound* 11, 149–157.

Godøy, R. I. (2008). "Reflections on chunking in music," in *Systematic and Comparative Musicology: Concepts, Methods, Findings*, ed. A. Schneider (Germany: Peter Lang), 117–132.

Godøy, R. I. (2010). "Gestural affordances of musical sound," in *Musical Gestures. Sound, Movement, and Meaning*, eds R. I. Godøy and M. Leman (New York, NY: Routledge).

Godøy, R. I. (2013). "Quantal elements in musical experience," in *Sound – Perception – Performance. Current Research in Systematic Musicology*, Vol. 1, ed. R. Bader (Berlin: Springer), 113–128. doi: 10.1007/978-3-319-00107-4_4

Godøy, R. I. (2014). "Understanding coarticulation in musical experience," in *Lecture Notes in Computer Science*, eds M. Aramaki, M. Derrien, R. Kronland-Martinet, and S. Ystad (Berlin: Springer), 535–547. doi: 10.1007/978-3-319-12976-1_32

Godøy, R. I. (2018). "Sonic object cognition," in *Springer Handbook of Systematic Musicology*, ed. R. Bader (Basingstoke: Springer Nature), 761–777. doi: 10.1007/978-3-662-55004-5_35

Godøy, R. I. (2019). "Musical shape cognition," in *The Oxford Handbook of Sound and Imagination*, eds M. Grimshaw, M. Walther-Hansen, and M. Knakkergaard (New York, NY: Oxford University Press), 2.

Grafton, S. T., and Hamilton, A. F. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Hum. Mov. Sci.* 26, 590–616. doi: 10.1016/j.humov.2007.05.009

Husserl, E. (1982). *Ideas Pertaining to a Pure Phenomenological Philosophy, First Book*. London: Kluwer Academic Publishers.

Husserl, E. (1991). *On the Phenomenology of the Consciousness of Internal Time, 1893 -1917. English translation by John Barnett Brough*. London: Kluwer Academic Publishers.

Klapp, S. T., and Jagacinski, R. J. (2011). Gestalt principles in the control of motor action. *Psychol. Bull.* 137, 443–462. doi: 10.1037/a0022361

Lartillot, O., and Toiviainen, P. (2007). "A matlab toolbox for musical feature extraction from audio," in *Proceeding of the International Conference on Digital Audio Effects*, (France), 2007.

Lashley, K. S. (1951). "The problem of serial order in behavior," in *Cerebral Mechanisms in Behavior*, ed. L. A. Jeffress (New York, NY: Wiley), 112–131.

Loram, I. D., Golle, H., Lakie, M., and Gawthrop, P. J. (2011). Human control of an inverted pendulum: is continuous control necessary? Is intermittent control effective? Is intermittent control physiological? *J. Physiol.* 2, 307–324. doi: 10.1113/jphysiol.2010.194712

Michon, J. (1978). "The making of the present: a tutorial review," in *Attention and Performance VII*, ed. J. Requin (Hillsdale), 89–111.

Miller, G. A. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97. doi: 10.1037/h0043158

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., and McAdams, S. (2011). The Timbre Toolbox: extracting audio descriptors from musical signals. *J. Acoust. Soc. Am.* 130, 2902–2916. doi: 10.1121/1.3642604

Petitot, J. (1990). *'Forme' in Encyclopædia Universalis*. Paris: Encyclopædia Universalis.

Pöppel, E. (1997). A hierarchical model of time perception. *Trends Cognit. Sci.* 1, 56–61.

Risset, J.-C. (1991). "Timbre analysis by synthesis: representations, imitations and variants for musical composition," in *Representations of Musical Signals*, eds G. De Poli, A. Piccialli, and C. Roads (Cambridge: The MIT Press), 7–43.

Schaeffer, P. (1952). *A la Recherche D'une Musique Concrète*. Paris: Éditions du Seuil.

Schaeffer, P. (1966). *Traité Des Objets Musicaux*. Paris: Éditions du Seuil.

Schaeffer, P. (2012). *In Search of a Concrete Music (English Translation by Christine North and John Dack)*. Oakland,CA: University of California Press.

Schaeffer, P. (2017). *Treatise on Musical Objects (English Translation by Christine North and John Dack)*. Oakland, CA: University of California Press.

Schaeffer, P., Reibel, G., and Ferreyra, B. (1998). *Solfège de L'objet Sonore*. Paris: INA/GRM.

Snyder, B. (2000). *Music and Memory. An Introduction*. Cambridge: The MIT Press.

Thom, R. (1983). *Paraboles et Catastrophes*. Paris: Flammarion.

Xenakis, I. (1992). *Formalized Music (Revised Edition)*. Stuyvesant: Pendragon 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 Psychology

ORIGINAL RESEARCH
published: 07 June 2021
doi: 10.3389/fpsyg.2021.673706

Repetition and Aesthetic Judgment in Post-tonal Music for Large Ensemble and Orchestra

Moe Touizrar[1†], Anna Lena Knoll[2] and Kai Siedenburg[2†]*

[1]*Finnish Centre for Interdisciplinary Music Research, Department of Music, Art and Culture Studies, University of Jyväskylä, Jyväskylä, Finland,* [2]*Department of Medical Physics and Acoustics, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany*

OPEN ACCESS

Edited by:
Mark Reybrouck,
KU Leuven, Belgium

Reviewed by:
Erkki Huovinen,
Royal College of Music in Stockholm, Sweden
Justin Christensen,
University of Saskatchewan, Canada

Correspondence:
Moe Touizrar
mohamed.n.touizrar@jyu.fi

†*These authors have contributed equally to this work and share first authorship*

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

Received: *28 February 2021*
Accepted: *29 April 2021*
Published: *07 June 2021*

Citation:
Touizrar M, Knoll AL and Siedenburg K (2021) Repetition and Aesthetic Judgment in Post-tonal Music for Large Ensemble and Orchestra. Front. Psychol. 12:673706. doi: 10.3389/fpsyg.2021.673706

Post-tonal music often poses perceptual and cognitive challenges for listeners, potentially related to the use of relatively uncommon and unfamiliar musical material and compositional processes. As a basic compositional device, repetition affects memory for music and is structured by composers in very different ways across tonal and post-tonal musical repertoires. Of particular concern is whether post-tonal music exhibits mnemonic affordances that allow listeners to experience a sense of global coherence, and whether repetition correlates strongly with aesthetic judgment. Although previous research suggests that repetition impacts aesthetic preference, empirical research has not mapped out the relationship between repetition and aesthetic judgments across a broad set of post-tonal music. Presenting 14 excerpts, grouped into three categories: tonal, modernist, and post-1970, we observed that indications of repetitions in the music for a group of 60 listeners, with and without musical training, showed significant periods of interindividual synchronization. Aesthetic judgments were assessed by means of ratings for the following parameters: familiarity with the piece, confidence of repetition responses, judgments of affordances for easy listening, coherence, similarity of moments, and recognition, as well as listeners' liking and interest. A principal component analysis (PCA) on the joint question data and the repetition responses suggested that two factors account for 92% of variance in the data. These factors were interpreted as dominated by aesthetic judgment and repetition strength. Linear mixed-effects regression indicated that repetition strength generally differed across excerpt category, with modernist excerpts featuring lower repetition strength compared to both post-1970 and tonal excerpts. Aesthetic preference, on the other hand, was lower for excerpts in both the modernist and post-1970 categories when compared to tonal excerpts. The analysis did not reveal difference in response behavior for repetition responses as a function of musical training, though it indicated higher preference of modernist excerpts with increasing levels of musical training. Overall, the results suggest that the two factors, aesthetic judgment and repetition strength, act as independent determinants in the experience of post-tonal music.

Keywords: repetition, post-tonal music, mnemonic affordance, musical coherence, psycho-aesthetics, aesthetic judgment

INTRODUCTION

In the present study, we wish to contribute to a better understanding of the relationship between the perception of repetition, comprehensibility, and aesthetic judgment in the chronically understudied domain of post-tonal music. Post-tonal compositions often pose perceptual and cognitive challenges to listeners, potentially related to their novel and unfamiliar musical material and formal processes. Of particular concern for music perception and cognition is whether music of this type exhibits mnemonic affordances that allow listeners to experience a sense of similarity and coherence. As a basic compositional device, repetition affects memory for music and is employed by composers in very different ways across the post-tonal musical repertoire.

Listening encompasses a complex set of dynamic perceptual-cognitive processes that are geared toward both attending to the immediate features of an auditory event and to the relative integration of past events to form a continuous understanding (Levinson, 1997; London et al., 1999; Huovinen, 2013). Everyday acts of listening to music are adaptive and rooted in embodied cognitive mechanisms (Kozak, 2019; Reybrouck, 2020), and the processes that underlie the formation of intelligible musical experiences, for which repetition is often a key factor, must be approached from both music theoretical as well as cognitive perspectives. Music theorists identify the contoured patterning of the motive as the basic unit of coherence in the Western classical tradition (Carpenter and Neff, 1995; Van den Toorn, 1996; Boss, 2000). Motives are usually presented as compact memorable musical ideas at or near the beginning of a piece and are subsequently re-presented in various novel configurations that preserve the general identity of the original, while making important developmental changes to its structuring. In this way, motives serve as formal mnemonic signposts that are experienced across the temporal unfolding of musical pieces. The mind's ability to establish and categorize connections between related motivic events engenders musical coherence and comprehension (Zbikowski, 1999).

Repetition in Music Listening

Lamont and Dibben (2001) identify three cognitive factors that contribute to a listener's understanding of motivic coherence: musical experience, familiarity – both with general styles of music and within specific pieces – and the degree of complexity in the surface features of a given auditory event. Repeated units of music accrue in experience and can produce habituation – a decline in responsiveness as the novelty of a stimulus recedes relative to exposure. The attentive focus of the mind relaxes as sonic patterns take on increasing familiarity. In a review of previous findings, Huron (2013) lists five factors that influence the speed with which habituation occurs. Firstly, the number of presentations of a given stimulus has a direct impact on the degree to which the mind assesses the importance of an event. Secondly, the rate of repetition helps to determine both the degree of imprint and the relative importance of the event; too few presentations spaced at too wide an interval will lessen the cognitive importance of a stimulus and the subsequent anticipation of its re-presentation. Thirdly, the relative predictability of a stimulus can affect the speed at which habituation takes effect. Fourthly, the relative prominence and energetic magnitude of the stimulus has a direct bearing on its integration and on how quickly its repetition leads to habituation. Energetic stimuli often resist habituation for a longer period than lower energy stimuli. Finally, the particular biographical history of the listener – whether they have a high degree of experience with a given type of musical stimulus or style – plays a determining role in their ability to both track and predict repetitions. Repeated exposure increases habituation.

During listening, similarity judgments depend primarily on the relative saliency of shared features perceived between auditory cues and on the listener's ability to organize representations, both present and past, into meaningful categories (Cambouropoulos, 2001). Within complex auditory scenes, similarity judgments for polyphonic textures are based on perceived differences in features such as amplitude, articulation, textural density, and gestural contour (Lamont and Dibben, 2001). Building upon early work on prototype theory by Rosch (1975) and Tversky (1977), and subsequent alternatives proposed by Murphy and Medin (1985), two primary modes of categorization have been proposed: perceptual equivalence (prototypical categorization) and theory-based classification (Cambouropoulos, 2001; Deliège, 2001a,b, 2007; Lamont and Dibben, 2001). Prototypical approaches to musical repetition hold that the presentation and initial few repetitions of a motive represent the privileged exemplar of a category, to which all subsequent similar repetitions, however, much altered, are understood to belong. In the context of music cognition, theory-based categorization involves prior knowledge of a given style of music acquired through acculturation. According to this view, categorization involves not only shared surface-level attributes between items (as in prototype models), but also an underlying conceptual knowledge that helps to select which items in a given scene should be attended to (Lamont and Dibben, 2001).

Echoing the theory-based view of categorization, Deliège (2001b) notes that listeners who are already familiar with a given piece of music employ functional categorization, where pre-existing concepts formed by previous exposure allow contextual knowledge to supersede surface-level similarity. Alternatively, in cases of initial exposure to unfamiliar music, Deliège (2001b) proposes that categorization occurs in real time as a prototypical comparison between present and past events, where similarity in features determines categorical grouping. Furthermore, the pairwise comparison of features relative to an initial prototype ascribes a highly mnemonic mode of cognition for music listening, at least where the repetition of material functions as the primary mode of musical organization. In these contexts, repetition engenders saliency by focusing attention on both what is at hand within an auditory scene as well as on the relationship of the present auditory event with those that have recently past. It is noteworthy here that Taher et al. (2016) demonstrate compelling evidence suggesting that when presented with tonal two-part contrapuntal textures, listeners are susceptible to a rapid type of habituation capable of guiding attention *away from* repeated motives in one voice and toward novel information in another after just a single repetition.

Aesthetic Judgment

Two primary factors have been shown to influence aesthetic judgments: familiarity through repeated exposure and the level of complexity displayed by a given stimulus (Hargreaves, 1984; North and Hargreaves, 1995). The "mere exposure" effect describes the positive increase in aesthetic judgment of music based on prior exposure, memory formation, and the accumulation of perceptual representation structures (Peretz et al., 1998). Repeated exposure to motives or other forms of recursion within musical works has been positively correlated to liking. Tillman and Bigand (1996) segmented two tonal piano pieces by Bach and Mozart, together with a post-tonal piece by Schoenberg into short chunks lasting approximately 6 s. Segments were linked into longer strands that either preserved the composer-intended order, or that reversed that order, but included all between-segment repetitions. The reversed order condition preserved the internal structure and ordering of each chunk, while obliterating the formally chained structure of the chunks as a larger ordered sequence. Non-musician participants rated the strands (either original or inverted) for musical expressiveness using 29 semantic scales, including ratings for coherence. Ratings for expressiveness and coherence across conditions showed no significant differences for the tonal music, and only minimal difference for the post-tonal music. Tan et al. (2006) had participants listen to 1-min excerpts of tonal piano music using two stimuli conditions: (1) intact excerpt and (2) patchworked hybrids of three 20-s excerpts from music by different composer linked together seamlessly without regard for similarity in thematic material, or structural parameters such as harmony, key, and tempo. Repeated hearings led to linear increase to cohesion and liking ratings for the patchwork compositions, while repeated exposure to intact stimuli led to decreased ratings.

Mnemonic Affordance and Aesthetic Intention in Post-tonal Music

The various perceptual and cognitive processes and schemata in operation during listening to tonal music have long been studied (Meyer, 1956; Lerdahl and Jackendoff, 1983; Krumhansl, 1990). Whereas tonal music is organized using structural features derived from the major-minor scale system that are shared across pieces and across historical style periods, post-tonal music lacks a similar degree of shared structuring and inter-stylistic uniformity. Early post-tonal music was freely structured from piece to piece and composer to composer (c. 1909–1919). Breaking from tonal conventions, where repetition and intelligibility are key factors in the perceived articulation of musical form, composers of the Second Viennese School, such as Arnold Schoenberg, Alban Berg, and Anton Webern (to name only the most prominent), re-imagined the very organizing principles of music. Their search for a shared structure that could replace tonal organization led to a subsequent period defined by the strict serial ordering of pitch classes, commonly referred to as "serialism" (c. 1919–1937).

Following the upheaval of the Second World War, a new approach to the post-tonal structuring of musical materials emerged, termed "total serialism" (c. 1945) for its serialization of additional musical parameters such as rhythm and dynamics. The tendency toward procedural evisceration of easily perceived repetition in order to create new forms of music was a key aesthetic goal of post-war modernist composers, many of whom pointed to a previous general precedent in the music of the Second Viennese School, and to the early post-tonal music of Anton Webern in particular (Erwin, 2021). However, several other key characteristics of post-war modernist music serve to shed light on the general renouncement of conspicuous repetition by composers of this period. Whether dogmatically serial in construction or not, modernist composers often eschew various cognitively salient facets of music that engender easy listening and a familiar sense of continuity and flow, including the use of octaves, consonance, and simple rhythmic grouping (Koivisto, 1996; Barton, 2012). Rather, modernist music is largely anti-thematic (and thus anti-repetitive) and places emphasis on the formal organization of various parameters of sound itself, such as pitch, duration, dynamics, and timbre.

For the post-war modernists, composition was seen as an experimental but highly structured process, one that underscored the acts of *making* and *analyzing* a work of music (Griffiths, 2011). That is to say, unity was privileged at the level of construction and made evident by the analysis of a work, but the music itself was composed intentionally to resist any easy-to-follow presentation of its organizing principles at the level of experience. Perhaps unsurprisingly, coherence as understood in tonal forms of music therefore becomes an easy and early casualty of the modernist artistic agenda to occult organization, wrenching musical variation away from a centralized and salient prototype such as a motive or theme. Difference was privileged over repetition (Campbell, 2013), resulting in the divorce of any strong sense of a work's internal unity from the act of listening. One conspicuous difference between music composed during the modernist period and, generally speaking, music composed post-1970, is the latter period's aesthetic reaction to and subversion of post-war modernism in the form of increased emphasis on experiential coherence (Hutchinson, 2016), often without forsaking structural rigor (Losada, 2009). Rather, the post-1970 era of post-tonal music articulates an affinity for music that affords novel forms of mnemonic salience (Pasler, 1993).

Empirical Studies on Post-tonal Music

The majority of empirical studies of similarity in post-tonal music have been conducted using music written for a single instrument. Krumhansl (1991) studied the function of memory for so-called "surface features" in the absence of the normative inter-opus cognitive schemata used for tonal music. She focused on the abstraction of features in post-tonal music and the ability to identify their characteristics, and to generalize surface features in subsequent passages. She found that listeners encoded and remembered a large amount of surface details in listening to a total serial piece for solo piano by Olivier Messiaen. Moreover, her results demonstrate that listeners abstract and retain knowledge regarding surface features and can accurately identify unfamiliar sections of the music as belonging to the same piece. Using eight 1-min long excerpts of pieces by Luciano Berio and Elliott Carter (each scored for a single instrument), Margulis (2013) had participants rate both original and modified recordings of post-tonal music. Modification involved artificially inserting repetitions into the original stimuli without specific regard for artistic or aesthetic considerations. Participants without previous experience of post-tonal music found the music to be more enjoyable and interesting when the excerpts contained added repetitions, regardless of whether the inserted repetition was immediate or placed later in the timeline of the excerpt. Importantly, as Margulis (2013, p. 54) notes,

> *End-of-session debriefings revealed that listeners were unaware that they had been exposed to the same excerpts in different conditions – they neither recognized that they had reheard particular examples in several forms, nor that the degree of internal repetition was varying from excerpt to excerpt. Thus, the differences in enjoyment ratings did not stem from conscious awareness of the relevant manipulations. Rather, higher degrees of repetition were associated with higher enjoyment ratings in such a way that listeners were unaware of this association.*

Although few studies to date examine post-tonal music composed for ensemble or orchestra, those that do have found evidence for the influence of timbre and changes to instrumentation on the identity of musical materials and the perception of similarity, the classification of related motives, and the experience of either repetition or segmentation in the large-scale unfolding of musical form (McAdams et al., 2004; Poulin-Charronnat et al., 2004; Taher et al., 2018).

The Present Study

Given the limited number of studies on post-tonal music and the direct relationship between repetition, musical form, and aesthetic preference, important questions remain unarticulated. How do the various strategies for repetition across styles of post-tonal music affect listening? How are non-verbatim repetition and aesthetic judgment related for a diverse population of listeners? And do previous findings with regard to aesthetic judgment hold up in ecologically valid paradigms? In the present experiment, 60 participants with diverse musical backgrounds were presented the first 3 min of all pieces in random order and were instructed to respond by pressing a button whenever they encounter a repetition in the music, where repetition is defined as something sounding familiar or self-similar within the piece. A second task asks participants to rate each excerpt for effects such as memorability, perceived coherence, and self-similarity. We predicted that (a) the total number of responses as well as the inter-participant agreement of responses would be high for classes I (tonal) and III (post-1970) and low for class II (modernist), and (b) that the total number of responses as well as the inter-participant agreement of responses would correlate with ratings for both overall memorability and coherence of the individual excerpts as well as ratings of aesthetic preference for these excerpts.

MATERIALS AND METHODS

Participants

Sixty listeners participated in the experiment, who were recruited as part of two groups: one group with and another group without experience in playing a musical instrument. The former group consisted of 30 participants (10 males, 20 females) who reported having played at least one musical instrument for more than 2 years. Participants in this group had a mean age of $M = 24.6$ years (STD = 3.3, range: 19–32) and had played their primary musical instruments for $M = 9.2$ years (STD = 6.3) and had received $M = 4.5$ years of music theoretical instruction (STD = 3.0). We measured musical training using the corresponding self-report inventory of the Goldsmiths Musical Sophistication Index (MSI, Müllensiefen et al., 2014), see the Appendix for details. The MSI yielded mean scores of 31.4 (STD = 20.3, range: 6–80) for the musician participants. One participant reported mild hearing loss, all other participants reported normal hearing. Another 30 participants (eight males, 22 females) reported not to have played a musical instrument for more than 2 years and had a mean age of M = 24.2 years (STD = 3.0, range: 19–33). In this group of participants, one participant reported moderate hearing loss. All participants received monetary compensation.

Stimuli

Fifteen musical excerpts were selected as stimuli for the present experiment. Excerpts were obtained from www.youtube.com. All stereo clips were peak-normalized in amplitude and converted to mp3 format (320 kbit/s). Of the 15 excerpts, three excerpts were from the tonal repertoire. Due to a technical problem, however, major portions of the data from one of the tonal excerpts (Manuel de Falla: *Ritual Fire Dance*) were lost. For that reason, only the data from the 14 other excerpts will be considered. **Table 1** provides information about the respective composers, titles, year, the performers and the duration and temporal placement of the excerpt used in the experiment.

The chosen excerpts were taken from ensemble and orchestra pieces that fall into one of three categories: *tonal*, *modernist*, and *post-1970*. Tonal pieces (by Pyotr Ilyich Tchaikovsky and Edvard Grieg) were selected for their clarity of repetition and comprehensibility. Modernist pieces (by Karlheinz Stockhausen, Iannis Xenakis, György Ligeti, Igor Stravinsky, Krzysztof Penderecki, and Pierre Boulez) were selected as representative pieces of the post-war era of musical modernism (the late

TABLE 1 | Excerpt information.

Composer	Title	Year	Performer	URL	Duration	Timing
I – Tonal						
Tchaikovsky	Violin Concerto (movement III)	1878	Janine Jensen, Deutsche Radio Philharmonie, Christoph Poppen	https://www.youtube.com/watch?v=KrVMmRWzRSM	3:06	0:00–3:06
Grieg	Holberg Suite (movement I)	1884	'A Far Cry' String Ensemble	https://www.youtube.com/watch?v=dFEBTbNs4yk	2:40	0:00–2:40
II – Modernist						
Stockhausen	Gruppen	1957	Berliner Philharmoniker, Friedrich Goldman (I) Claudio Abbado (II) Marcus Creed (III)	https://www.youtube.com/watch?v=CZ7jpKh_UF0	3:06	0:00–3:06
Xenakis	Achorripsis	1957	Luxembourg Philharmonic Orchestra, Arturo Tamayo	https://www.youtube.com/watch?v=rEyqJPW3Hi8	2:55	0:00–2:55
Ligeti	Apparitions	1959	Berlin Philharmonic Orchestra, Jonathan Nott	https://www.youtube.com/watch?v=pCS8DJJnxOE	3:07	0:00–3:07
Stravinsky	The Flood, (movement III)	1962	London Sinfonietta, Oliver Knussen	https://www.youtube.com/watch?v=ISCnHvosib4&list=OLAK5uy_klRrg8cPiONrF_ekQtLOIf5P0Xv2pC3Zg&index=6	2:34	0:00–2:34
Penderecki	Fluorescences	1962	Polish National Radio Symphony Orchestra, Antoni Wit	https://www.youtube.com/watch?v=DBbSZD2lkJI	3:01	0:00–3:01
Boulez	Figures-Doubles-Prismes	1968	BBC Symphony Orchestra, Pierre Boulez	https://www.youtube.com/watch?v=SKEBBKQ82_8&t=0	3:00	0:00–3:00
III – Post-1970						
Grisey	Partiels	1975	Asko Ensemble, Stefan Asbury	https://www.youtube.com/watch?v=1v7onrjN6RE&list=RD1v7onrjN6RE&start_radio=1&t=0	3:06	0:00–3:06
Dutilleux	L'arbre des songes	1985	Olivier Charlier, BBC Philharmonic, Yan Pascal Tortelier	https://www.youtube.com/watch?v=EDVFNh7MDQk	3:14	0:00–3:14
Pärt	Fratres	1991	Antal Eisrich and Miklós Kovács, Strings of Hungarian State Opera Orchestra, Tamás Benedek	https://www.youtube.com/watch?v=UlelRghsD_k	3:07	0:00–3:07
Hurel	Six miniatures en trompe-l'œil, (movement III)	1991	Ensemble Intercontemporain, Pierre Boulez	https://www.youtube.com/watch?v=lT4jQilFq8o	2:50	7:12–10:02
Boulez	Sur Incises	1998	Ensemble Intercontemporain, Matthias Pintscher	https://www.youtube.com/watch?v=HCQl6Wu3QxE	3:08	0:00–3:08
Romitelli	Flowing down too slow	2001	Musiques Nouvelles, Jean-Paul Dessy	https://www.youtube.com/watch?v=Xg5UQVa5CBA	3:02	0:00–3:02

1940's, the 1950's, and the 1960's). These pieces are generally characterized as difficult to follow by non-specialist audiences. Modernist music generally lacks clear repetitive structuring and do not often utilize a tonally-centered or consonant musical language. While the selected pieces for the post-1970 category (by Gérard Grisey, Henri Dutilleux, Arvo Pärt, Philippe Hurel, Pierre Boulez, and Fausto Romitelli) do not themselves constitute a particular style-grouping, they do share two general structural affinities relative to the modernist pieces: a more pronounced use of repetition and a noticeably more consonant musical language. The authors acknowledge that the selection of pieces is not, and indeed could not be representative of the many inter- and intra-stylistic idiosyncrasies, artistic agendas, and other genre-oriented determinants of music within and between the chosen categories.

Procedure

The experiment was implemented using the test platform www.testable.org. Participants were recruited from the online job board of the University of Oldenburg and received a private link that provided access to the experiment. They were instructed to listen with headphones to the presented experimental stimuli. For every stimulus, participants were first asked to indicate repetitions in the music as the music unfolded and second to respond to a set of eight questions. Specifically, they received the following instruction: "Please press the spacebar whenever you have the impression that some aspect of the music is repeating. This repetition does not need to be exact. Rather, you should press the spacebar whenever you feel a sense of repetition in the music. You should indicate a repetition at least once per excerpt." After the music had ended, participants received the following set of questions that were to be answered on scales from one to five: (Q1) Have you heard the piece before? (certainly not – certainly yes); (Q2) How confident are you in your ability to identify repetitions in this excerpt? (highly confident – highly unconfident); (Q3) How easy was it to follow this music? (very difficult – very easy); (Q4) How coherent was the piece? (highly coherent – highly incoherent); (Q5) How similar were individual moments of the piece? (highly similar – highly dissimilar); (Q6) Do you think you will recognize this piece if you hear it a week from now? (certainly yes – certainly no); (Q7) Did you like the piece? (certainly yes – certainly no); (Q8) Did you find the piece interesting? (certainly yes – certainly no). At the end of the experiment, there was a demographic questionnaire. The research reported in this manuscript was carried out according to the principles expressed in the Declaration of Helsinki and was approved by the ethics board of the University of Oldenburg.

Data Analysis

The relative timings of the spacebar taps were represented at a sampling rate of 1,000 Hz. Each tap was converted into a rectangular function of height one and with a width of 1 s, centered at the timepoint of the original tap. That is, per participant and per piece, repetition responses consisted of a sequence of zeros and ones that encoded indicated repetitions with a temporal granularity of 1 s. In order to obtain a measure of inter-participant synchronization, these series were averaged across participants. The resulting time-series of inter-participant synchronization corresponds to the proportion of participants that simultaneously indicated a repetition at a given point in time with a tolerance of plus/minus 1 s. Note that these data capture more fine-grained patterns of synchronization compared to histograms with, say, 1-s bins, because the latter approach does not appropriately represent closely spaced patterns at time points with non-integer periodicity. Significant periods of synchronization were assessed *via* bootstrapping: for every stimulus, synchronization time-series were computed for 60 randomly selected participants (drawn 1,000 times with replacement). If for a given time point, the first percentile of the bootstrapped distribution exceeded listeners' mean response rate across the entire excerpt, the time point was considered to exhibit significant inter-participant synchronization.

Two summary measures were computed from the repetition data: First, the raw number of indicated repetitions per participant and piece was computed ("#Rep"). Next, the duration of segments with significantly synchronized responses relative to the overall stimulus duration was considered ("Prop. Sync."). The data from questions Q1–Q8 were analyzed descriptively and by providing 95% CIs for means across participants. In order to explore the major factors underlying participants' responses, the data from variable (i.e., questions) Q1–Q8 together with the variable #Rep were z-normalized by participant and by variable, before a principal component analysis (PCA) with factor rotation was computed. The Prop. Sync. variable was not used in the PCA, because it yielded data on a group level and was not specific to individual participants. Finally, a linear mixed model (LME) was run to confirm effects of stimulus category (tonal, modernist, post-1970) and musical training on the two major factors identified by the PCA. The data analysis was implemented in MATLAB.[1] The LME was implemented in R using the lme4 package (Bates et al., 2014).

RESULTS

Figure 1 shows the repetition responses (as indicated by individual depressions of the space bar) for all 60 participants for four selected examples (blue and white dots correspond to participants with and without music training, respectively). As indicated by the figure, individual excerpts are characterized by different response rates; for instance, responses to Xenakis' *Achorripsis* are sparser compared to the other three examples. Whereas Xenakis does not visually exhibit points where responses appear to be synchronized across participants, Grieg's *Holberg Suite* and Pärt's *Fratres* show multiple timepoints where this appears to be the case. Even more clearly, Grisey's *Partiels* exhibits a quasi-periodic structure of strong synchronization across almost all participants. This is not to say that there is no individual variability. In fact, there are substantial differences in response rates across participants: 29 participants had average

[1]www.mathworks.com

response rates of less than five responses per piece, whereas the other half of participants had mean rates of 12.2 responses per piece (max = 35).

However, as is already visible in the displayed individual data in **Figure 1**, for the tapping data there were no indicative differences between the two groups of participants with and without musical training. Time series of synchronous responses were highly similar for both groups such that the distribution of differences between groups significantly deviated from zero only for less than 1 % of time points (M = 0.5%, STD = 0.4%, max = 1,1%).

Figure 2 displays the proportion of synchronous responses across participants and the time points (in red), where significant synchronization across participants is indicated (at an alpha

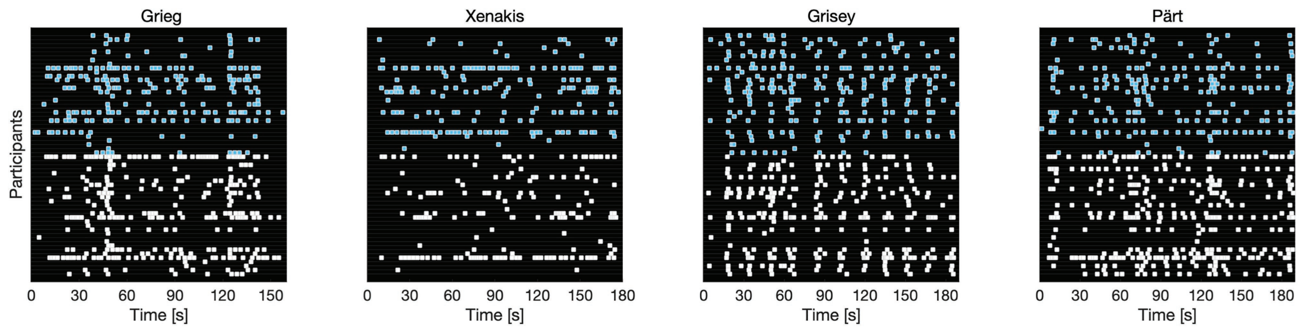

FIGURE 1 | Repetition response data from all 60 individual participants for four selected stimuli. White and blue dots correspond to participants with and without musical training, respectively.

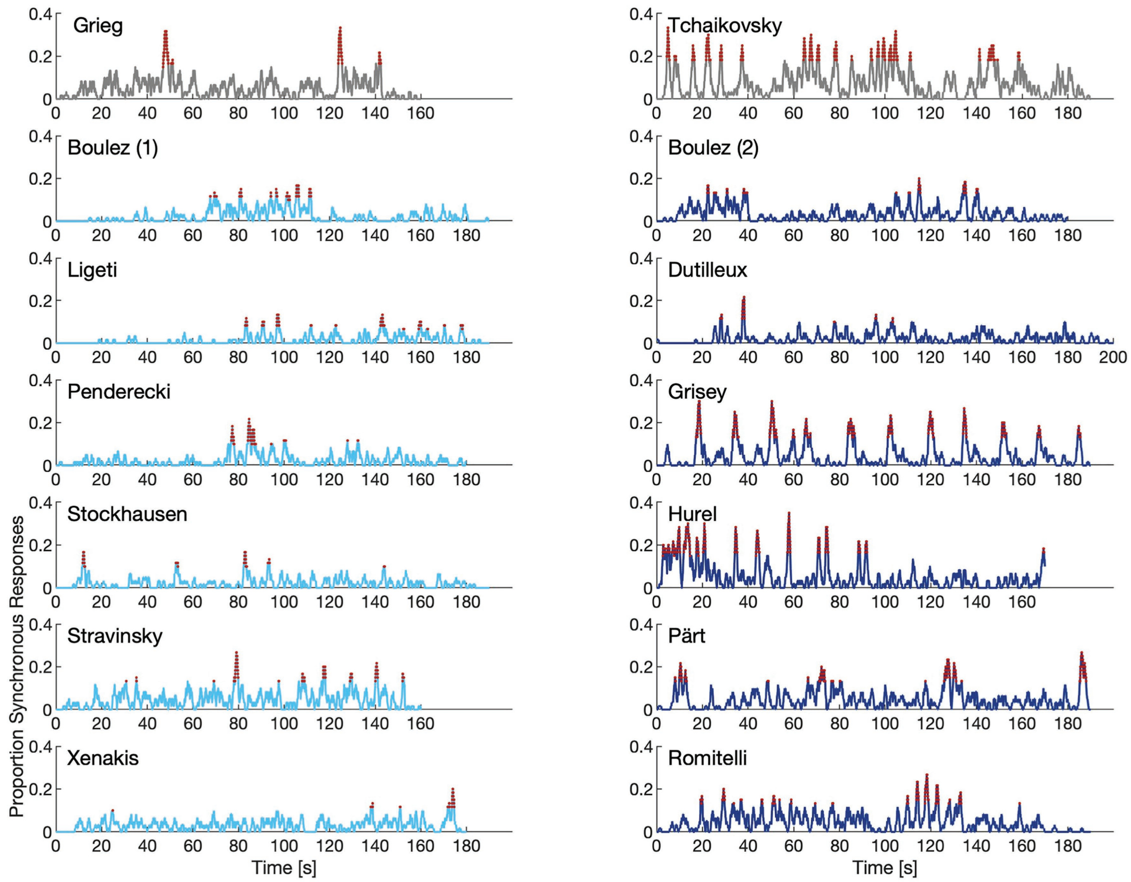

FIGURE 2 | Proportion synchronous responses across participants (with a tolerance of plus/minus 1 s). Red portions of the graphs indicate time points where the synchronization significantly differed from chance.

level of 0.01). These response signals show that for all excerpts in all categories, there are time-points of significant inter-participant synchronization. The excerpts of the modernist stimulus set tend to exhibit a sparser distribution of synchronization periods compared to post-1970 excerpts. Furthermore, individual excerpts show distinct response profiles, which becomes particularly pronounced when comparing response profiles within categories of excerpts, such as between Grieg and Tchaikovsky or Grisey and Hurel.

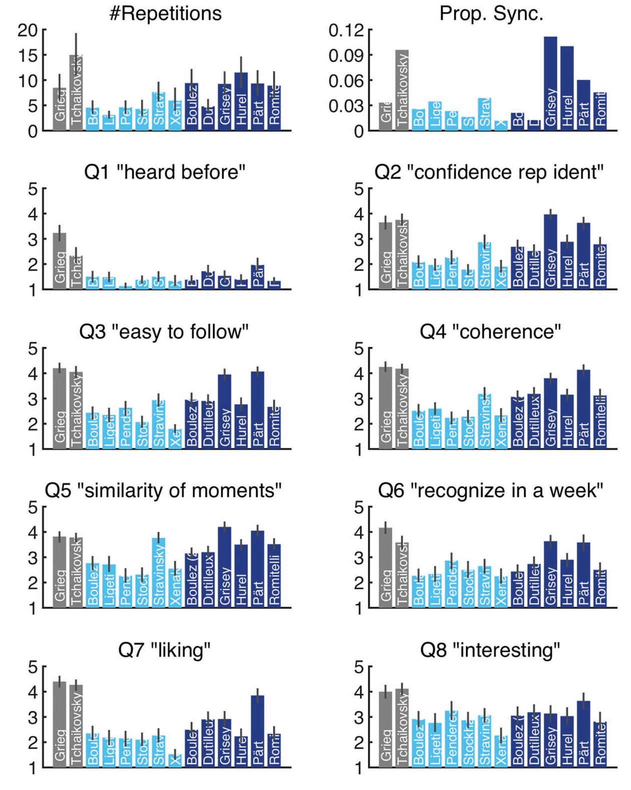

FIGURE 3 | Overview of responses in terms of the number of repetitions (#Repetitions), the proportion of time that a stimulus generated significant synchronization (Prop. Sync.), and responses to the eight questions Q1–Q8. Error bars correspond to 95% CIs.

Two indices of the repetition data (average #repetition responses, proportion of significant synchronization) plus the responses to questions Q1–Q8 are given in **Figure 3**. Indices of repetition responses show substantial differences across stimuli. Whereas there is a medial number of repetition responses and rather low proportion of synchronized responses for Grieg, Tchaikovsky yields many more as well as more strongly synchronized responses. The proportion of synchronization is relatively low for modernist excerpts but varies strongly for post-1970 excerpts with the highest values for Grisey, followed by Hurel, Pärt, and Romitelli. Considering the questions Q1–Q8, all but the two tonal excerpts were rated as unfamiliar in Q1. Within the modernist category of excerpts, Stravinsky stands out by receiving particularly high scores in Q2 (confidence rep. ident.), Q4 (coherence), and Q5 (similarity of moments). From the category of post-1970 excerpts, Grisey and Pärt receive particularly high scores in questions Q2–Q6. When it comes to Q7 (liking), however, only Pärt receives scores that approach those of the tonal excerpts by Grieg and Tchaikovsky.

To explore the underlying structure of these 10 variables (see **Figure 4A**, for a correlation matrix), a PCA on the data averaged across participants was computed. The first two components accounted for 92% of the variance in the data and exhibited a clear knee point in the scree plot, which is the reason why a two-dimensional representation was adopted. To increase interpretability, components were rotated using the varimax rotation. The resulting rotated factors are displayed in **Table 2**. Results indicate strong loadings of questions Q1

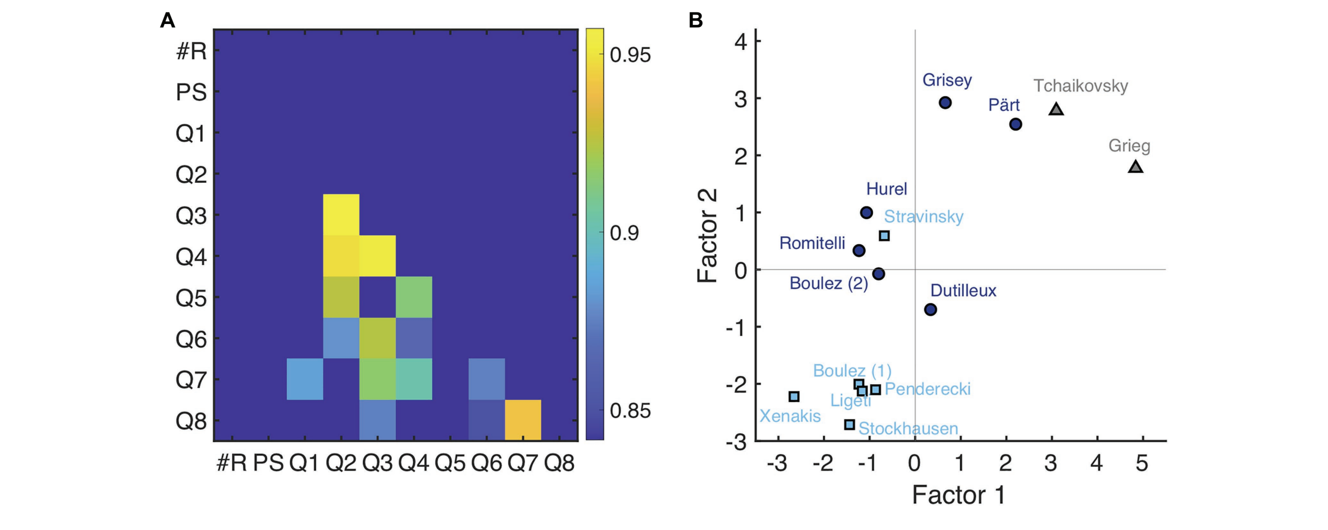

FIGURE 4 | (A) Linear correlation matrix displaying significant correlations according to Bonferroni-Holm correction for multiple comparisons. **(B)** Coordinates of individual excerpts according to Factor 1 (aesthetic preference) and Factor 2 (repetition strength) derived by principal component analysis (PCA) and varimax rotation.

(familiarity), Q7 (liking), and Q8 (interest) on Factor 1, highlighting that this first factor is dominated by aspects of familiarity and aesthetic preference. The number of repetitions together with Q5 (similarity of individual moments) strongly load on Factor 2, suggesting that this second factor reflects the perceived strength and frequency of repetitions in the music. Note that Factor 2 (repetition strength) robustly correlated with the proportion of synchronous responses, r = 0.83, CI: [0.54, 0.96], $p < 0.001$ [whereas there was only a marginal correlation for Factor 1, r = 0.51, CI: (−0.03, 0.81), p = 0.065]. This confirms that excerpts with higher scores of repetition strength also featured more moments of inter-subjective agreement about the presence or absence of repetitions. Taken together, the analysis indicates that the present data are determined by the two major factors of aesthetic preference and perceived repetition strength.

Figure 4B provides the coordinates of the mean responses for each piece with respect to Factors 1 and 2. Modernist excerpts tended to cluster according to both low aesthetic preference and repetition strength. An exception is the excerpt by Stravinsky, which scored higher in terms of repetition strength. Post-1970 excerpts had higher scores with respect to repetition strength (Factor 2). Most notably, Grisey and Pärt had similar scores in terms of repetition strength compared to the tonal excerpts. In terms of aesthetic preference, however, Tchaikovsky and Grieg yielded higher scores. Overall, three clusters emerged from the PCA: modernist excerpts with low aesthetic scores and low repetition strength, post-1970 excerpts with low aesthetic preference scores but medial repetition strength scores, and tonal or post-1970 excerpts with rather high aesthetic preference scores and high repetition strength.

TABLE 2 | First two components of the PCA after varimax rotation.

Variable	Factor 1	Factor 2
#Rep	−0.20	0.59
Q1	0.54	−0.13
Q2	0.09	0.42
Q3	0.26	0.26
Q4	0.21	0.31
Q5	0.06	0.52
Q6	0.34	0.15
Q7	0.45	−0.03
Q8	0.47	−0.02

The top three highest loadings per factor are marked in bold font. "#Rep" corresponds to the average number of repetitions per excerpt.

In a final step, the observation regarding the effects of stimulus category on aesthetic preference and perceived repetition strength was sought to be confirmed by means of regression modeling. Two separate LME were set up to test the fixed effects of stimulus category (I: tonal, II: modernist, III: post-1970) and MSI scores on aesthetic preference (Factor 1) and repetition strength (Factor 2). The data for these two dependent variables was derived by projecting the data of individual participants on the rotated factors that were derived from the PCA (see the **Supplementary Material** for a visualization). The random effects structure of the LME consisted of by-participant intercepts and slopes for the stimulus category and by-item (i.e., stimulus) intercepts.

Considering the aesthetic preference factor, estimated marginal means (95% CIs in square brackets) of the three categories were 1.38, CI: [0.79, 1.97], for tonal excerpts, −0.47, CI: [−0.80, −0.13], for the modernist excerpts, and 0.01, CI: [−0.33, 0.34], for the post-tonal excerpts. With aesthetic preference as dependent variable in the LME, there were significant differences between modernist and tonal excerpts [β = −1.85, CI: (−2.45, −1.24), $p < 0.001$] and between post-1970 and tonal excerpts [β = −1.38, CI: (−2.00, −0.76), p = 0.001], but only a marginal effect of musical training as measured by the MSI [β = −0.16, CI: (−0.34, 0.02), p = 0.091]. There was an interaction of the factor of musical training and the factor contrasting tonal and modernist excerpts [β = 0.24, CI: (0.03, 0.44), p = 0.028], see the **Supplementary Material** for the full set of statistics. As indicated in **Figure 5**, this interaction effect is also visible in a trend toward a negative correlation between participants' MSI scores and their raw scores along Factor 1 (averaged across excerpts) for tonal excerpts, r = −0.23, p = 0.08, together with a weak positive correlation for modernist excerpts, r = 0.31, p = 0.02. That is, in general the stimulus categories of modernist and post-1970 excerpts were less preferred compared to the tonal excerpts, but participants with more formal musical training appeared to show slightly higher aesthetic preference for modernist excerpts compared to participants with less musical training.

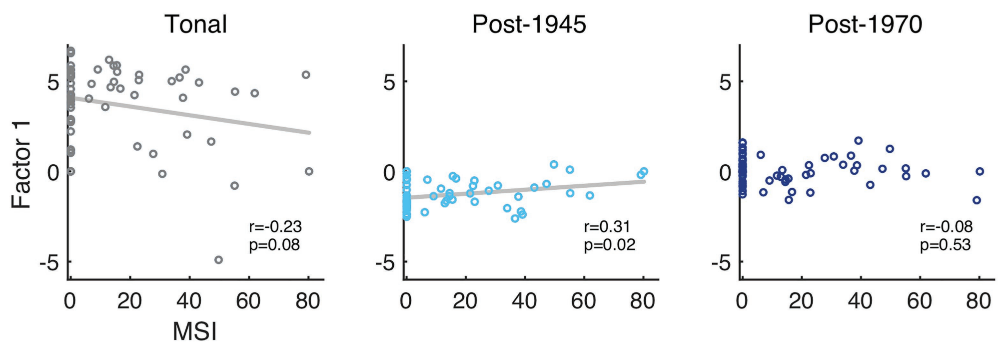

FIGURE 5 | Scatter plot of the Gold-Musical Sophistication Index (MSI) of musical training (Müllensiefen et al., 2014) and Factor 1 (aesthetic preference).

Considering the repetition strength factor, estimated marginal means were 0.82, CI: [0.10, 1.54], for the tonal excerpts, −0.64, CI: [−1.05, −0.22], for the modernist excerpts, and 0.36, CI: [−0.05, 0.78], for the post-1970 excerpts, indicating least repetition strength for the modernist excerpts. With repetition strength as dependent variable in the LME, there was a significant effect of modernist excerpts compared to tonal excerpts [$\beta = -1.45$, CI: (−2.20, −0.71), $p = 0.003$], but repetition strength did not differ for post-1970 excerpts compared to tonal excerpts [$\beta = -0.46$, CI: (−1.20, 0.28), $p = 0.25$] and there also was no effect of musical training [$\beta = 0.02$, CI: (−0.17, 0.12), $p = 0.74$] and no interaction effect, see the **Supplementary Material** for the full set of statistics. Hence, modernist excerpts showed significantly less repetition strength compared to tonal and post-1970 excerpts.

DISCUSSION

Presenting excerpts from the post-tonal repertoire, we observed that indications of repetitions in the music in a group of 60 listeners with and without musical training showed significant periods of interindividual synchronization. Ratings of a set of eight questions that probed aspects listeners' (Q1) familiarity with the piece, (Q2) confidence of repetition responses, and judgments of affordances for (Q3) easy listening, (Q4) coherence, (Q5) similarity of moments, and (Q6) recognition, as well as their (Q7) liking, and (Q8) interest in the piece indicated particularly strong correlations between questions Q2, Q3, and Q4. A PCA on questions Q1 – Q8 and the number of repetition responses as well as the proportion of timepoints with significant inter-participant synchronization of responses suggested that two major factors account for 92% of variance in the data. We interpreted these factors in terms of aesthetic judgment (dominated by Q1, Q7, and Q8) and repetition strength (dominated by the number of repetitions and Q5).

Regression modeling indicated that repetition strength generally differed across excerpt category, with modernist excerpts featuring lower repetition strength compared to post-1970 and tonal excerpts. Aesthetic preference, on the other hand, was significantly lower both for excerpts from modernist and post-1970 pieces compared to excerpts from tonal pieces. The analysis did not reveal any differences in response behavior as a function of musical training with regards to the perception of repetitions, neither as measured by the repetition strength factor, nor for the time series based on the synchronous responses. Note that a comparable independence of musical training and response behavior has been observed in the segmentation literature (Hartmann et al., 2016; Popescu et al., 2021). With regards to Factor 1 (aesthetic preference), however, the regression model indicated a (comparatively weak) interaction effect of excerpt category and musical training, which was based on an association between preference scores for modernist excerpts and the level of musical training of participants, demonstrating a slight rise in preference for post-tonal music by listeners with previous musical training.

Repetition and Aesthetic Preference as Two Determinants of Post-tonal Music

Previous research in the psycho-aesthetics of music make persistent mention of the co-dependency between complexity and exposure (also sometime called familiarity) and their mediating influence on aesthetic judgments. Based on a hypothesis proposed by Berlyne (1971), the inverted-U model holds that aesthetic preference ratings are likely to be highest when the stimulus is within an optimal intermediate range that lies between the extremities of underdetermined and overly determined complexity. Several studies that examine the effects of familiarity, coherence, and repetition on aesthetic preference judgments have appealed to the inverted-U model to help explain their results after repeated exposures to music with an intervening period of time (Hargreaves, 1984; North and Hargreaves, 1995; Peretz et al., 1998; Tan et al., 2006). Moreover, Margulis (2013) suggests that given the complexity of post-tonal musical structure, and the relative unfamiliarity most participants have with modernist music, the inverted-U model might also help to account for within-composition repetition, as repeated exposure within a limited timeframe may attenuate the perceived complexity of the repeated events. This suggestion conforms to the habituation theory proposed by Huron (2013). A recent review of the music psychology literature asserts the inverted-U model's efficacy in accounting for a large amount of data (Chmiel and Schubert, 2017).

Taken together, the findings of the present study both complement and challenge earlier work that suggests a link between repetition and aesthetic preference in music listening. In what may be the only previous study devoted exclusively to repetition in post-tonal music, Margulis (2013) showed that aesthetic judgments skew upward when repetitions are inserted into excerpts artificially. It is important to note that participants in study of Margulis (2013) were not asked to identify repetitions, and in many cases were not aware that they had heard multiple versions of the same music containing differing degrees of repetitive materials. Her findings signal the important role repetition plays in listening to music that lacks easily cognized recursion at the structural level and demonstrate that post-tonal music can be made more aesthetically pleasing with the additional perceptual parsing that the repetition of surface events provides.

However, our results may be interpreted to indicate that repetitiveness *per se* is not as strongly correlated to aesthetic preference as the previous literature on both tonal and post-tonal music would seem to indicate. Rather, our data suggests a two-factor model wherein aesthetic preference and repetition strength may be considered to some degree separate determinants of the listening experience. Although none of the excerpts presented within this study both scored highly on the aesthetic preference factor, while at the same time scoring low on the repetition strength factor – a fact that conforms to findings of Margulis (2013) – several observations gleaned from our data converge to demonstrate a degree of independence between repetition and aesthetic preference.

First, there are important discrepancies between the number of repetitions recorded for a given excerpt and the proportion of intersubjective synchronizations for these repeats, demonstrating that not all repetitions are equally perceptually salient (see **Figure 2**). This finding may also indicate that certain repetitions are perceptually more prominent, and therefore more important to the formal unfolding of a given excerpt than others. Second, synchronous responses do not always correlate with aesthetic qualities (see **Figures 2, 3**). For instance, "liking" ratings for Tchaikovsky and Grieg were the highest recorded, while the proportion of synchronous responses for the Grieg excerpt are substantially curtailed relative to the Tchaikovsky. Curiously, the confidence self-report for identifying repetitions in the Grieg excerpt is among the highest across excerpts. It is important to note that both of these pieces belong to the tonal category, and therefore the presence of tonal structure cannot account for differences in the results.

Similar discrepancies can be observed for the post-1970 category. For example, the Hurel excerpt displays a large amount and high degree of synchronous responses, but scores relatively low for liking, high for similarity of moments and coherence, and low for ease of following. Conversely, the Dutilleux excerpt displays an extremely low number of repetitions and low participant synchronicity, yet scored moderately well in liking, interest, and ease of following. Although the variety of considered composers and excerpts was naturally limited in this study, these findings may generally be taken to indicate that repetition strength exists somewhat independently of aesthetic preference. As our cross-stylistic results demonstrate, the mere presence of strongly rated repetition does not in itself guarantee aesthetic preference, but in its absence, at least in most cases, aesthetic preference appears to be clearly diminished. Further comparison of the response data together with the quantitative measures for repetition within individual excerpts suggests that additional factors contribute to preference judgments. For example, repetition strength and frequency for the Hurel excerpt would seem to suggest that the music is easy to attend to; repetitions are both prominent and obvious to listeners, at least for the first half of the excerpt. Moreover, ratings for Q5 (similarity) suggest that in general, participants felt the music to be coherent. However, participant ratings for Q3 (easy to follow) and Q4 (coherence) display slightly diminished ratings, and the score for Q7 (liking) reveals the excerpt to not only be the least liked within its category, but also rated on part with the modernist excerpts.

A more pronounced discrepancy is evident when we compare results for the Grisey and Pärt excerpts with those for Grieg and Tchaikovsky, the four excerpts that occupy the upper right quadrant of **Figure 4B**. First, participant agreement and synchronization are both high and consistent across the Grisey excerpt (**Figure 2**). From a listener's perspective, the repeating event can be characterized as low in complexity and evolves only minimally across the excerpt. Repetitions across the excerpt are the most periodic of all the stimuli. For the most part, qualitative response data for Grisey are scored quite high. Questions 2 (repetition confidence), 3 (easy to follow), 4 (coherence), 5 (similarity), and 6 (memorability) all suggest a deeply engaging piece of music. However, relative to the other members of the post-1970 category, interest is flat. And, perhaps more importantly, compared to the two tonal pieces and fellow category member Pärt, Q7 (liking) is rated surprisingly low. Taken together with the lack of development across repetitions, the regularity of the period points to high predictability and rapid habituation as a potential explanation. Excerpts for Pärt and Grieg scored disproportionately high on Q7 (liking) despite receiving far fewer synchronous responses relative to Tchaikovsky, Grisey, and Hurel. In the case of Grieg, the proportion of synchronous is low relative not only to the Tchaikovsky excerpt, but also to Grisey, Hurel, and Pärt. Despite this finding, relatively high scores persist for the Grieg excerpt across all other qualitative measures and culminate in the highest score for liking.

Our data leaves several important questions open to further investigation. The general across-category results would seem to suggest that further underlying parameters play into determinations of aesthetic preference beyond the repetition-liking paradigm hitherto acknowledged by previous literature and supported by the present experiment. Given the structural differences between tonal and post-tonal music, our data suggests that across category differences in responses involves both stylistic considerations and may also involve structural-harmonic differences. However, the relatively consistent categorical differences reported here between what we have termed modernist and post-1970 forms of post-tonal music suggest that further empirical delineation of stylistic, structural, and cognitive elements of post-tonal music are required in order to develop a more complete explanation of aesthetic preference.

With a few important and noteworthy exceptions (McAdams et al., 2004; Poulin-Charronnat et al., 2004), previous research on similarity, repetition, and coherence in Western classical music has limited its stimuli to single instruments. While stimuli for the present study were drawn from the repertoire for ensemble and orchestra, where multiple instrumental parts contribute to complex textures, our experimental design does not take into account the important and intricate sonic differences between simple and complex textures. Moreover, ecological validity of the excerpts was a crucial element in the experimental design that prevents us from testing additional contributing variables such as the differences in harmonic configuration and syntax between tonal and post-tonal music. Although both modernist and post-1970 music are structured using post-tonal configurations, modernist excerpts contain a low number of internal repetitions and score lowest for aesthetic preference. However, post-1970 excerpts, while containing a proportionately larger number of repetitions, score disproportionally lower than tonal excerpts. With the caveat that more research is desirable in order to more comprehensively understand the relationship between repetition and aesthetic preference, and a wider array of compositions inclusive of the plethora of styles that have developed in contemporary music need to be considered, our findings suggest that in addition to repetition, factors including tonality and harmonic configuration, the distribution of repetition, and the specific nature of repeated events, need to be accounted for in research on the perception and cognition of post-tonal 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 Kommission für Forschungsfolgenabschätzung und Ethik. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MT and KS designed the experiment and co-wrote the manuscript. KS and AK collected and analyzed the data. MT, AK, and KS edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by a Freigeist Fellowship of the Volkswagen Foundation to KS.

ACKNOWLEDGMENTS

The authors wish to thank the two reviewers for their insightful comments, which helped to improve the manuscript, and the editors of the research topic on Neuroscience and New Music. The authors further thank Merle Schlender for data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.673706/full#supplementary-material

REFERENCES

Barton, S. D. (2012). Understanding Musical Discontinuity. Dissertation. Charlottesville, VA: University of Virginia.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv [Preprint]. doi: 10.18637/jss.v067.i01

Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. New York: Appleton-Century-Crofts.

Boss, J. (2000). The "musical idea" and global coherence in Schoenberg's atonal and serial music. *Intégral* 14, 209–264.

Cambouropoulos, E. (2001). Melodic cue abstraction, similarity, and category formation: a formal model. *Music. Percept.* 18, 347–370. doi: 10.1525/mp.2001.18.3.347

Campbell, E. (2013). *Music After Deleuze*. New York: Bloomsbury Academic.

Carpenter, P., and Neff, S. (1995). "Commentary," in *The Musical Idea and the Logic, Technique, and Art of Its Presentation*. ed. A. Schoenberg (New York: Columbia University Press), 1–74.

Chmiel, A., and Schubert, E. (2017). Back to the inverted-U for music preference: a review of the literature. *Psychol. Music* 45, 886–909. doi: 10.1177/0305735617697507

Deliège, I. (2001a). Similarity perception↔ categorization↔ cue abstraction. *Music Percept.* 18, 233–243. doi: 10.1525/mp.2001.18.3.233

Deliège, I. (2001b). Prototype effects in music listening: an empirical approach to the notion of imprint. *Music Percept.* 18, 371–407. doi: 10.1525/mp.2001.18.3.371

Deliège, I. (2007). Similarity relations in listening to music: how do they come into play? *Music Sci.* 11, 9–37. doi: 10.1177/1029864907011001021

Erwin, M. (2021). An apprenticeship and its stocktakings: Leibowitz, Boulez, Messiaen, and the discourse and practice of new music. *Twentieth Century Music* 18, 71–93. doi: 10.1017/S1478572220000213

Griffiths, P. (2011). *Modern Music and After*. New York: Oxford University Press.

Hargreaves, D. J. (1984). The effects of repetition on liking form. *J. Res. Music Educ.* 32, 35–47. doi: 10.2307/3345279

Hartmann, M., Lartillot, O., and Toiviainen, P. (2016). Multi-scale modelling of segmentation: effect of music training and experimental task. *Music Percept.* 34, 192–217. doi: 10.1525/mp.2016.34.2.192

Huovinen, E. (2013). Concatenationism and anti-architectonicism in musical understanding. *J. Aesthet. Art Critic.* 71, 247–260. doi: 10.1111/jaac.12018

Huron, D. (2013). A psychological approach to musical form: the habituation-fluency theory of repetition. *Curr. Musicol.* 96, 7–36. doi: 10.7916/D8KP81FG

Hutchinson, M. (2016). *Coherence in New Music: Experience, Aesthetics, Analysis*. New York: Routledge.

Koivisto, T. T. H. (1996). The Moment in the Flow: Understanding Continuity and Coherence in Selected Atonal Compositions. Dissertation. Ann Arbor, MI: University of Michigan.

Kozak, M. (2019). *Enacting Musical Time: The Bodily Experience of New Music.* New York: Oxford University Press.

Krumhansl, C. L. (1990). Tonal hierarchies and rare intervals in music cognition. *Music Percept.* 7, 309–324. doi: 10.2307/40285467

Krumhansl, C. L. (1991). Memory for musical surface. *Mem. Cogn.* 19, 401–411. doi: 10.3758/BF03197145

Lamont, A., and Dibben, N. (2001). Motivic structure and the perception of similarity. *Music Percept.* 18, 245–274. doi: 10.1525/mp.2001.18.3.245

Lerdahl, F., and Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Percept.* 1, 229–252. doi: 10.2307/40285257

Levinson, J. (1997). *Music in the Moment*. Ithaca: Cornell University Press.

London, J., Cox, A., Morrison, C. D., Maus, F. E., Repp, B. H., and Levinson, J. (1999). Music in the moment: a discussion. *Music Percept.* 16, 463–494. doi: 10.2307/40285805

Losada, C. C. (2009). Between modernism and postmodernism: strands of continuity in collage compositions by Rochberg, Berio, and Zimmermann. *Music Theor. Spect.* 31, 57–100. doi: 10.1525/mts.2009.31.1.57

Margulis, E. H. (2013). Aesthetic responses to repetition in unfamiliar music. *Empir. Stud. Arts* 31, 45–57. doi: 10.2190/EM.31.1.c

McAdams, S., Vieillard, S., Houix, O., and Reynolds, R. (2004). Perception of musical similarity among contemporary thematic materials in two instrumentations. *Music Percept.* 22, 207–237. doi: 10.1525/mp.2004.22.2.207

Meyer, L. (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

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

Murphy, G. L., and Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychol. Rev.* 92:289.

North, A. C., and Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicol. Music Mind Brain* 14, 77–93. doi: 10.1037/h0094090

Pasler, J. (1993). Postmodernism, narrativity, and the art of memory. *Contemp. Music Rev.* 7, 3–32. doi: 10.1080/07494469300640011

Peretz, I., Gaudreau, D., and Bonnel, A.-M. (1998). Exposure effects on music preference and recognition. *Mem. Cogn.* 26, 884–902. doi: 10.3758/BF03201171

Popescu, T., Widdess, R., and Rohrmeier, M. (2021). Western listeners detect boundary hierarchy in Indian music: a segmentation study. *Sci. Rep.* 11:3112. doi: 10.1038/s41598-021-82629-y

Poulin-Charronnat, B., Bigand, E., Lalitte, P., Madurell, F., Vieillard, S., and McAdams, S. (2004). Effects of a change in instrumentation on the recognition of musical materials. *Music Percept.* 22, 239–263. doi: 10.1525/mp.2004.22.2.239

Reybrouck, M. (2020). Music listening as adaptive behavior: enaction meets neuroscience. *J. Interdiscip. Music Stud.* 10, 34–58. doi: 10.25364/24.10:2020.1.3

Rosch, E. (1975). Cognitive representations of semantic categories. *J. Exp. Psychol. Gen.* 104, 192–233. doi: 10.1037/0096-3445.104.3.192

Taher, C., Hasegawa, R., and McAdams, S. (2018). Effects of musical context on the recognition of musical motives during listening. *Music Percept.* 36, 77–97. doi: 10.1525/mp.2018.36.1.77

Taher, C., Rusch, R., and McAdams, S. (2016). Effects of repetition on attention in two-part counterpoint. *Music Percept.* 33, 306–318. doi: 10.1525/mp.2016.33.3.306

Tan, S.-L., Spackman, M. P., and Peaslee, C. L. (2006). The effects of repeated exposure on liking and judgments of musical unity of intact and patchwork compositions. *Music Percept.* 23, 407–421. doi: 10.1525/mp.2006.23.5.407

Tillman, B., and Bigand, E. (1996). Does formal musical structure affect perception of musical expressiveness? *Psychol. Music* 24, 3–17. doi: 10.1177/0305735696241002

Tversky, A. (1977). Features of similarity. *Psychol. Rev.* 84, 327–352. doi: 10.1037/0033-295X.84.4.327

Van den Toorn, P. (1996). What's in a motive? Schoenberg and Schenker reconsidered. *J. Musicol.* 14, 370–399. doi: 10.1525/jm.1996.14.3.03a00050

Zbikowski, L. M. (1999). Musical coherence, motive, and categorization. *Music Percept.* 17, 5–42. doi: 10.2307/40285810

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: 27 August 2021
doi: 10.3389/fnins.2021.674050

Suppression, Maintenance, and Surprise: Neuronal Correlates of Predictive Processing Specialization for Musical Rhythm

Ulvhild Færøvik[1]*, Karsten Specht*[1,2,3] *and Kjetil Vikene*[1,3]*

[1] Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway, [2] Department of Education, The Arctic University of Norway, Tromsø, Norway, [3] Mohn Medical Imaging and Visualization Centre, Haukeland University Hospital, Bergen, Norway

OPEN ACCESS

Edited by:
Thomas James Lundy, Cuttlefish Arts, United States

Reviewed by:
Daya Shankar Gupta, Camden County College, United States
Lu Zhang, Georgia Institute of Technology, United States

***Correspondence:**
Kjetil Vikene
kjetil.vikene@uib.no

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

Received: 28 February 2021
Accepted: 17 June 2021
Published: 27 August 2021

Citation:
Færøvik U, Specht K and Vikene K (2021) Suppression, Maintenance, and Surprise: Neuronal Correlates of Predictive Processing Specialization for Musical Rhythm. Front. Neurosci. 15:674050. doi: 10.3389/fnins.2021.674050

Auditory repetition suppression and omission activation are opposite neural phenomena and manifestations of principles of predictive processing. Repetition suppression describes the temporal decrease in neural activity when a stimulus is constant or repeated in an expected temporal fashion; omission activity is the transient increase in neural activity when a stimulus is temporarily and unexpectedly absent. The temporal, repetitive nature of musical rhythms is ideal for investigating these phenomena. During an fMRI session, 10 healthy participants underwent scanning while listening to musical rhythms with two levels of metric complexity, and with beat omissions with different positional complexity. Participants first listened to 16-s-long presentations of continuous rhythms, before listening to a longer continuous presentation with beat omissions quasi-randomly introduced. We found deactivation in bilateral superior temporal gyri during the repeated presentation of the normal, unaltered rhythmic stimulus, with more suppression of activity in the left hemisphere. Omission activation of bilateral middle temporal gyri was right lateralized. Persistent activity was found in areas including the supplementary motor area, caudate nucleus, anterior insula, frontal areas, and middle and posterior cingulate cortex, not overlapping with either listening, suppression, or omission activation. This suggests that the areas are perhaps specialized for working memory maintenance. We found no effect of metric complexity for either the normal presentation or omissions, but we found evidence for a small effect of omission *position*—at an uncorrected threshold—where omissions in the more metrical salient position, i.e., the first position in the bar, showed higher activation in anterior cingulate/medial superior frontal gyrus, compared to omissions in the less salient position, in line with the role of the anterior cingulate cortex for saliency detection. The results are consistent with findings in our previous studies on Parkinson's disease, but are put into a bigger theoretical frameset.

Keywords: musical rhythm and beat processing, temporal cortex, predicitve maintenance, supression, surprisal

INTRODUCTION

New musical forms and experiments often challenge established structural forms in composition or reshape old ones into new uses, reordering fundamental musical building blocks to challenge the perception of music. The challenge of studying such diverse music in a principled and systematic manner within a context of neuroscience demands first to establish some fundamental and perhaps common mechanisms in music perception, a work that has been flourishing in the last decades. In this article, we focus on rhythm, one of these fundamental building blocks. Within new music, some artists work with rhythmic entropy while others try to dispel rhythm altogether. Insights into some basic mechanisms of the "listening apparatus" in our nervous system as they pertain to perception of musical rhythms might perhaps still be of use to both researchers and artists, either as a starting point for more advanced research or as a starting point of artistic defiance.

In short, auditory repetition suppression and omission activation, which this study address, are opposite neural phenomena and manifestations of principles of predictive processing. Repetition suppression is the reduction of neuronal activity during listening to a repeated sound that is *present*, omission activation is the increased neuronal activity that occurs when an expected sound is *not* present. Listening activates bilateral superior temporal cortices, independent of whether we are exposed to tones, words, animal, and instrumental sounds (Specht and Reul, 2003), while prolonged listening to an unchanging sequence of sounds quickly leads to a deactivation of neural activity, a phenomenon known as repetition suppression (Grill-Spector et al., 2006). Auditory repetition suppression is a robust, experience-dependent adjustment of neural functions (Grotheer and Kovacs, 2016), predictability (Costa-Faidella et al., 2011; Cacciaglia et al., 2019), and prior expectation (Summerfield et al., 2008; Todorovic et al., 2011). It is modulated by a range of factors, such as time scales of presentation rates, sequence position and stimuli similarities (Kovács and Schweinberger, 2016), and stimuli-specific characteristics (Linke et al., 2011), and is also task-dependent (Arnott et al., 2005). Repetition suppression also neatly demonstrates principles of predictive coding (Friston, 2005; Baldeweg, 2006; Auksztulewicz and Friston, 2016), where repetition suppression biases the activity in sensory cortices (Sreenivasan et al., 2014). For auditory stimuli, repetition suppression can be seen in the temporal cortices, including Herschel's gyrus (HG), superior temporal gyrus (STG), and middle temporal gyrus (MTG) (Auksztulewicz and Friston, 2016; Cacciaglia et al., 2019).

Repetition suppression is related to working memory mechanisms, where attenuation of neural activity can be interpreted as a minimization of activity needed in a working memory maintenance stage (Kumar et al., 2016) through the suppression of irrelevant information (Linke et al., 2011; Ahveninen et al., 2017). It is, however, unclear whether working memory is dependent on persistent neural activity (for maintaining the information) (Huang et al., 2016) or on transient reorganization of synaptic weights (D'Esposito and Postle, 2015) in representational states (Myers et al., 2017; Sreenivasan and D'Esposito, 2019) or a combination of the two (Sreenivasan et al., 2014). A potential theoretical (or actual) difference between these two descriptions—working memory as either persistent or transient neural states [or what within a predictive coding framework can be called "representation units" (Clark, 2013)]—could lie in a distinction between auditory sensory memory (shorter low-level sensory cortical retention intervals) and higher-level working memory network organization (Nees, 2016). Different cortical activation for simple sequence processing and more complex, task-specific working-memory maintenance could point to such nuances (Brechmann et al., 2007). Neural correlates for auditory working memory have been shown in temporal cortices, including STG, HG, and planum temporale (PT) (Brechmann et al., 2007; Kumar et al., 2016). Furthermore, distinct neural differences between perceptual processing and active working memory tasks for melody and pitch (Zatorre et al., 1994), separate neural correlates for duration-based and beat-based auditory timing (Teki et al., 2011), and differentiations for melody and rhythm have been shown—with the right inferior frontal gyrus and insula particularly involved (Jerde et al., 2011).

Within a predictive coding framework, auditory prediction errors (Friston, 2005) must depend on working memory or sensory memory mechanisms, since they occur when an unchanging and predictable sequence of sounds suddenly changes, i.e., when predictions and expectations are breached, or when the incoming sensory signal does not match the "representation unit" (Clark, 2013); what we will henceforth call the *representational maintenance*. The concept of prediction errors [or "surprisal" (Clark, 2013)] draws on findings in EEG/ERP-studies on mismatch negativity (MMN) (Naatanen et al., 1978; Kompus et al., 2015). MMN potentials are measurable neural spikes triggered by deviant and rare stimuli in a chain of standard stimuli, where the difference between the deviant and the standard stimuli are proportional to the deviance [see (Näätänen, 1992) for a review]. This difference is the *mismatch* or prediction error (Friston, 2005). In fMRI, as in EEG/ERP studies, the size of activation reflects the magnitude of the MMN deviant (Mathiak et al., 2002; Liebenthal et al., 2003; Eichele et al., 2008). Omissions are a particularly interesting type of deviant stimuli, where omission activation (Raij et al., 1997; Wacongne et al., 2011) describes cortical activation as a result of *missing* stimuli in a predictable sequence of sounds, and can therefore be assumed to be generating internal responses based solely on expectancy or prediction, and not by a change in the deviant characteristics in the stimuli itself (Jongsma et al., 2005). As with repetition suppression, MMNs or prediction error magnitude is modulated by numerous factors (Näätänen, 1992), and as with repetition suppression and auditory working memory, specific parts of the temporal cortices have repeatedly been shown to be involved in the reporting of such prediction errors (Friston, 2005). Imaging and electrophysiological studies have consistently shown that a main source of MMN potentials is located in the intersection of STG, PT, and HG (Recasens et al., 2014), predominantly right lateralized (Tervaniemi et al., 2000; Opitz et al., 2002; Doeller et al., 2003; Rinne et al., 2005), also for omissions (Mustovic et al., 2003; Voisin et al., 2006;

SanMiguel et al., 2013a,b), although one study has found omission activation predominantly on the left (Nazimek et al., 2013). Pertaining to our study, these cortical areas are sensitive to beat and pattern deviations, as shown in both EEG/ERP and fMRI studies (Tervaniemi et al., 2000; Opitz et al., 2002; Doeller et al., 2003; Mustovic et al., 2003; Rinne et al., 2005; Voisin et al., 2006; Nazimek et al., 2013; Recasens et al., 2014). Beat omission and positional saliency have been used to investigate rhythm and pattern-related phenomena with several imaging and neurophysiological techniques, indicating different levels of magnitude for salient and less salient metric beat positions, although findings are somewhat ambiguous (Winkler and Schroger, 1995; Jongsma et al., 2003, 2005, 2006; Ladinig et al., 2009; Wacongne et al., 2011; Salisbury, 2012; Bouwer et al., 2014; Damsma and van Rijn, 2016).

In the current study, we wanted to examine repetition suppression, representational maintenance, and omission activation during the perceptual processing of musical rhythms. In short, these three phenomena can be seen as manifestations of key principles in predictive processing frameworks, and musical rhythms are ideal stimuli to operationalize and demonstrate these principles because of their predictive, temporal nature (Vuust and Witek, 2014; Koelsch et al., 2019). Musical rhythms also facilitate operationalization of modulating factors such as contextual characteristics (simple or complex rhythms) and saliency (position of omission).

In addition, the neuronal mechanisms involved in the perception of musical rhythms are partly known, which makes it possible to compare our results with existing literature. Listening to rhythms activates cortical motor areas, such as premotor cortex and supplementary motor area (SMA), the basal ganglia, as well as large-scale networks across the brain (Grahn and Brett, 2007; Chen et al., 2008a,b; Bengtsson et al., 2009; Geiser et al., 2009, 2012; Grahn, 2009; Trost et al., 2014; Large et al., 2015).

We also wanted to examine the effect of pattern *complexity* on repetition suppression and omission activation and the effect of *positional saliency* on omission activation. To this end, two musical rhythms, one simple and one complex, were presented several times to the participants during scans. The first part of the stimuli presentation consisted of a short presentation (16 s) of continuous, unperturbed rhythmic repetition to examine repetition suppression, segueing into a longer continuous presentation of the same rhythm. In this second part, an overt target-detection task (with a quasi-randomly distributed deviant tone) was introduced to keep the participant attending to the rhythm, while quasi-randomly distributed beat omissions were used to covertly examine omission activation (see section "Materials and Methods" for more details on the stimuli and the paradigm).

Based on previous literature, we expected to see listening, repetition suppression, and omission activation in largely overlapping areas in the temporal cortices, with omission activation occurring in more posterior areas than suppression and maintenance. We hypothesized that representation maintenance would also occur in the temporal cortices, with additional activation of inferior frontal areas, insula, and premotor cortices. We hypothesized that complexity would differentially affect suppression, as the encoding stage presumably would be affected by a higher cognitive load for the complex rhythm. We also hypothesized that omission activation would be modulated by rhythmic context where the cognitive demand (i.e., higher neural activity in a representational maintenance) of the more complex rhythm would result in smaller omission sizes, and furthermore that *positional* saliency would modulate omission activation, where a more salient position (beat position number one) would show higher activation than the less salient position (beat position number two). Finally, we hypothesized that there would be an interaction between pattern complexity and beat position for omission activation.

MATERIALS AND METHODS

Participants

Participants were recruited among Norwegian-speaking students enrolled at the University of Bergen (UiB). Fourteen participants underwent scanning with fMRI, but four were excluded due to head movement in the scanner. Analyses were done on the remaining 10 (6 females, mean age = 24.4). Eight were right-handed by self-report. The number of years participants had played instruments (outside mandatory music lessons in public school) was done by self-report. A participant was labeled as a musician if s/he had 5 or more years of consistent instrument practice, and 5 out of the 10 participants reached this target. Personal data were coded and stored offline and anonymity was assured. All procedures were approved by the Regional Committee for Medical and Health Research Ethics (REK no 2014/1915) and carried out in accordance with the code of Ethics of the World Medical Association, Declaration of Helsinki. Upon enrollment in the study, all participants gave written informed consent to participate in the study and were rewarded 50NOK for their participation.

Stimuli

We used two musical rhythmical stimuli of different rhythmic complexity described elsewhere (Vikene et al., 2018, 2019). The stimuli consisted of deep, multilayered synthesizer bass sounds in two octaves and a sampled bass drum sound, to place the general character of the stimuli in a different frequency range than the Eigenfrequency of the scanner during the echo planar imaging (EPI) sequence. The first 8 bars (16 s) of each stimulus contained an alternating piano chord, at the first position to clearly mark the beginning of the bar. For each rhythm (simple/complex), these chords were composed in one major and one minor mode for listening variation (no tests where planned for the effect of mode). The remaining 44 bars/88 s of each stimuli were constructed with quasi-distributed overt deviant probe tones (consisting of a six-note up-shift of tonality, to keep the participants attending to the musical rhythms), and covert beat omissions. Probe tones were always placed on the first position of the bar, while omissions were placed in equal numbers on first and second positions. Each of the four versions of the stimuli blocks (simple/complex vs. major/minor) had six omissions. The omissions were either at the first or second position of the rhythmic patterns. In two versions of the blocks, three omissions were at the first position and three omissions were at the second position. In one version of the blocks, four omissions were at the first position and two were at the second position. In one version of the blocks, two omissions were at the first position and four were at the second position. The smallest time between two consecutive omissions was 8.5 s; the longest was 17.5 s. Because we were interested in the covert, or passive, detection of omission, and not active detection success, and because target detection of the probe tones involved button-pushing on a hand grip leading to motor area activity, no analysis was planned for the overt probe tone detection. All participants did, however, correctly detect all target tones. Stimuli were created using Steinberg Cubase 7 and presented in the scanner with EPrime 2.0 (Ver 2.3 Professional), which was also used to collect responses to the overt task.

Experimental Design

Participants were given earplugs and were placed comfortably in the scanner. They were given fMRI-compatible headphones with additional physical noise cancelation foamed ear plugs. Participants were also fitted with fMRI-compatible video goggles and a handgrip with buttons to respond to the overt target-detection attentional task. After initial structural scans and a 5-min-long resting-state fMRI scan (not part of this report), goggles were turned on and participants were given instructions for the study. Participants were told to keep eyes open and look at a cross in the middle of the screen and asked to press a button on the hand grip when the probe tone was detected. Instructions were followed by a short test run before scanning started. Before each trial, the same written instructions were repeated in the goggles (4.5 s), followed by a blank screen and silences ranging from 13 to 19 s. When the music stimuli began playing, a cross was presented in the goggles as a focus point to minimize head movement. Each sound file was presented twice during the scan, in randomized order between subjects. Total scan time for the paradigm was 33 min. See **Figure 1** for an overview of the paradigm.

Data Acquisition and Pre-processing

fMRI images were acquired using a 3-T scanner (GE Signa Excite 750) with a 32-channel coil. Repetition time (TR) for the EPI sequence was 1.5 s, echo time (TE) was 30 ms, voxel size was $3.44 \times 3.44 \times 5$ with 28 slices interleaved, for 1325 volumes. Pre-processing steps included realignment (0.9 quality, 5 mm smoothing kernel, registered to first image with second-degree B-spline), unwarping (using 12×12), resliced to mean image, normalized to ICBM template (with $2mm^3$ voxel size), and smoothing with Gaussian kernel ($5\ mm^3$).

First-Level Analysis

We aligned the onset of stimuli epochs (below) to 13th of the 28 interlaced slices. A high-pass filtering threshold was set at 1/249 Hz cutoff (calculated as the mean between onsets of the stimuli blocks). Single-subject data were analyzed by specifying a general linear model, and for the whole scan, movement-related variance (realignment parameters) was included in the model as six covariates of no interest.

Each block was modeled as follows: 4.5 s of on-screen instructions were labeled as "READ." Silence periods (randomly assigned between 13 and 19 s) between each block were not segmented and thus served as contrast for all other epochs ("REST"). The whole block was segmented into 2-s bins, i.e., the total length of one whole rhythmic pattern. The first 4 s of each block were labeled "LEARNING"; the next 12 s, "SUPPRESSION." In the following 88 s of each block, bins containing a probe tone were labeled "PROBE," bins containing the normal, unperturbed version of the rhythmic pattern were labeled "MAINTENANCE," while 2-s bins containing an omission were labeled according to the position of the omission (i.e., "OMISSION1" for the first position-omission). The 2-s bin immediately following an omission was labeled "NOT OF INTEREST" to avoid any secondary effects of the omission spilling into the segments labeled "MAINTENANCE."

All blocks were divided into simple and complex rhythm, and segmentation of different epochs were labeled "SIMPLE" or "COMPLEX." For example, the "LEARNING" epochs in the simple rhythm were called "SIMPLE_LEARNING"; for the omission in position 2 in the complex rhythm: "COMPLEX_OMISSION2." First-level analysis produced 10 contrasts for "SIMPLE/COMPLEX" blocks, with "LEARNING/SUPPRESSION/MAINTENANCE/OMISSION (1/2)" epochs, all with REST epochs subtracted for the contrasts.

Second-Level Analysis (Initial)

A full factorial analysis—akin to a repeated measures ANOVA—was conducted, with MUSICIAN (musician/non-musician), RHYTHM ("SIMPLE/COMPLEX"), and TYPE ["LEARNING/SUPPRESSION/MAINTENANCE/OMISSION (1/2)] as dependent factors. The analysis was examined with a threshold of family-wise error (FWE) correction for multiple comparisons at $p < 0.05$, with a minimal voxel-cluster size of at least 10 voxels. No main effect or interaction effects were found for neither MUSICIAN nor RHYTHM, but as expected, a main effect for TYPE was found. Since we had a clear hypothesis on omission complexity and position, we did, however, probe these comparisons through t-tests, but did not find any significant differences on omission, neither between rhythms nor positions at a FWE-corrected level. For completeness, we would nonetheless mention that at an uncorrected level ($p < 0.001$, cluster size of 100 voxels), we found higher activation in an area in the anterior cingulate cortex (ACC)/medial prefrontal gyrus (see **Supplementary Figure 1** and **Supplementary Table 1**).

Second-Level Analysis (Reduced Model)

Based on the lack of main and interaction effects for RHYTHM, and lack of significant t-test results on OMISSION, we decided to re-segment the data, dispelling of the division into two rhythms, as well as omission position. Furthermore, based on the lack of main and interaction effects of the categorical MUSICIAN group division, we instead included the number of years playing an instrument as a covariate in the analysis. A new full factorial analysis was

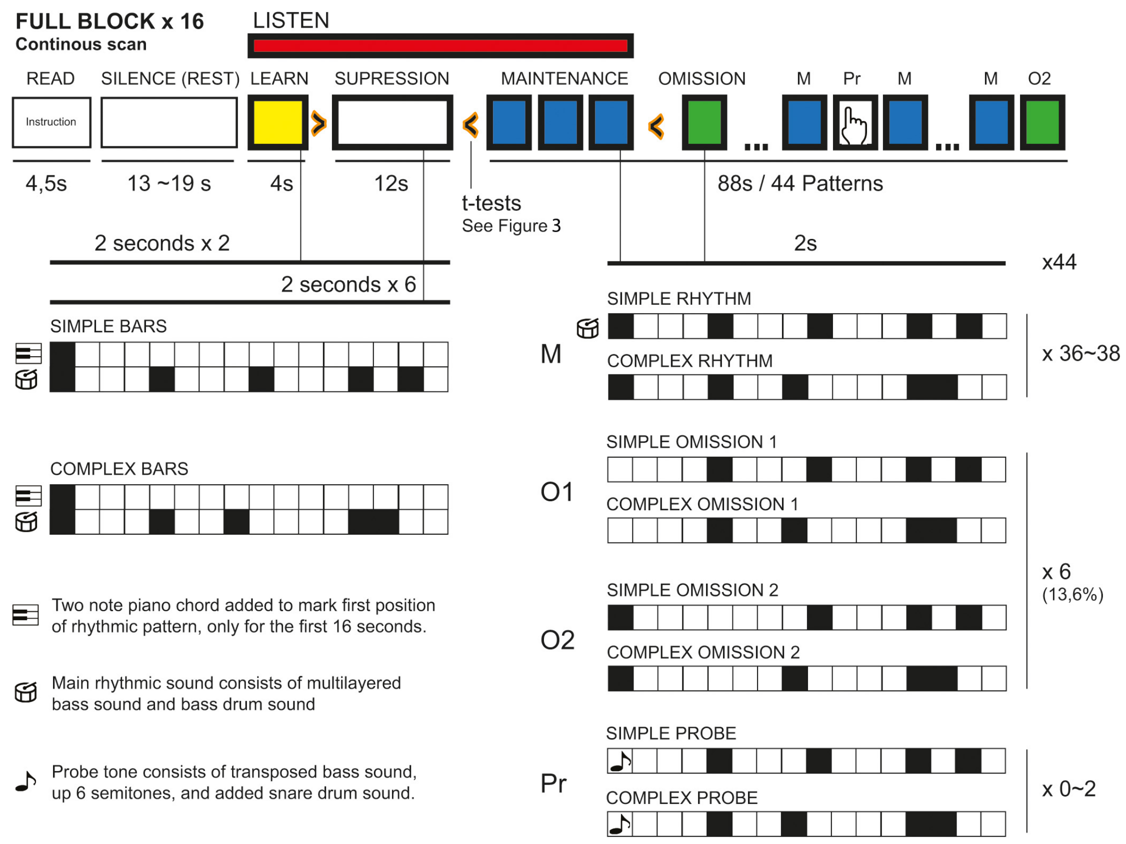

FIGURE 1 | Construct of fMRI paradigm.

therefore conducted using only TYPE as dependent factors, i.e., "LEARNING/LISTENING/MAINTENANCE/OMISSION." A main effect for TYPE was found [$F(1,31) = 18.36$, $p < 0.001$], and we proceed to do t-tests for our planned comparisons.

RESULTS

All results are reported with a FWE-corrected threshold of $p < 0.05$ and at least 10 voxels per cluster. **Figure 2** shows a detailed excerpt of the frontal right hemisphere of the findings listed below. **Figure 3** shows a more detailed overview of the findings across the while brain.

Overall Listening

For the combined LEARNING, SUPPRESSION, and MAINTENANCE bins, we found activations in large parts of bilateral superior and middle temporal gyri (STG/MTG), as well as bilateral angular gyri, SMA, cerebellum, and posterior areas [fusiform, occipital, posterior cingulate cortex (PCC)] (**Table 1**).

Learning/Suppression

We contrasted the first 4 s of the introduction of music, LEARNING, by subtracting the following 12 s of SUPPRESSION. This showed more bilateral activations across many areas, including STG, angular gyri, and cerebellum (crus 1 and vermis 6), the basal ganglia (caudate nucleus and putamen), and thalamus, in the first 4 s (**Table 2**).

Maintenance

We subtracted the last of the 12 s of the initial listening (SUPPRESSION) from MAINTENANCE to examine which brain areas, after the initial listening period, showed more activation during maintenance of the rhythm throughout the remaining 88 s. The rationale behind this contrast was that after the first 4 s (LEARNING), the next 12 s (SUPPRESSION) consolidates the (reduced) neural activation related to the establishment of a predictive model of the rhythmic pattern, while MAINTENANCE represents additional areas needed to keep the rhythmic pattern in working memory pertaining to the overt task. Areas including SMA; caudate nucleus; anterior insula (AIN); superior, middle, and inferior frontal gyrus (SFG/MFG/IFG); and middle and posterior cingulate cortex, as well as parts of the occipital cortex, showed more activation during MAINTENANCE than during SUPPRESSION (**Table 3**).

Omission

We contrasted OMISSION by subtracting MAINTENANCE to examine which brain areas were activated during beat omissions.

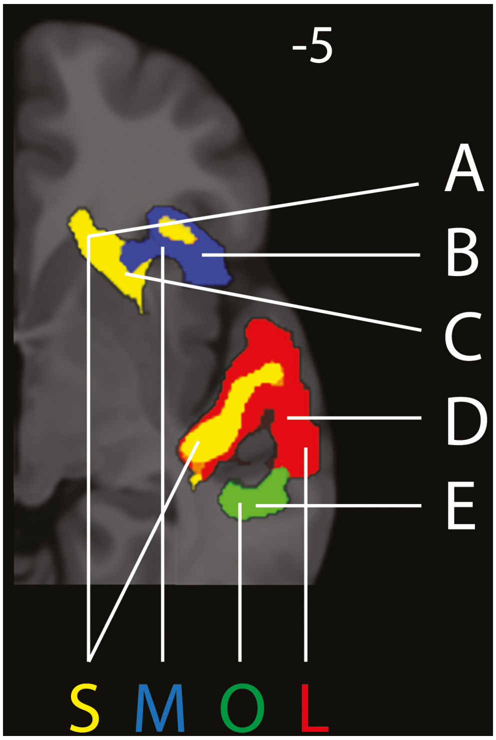

FIGURE 2 | Excerpt from right hemisphere of separation of (in red) LISTENING (L), (in yellow) SUPRESSION (S), (in blue) MAINTENANCE (M) and (in green) OMISSION (O). A = Caudate Nucleus, B = Anterior Insula, C = Putamen, D = Superior Temporal Gyrus, E = Middle Temporal Gyrus. Panel at MNI z = –5 on the axial plane. (Figure made partly with MRIcroGL, with small cluster removal).

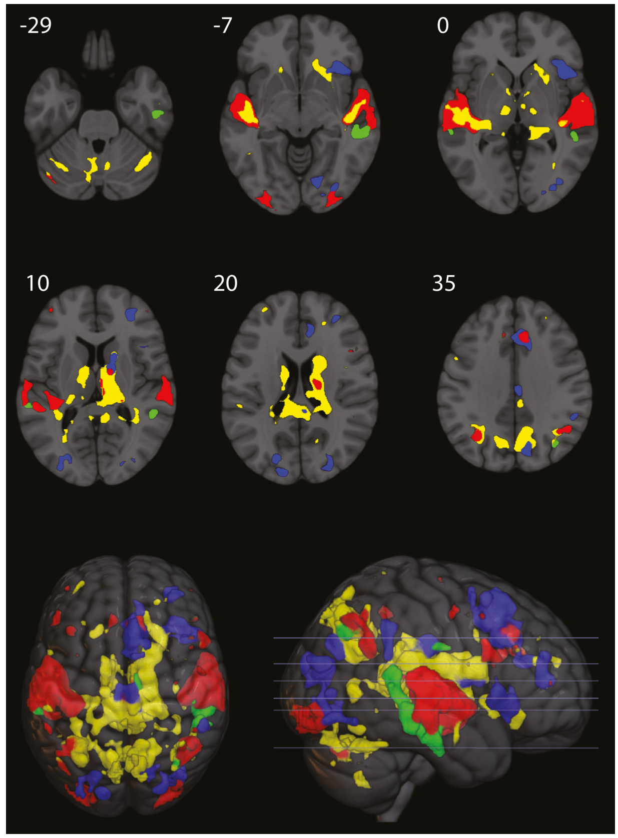

FIGURE 3 | Red areas, overall activity during listening for the first 16 s. Yellow areas, where activity decreased in the last 12 s, compared to the first 4 s of listening. Blue areas, persistent activity for the remaining 88 s of listening (only for epochs of normal presentation of the rhythm). Green areas, beat omission activity in the middle temporal gyrus. Numbers refer to MNI coordinate on the axial plane (figure made partly with MRIcroGL, with small cluster removal).

The rationale behind this contrast was that MAINTENANCE, as stated above, represents areas needed to keep the rhythmic pattern in working memory, while OMISSION represents the neural activation related to the prediction error triggered by the missing beat.

For OMISSION, more activation was seen, predominantly in the right MTG, extending from the inferior to the superior temporal gyrus. In addition, a smaller activation was seen in the left MTG as well as in the right angular gyrus (**Table 4**).

DISCUSSION

Listening to musical rhythms predictably activated the bilateral STG as well as bilateral angular gyrus. Areas in the STG attenuated after the first 4 s of repetition of the rhythmic patterns overlapped exclusively with these areas, with the size of deactivation being larger in the left STG. In addition, activity in bilateral cerebellum, and the (predominantly right) basal ganglia, including caudate nucleus, putamen, and thalamus, decreased after the first 4 s. Since processing of music (Zatorre and Zarate, 2012) and particularly rhythm (Thaut et al., 2014) has been found to be right lateralized (Large et al., 2015), this larger deactivation in the left STG might reflect an asymmetric allocation of resources, where—after initial processing in sensory cortices—the processing of musical rhythms is predominantly done in the right STG. The rapid deactivation of the basal ganglia and cerebellar areas points to a role for these areas in initial beat detection (Peretz and Zatorre, 2005; Grahn, 2009).

During the prolonged listening to rhythms after the initial 16 s of encoding, larger activity was found in the SMA and the caudate nucleus, areas well known to be activated by rhythm (Grahn and Brett, 2007; Bengtsson et al., 2009). Areas related to attention, such as the anterior insula and frontal areas (the inferior frontal gyrus in particular, but also middle and superior frontal areas), were also activated. These areas have been directly implemented in rhythm perception (Chapin et al., 2010; Heard and Lee, 2020), in particular as crucial in working memory for rhythm (Jerde et al., 2011). Activity in the precuneus, the posterior part of the cingulate cortex, and middle prefrontal cortex has in addition been found to play a particular role in the maintenance of musical beats with high beat salience (Toiviainen et al., 2020), which the rhythms in the current study must be characterized as. The areas found to be more activated during maintenance (anterior insula, frontal areas, SMA, posterior cingulate, and precuneus), with more activity in the right hemisphere, were clearly distinct compared to the other conditions and could indicate that working memory mechanisms of representational maintenance are allocated in different areas of the brain, separate from primary cortices activated during initial sensory processing.

Omission activation was distinctly more posterior than listening, suppression, and maintenance, located predominantly in the MTG, and, as expected, omission activation was significantly bigger in the right MTG (Raij et al., 1997) (see **Figure 3** and **Table 4**), with coordinates closely matching those found in previous fMRI (Mustovic et al., 2003; Voisin et al., 2006) and EEG (SanMiguel et al., 2013a,b) studies on omissions and silences in healthy controls and in our previous fMRI study on persons with Parkinson's disease (Vikene et al., 2019).

In addition, at uncorrected levels, the more salient position of the omission showed a higher activation in the ACC, consistent with previous research implicating the ACC in saliency detection (Menon, 2011) (**Supplementary Figure 1** and **Supplementary Table 1**).

On a theoretical level, our findings can be interpreted as manifestations of crucial principles in predictive processing frameworks, where repetition suppression (Friston, 2005; Baldeweg, 2006; Auksztulewicz and Friston, 2016) can be interpreted as model building; maintenance (Brechmann et al., 2007; Kumar et al., 2016) as a "representation unit" (Clark, 2013); and omission activation (Raij et al., 1997; Wacongne et al., 2011)—an internal response based solely on expectancy or prediction (Jongsma et al., 2005)—as a prediction error (Friston, 2005). Despite the limitations in this study, we will claim that it robustly shows results in line with previous findings on repetition suppression and omission activation—and the perception of musical rhythms in general—and that our paradigm, concretely or abstractly, can be used as a starting point for more refined studies of predictive processing mechanisms for the perception of musical rhythms.

TABLE 1 | General effect of listening (LEARNING + LISTENING + MAINTENANCE).

		MNI				
	Region	***X***	***Y***	***Z***	**Size**	***t***
R	Superior temporal gyrus	42	−20	−6	1146	11.97
	Extending to	50	−10	−10		10.75
	Middle temporal gyrus	56	−4	−8		10.32
L	Superior temporal gyrus	−46	−8	−6	1518	11.73
	Extending to	−44	−16	−8		11.34
	Middle temporal gyrus	−52	−16	−4		10.71
R	Angular gyrus	34	−84	−16	61	8.70
		32	−94	−16		6.20
		28	−94	−8		6.12
L	Cerebellum crus 1	−46	−72	−34	22	8.28
R	Angular gyrus	42	−56	28	50	7.86
L	Fusiform	−24	−96	−20	81	7.48
L	Inferior occipital cortex	−28	−88	−16		7.47
		−42	−88	−10		6.78
R	Supplementary motor area	0	18	48	29	7.32
		2	24	42		6.29
R	Posterior cingulate cortex	4	−22	28	49	7.22
L	Angular gyrus	−34	−64	36	49	7.19
		−36	−56	40		6.62
R	Supplementary motor area	10	24	36	24	6.81
R	Thalamus	4	−14	16	10	6.76
L	Middle temporal gyrus	−48	32	26	12	6.75
R	Inferior frontal gyrus	56	14	26	20	6.59
R	Inferior occipital cortex	28	−84	−4	10	6.52

All results reported as t-tests with family-wise error (FWE) correction at $p < 0.05$, cluster size of 10 voxels. Coordinates listed without cluster size belong to the previous listed cluster. Size = number of voxels activated in the cluster.

TABLE 2 | Effect of Time (LEARNING > SUPPRESSION).

		MNI				
	Region	***X***	***Y***	***Z***	**Size**	***t***
R	Thalamus	6	−22	12	1650	10.38
	Caudate nucleus	12	−2	14		10.36
	Hippocampus	12	−28	8		9.7
R	Precuneus	8	−66	34	357	8.86
		10	−72	48		7.62
		12	−70	26		7.55
L	Precuneus	−10	−68	40	94	8.56
		−8	−76	44		6.26
R	Superior temporal gyrus	44	−20	−6	78	8.46
	Middel temporal gyrus	50	−12	−10		7.26
L	Angular gyrus	−32	−68	44	180	8.42
	Inferior parietal lobule	−32	−62	36		8.24
L	Thalamus	−12	−6	12	114	7.98
	Caudate nucleus	−14	0	24		6.11
L	Cerebellum crus 1	−4	−80	−20	26	7.8
L	Cerebellum crus 1	−44	−72	−34	16	7.72
L	Cerebellum 6	−30	−66	−26	38	7.62
L	Superior temporal gyrus	−44	−22	−4	182	7.62
	Inferior temporal gyrus	−40	−26	−18		7.6
	Middle temporal gyrus	−52	−24	0		7.22
R	Putamen	22	22	−4	57	7.61
R	Cerebellum crus 1	48	−48	−28	52	7.59
		38	−60	−28		7.46
L	Hippocampus	−28	−28	0	101	7.32
		−34	−32	8		6.76
	Thalamus	−22	−28	14		6.19
L	Superior temporal gyrus	−66	−22	10	39	7.26
		−60	−26	6		7.07
L	Planum temporale	−34	−44	8	25	7.26
L	Thalamus	−10	−14	2	24	7.12
L	Precuneus	−8	−78	54	17	6.69
R	Rectus	10	30	−12	15	6.56
R	Angular gyrus	40	−62	38	10	6.17

All results reported as t-tests with family-wise error (FWE) correction at $p < 0.05$, cluster size of 10 voxels. Coordinates listed without cluster size belong to the previous listed cluster. Size = number of voxels activated in the cluster.

Limitations

Due to the small sample of participants ($n = 10$) taken from a homogenous population of Western Educated Industrialized

TABLE 3 | Maintenance of rhythmic patterns (MAINTENANCE > SUPPRESSION).

		MNI				
	Region	***X***	***Y***	***Z***	**Size**	***t***
R	Supplementary motor area	4	12	56	366	8.97
		2	20	54		8.88
	Middle cingulum cortex	10	24	36		8.03
R	Precuneus	12	−74	32	61	8.23
R	Anterior insula	34	28	−4	97	8.23
		42	16	−2		7.02
R	Posterior cingulum cortex	4	−28	26	147	8.16
		−4	−20	28		7.17
		12	−36	24		6.49
R	Caudate nucleus	14	−2	14	53	7.91
		16	8	12		6.54
R	Inferior frontal gyrus	50	20	4	21	7.39
R	Lingual	16	−78	−8	17	7.28
L	Superior occipital	−14	−92	20	26	7.21
R	Superior frontal	30	48	10	22	7.02
L	Middle occipital	−20	−78	20	30	6.91
R	Inferior occipital	34	−80	−14	31	6.77
		30	−82	−6		6.55
R	Middle frontal gyrus	46	18	42	10	6.77
R	Middle occipital	28	−84	20	17	6.49

All results reported as t-tests with family-wise error (FWE) correction at p < 0.05, cluster size of 10 voxels. Size = number of voxels activated in the cluster.

TABLE 4 | Omission activation (OMISSION > MAINTENANCE).

		MNI				
	Region	***X***	***Y***	***Z***	**Size**	***t***
R	Middle temporal gyrus	60	−30	−12	197	7.41212416
		50	−26	−12		7.2957921
	Inferior temporal gyrus	52	−16	−30		6.50495815
L	Middle temporal gyrus	−56	−20	−20	23	7.19963741
R	Middle temporal gyrus	50	−42	6	17	6.6976552
R	Angular gyrus	36	−72	38	13	6.42334318

All results reported as t-tests with family-wise error (FWE) correction at p < 0.05, cluster size of 10 voxels. Coordinates listed without cluster size belong to the previous listed cluster. Size = number of voxels activated in the cluster.

Rich Democratic (WEIRD) students of psychology, our findings are difficult to generalize. During a continuous scanning paradigm, the scanner is never silent, which makes the study of omissions questionable, although our results are consistent with previous research. Previous studies have also shown differences in omission detection between musicians and non-musicians (Ono et al., 2013, 2015), but we did not find a difference between them in this study. This might be a result of a low number of participants in the study. Using musical stimuli at "ecological" tempo (120 bpm), with 250 ms ISI in the isochronous metric framework, meant that omissions had to be distributed at fairly long intervals (between 8 and 17.5 s), and to amass sufficient instances of the omissions for statistical analysis, the total scan time for the paradigm was long (33 min). The lengthy paradigm could have affected levels of vigilance and attention during the scan and, as a consequence, could have influenced the results. We did try to remedy this by adding an overt target-detection task, which all participants performed correctly. Furthermore, we only used two rhythms, both of which were repeated several times during the scans. The consequence of longer-term habituation and learning effects could therefore also have influenced the results. In addition, the lack of effects on complexity in our study could indicate that the two rhythms chosen for the study did not differ (enough) in their level of complexity to yield such differences. Finally, due to the poor temporal resolution of fMRI, we might not have been able to pick up finer details of the mechanisms we have tried to describe. Future studies should try to limit paradigm length and use more varied and perhaps "real" musical samples, also with more levels of complexity. Factors such as musical aptitude, level of vigilance, and general working memory capacity should also be taken into consideration in future studies.

CONCLUSION

Our study successfully replicated previous findings for repetition suppression and omission activation and shows that tailored musical stimuli can be used in an fMRI setting to robustly investigate such neural phenomena, even with a limited number of participants ($n = 10$). Importantly, our findings show a clear separation between repetition suppression and prediction error activation, and additionally indicate that representational maintenance activates areas different from those deactivated during repetition suppression. While listening and subsequent repetition suppression were located mainly in anterior parts of the superior temporal gyrus, prediction errors (omission activation) were clearly separated from these areas and located mainly in posterior parts of the MTG. Representational maintenance activated SMA, caudate nucleus, anterior insula, frontal areas, and middle and posterior cingulate cortex, potentially showing persistent representational maintenance activity in areas separate from initial listening (encoding), repetition suppression (attenuation), and prediction error (omission).

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 Regional Committee for Medical and Health Research Ethics, REK-VEST (2014/1915). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

UF, KS, and KV contributed in the conception of the research project, did statistical analysis, and contributed to the final version of the manuscript. UF recruited the participants, organized the study, and wrote the first draft. KV designed the paradigm. All authors contributed to the article and approved the submitted version.

FUNDING

This project was financed through a grant from the Research Council of Norway (Grant Number: 217932/F20) awarded to KS. The salary of UF was covered in part by a master student scholarship at the University of Bergen, and in part by a grant from the Research Council of Norway (Grant Number: 260576) awarded to Stefan Koelsch. KV has received traveling grants from Meltzer's Fund, University of Bergen, Norwegian Research School of Neuroscience, and Department of Biological and Medical Psychology, University of Bergen, and financial support through GC Rieber Foundations and the Research Council of Norway (Grant Number: 217932/F20).

ACKNOWLEDGMENTS

The authors wish to thank the participants who took part in this study. A particular thank you to the staff at the Department of Radiology, University Hospital of Haukeland, Bergen, for their invaluable help with scanning the participants.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.674050/full#supplementary-material

Supplementary Figure 1 | Visualization of findings in **Supplementary Table 1**.

Supplementary Table 1 | Omission Position 1 > Position 2 across rhythms. Results reported as *t*-tests uncorrected at $p < 0.001$, cluster size of 100 voxels. ACC, anterior cingulate cortex; MPFC, medial prefrontal cortex.

REFERENCES

Ahveninen, J., Seidman, L. J., Chang, W. T., Hamalainen, M., and Huang, S. (2017). Suppression of irrelevant sounds during auditory working memory. *Neuroimage* 161, 1–8. doi: 10.1016/j.neuroimage.2017.08.040

Arnott, S. R., Grady, C. L., Hevenor, S. J., Graham, S., and Alain, C. (2005). The functional organization of auditory working memory as revealed by fMRI. *J. Cogn. Neurosci.* 17, 819–831. doi: 10.1162/0898929053747612

Auksztulewicz, R., and Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex* 80, 125–140. doi: 10.1016/j.cortex.2015.11.024

Baldeweg, T. (2006). Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends Cogn. Sci.* 10, 93–94. doi: 10.1016/j.tics.2006.01.010

Bengtsson, S. L., Ullen, 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

Bouwer, F. L., Van Zuijen, T. L., and Honing, H. (2014). Beat processing is pre-attentive for metrically simple rhythms with clear accents: an ERP study. *PLoS One* 9:e97467. doi: 10.1371/journal.pone.0097467

Brechmann, A., Gaschler-Markefski, B., Sohr, M., Yoneda, K., Kaulisch, T., and Scheich, H. (2007). Working memory–specific activity in auditory cortex: potential correlates of sequential processing and maintenance. *Cereb. Cortex* 17, 2544–2552. doi: 10.1093/cercor/bhl160

Cacciaglia, R., Costa-Faidella, J., Zarnowiec, K., Grimm, S., and Escera, C. (2019). Auditory predictions shape the neural responses to stimulus repetition and sensory change. *Neuroimage* 186, 200–210. doi: 10.1016/j.neuroimage.2018.11.007

Chapin, H. L., Zanto, T., Jantzen, K. J., Kelso, S. J., Steinberg, F., and Large, E. W. (2010). Neural responses to complex auditory rhythms: the role of attending. *Front. Psychol.* 1:224. doi: 10.3389/fpsyg.2010.00224

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. (2008a). Listening to musical rhythms recruits motor regions of the brain. *Cereb. Cortex* 18, 2844–2854. doi: 10.1093/cercor/bhn042

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

Costa-Faidella, J., Baldeweg, T., Grimm, S., and Escera, C. (2011). Interactions between "What" and "When" in the auditory system: temporal predictability enhances repetition suppression. *J. Neurosci.* 31, 18590–18597. doi: 10.1523/jneurosci.2599-11.2011

Damsma, A., and van Rijn, H. (2016). Pupillary response indexes the metrical hierarchy of unattended rhythmic violations. *Brain Cogn.* 111, 95–103. doi: 10.1016/j.bandc.2016.10.004

D'Esposito, M., and Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* 66, 115–142.

Doeller, C. F., Opitz, B., Mecklinger, A., Krick, C., Reith, W., and Schroger, E. (2003). Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *Neuroimage* 20, 1270–1282. doi: 10.1016/s1053-8119(03)00389-6

Eichele, T., Calhoun, V. D., Moosmann, M., Specht, K., Jongsma, M. L., Quiroga, R. Q., et al. (2008). Unmixing concurrent EEG-fMRI with parallel independent component analysis. *Int. J. Psychophysiol.* 67, 222–234. doi: 10.1016/j.ijpsycho.2007.04.010

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836.

Geiser, E., Notter, M., and Gabrieli, J. D. (2012). A corticostriatal neural system enhances auditory perception through temporal context processing. *J. Neurosci.* 32, 6177–6182. doi: 10.1523/jneurosci.5153-11.2012

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

Grahn, J. A. (2009). "The role of the basal ganglia in beat perception neuroimaging and neuropsychological investigations," in *Neurosciences and Music Iii: Disorders and Plasticity*, eds S. DallaBella, N. Kraus, K. Overy, C. Pantev, J. S. Snyder, M. Tervaniemi, et al. (New York, NY: Annals of the New York Academy of Sciences) 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

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10, 14–23. doi: 10.1016/j.tics.2005.11.006

Grotheer, M., and Kovacs, G. (2016). Can predictive coding explain repetition suppression? *Cortex* 80, 113–124. doi: 10.1016/j.cortex.2015.11.027

Heard, M., and Lee, Y. S. (2020). Shared neural resources of rhythm and syntax: an ALE meta-analysis. *Neuropsychologia* 137:107284. doi: 10.1016/j.neuropsychologia.2019.107284

Huang, Y., Matysiak, A., Heil, P., König, R., and Brosch, M. (2016). Persistent neural activity in auditory cortex is related to auditory working memory in humans and nonhuman primates. *eLife* 5:e15441.

Jerde, T. A., Childs, S. K., Handy, S. T., Nagode, J. C., and Pardo, J. V. (2011). Dissociable systems of working memory for rhythm and melody. *Neuroimage* 57, 1572–1579. doi: 10.1016/j.neuroimage.2011.05.061

Jongsma, M. L., Desain, P., Honing, H., and van Rijn, C. M. (2003). Evoked potentials to test rhythm perception theories. *Ann. N. Y. Acad. Sci.* 999, 180–183. doi: 10.1196/annals.1284.025

Jongsma, M. L., Eichele, T., Quian Quiroga, R., Jenks, K. M., Desain, P., Honing, H., et al. (2005). Expectancy effects on omission evoked potentials in musicians and non-musicians. *Psychophysiology* 42, 191–201. doi: 10.1111/j.1469-8986.2005.00269.x

Jongsma, M. L., Eichele, T., Van Rijn, C. M., Coenen, A. M., Hugdahl, K., Nordby, H., et al. (2006). Tracking pattern learning with single-trial event-related potentials. *Clin. Neurophysiol.* 117, 1957–1973. doi: 10.1016/j.clinph.2006.05.012

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

Kompus, K., Westerhausen, R., Craven, A. R., Kreegipuu, K., Põldver, N., Passow, S., et al. (2015). Resting-state glutamatergic neurotransmission is related to the peak latency of the auditory mismatch negativity (MMN) for duration deviants: an 1H-MRS-EEG study. *Psychophysiology* 52, 1131–1139. doi: 10.1111/psyp.12445

Kovács, G., and Schweinberger, S. R. (2016). Repetition suppression-An integrative view. *Cortex* 80, 1–4. doi: 10.1016/j.cortex.2016.04.022

Kumar, S., Joseph, S., Gander, P. E., Barascud, N., Halpern, A. R., Griffiths, T. D., et al. (2016). A brain system for auditory working memory. *J. Neurosci.* 36, 4492–4505.

Ladinig, O., Honing, H., Haden, G., and Winkler, I. (2009). Probing attentive and preattentive emergent meter in adult listeners without extensive music training. *Music Percept.* 26, 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

Liebenthal, E., Ellingson, M. L., Spanaki, M. V., Prieto, T. E., Ropella, K. M., and Binder, J. R. (2003). Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. *Neuroimage* 19, 1395–1404. doi: 10.1016/s1053-8119(03)00228-3

Linke, A. C., Vicente-Grabovetsky, A., and Cusack, R. (2011). Stimulus-specific suppression preserves information in auditory short-term memory. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12961–12966. doi: 10.1073/pnas.1102118108

Mathiak, K., Rapp, A., Kircher, T. T. J., Grodd, W., Hertrich, I., Weiskopf, N., et al. (2002). Mismatch responses to randomized gradient switching noise as reflected by fMRI and whole-head magnetoencephalography. *Hum. Brain Mapp.* 16, 190–195. doi: 10.1002/hbm.10041

Menon, V. (2011). Large-scale brain networks and psychopathology: a unifying triple network model. *Trends Cogn. Sci.* 15, 483–506. doi: 10.1016/j.tics.2011.08.003

Mustovic, H., Scheffler, K., Di Salle, F., Esposito, F., Neuhoff, J. G., Hennig, J., et al. (2003). Temporal integration of sequential auditory events: silent period in sound pattern activates human planum temporale. *Neuroimage* 20, 429–434. doi: 10.1016/s1053-8119(03)00293-3

Myers, N. E., Stokes, M. G., and Nobre, A. C. (2017). Prioritizing information during working memory: beyond sustained internal attention. *Trends Cogn. Sci.* 21, 449–461. doi: 10.1016/j.tics.2017.03.010

Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Naatanen, R., Gaillard, A. W., and Mantysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol. (Amst.)* 42, 313–329. doi: 10.1016/0001-6918(78)90006-9

Nazimek, J. M., Hunter, M. D., Hoskin, R., Wilkinson, I., and Woodruff, P. W. (2013). Neural basis of auditory expectation within temporal cortex. *Neuropsychologia* 51, 2245–2250. doi: 10.1016/j.neuropsychologia.2013.07.019

Nees, M. A. (2016). Have we forgotten auditory sensory memory? Retention intervals in studies of nonverbal auditory working memory. *Front. Psychol.* 7:1892. doi: 10.3389/fpsyg.2016.01892

Ono, K., Altmann, C. F., Matsuhashi, M., Mima, T., and Fukuyama, H. (2015). Neural correlates of perceptual grouping effects in the processing of sound omission by musicians and nonmusicians. *Hear. Res.* 319, 25–31. doi: 10.1016/j.heares.2014.10.013

Ono, K., Matsuhashi, M., Mima, T., Fukuyama, H., and Altmann, C. F. (2013). Effects of regularity on the processing of sound omission in a tone sequence in musicians and non-musicians. *Eur. J. Neurosci.* 38, 2786–2792. doi: 10.1111/ejn.12254

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D. Y., and Schroger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *Neuroimage* 15, 167–174. doi: 10.1006/nimg.2001.0970

Peretz, I., and Zatorre, R. J. (2005). Brain organization for music processing. *Annu. Rev. Psychol.* 56, 89–114. doi: 10.1146/annurev.psych.56.091103.070225

Raij, T., McEvoy, L., Mäkelä, J. P., and Hari, R. (1997). Human auditory cortex is activated by omissions of auditory stimuli. *Brain Res.* 745, 134–143. doi: 10.1016/s0006-8993(96)01140-7

Recasens, M., Grimm, S., Capilla, A., Nowak, R., and Escera, C. (2014). Two sequential processes of change detection in hierarchically ordered areas of the human auditory cortex. *Cereb. Cortex* 24, 143–153. doi: 10.1093/cercor/bhs295

Rinne, T., Degerman, A., and Alho, K. (2005). Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: an fMRI study. *Neuroimage* 26, 66–72. doi: 10.1016/j.neuroimage.2005.01.017

Salisbury, D. F. (2012). Finding the missing stimulus mismatch negativity (MMN): emitted MMN to violations of an auditory gestalt. *Psychophysiology* 49, 544–548. doi: 10.1111/j.1469-8986.2011.01336.x

SanMiguel, I., Saupe, K., and Schroger, E. (2013a). I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted "what" but not "when". *Front. Hum. Neurosci.* 7:403. doi: 10.3389/fnhum.2013.00407

SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., and Schröger, E. (2013b). Hearing silences: human auditory processing relies on preactivation of sound-specific brain activity patterns. *J. Neurosci.* 33, 8633–8639. doi: 10.1523/jneurosci.5821-12.2013

Specht, K., and Reul, J. (2003). Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: an auditory rapid event-related fMRI-task. *Neuroimage* 20, 1944–1954. doi: 10.1016/j.neuroimage.2003.07.034

Sreenivasan, K. K., Curtis, C. E., and D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends Cogn. Sci.* 18, 82–89. doi: 10.1016/j.tics.2013.12.001

Sreenivasan, K. K., and D'Esposito, M. (2019). The what, where and how of delay activity. *Nat. Rev. Neurosci.* 20, 466–481. doi: 10.1038/s41583-019-0176-7

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., and Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* 11, 1004–1006. doi: 10.1038/nn.2163

Teki, S., Grube, M., Kumar, S., and Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *J. Neurosci.* 31, 3805–3812. doi: 10.1523/jneurosci.5561-10.2011

Tervaniemi, M., Schroger, E., Saher, M., and Naatanen, R. (2000). Effects of spectral complexity and sound duration on automatic complex-sound pitch processing in humans - a mismatch negativity study. *Neurosci. Lett.* 290, 66–70. doi: 10.1016/s0304-3940(00)01290-8

Thaut, M. H., Trimarchi, P. D., and Parsons, L. M. (2014). Human brain basis of musical rhythm perception: common and distinct neural substrates for meter, tempo, and pattern. *Brain Sci.* 4, 428–452. doi: 10.3390/brainsci4020428

Todorovic, A., van Ede, F., Maris, E., and de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *J. Neurosci.* 31, 9118–9123. doi: 10.1523/jneurosci.1425-11.2011

Toiviainen, P., Burunat, I., Brattico, E., Vuust, P., and Alluri, V. (2020). The chronnectome of musical beat. *Neuroimage* 216:116191. doi: 10.1016/j.neuroimage.2019.116191

Trost, W., Fruhholz, S., Schon, D., Labbe, 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

Vikene, K., Skeie, G. O., and Specht, K. (2018). Abnormal phasic activity in saliency network, motor areas, and basal ganglia in Parkinson's disease during rhythm perception. *Hum. Brain Mapp.* 40, 916–927. doi: 10.1002/hbm.24421

Vikene, K., Skeie, G. O., and Specht, K. (2019). Compensatory task-specific hypersensitivity in bilateral planum temporale and right superior temporal gyrus during auditory rhythm and omission processing in Parkinson's disease. *Sci. Rep.* 9:12623.

Voisin, J., Bidet-Caulet, A., Bertrand, O., and Fonlupt, P. (2006). Listening in silence activates auditory areas: a functional magnetic resonance imaging study. *J. Neurosci.* 26, 273–278. doi: 10.1523/jneurosci.2967-05.2006

Vuust, P., and Witek, M. A. (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

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., and Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20754–20759. doi: 10.1073/pnas.1117807108

Winkler, I., and Schroger, E. (1995). Neural representation for the temporal structure of sound patterns. *Neuroreport* 6, 690–694. doi: 10.1097/00001756-199503000-00026

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

Zatorre, R. J., and Zarate, J. M. (2012). "Cortical processing of music," in *The Human Auditory Cortex Springer Handbook of Auditory Research*, eds D. Poeppel, T. Overath, A. Popper, and R. Fay (New York, NY: Springer), 261–294. doi: 10.1007/978-1-4614-2314-0_10

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: 24 September 2021
doi: 10.3389/fpsyg.2021.707090

Temporal Coordination in Piano Duet Networked Music Performance (NMP): Interactions Between Acoustic Transmission Latency and Musical Role Asymmetries

Auriel Washburn[1]*, Matthew J. Wright[1], Chris Chafe[1,2] and Takako Fujioka[1,2]

[1] Center for Computer Research in Music and Acoustics, Department of Music, Stanford University, Stanford, CA, United States, [2] Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, Untied States

OPEN ACCESS

Edited by:
Thomas James Lundy,
Cuttlefish Arts, United States

Reviewed by:
Thibault Chabin,
INSERM CIC 1431 Centre d'Investigation Clinique de Besançon, France
Anna Zamm,
Central European University, Hungary
Melinda Alicia Mende,
University of Potsdam, Germany
Cory DeWitt Meals,
University of Houston, United States

****Correspondence:***
Auriel Washburn
awashburn@ucdavis.edu

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

Received: *08 May 2021*
Accepted: *16 August 2021*
Published: *24 September 2021*

Citation:
Washburn A, Wright MJ, Chafe C and Fujioka T (2021) Temporal Coordination in Piano Duet Networked Music Performance (NMP): Interactions Between Acoustic Transmission Latency and Musical Role Asymmetries. Front. Psychol. 12:707090. doi: 10.3389/fpsyg.2021.707090

Today's audio, visual, and internet technologies allow people to interact despite physical distances, for casual conversation, group workouts, or musical performance. Musical ensemble performance is unique because interaction integrity critically depends on the timing between each performer's actions and when their acoustic outcomes arrive. Acoustic transmission latency (ATL) between players is substantially longer for networked music performance (NMP) compared to traditional in-person spaces where musicians can easily adapt. Previous work has shown that longer ATLs slow the average tempo in ensemble performance, and that asymmetric co-actor roles and empathy-related traits affect coordination patterns in joint action. Thus, we are interested in how musicians collectively adapt to a given latency and how such adaptation patterns vary with their task-related and person-related asymmetries. Here, we examined how two pianists performed duets while hearing each other's auditory outcomes with an ATL of 10, 20, or 40 ms. To test the hypotheses regarding task-related asymmetries, we designed duets such that pianists had: (1) a starting or joining role and (2) a similar or dissimilar musical part compared to their co-performer, with respect to pitch range and melodic contour. Results replicated previous clapping-duet findings showing that longer ATLs are associated with greater temporal asynchrony between partners and increased average tempo slowing. While co-performer asynchronies were not affected by performer role or part similarity, at the longer ATLs starting performers displayed slower tempos and smaller tempo variability than joining performers. This asymmetry of stability vs. flexibility between starters and joiners may sustain coordination, consistent with recent joint action findings. Our data also suggest that relative independence in musical parts may mitigate ATL-related challenges. Additionally, there may be a relationship between co-performer differences in empathy-related personality traits such as locus of control and coordination during performance under the influence of ATL. Incorporating the emergent coordinative dynamics between performers could help further innovation of music technologies and composition techniques for NMP.

Keywords: perceptual-motor coordination, interpersonal coordination, joint action, technology-mediated interaction, acoustic transmission latency, role asymmetries, time-delayed coupling synchronization, musical turn-taking

INTRODUCTION

Cyber-based social interaction has become ubiquitous in our society due to rapid advances in interactive technologies including video-conferencing, online social networking, and even multiplayer gaming platforms and virtual reality. This has generated a wide range of opportunities for collaborative interaction across distinct geographic locations. As with in-person human interaction, many cyber-based collaborations require individuals to effectively coordinate their actions with others. However, information transmission latencies are present at greater lengths and with increasing frequency during social interactions. This can substantially impact behavioral coordination. For instance, when auditory and visual feedback from listeners in response to a speaker is absent or delayed during videoconferencing this can disrupt conversational turn-taking behaviors (O'Conaill et al., 1993). Such effects are especially poignant in ensemble musical performance. Here the timing information about a performer's actions is extremely critical to their co-performers' actions, all of which have an immediate effect on the quality of musical outcomes (Goebl and Palmer, 2009). This also affects audiences' perceptions. When a temporal offset is introduced between the video of a conductor and the audio of the ensemble performance, audiences perceive conducting quality to be lower (Meals et al., 2019). As a result, while "networked" music performance (NMP) offers intriguing opportunities for experimental music and novel compositional techniques (e.g., Chafe, 2011), it also poses significant challenges to the achievement of robust, gratifying ensemble performance.

Scarce work exists on how musicians might maintain collective temporal coordination and sustain ensemble performance during NMP. Past empirical work has, however, demonstrated that acoustic transmission latencies (ATLs) disrupt the temporal dynamics of many perceptual-motor tasks. Notably, ATLs in the action-feedback an individual receives can disrupt behavioral production of tapping, speech, and musical performance (e.g., Fairbanks and Guttman, 1958; Robinson, 1972; Gates et al., 1974; Howell et al., 1983; Finney, 1997; Finney and Warren, 2002; Pfordresher and Palmer, 2002). For example, Pfordresher and Palmer (2002) found that for a pianist playing a rhythmically isochronous melody, increases in the ATL between their keypresses and the resulting sound is associated with increased temporal variability in their playing.

ATLs and Music Ensemble Coordination

For ensemble musical performance, the effects of ATL show more complex patterns characterizing the temporal relationships between musicians' behaviors (Delle Monache et al., 2019; see Rottondi et al., 2016 for review). Chafe et al. (2010) asked pairs of individuals to perform a coordinated clapping task to establish an initial empirical understanding of the limitations and implications of NMP. ATLs were introduced bidirectionally between the duet partners (i.e., each performer heard their co-performers' clapping at the same, fixed latency during a given trial). There was no pre-designated "leader" in this clapping duet like an orchestra conductor. Instead, one individual started first as a solo performer (henceforth "starter"), repeating a rhythmic pattern ["dum-(rest)-da-da"], before the other joined (henceforth "joiner") repeating the same rhythmic pattern but in a temporally staggered manner ["da-da-dum-(rest)"]. The collective rhythm was the two unison notes (clapped by both) always interleaved with a solo note (clapped by a starter or joiner).

The researchers observed the emergence of four distinct coordinative regimes, each corresponding to a different range of ATL. Most notably, when performers exhibited stable synchronized behavior at latencies of 10–25 ms they appeared to maintain symmetrical behavioral roles. Contrastingly, the emergence of a starter-joiner dynamic for latencies between 25 and 60 ms involved distinctly different, or asymmetric roles. Different ATL ranges can therefore lead to distinct, emergent temporal relationships that can support the maintenance of ongoing coordination. The effect of a given ATL may also depend on training and the instrument being played. For example, percussionists appear able to maintain tempo when experiencing both moderate and extreme delays while other instrumentalists (e.g., harpists and flutists) are potentially more affected due to melodic and agogic constraints (Jack et al., 2018; Delle Monache et al., 2019). Past work has also revealed differences in the effects of ATL among instruments with melodic constraints (Bartlette et al., 2006), as well as variations due to instrument entropy (Rottondi et al., 2015) and reverberation (Carôt et al., 2009; Farner et al., 2009).

Asymmetries in Musical Interaction

The "functional asymmetry" in co-actor roles observed in Chafe et al. (2010) is similar to those that often emerge over the course of general, non-musical, interpersonal interaction without explicit instruction or intention (Richardson et al., 2016). Other factors that support functional asymmetries are primarily related to (1) task structure, and (2) individual traits.

Task-Related Asymmetries

Individuals commonly adopt functionally asymmetric behaviors based on their distinct task roles. For example, when two people move a table together one might support the table while moving backward while the other moves forward and guides the direction. There is evidence of asymmetries in musical interaction as well. Ensemble performers often have distinct roles corresponding to their part/instrument that are understood in a hierarchical fashion (i.e., the "first" violinist is the predetermined leader of a string quartet). Timmers et al. (2014) identified a leader-follower relationship in performance timing between Viola and Violin 1 and between Violin 1 and Cello, as well as mutually adaptive relationships between Violin II and Viola and between Violin II and Cello in a professional quartet. Other studies have shown that body sway movements of the individual playing the leading melody precede those of their co-performers (Goebl and Palmer, 2009; Badino et al., 2014). Interestingly, Chang et al. (2017) found that if a leader other than Violin 1 is explicitly designated during string quartet performance then the other members' body movements tend to follow theirs. This suggests that explicit role recognition influences the overall ensemble coordination behaviors.

Such task-related movement patterns are further associated with differences in underlying neural activities. Vanzella et al. (2019) used functional near-infrared spectroscopy (fNIRS) to identify greater activation in temporo-parietal and somatomotor regions for individuals performing the second violin part vs. the first violin part in a duet. This indicates that collective goals for ensemble performance may especially shape the behavior of the follower musician. Furthermore, a functional magnetic resonance imaging (fMRI) study of dancers revealed that the level of expertise in one role compared to the other (e.g., leader role or follower role) in couple-style dances enhanced brain activation specific to the trained role during joint hand movement coordination (Chauvigné and Brown, 2018). Thus, the two types of behavior may correspond to distinct synchronizing strategies, where a greater focus on one's own behavioral outcomes or those of the interaction can be emphasized. Electroencephalography (EEG) hyperscanning has also revealed theta- and delta-band oscillatory coupling between musicians' cortical activity during coordination (Lindenberger et al., 2009), and stronger alpha- and beta-band oscillatory coupling for follower to leader behavior than vice versa during guitar duet performance (Sänger et al., 2013).

In addition to explicit and predetermined roles, differences in the musical content between performers' parts may also contribute to the emergence of asymmetries in coordination. The difference in the number of notes between players' parts can actually affect patterns of relative adaptation between piano duet performers more than assigned leadership (Goebl and Palmer, 2009). Here, the pianist whose part had twice as many notes to play was more likely to exhibit a temporal lead, regardless of who had been assigned the "leader" role. Interestingly, this behavior may be accounted for by the widely observed "more is up" phenomenon in which larger quantitative magnitudes are associated with higher physical space (Fischer, 2012; Shaki and Fischer, 2018). Thus the performer with a greater number of notes, and larger workload, may feel hierarchically higher relative to their co-performer and more responsible for leading. Further, Loehr and Palmer (2011) observed that the melodic and harmonic complexity of the accompaniment part within a piano duet resulted in temporal grouping coinciding with the melodic structure, showing increased adaptation in the accompaniment role. Relatedly, Bishop and Goebl (2020) suggest that asymmetric leader-follower relationships can make performance more predictable for performers by decreasing variability in interpretation and increasing predictability. This potentially unconscious strategy is similar to instances of non-musical joint action where co-actors make their behavior more predictable, and consequently achieve greater coordination (Vesper et al., 2011).

Person-Related Asymmetries

Loehr and Palmer (2011) showed that a greater difference between partners in individual preferred tempo led to an increase in their temporal asynchronies. Pianists matched for preferred tempo achieved greater interpersonal synchrony and synchronization stability during duet performance than those who were not (Zamm et al., 2016). Wing et al. (2014) examined temporal relations between musicians in two separate professional quartets performing the same piece, finding that the two groups exhibited unique patterns of symmetry and asymmetry at the beat-to-beat timescale. Thus, functional asymmetries in temporal coordination between musicians are shaped not just by differences in musical role and the musical content of their parts, but also by the performers themselves.

Volpe et al. (2016) emphasize that coordination and leadership within music ensembles can also be related to the social dynamics between individual performers. The personality traits of each performer, including locus of control and empathy, are relevant to the emergence of co-performer asymmetries in music. In particular, individual differences may at least partially contribute to the asymmetries observed in behavioral or neurophysiological patterns. For example, during a synchronized tapping task with a virtual partner, individuals with a latent internal locus of control (i.e., perceive events to be the consequences of their own actions) showed less adaptive behavior and more leader characteristics, while those with a latent external locus of control (i.e., perceive events to be caused by external forces) adopted more of a follower role, exhibiting more frequent corrective behavior (Fairhurst et al., 2013). Further, Fairhurst et al. (2014) and Konvalinka et al. (2014) have found distinct neural activity associated with leading vs. following behavior and corresponding personality traits. For example, using fMRI during a finger tapping task with a virtual adaptive partner, Fairhurst et al. (2014) identified a correlation between internal locus of control score and the tapping interval stability in those who showed "leader" behavior (e.g., less adaptive). The score was further positively related to subjective perception of the leadership role as well. Similarly in Konvalinka et al. (2014) only individuals who showed leadership behavior during synchronized tapping exhibited alpha suppression as captured *via* EEG.

Individual differences in the perspective taking dimension of empathy also affect synchronized tapping, with higher scores linked to greater anticipation of tempo-changing metronome sequences (Pecenka and Keller, 2011). Such individual differences appear to further interact with the task asymmetry between co-performers. In our own recent piano-duet work (Washburn et al., 2019), pianists with high empathy scores exhibited greater variability in temporal coordination when the performers' parts were melodically dissimilar as compared to when they were similar, even though duet parts were rhythmically identical. Together, these studies consistently demonstrate that a person with increased empathy or perspective taking tends to be more influenced by the activity of an external stimulus, or co-performer behavior. This point is further supported by two other studies, both requiring pianists to perform the right-hand part of a piece with a pre-recorded left-hand part (Novembre et al., 2012, 2013). During the performance, researchers used transcranial magnetic stimulation (TMS) to facilitate or inhibit excitability of the right motor cortex (thus, modulating the person's perceptual-motor experience of the left hand). The effects of stimulation were associated with increased impairment of tempo adaptation accuracy for more empathic individuals.

Current Study

Acoustic transmission latencies clearly influence the stable coordinative states available to co-performers. However, there is little existing evidence on how these ATLs interact with the aforementioned task and personality-related factors. Direct manipulation of ATL between co-performers may systematically influence how each type of asymmetry plays its role. In the current study we assessed this possibility by examining how two pianists, playing in separate rooms without seeing each other, adapt their coordination patterns during naturalistic, simple, rhythmic duet tasks at a given ATL. We composed eight original duets employing the same interlocking rhythmic pattern with two independent but equal parts previously studied with clapping (Chafe et al., 2010). These duets included two forms of task-related asymmetry: performer role asymmetry (starting vs. joining roles), and musical part asymmetry (similar vs. dissimilar musical parts with respect to pitch range and melodic complexity). To examine the effects of person-related asymmetry we evaluated perspective taking and locus of control. During duet performance, we introduced three ATLs (10, 20, and 40 ms), allowing us to evaluate how ATL interacts with both musical task-based and person-related asymmetries in shaping coordination. In the past decade standard internet latencies have typically ranged from 20 to 100 ms (Cáceres et al., 2008; Cáceres and Chafe, 2010), with current 5G networks producing latencies on the order of tens of milliseconds (Landström et al., 2016). The ATL values for the current study therefore fall within typical internet latencies while targeting three characteristic effects. These include a speeding up, steady maintenance, or slowing down of the average tempo for duet performance when initial performance tempo is around 90 beat-per-minute (bpm) (e.g., Farner et al., 2009; Chafe et al., 2010; Rottondi et al., 2015).

Our primary coordination measures were the magnitude and variability of note-onset asynchronies between co-performers at unison points in each duet as well as the magnitude and variability of tempo. Based on previous work, we predicted that starting performers would exhibit greater note-onset asynchrony leads, and less variability in both note-onset asynchronies and tempo. We also expected possible interactions between performer role and musical part asymmetry, with greater differences between starters and joiners in both asynchrony and tempo when the duet parts were dissimilar. We predicted that ATL would moderate these effects, with greater ATL leading to greater differences in starting vs. joining behavior. This study's overarching goal was to provide a foundation for understanding how key aspects of the music ensemble setting shape co-performer interaction within cyber-mediated environments such as NMP.

MATERIALS AND METHODS

Participants

We conducted a power analysis for the number of participant pairs in the current study. We based this on the Pearson correlation coefficient (R = 0.96) for the relationship between ATL and collective lead/lag from Chafe et al. (2010) (we refer to this measure as "cycle asynchrony," see section "Measures and Analyses"). Using a significance level of alpha = 0.05, an intended power of 0.9, and a directional analysis without bias-correction, we obtained a sample size of N = 6 pairs of participants. We recruited 12 pairs of performers, satisfying the minimum sample size supported by the power analysis.

Twenty-four pianists (12 pairs) ranging from 18 to 47 years old were recruited from the Stanford University community. All had at least 4 years of piano-playing experience and all but one were active musicians, playing an instrument or singing at least 2 h a week. No one reported hearing problems relevant to their musical pursuits.[1]

Of the pairs recruited, three pairs knew each other and had played music together prior to the experiment, ranging between one and six occasions. These three pairs did not exhibit particular advantages against the ATL effects compared to the other nine pairs, or consistent outlying values for within-pair differences in the perspective taking and locus of control scores compared to the other eight pairs included in the correlation analyses. The study was approved by the Stanford University Institutional Review Board. All participants provided informed consent *via* a signed form and were paid $20/h for participating.

Apparatus

Two Yamaha Axiom-61 digital keyboards were located in two adjacent rooms, where the smaller room included sound shielding for use during EEG studies (see **Figure 1**). Each room was equipped with a pair of AKG K271 MKII Closed-Back headphones. We made a custom program using the Max/MSP 7.0.1 platform to not only synthesize and control all acoustic feedback but also to monitor accuracy of the performers' keypresses and record all the keypress timing data explicitly associated with the individual notes of the duet compositions throughout the study *via* a Macintosh computer (OSX 10.9.5). Sounds recorded from the built-in OSX MIDI sound synthesizer, AU DLS Synth, were precisely triggered to create the piano timbre used throughout and the snare drum "cross-stick" timbre used for introductory metronome count-ins. The experimenter sat with this computer in the larger room. Play-back loudness was set constant regardless of the MIDI keypress velocity with pianists unable to produce changes in dynamics during performance. Note duration was fixed at 200 ms regardless of performer key-offset timing. These settings were controlled so that we could best examine the effects of the experimental manipulations of interest on temporal coordination.

As demonstrated by Wright et al. (2004), we evaluated the latency inherent to the experimental apparatus by obtaining audio recording for the following events simultaneously: (1) the acoustic sound of each keyboard keypress (captured with an AKG C 414 B-ULS microphone from approximately 5 cm distance) and (2) the resulting acoustic feedback (a piano tone emitted by the apparatus), *via* a direct electrical connection from the output of the sound generator into the input of the audio interface connected to Audacity software. From these recorded audio files

[1] One participant reported a hearing impairment at high frequencies necessitating the use of a hearing aid in a classroom but not while playing music.

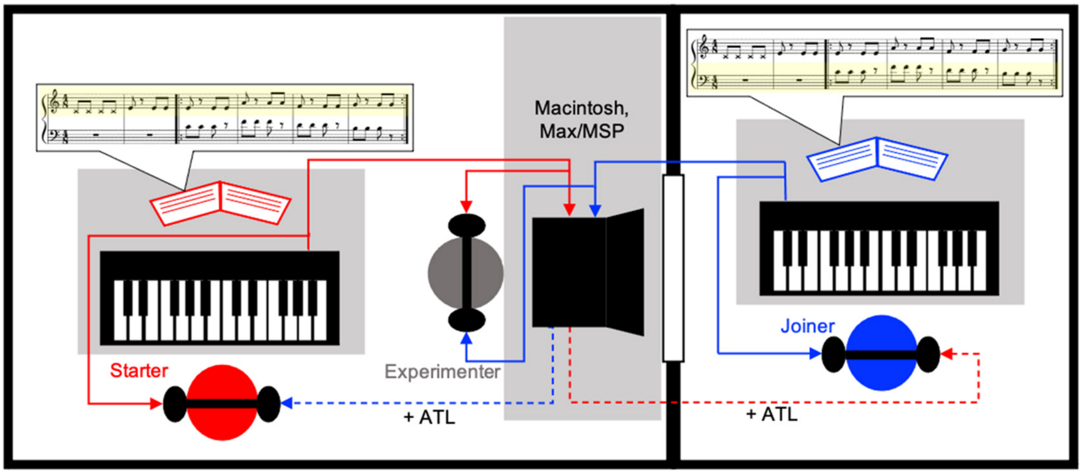

FIGURE 1 | The experimental set-up. The experimenter (center, gray circle) was able to see both pianists (red and blue circles) while they were located in separate rooms and could not see each other. The keyboards were controlled by the experimenter's Macintosh computer with Max/MSP software such that the pianists always heard their own playing at the base latency inherent within the apparatus (solid red and blue lines) but heard their partner's playing with an additional ATL of 10, 20, or 40 ms (dashed red and blue lines) during experimental trials. In this figure the left pianist (red) is performing the starter role (playing the associated yellow highlighted score part), with the right pianist (blue) performing the joiner role (playing the associated yellow highlighted score part). All pianists played both parts of all musical compositions; this role distribution occurred for half of the trials, while for the other half the roles were reversed.

we detected the instantaneous onset time of each event using a simple amplitude threshold (set to 10 times the maximum amplitude of the recorded background noise) and the rule that after one onset the amplitude must remain below the threshold for ∼2 ms before the next onset can be detected. Subtracting the keypress onset time from the corresponding piano tone onset time gives the latency for each keypress. Following outlier removal the average keypress to acoustic feedback onset for Keyboard 1 (larger room) press to Keyboard 1 audio was 33.5 ms (SD = 3.3 ms), for Keyboard 2 (smaller room) press to Keyboard 2 audio was 25.4 ms (SD = 2.0 ms). The base latency for Keyboard 1 press to Keyboard 2 audio was 27.7 ms (SD = 2.5 ms), and for Keyboard 2 press to Keyboard 1 audio was 35.3 ms (SD = 3.6 ms). Kolmogorov–Smirnov tests were used to assess whether there were significant differences between the distributions of the mean-centered latency for: (1) self-feedback at each of the two keyboards (i.e., K1-K1 vs. K2-K2), (2) self vs. other feedback at Keyboard 1 (i.e., K1-K1 vs. K2-K1), (3) self vs. other feedback at Keyboard 2 (K2-K2 vs. K1-K2), or (4) other-feedback at each of the two keyboards (i.e., K1-K2 vs. K2-K1). All tests were not significant.

Importantly, the latency differences between self-produced sound and the other keyboard sound were compatible (Keyboard 1: 33.5 vs. 35.3 ms; Keyboard 2: 25.4 vs. 27.7 ms). Performers had the opportunity to calibrate to the base latency at their own keyboard during the initial practice and test trials in the same way that musicians regularly adjust to the inherent base latency of a given environment. This means that, in our recorded timing data, the effect of the additional ATL from the partner's keyboard sound in the experimental condition was only compared to each respective self-produced action-to-playback latency.

The Max/MSP program was used to introduce additional bidirectional ATLs of 10, 20, or 40 ms between performers during experimental trials as described below in section "Procedure." In these trials both pianists experienced a given latency of experimentally induced delay in the acoustic outcomes of their partner's keypresses, but not their own. Throughout every trial the Max/MSP program tracked the pianists' performance (the MIDI inputs from the two keyboards) for accuracy in note pitch and sequence compared to the musical score to identify errors. With the introduction of ATLs we expected tempo changes, and therefore allowed all timing distortions as long as the duet collectively played the correct alternation of solo notes and "unison" notes (regardless of the timing difference in note-onset). The program also controlled the order of the ATL conditions and metronome tempo, and generated codes for events associated with experimental conditions to store these with timepoints for the data analyses.

Stimulus and Task

Eight duets in C Major were composed where each performer was meant to play either the top or bottom line (see **Figure 2**). All duets employed the same interlocking rhythmic pattern, with two independent but equal parts such that temporally synchronized performance results in unison just once every two notes, and the analysis of a lead/lag relationship between parts is possible (i.e., metrical phase for each part can be assessed individually) (Chafe et al., 2010). All parts were to be played with the right hand. Each duet included four full bars in which both performers played, along with a preceding pickup bar to be played by the "starting" performer. Both performers played the same rhythmic pattern consisting of three consecutive eighth notes followed by one eighth rest. However, the parts were temporally staggered such that when one player had the rest, the other player had one eighth note to play. The composition's repeat sign meant that the players had to repeat

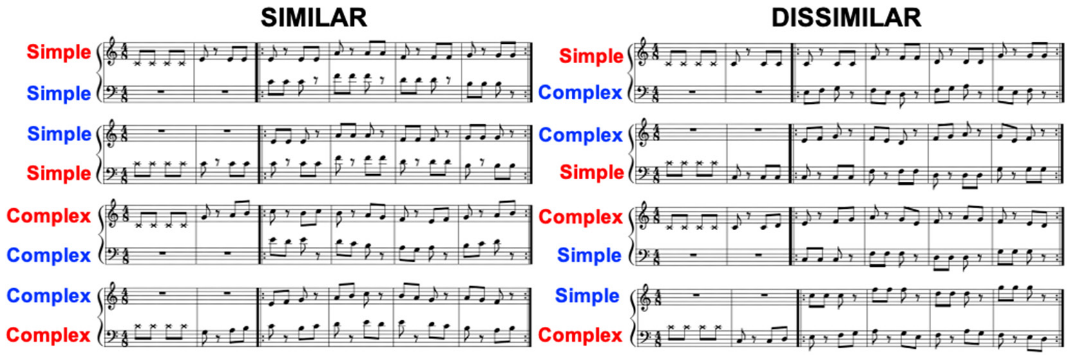

FIGURE 2 | Music scores of eight piano duets composed for this study. Each duet consisted of two parts (e.g., lines) that were either *similar* or *dissimilar*, as defined by pitch range and melodic complexity. In *similar* duets (left panel) the pitch ranges of the two parts overlapped while in *dissimilar* compositions they did not (right panel). *Similar* duets also included parts that were either both melodically simple or complex, while *dissimilar* duets included one of each kind of part. For half of the duets the starting part (red labels) was placed in the top line with the joining part (blue labels) in the bottom. For the other half the line placement was reversed. Pianists had to repeat the four bars of each duet indicated by the repeat signs a total of four times. The metronome count-ins for the starting pianist were added to the starter's line.

their part four times before ending the trial at the first note of the fifth repeat.

We employed two factors to define partners' musical part similarity for the duets (**Figure 2**). The first of these was the pitch range of the co-performers' melodic parts; the two parts had either overlapping pitch ranges (sharing at least one pitch), or were separated into two distinct ranges (at least three semitones apart). In our "similar" conditions, the unison notes had an average interval of 6.16 semitones (range 3–10) while for the "dissimilar" conditions it was 14.1 semitones (5–22). The second factor contributing to musical part similarity was melodic complexity (simple vs. complex), with players assigned to the same or different types. In our melodies, "simple" ones always used three-note phrases with one-pitch repetition, while "complex" melodies had three-note phrases out of three different pitched notes, leading to a more dynamic melodic contour change over the course of the composition.

In the *similar* duets the performers' parts were either both simple or both complex. By also requiring that each condition included one duet that started on the top line and one that started on the bottom line, four *similar* duets were generated. The *dissimilar* duets had distinct pitch ranges and always combined one melodically simple part and one melodically complex part. This resulted in eight total compositions: four *similar* and four *dissimilar*. We employed the pitch range distance and melodic complexity factors to manipulate melodic similarity between the duet parts based on previous research findings. Halpern (1984) found that musicians and non-musicians alike rated melodies least similar when they have different pitch contour and rhythm. Further, when two melodies are presented in an interleaved manner (e.g., alternate one note between the melodies), listeners can identify melodies only when their pitch ranges do not overlap (Dowling, 1973), closely related to Gestalt principles of similarity resulting in auditory streaming (Bregman, 1990). In a statistical learning study which also used alternating tones, listeners learned the statistical regularity within each melodic stream much better when the two melodies were perceptually well segregated by a contrasting grouping cue such as timbre or pitch range (Creel et al., 2004). Our duet compositions follow these principles straightforwardly. Each performer was given the opportunity to act as both the starter and joiner on each duet, for a total of 16 compositions. Duet order was organized pseudo-randomly by pair with at least one *similar* duet and at least one *dissimilar* duet occurring within both the first four compositions and the last four compositions. There were never more than three of either the *similar* or *dissimilar* duets in a row within the full presentation of the eight duets. Similarly, at least one top-line starting duet and one bottom-line starting duet occurred in both the first four duets and the last four duets and no more than three same-line starting duets ever occurred in a row. For a given duet, pianists played alternating starter/joiner roles one after another meaning that no more than two compositions in a row had the same starting vs. joining player assignment.

Procedure

Paired participants received instructions together before being randomly assigned to separate rooms, within which each had a USB/MIDI keyboard controller and headphones (see "Apparatus" section above and **Figure 1** for details). One pianist sat in the same room as the experimenter (larger room), while the other pianist was alone (smaller room). A window between the rooms allowed the experimenter to see both individuals, and vice versa, but the participants were not able to see each other when seated in front of their respective keyboards. With the headphones participants were able to hear acoustic output corresponding to their own playing, as well as their partner's. Overall volume levels were adjusted to performers' comfort.

Performers received binders with copies of the 16 duet compositions with the part they were to play in each highlighted for them (i.e., the starting vs. joining line). As the starter, performers were to follow the metronome tempo they heard *via* their headphones to begin the pickup measure, while as the joiner they would take tempo from the starter. Only after completing seven successful performances of a given composition would the pair move on to the next (as described below, these seven performances included one test trial without ATL, and six experimental trials with two trials at each of the three ATLs). Performers were not given any specific information about the ATL conditions but asked to play the notes of their part accurately while maintaining coordinated playing with their partner and also aiming to maintain the tempo of the metronome heard by the starting player as best as possible. A successful trial required them to play all notes of the pickup and four repeats of the subsequent four bars of a composition accurately and in the correct order, ending on the first note of what would be the fifth repeat of the composition (i.e., 52 total notes for the starting part and 49 total notes for the joining part). If any notes were missed or incorrect pitches were played the trial would abort immediately and they would start again (cued by a new metronome count-in).

Each time a new composition was presented, the performers could practice their part for a few minutes to be able to play successfully with comfort. During this practice period both performers could hear each other (at the base latency for the apparatus with no additional ATL introduced). Once they were comfortable, performers would indicate to the experimenter that they were ready to begin the recorded trials. At the beginning of each recorded trial a metronome consisting of four eighth note beats was presented exclusively to the starter. The first recorded trial following the practice period was a test trial to establish that the pair could successfully play the duet under normal acoustic conditions (i.e., with the base latency for the apparatus). The tempo of the metronome in these trials was always 90 bpm (e.g., one eighth note = 666.67 ms) where in 4/8 time the beat occurs at the eighth note level, consistent with the average tempo used in the clapping duet study by Chafe et al. (2010).

In the six experimental trials for each composition, bidirectional ATLs were introduced between performers (i.e., performers heard their own playing at the base latency but heard their partner's playing with an additional ATL). Three different ATLs were used: 10, 20, and 40 ms. Latencies around or below 10 ms are associated with persistent anticipatory behavior between co-performers, resulting in a progressive increase in playing tempo over the course of a performance (Chafe et al., 2010). Interestingly, this range overlaps with the acoustic latencies typically experienced by performers in small chamber music ensembles, which have previously been reported as 6–9 ms (Chafe et al., 2010) or 5–10 ms (Bartlette et al., 2006).[2] Latencies between 10 and 25 ms are found to support a high incidence of synchronous behavior between performers and stable tempo. At latencies of 25–60 ms performers begin to show either decreases in tempo or the formation of a new strategy for maintaining synchronization, namely a consistent starter-joiner dynamic. Beyond 60 ms ATL coordination generally deteriorates until performance is no longer sustainable. Each of the three ATLs employed in the current study were meant to elicit one of each of the distinct coordinative states preceding coordination deterioration.

[2]The speed of sound in air is 343 m/s, at 68°F (20°C) and 1 atmosphere pressure. For example, a distance of 4 m would produce 11.6 ms delay.

An example of the organization of ATL and tempo conditions for the set of seven required trials associated with a single composition is provided in **Table 1**. The first of these trials was always the test trial (base apparatus latency with no added ATL, 90 bpm). Unlike the test trial, the six experimental trials for a given musical composition used two different tempo setups, either 84 or 96 bpm. While the average of these two tempi is 90 bpm, this variation was introduced to engage participants and maintain attention over the course of the session as it required them to adjust their internal tempo frequently between trials and respond to a given ATL rather than relying on memory. Within the six experimental trials each ATL was introduced twice and each tempo was presented three times. Presentation of the ATL and tempo conditions was organized into two blocks of three trials each. Within a three-trial block ATLs were presented pseudo-randomly such that each ATL was experienced once before any ATL was repeated. Tempo order for the six experimental trials associated with a given composition was also pseudo-random with one tempo experienced just once in the first three trials and the other experienced just once in the remaining three trials. Importantly, the tempo associated with the second presentation of a given ATL within these trials was always different from the one used for the first presentation. As noted in the "Apparatus" section above, all tempo changes during a trial were allowed as long as the duet collectively played the correct alternation of solo notes and "unison" notes (regardless of the timing difference in note-onset).

Each successful trial took between 18 and 48 s to complete, as performances with the longer ATLs took a longer time to finish due to the collective slowing described in the results below. The six experimental trials associated with each of the 16 compositions resulted in a total of 96 trials for each pair. Following the duet-playing task, individual participants were asked to stay in their separate rooms to complete Davis's (1980) Interpersonal Reactivity Index and the Internality, Powerful Others, and Chance (IPC) Scales (Levenson, 1981) before they were debriefed about the purpose of the study together. Each session took approximately 1.5 h.

Measures and Analyses

Performers' coordination behavior was examined using MIDI keypress timing for the notes to be played in unison based on the interlocking rhythmic pattern underlying all compositions (i.e., the first and third eighth note of every measure following

TABLE 1 | Example ATL and tempo condition organization for the seven required trials associated with a single composition.

Trial	ATL	Tempo
1 (test)	0 ms (base)	90 bpm
2 (experimental)	20 ms	94 bpm
3 (experimental)	10 ms	94 bpm
4 (experimental)	40 ms	86 bpm
5 (experimental)	20 ms	86 bpm
6 (experimental)	40 ms	94 bpm
7 (experimental)	10 ms	86 bpm

the pickup). There were 33 of these unison points in every trial. In total, four timing measures were extracted. Two measures of collective temporal dynamics, *cycle asynchrony*, and *collective tempo* allowed comparison to previously reported effects of ATLs on rhythmic clapping (Chafe et al., 2010). Two additional timing measures, *note-onset asynchronies* and *individual tempo*, allowed us to further investigate the effect of ATLs on inter- and intrapersonal musical timing, respectively.

Questionnaire data from the *Interpersonal Reactivity Index* (Davis, 1980) and the *IPC Scales* (Levenson, 1981) were used to gather a perspective taking score and locus of control score for each performer, respectively. This allowed us to identify relationships between performer personality characteristics, performer role (i.e., starting vs. joining), and environmental ATLs in shaping temporal coordination. Details of each of these measures are described below.

Cycle Asynchrony

We can define a "cycle" as a single notated measure of an interlocking duet composition (shown in **Figure 2**). This consists of a "unison" note (both players synchronously), one player's solo note, a second unison note, and the other player's solo note. Each unison note supposed to be played synchronously was, in reality, associated with two individual keypress timepoints. The total amount of time disparity between performers within this one cycle would express the collective anticipation (lead) or lateness (lag) of performers with respect to each other's playing at that moment (Chafe et al., 2010) [note that Chafe et al. (2010) referred to this measure as "collective lead/lag" and calculated it as a percentage rather than a proportion as we have here]. Effectively, this measure captures whether performers are both consistently leading or lagging each other to display acceleration or deceleration in a given cycle. For example, if the starter leads the joiner at one unison note, and the joiner leads the starter at the next then the duet is displaying acceleration. This was calculated for a pair of unison points as in Eq. 1,

$$\text{lead/lag}\ (k) = \left(start_{\text{unison}}\ [n]\right) - \left(join_{\text{unison}}\ [n]\right) + \left(join_{\text{unison}}[n+1]\right) - \left(start_{\text{unison}}[n+1]\right) \tag{1}$$

where $start_{unison}[n]$ and $join_{unison}[n]$ correspond to the onset timing of the n-th unison note played by the starting player and joining player, respectively, while $2k$ equals to n (i.e., n is an even number). This method of evaluating asynchronies preserves the sign of the difference, thus maintaining information about the potentially dynamic relationship between performers occurring over the course of a trial.

Since a given trial would contain 16 cycles in total, the 16 lead/lag values obtained were then averaged to provide a measure of the overall acceleration or deceleration exhibited by the pair across all cycles within a trial. Each pair produced 16 trials in each of the part similarity (similar vs. dissimilar) × ATL (10, 20, and 40 ms) condition combinations. We used these trials to establish an average cycle asynchrony per pair for each part similarity × ATL condition combination. These pair averages per ATL were then averaged to identify the characteristic cycle asynchrony for each ATL condition across all pairs.

Collective Tempo

This measure expresses a momentary tempo estimate of the performance as it evolved over the course of a trial. First, we determined a collective unison time for each of the 33 unison points as the midpoint between player onset times. We then found the inter-unison intervals (IUIs) between the collective unison times for a given trial, resulting in a total of 32 IUIs. At the n-th unison note, the interval value was then expressed as a tempo value (in bpm; beat per minute) by Eq. 2,

$$\text{collective tempo}\ (n) = 60/\left(collective\ IUI\ [n]\right) \tag{2}$$

The collective tempo values for each trial were used to visualize tempo drift occurring over the course of a single performance. We averaged these tempo drift series associated with each part similarity × ATL condition combination to generate an average tempo drift series per pair for each condition combination. We then averaged these tempo drift series within condition combinations to provide a characteristic tempo drift series for each condition combination across all pairs.

Note-Onset Asynchronies

Note-onset asynchronies were also evaluated with respect to the performers' starting and joining roles in a given trial. This allowed us to establish the frequency with which each player led or lagged the other within the trial, the average magnitude of temporal lead for each player when they played first at unison points ("asynchrony lead"), and the standard deviation of temporal lead at these points ("asynchrony variability"). The magnitude of a single unison point asynchrony in this context was calculated as the proportion of a beat based on the starting tempo for the given trial as

$$\text{asynchrony}\ (n) = \left(\text{tempo}/60\right) \times \left(start_{\text{unison}}\ [n] - join_{\text{unison}}\ [n]\right) \tag{3}$$

Note that if this value is negative, that means that the starter played first at the unison. Thus, the negative values that occurred within a given trial were used to identify the frequency of starter leading, the average magnitude of leading, and the standard deviation of leading magnitude during a trial. For the positive values from (Eq. 3) the inverse was taken to calculate the frequency, average lead, and standard deviation of joiner-led asynchronies during a trial.

Ultimately, a single pair produced 16 trials in each of the part similarity (similar vs. dissimilar) × ATL (10, 20, and 40 ms) condition combinations. From those trials, we identified the average frequency of leading for each participant when they were assigned the starter role and when they were assigned the joiner role. We used these values to identify the within-person difference in frequency of leading for the starter vs. joiner role. This within-person difference was averaged per pair for each of the condition combinations, and then across pairs to give a characteristic within-person difference in frequency of leading for the starter vs. joiner role associated with each of the condition combinations. For each of the trial-wise measures of asynchrony lead and asynchrony variability we calculated the unique starter and joiner averages for a pair across the 16 trials in each of the part similarity × ATL condition combinations. These pair averages were ultimately used to establish the unique average asynchrony lead and asynchrony variability for starters and joiners in each of the noted condition combinations across all pairs.

The cycle asynchrony measure depicted the collective leading vs. lagging behavior exhibited by both members of a pair over the course of a "cycle" including two unison notes and two solo notes per performer. The asynchrony lead measure revealed the magnitude of leading displayed by the performer who played first at each unison timepoint. While these measures are related, cycle asynchrony can be understood as representing the mutual temporal disconnect between performers over the course of an exchange in solo behavior. Alternatively, asynchrony lead establishes the absolute temporal lead exhibited by whichever performer plays first at each unison timepoint, and allows for a comparison between starter vs. joiner behavior. Asynchrony variability also provides further opportunity to evaluate possible differences in starter vs. joiner asynchrony behavior during performance.

Individual Tempo

We also evaluated individual tempo through the IUIs derived for the starting and joining player separately in each trial, as in Eq. 4,

$$\text{player tempo}\,(n) = 60/\left(\mathit{player\ IUI}\,[n]\right). \tag{4}$$

These series were used to determine the average and standard deviation of starting and joining player tempo for each trial.

For each of the average individual tempo and standard deviation of individual tempo measures we calculated the unique starter and joiner averages for a pair across the 16 trials in each of the part similarity (similar vs. dissimilar) × ATL (10, 20, and 40 ms) condition combinations. We then used these pair averages to determine the overall average individual tempo and individual tempo variability across all pairs in each condition combination.

Perspective Taking and Locus of Control

The Interpersonal Reactivity Index (Davis, 1980) contains a Perspective Taking Subscale with seven questions. A higher score on this subscale is indicative of more frequent perspective taking, with the maximum score being 28 and the minimum being zero. One participant did not provide responses to this index, resulting in data for a total of 23 participants for this measure.

The IPC scales (Levenson, 1981) contain the Internal Locus of Control Subscale with eight questions. The maximum score on this scale is 48 and the minimum is zero, with higher scores associated with a stronger sense of internal locus of control possessed by a person. The participant who did not provide data on the Perspective Taking Subscale also did not provide responses to this subscale, resulting in data for the same total of 23 participants for this measure.

We analyzed correlations between each of the perspective taking and locus of control scores and our behavioral measures at both individual and pair levels, as described below in section "Results." Correlations based on individual behavioral measures included 23 data points, while those based on pairwise behavioral measures included 11 data points.

Statistical Analyses

For all ANOVAs conducted as described in the Results Greenhouse–Geisser correction was applied for the degrees of freedom (dfs) when the assumption of sphericity was not met. Corrected dfs are reported. *Post hoc* tests were performed using either additional ANOVAs for interactions or Fisher's least significant difference (LSD) pairwise comparisons with Bonferroni correction for main effects.

For all statistical tests, the significance level was set at alpha = 0.05. Analyses were performed using SPSS (ver. 20, IBM Inc.).

RESULTS

Pairs performed between 110 and 174 trials to achieve the 96 required successful experimental trials. An average of 73% (SD = 9%) of the total trials performed by each pair were deemed successful and included in our analyses.

Cycle Asynchrony

An initial simple regression model including both ATL and part similarity (IVs: part similarity and ATL; DV: cycle asynchrony) revealed no effect of part similarity, so the model for the collapsed data is reported (IV: ATL; DV: cycle asynchrony). This established a strong relationship between ATL and cycle asynchrony. Specifically, pairs exhibited a small degree of leading with respect to each other's behavior at the 10 ms ATL, and increasing amounts of lagging at the 20 and 40 ms ATLs (see **Figure 3A**).[3] ATL significantly predicted cycle asynchrony, $b = 0.93$, $t(34) = 14.56$, $p < 0.001$. ATL also explained a significant proportion of variance in cycle asynchrony, $R^2 = 0.86$, $F(1,34) = 211.88$, $p < 0.001$.

Collective Tempo

Collective tempo curves for each ATL condition illustrated characteristic patterns of tempo change over the course of a trial (see **Figure 3B**). On average, pairs accelerated slightly in the 10 ms condition, maintained the starting tempo in the 20 ms condition, and exhibited a substantial, progressive decrease in tempo in the 40 ms condition. An initial linear regression model for collective tempo used part similarity, unison note position, and ATL (IVs: part similarity, unison

[3]We are presenting cycle asynchrony and note-onset asynchrony results following the simple sign. This means that the individual who plays their note first during a given note-onset asynchrony will have a smaller time value than the individual who plays second, resulting in a negative value associated with leading behavior during asynchrony. This approach is consistent with many coordination researchers but is the opposite of how results were presented by Chafe et al. (2010). Ultimately this difference in approach does not affect the interpretation of the results.

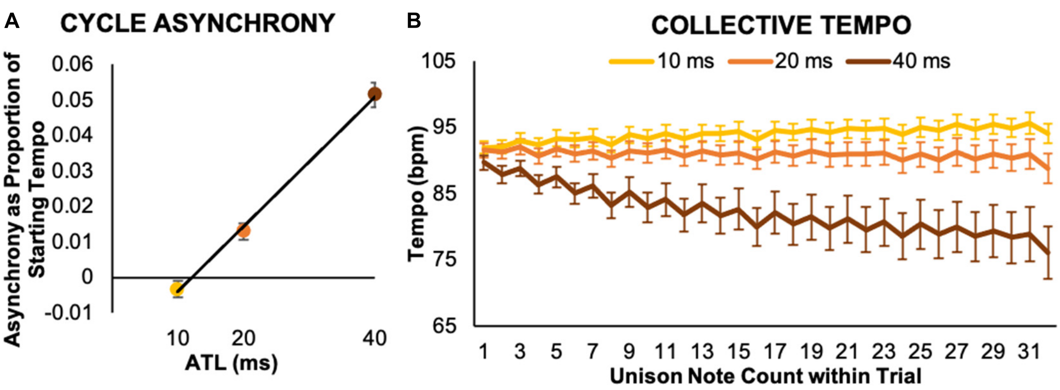

FIGURE 3 | Collective temporal behaviors. **(A)** Cycle asynchrony indicates how much lead or lag performers produced in one cycle of the collective rhythmic pattern, expressed as a ratio to the initial metronome tempo. It revealed that both performers' keypress timing preceded each other's keypress timing at the 10 ms ATL and then lagged each other's behavior by increasing amounts at the longer 20 and 40 ms ATLs. **(B)** Collective tempo indicates the tempo estimated at each unison note position. Over the course of a trial performers sped up in the 10 ms ATL condition, maintained a consistent tempo in the 20 ms ATL condition, and got substantially slower in the 40 ms ATL condition. Error bars show standard error.

note position, and ATL; DV: collective tempo). As for cycle asynchrony, the initial model revealed no effect of part similarity and the model for the collapsed data is reported (IVs: unison note position and ATL; DV: collective tempo). This model confirmed our observations based on visual inspection of the tempo curves, with both ATL, $b = -0.92$, $t(93) = -25.35$, $p < 0.001$, and unison note position, $b = -0.17$, $t(93) = -4.53$, $p < 0.001$, significantly predicting collective tempo. This model also explained a significant proportion of variance in collective tempo, $R^2 = 0.88$, $F(2,93) = 331.62$, $p < 0.001$.

Note-Onset Asynchronies

We conducted a 2 (part similarity: similar, dissimilar) × 3 (ATL: 10, 20, and 40 ms) repeated measures analysis of variance (ANOVA) on the within-person difference between the frequency of leading behavior in the starter vs. joiner role (IVs: part similarity and ATL; DV: frequency of leading). This revealed no significant interactions or main effects (overall $M = 0.32$, SD = 1.1).

We conducted a 2 (performer role: starter, joiner) × 2 (part similarity: similar, dissimilar) × 3 (ATL: 10, 20, and 40 ms) repeated measures ANOVA on asynchrony lead (IVs: performer role, part similarity, and ATL; DV: asynchrony lead). Specifically, this measure captured the magnitude of average temporal lead for each player when they played first at unison points (see **Figure 4A**). The ANOVA demonstrated a significant main effect of ATL, $F(1.13,12.37) = 55.46$, $p < 0.001$, $\eta_p^2 = 0.83$, but no other main effects or interactions between variables. Fisher's LSD *post hoc* comparisons revealed a significantly larger asynchrony lead at 40 ms ATL compared to 10 and 20 ms ATL, $ps < 0.001$.

We also conducted a 2 (performer role) × 2 (part similarity) × 3 (ATL) repeated measures ANOVA on the standard deviation of asynchrony lead (IVs: performer role, part similarity, and ATL; DV: asynchrony variability). This allowed us to determine the effect of the current experimental conditions on the variability of asynchronies between performers (see **Figure 4B**). Like the ANOVA for average asynchrony lead, this analysis illustrated a significant main effect of ATL, $F(1.16,12.8) = 40.86$, $p < 0.001$, $\eta_p^2 = 0.79$, but no other main effects or interactions between variables. Fisher's LSD *post hoc* comparisons revealed significantly larger variability at 40 ms ATL compared to 10 ms ATL and to 20 ms ATL, $ps < 0.001$.

Individual Tempo

We conducted a 2 (performer role) × 2 (part similarity) × 3 (ATL) repeated measures ANOVA on individual tempo (IVs: performer role, part similarity, and ATL; DV: individual tempo). This allowed us to identify the effect of experimental condition on performance tempo (see **Figure 5A**). The omnibus ANOVA revealed a significant interaction between performer role and ATL, $F(2,22) = 5.85$, $p = 0.009$, $\eta_p^2 = 0.35$, as well as a significant interaction between part similarity and ATL, $F(1.22,13.45) = 10.53$, $p = 0.004$, $\eta_p^2 = 0.49$, and significant main effects of performer role, $F(1,11) = 7.21$, $p = 0.021$, $\eta_p^2 = 0.40$, and ATL, $F(1.03,11.28) = 318.40$, $p < 0.001$, $\eta_p^2 = 0.97$. The interactions are detailed below and in **Figure 5A**; the main effect of performer role revealed a greater individual tempo for the joiner compared to the starter while the main effect of ATL revealed significant differences in individual tempo between all three conditions ($ps < 0.001$). The fastest individual tempo was observed at 10 ms ATL and progressively slower tempos were observed at 20 and 40 ms ATL, respectively.

To explore the interaction between performer role and ATL we collapsed the individual tempo data across the similar vs. dissimilar part similarity conditions and conducted a simple effects analysis evaluating the effect of performer role in each of the ATL conditions. We found a significant effect of performer role in the 20 ms, $F(1,11) = 5.01$, $p = 0.047$, $\eta_p^2 = 0.31$, and 40 ms conditions, $F(1,11) = 8.37$, $p = 0.015$, $\eta_p^2 = 0.43$, but not in the 10 ms condition. These analyses established that in both the 20 and 40 ms ATL conditions performers with the joiner role played faster than those with the starter role. To elucidate the interaction between part similarity and ATL we collapsed the

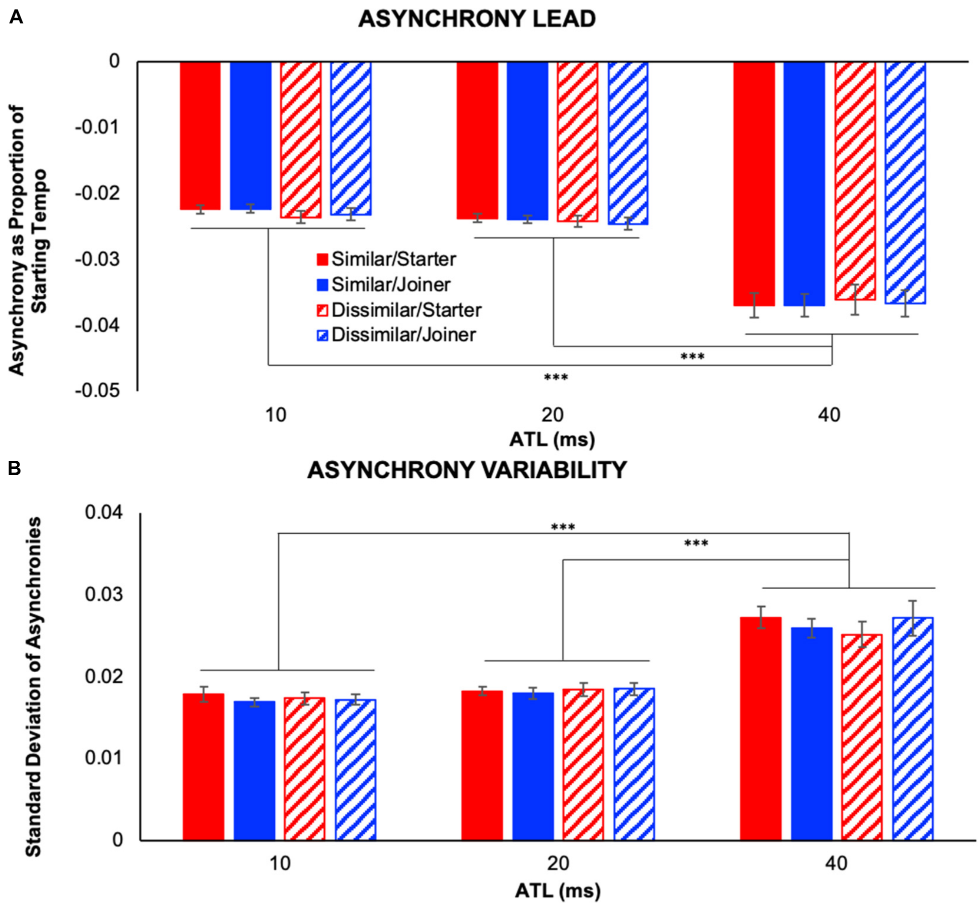

FIGURE 4 | Note asynchrony between partners. **(A)** Asynchrony lead indicates the magnitude of temporal lead exhibited when the performer played first at unison note positions. **(B)** Asynchrony variability illustrates the standard deviation of asynchrony lead that occurred at unison note positions. Both measures were significantly greater in the 40 ms ATL condition than in either of the other two ATL conditions. Error bars show standard error. ***$p < 0.001$.

individual tempo data across the starter vs. joiner performer role conditions and conducted a simple effects analysis evaluating the effect of part similarity in each of the ATL conditions. We found a significant effect of part similarity in the 10 ms, $F(1,11) = 14.0$, $p = 0.003$, $\eta_p^2 = 0.56$, and 40 ms conditions, $F(1,11) = 8.76$, $p = 0.013$, $\eta_p^2 = 0.44$, but not in the 20 ms condition. These analyses established that at 10 ms ATL performers played slower when their parts were dissimilar compared to when they were similar, but at 40 ms ATL the opposite was true as they played faster when their parts were dissimilar.

We also conducted a 2 (performer role) × 2 (part similarity) × 3 (ATL) repeated measures ANOVA on the standard deviation of individual tempo (IVs: performer role, part similarity, and ATL; DV: individual tempo variability). This allowed us to identify differences in individual tempo variability between the experimental conditions (see **Figure 5B**). The omnibus ANOVA revealed a significant interaction between performer role and ATL, $F(2,22) = 5.22$, $p = 0.014$, $\eta_p^2 = 0.32$, as well as a significant interaction between part similarity and ATL, $F(2,22) = 31.86$, $p < 0.001$, $\eta_p^2 = 0.74$, and significant main effects of performer role, $F(1,11) = 7.85$, $p = 0.017$, $\eta_p^2 = 0.42$, and ATL, $F(1.27,13.96) = 70.1$, $p < 0.001$, $\eta_p^2 = 0.86$. The interactions are detailed below and in **Figure 5B**; the main effect of performer role revealed greater individual tempo variability for the joiner compared to the starter while the main effect of ATL revealed significantly greater individual tempo variability in the 40 ms ATL condition than in either of the other two ATL conditions.

To further evaluate the interaction between performer role and ATL we collapsed the individual tempo variability data across the similar vs. dissimilar part similarity conditions and conducted a simple effects analysis evaluating the effect of performer role in each of the ATL conditions. We found a significant effect of performer role in the 20 ms, $F(1,11) = 6.35$, $p = 0.028$, $\eta_p^2 = 0.37$, and 40 ms conditions, $F(1,11) = 7.69$, $p = 0.018$, $\eta_p^2 = 0.41$, but not in the 10 ms condition. These analyses established that in the 20 and 40 ms ATL conditions performers in the joiner role exhibited greater individual tempo variability than performers in the starter role. To better understand the interaction between part similarity and ATL for standard deviation of individual tempo we collapsed individual tempo variability data across the starter vs. joiner performer role conditions and conducted a simple effects analysis evaluating the effect of part similarity in each of the

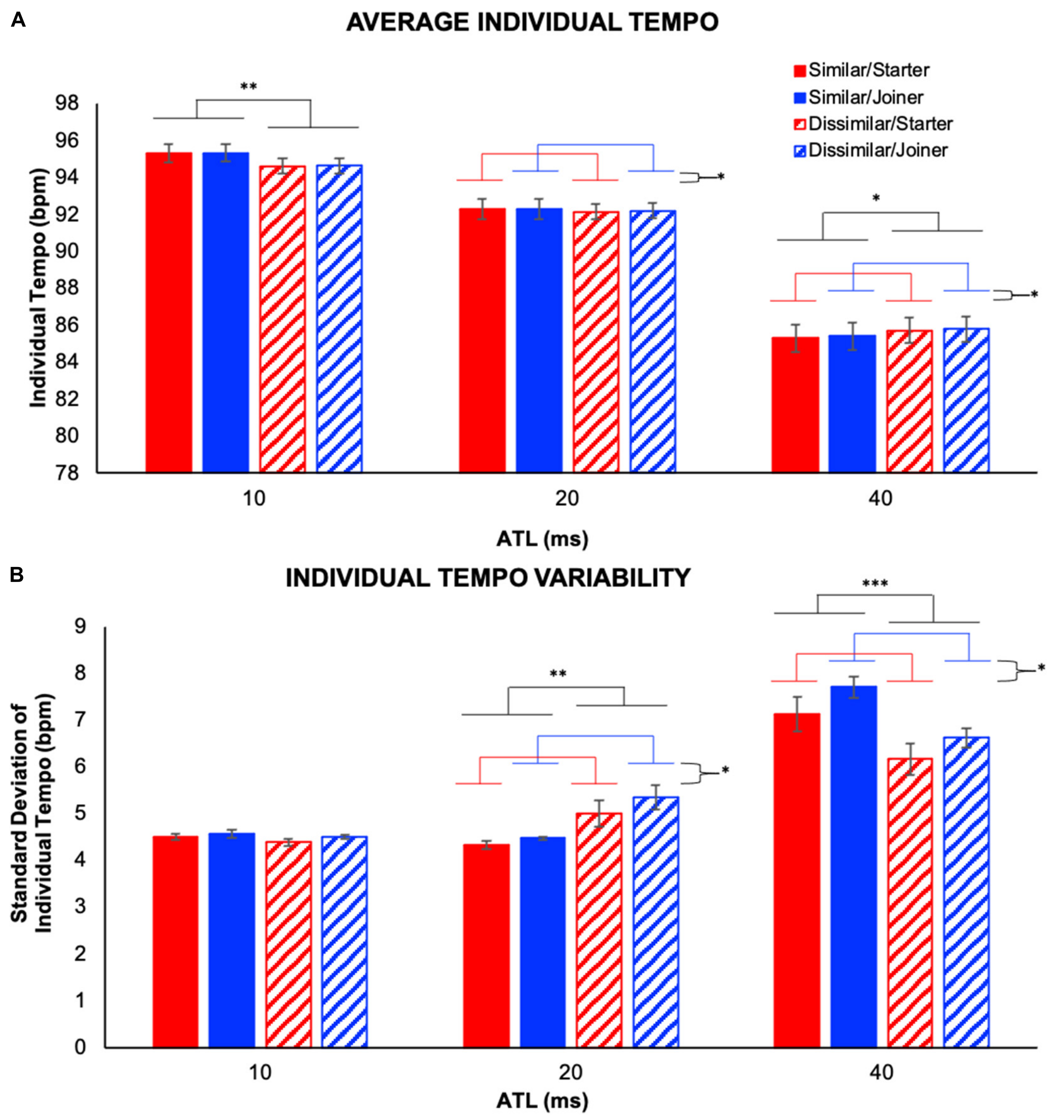

FIGURE 5 | Individual temporal behavior. **(A)** Average individual tempo illustrates the tempo produced by each individual averaged over the trials. Note that the 10- and 20-ms ATL resulted in a faster tempo than the initial tempo (90 bpm on average) whereas the 40-ms ATL made performers play slower. **(B)** Individual tempo variability is computed as the standard deviation of each individual's tempo values over the course of a trial. Significant interactions were observed between part similarity and ATL, as well as between performer role and ATL for average individual tempo **(A)**, and individual tempo variability **(B)**. Error bars show standard error. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

ATL conditions. We found a significant effect of part similarity in the 20 ms, $F(1,11) = 9.85$, $p = 0.009$, $\eta_p^2 = 0.47$, and 40 ms conditions, $F(1,11) = 65.58$, $p < 0.001$, $\eta_p^2 = 0.86$, but not in the 10 ms condition. These analyses established that while there was no effect of part similarity on individual tempo variability at 10 ms ATL, at 20 ms ATL performers were more variable when their parts were dissimilar and at 40 ms ATL performers were more variable when their parts were similar.

Perspective Taking and Locus of Control

Our results regarding the perspective taking and locus of control measures are based on a total of 23 individual participants or 11 duet pairs as one participant did not provide responses. Perspective taking scores in this group ranged from 15 to 26 ($M = 20$; SD = 3.37) and locus of control scores ranged from 22 to 41 ($M = 32.61$; SD = 5.96).

We first evaluated individual traits in relation to timing behavior. Specifically, we examined Pearson correlations between each performer's perspective taking score and their own asynchrony lead, asynchrony variability, individual tempo, and individual tempo variability performing (1) the starter role, and (2) the joiner role (IV: performer perspective taking score; DVs: asynchrony lead as starter, asynchrony variability as starter, individual tempo as starter, individual tempo variability as starter, asynchrony lead as joiner, asynchrony variability as joiner, individual tempo as joiner, and individual tempo variability as joiner). We performed the same set of correlations for each performer's locus of control score (IV: performer locus of control score; DVs: asynchrony lead as starter, asynchrony variability as starter, individual tempo as starter, individual tempo variability as starter, asynchrony lead as joiner, asynchrony variability as joiner, individual tempo as joiner, and individual tempo variability as joiner).

Second, we calculated the correlation between each performer's perspective taking score and the difference in each of the asynchrony and tempo measures between the participant's behavior when they performed the starter role vs. when they performed the joiner role (IV: performer perspective taking score; DVs: difference in asynchrony lead as starter vs. joiner, difference in asynchrony variability as starter vs. joiner, difference in individual tempo as starter vs. joiner, and difference in individual tempo variability as starter vs. joiner). We also performed the same correlations using each performer's locus of control score (IV: performer locus of control score; DVs: difference in asynchrony lead as starter vs. joiner, difference in asynchrony variability as starter vs. joiner, difference in individual tempo as starter vs. joiner, and difference in individual tempo variability as starter vs. joiner).

Lastly, we examined possible associations between within-pair differences in co-performer trait scores and interactive behavior during performance. Specifically, we ran two additional sets of correlations. In the first set we calculated the correlation between co-performer perspective taking score differences and differences in the average starter vs. joiner behavior exhibited collectively by both co-performers for the asynchrony and tempo measures (IV: co-performer difference in perspective taking scores; DVs: difference in pair asynchrony lead for starter vs. joiner roles, difference in pair asynchrony variability for starter vs. joiner roles, difference in pair individual tempo for starter vs. joiner roles, and difference in pair individual tempo variability for starter vs. joiner roles). We did the same for co-performer locus of control score differences (IV: co-performer difference in locus of control scores; DVs: difference in pair asynchrony lead for starter vs. joiner roles, difference in pair asynchrony variability for starter vs. joiner roles, difference in pair individual tempo for starter vs. joiner roles, and difference in pair individual tempo variability for starter vs. joiner roles). In the second set of correlations involving within-pair differences in co-performer scores, we assessed the correlations between co-performer perspective-taking score differences and co-performer differences in the asynchrony and tempo measures (IV: co-performer difference in perspective taking scores; DVs: co-performer difference in asynchrony lead, co-performer difference in asynchrony variability, co-performer difference in individual tempo, and co-performer difference in individual tempo variability). We also performed equivalent correlations for locus of control scores (IV: co-performer difference in locus of control scores; DVs: co-performer difference in asynchrony lead, co-performer difference in asynchrony variability, co-performer difference in individual tempo, and co-performer difference in individual tempo variability).

We found no correlations between perspective taking scores and any of the individual performance measures or within-pair differences. However, data from the 11 pairs we analyzed revealed an interesting pattern of preliminary correlations between within-pair differences in co-performer locus of control and co-performer differences in average asynchrony lead. Specifically, there was a significant negative correlation between co-performer locus of control difference and co-performer average asynchrony lead difference in the 10 ms ATL condition, $r(11) = -0.62$, $p = 0.04$ ($R^2 = 0.38$; see **Figure 6A**). This means that the larger the trait score difference between duet partners, the larger the timing discrepancy between co-performers as reflected in the average asynchrony. While there was no association between these measures in the 20 ms ATL condition (see **Figure 6B**),

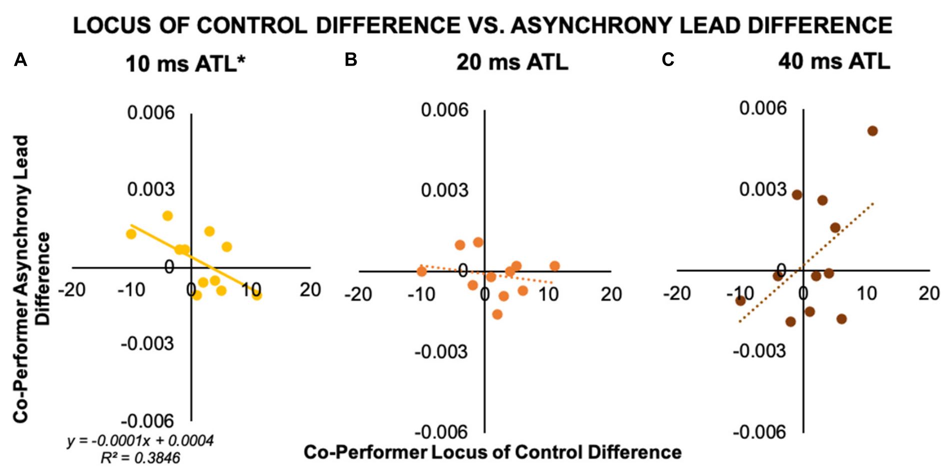

FIGURE 6 | Correlation of co-performer measures between asynchrony and locus of control score. **(A) 10 ms ATL**. Within a pair, the performer with a more internal locus of control exhibited more anticipatory behavior than their co-performer in the 10 ms ATL condition. **(B) 20 ms ATL**. There was no association between within-pair differences in locus of control and asynchrony lead in the 20 ms ATL condition. **(C) 40 ms ATL**. There was a trending positive association in the 40 ms ATL condition, with relatively less anticipatory behavior by the co-performer with the more internal locus of control. $^*p < 0.05$.

in the 40 ms ATL condition the data trended toward a positive association, $r(11) = 0.51$, $p = 0.11$ ($R^2 = 0.26$; see **Figure 6C**). As shown in **Figure 6**, the regression slopes ($\beta_{10} = -0.0001$, $\beta_{40} = 0.0002$) indicated similar changes in asynchrony lead difference vs. locus of control difference in the 10 and 40 ms ATL conditions, with opposite directions. Interestingly the intercepts were close to zero, showing that although this analysis was based on a relatively small number of pairs, both sets of co-performer differences made by our convenient sample of pianist pairs produced distributions around zero without skews.

DISCUSSION

Our findings employing piano duets are largely consistent with past work examining the effects of ATL on coordination and temporal stability during music ensemble performance (e.g., Farner et al., 2009; Chafe et al., 2010; Rottondi et al., 2015). We replicated three types of ATL effects on collective tempo such that with an ATL of 10 ms performers gradually accelerated, with an ATL of 20 ms they maintained tempo, and with an ATL of 40 ms they exhibited a progressive deceleration over the course of a trial. Our data also demonstrated that the increased asynchrony between performers at unison points was accompanied by a significant increase in the variability of asynchronies at 40 ms ATL as compared to the other two ATL conditions. This collective slowing and increased asymmetry were further confirmed in the average individual tempo and greater tempo variability. Altogether, our results demonstrate that the previously observed coordinative regime of mutual co-performer lagging and progressive tempo decline associated with ATLs at and above 40 ms is linked to a high degree of instability in both interpersonal asynchrony lead and intrapersonal performance tempo.

More importantly, however, the current study reveals novel findings on how task-related asymmetries in performer role (starting vs. joining) and musical part similarity (similar vs. dissimilar parts) shape temporal coordination, extending the replication of the overall ATL effects. Regarding performer role, it is especially remarkable that in both the 20 and 40 ms ATL conditions we saw clear differences between starters and joiners; the starters exhibited significantly lower average individual tempo and lower individual tempo variability compared to joiners. This indicates that while starters may have prioritized maintaining a stable tempo, joiners may have taken on a more adaptive role in which they adjusted their own behavior frequently to maintain coordination with their co-performer. This is consistent with the kind of functional asymmetry often exhibited by individuals assuming “leader” vs. “follower” roles in musical tasks (e.g., Fairhurst et al., 2014; Timmers et al., 2014; Chang et al., 2017) and non-musical tasks (e.g., Schmidt et al., 1994). Notably, our paradigm did not explicitly designate musical leadership to either performer. However, the starting pianist, who was given the metronome counts and started a trial as a solo, appears to have assumed leadership nonetheless. Thus, our results point to the importance of emergent leadership in temporal stability and complementary roles of musicians in duet performance.

Musical part similarity affected average individual tempo differently according to the ATL; at 10 ms ATL performers exhibited a slower average individual tempo when their parts were dissimilar compared to when they were similar. At 40 ms this was reversed such that performers displayed a faster average individual tempo when their parts were dissimilar. There was no effect of musical part similarity on average individual tempo at 20 ms ATL. This illustrates that having dissimilar parts may actually mitigate the adverse effects of ATL on individual tempo to some degree. Specifically, the acceleration observed at 10 ms ATL and the deceleration observed at 40 ms ATL appear to be somewhat diminished and there is less progressive change in tempo over time. This benefit of task complementarity is particularly interesting as previous joint action studies mostly focus on the benefits of “action simulation” for representing self and other’s action similarity through a shared coding scheme and anticipating its outcome efficiently. For example, in a joint Simon task, incongruent stimulus-response geometric mapping resulted in interference (e.g., slower reaction time) for the joint setting, and an enhanced event-related potential (ERP) component related to motor preparation (Sebanz et al., 2003; Tsai et al., 2006). Further, Novembre et al. (2014) suggested that in a piano duet task, knowledge of the other’s action represented in the motor system is key to successful coordination. This was based on the finding that TMS caused interference only for coordinating with the learned left-hand part, pre-recorded by other pianists. In our paradigm, all parts were ultimately played by all pianists, as the starting and joining roles were assigned in a counterbalanced order. This would mean that only a momentarily more active status would be given to the currently assigned duet part, compared to the partner’s part for the consecutive trials. Thus, based on the shared representation scheme, the more musically similar the parts, the more successfully the temporal coordination would have been predicted to counter the adverse effect of the ATLs on tempo drift. The opposite was observed here.

In fact, no theory accounts for how the ATL between co-actors’ actions would affect respective action representations. At least, the delayed auditory feedback for a single agent task such as auditory paced tapping is known to cause increased stimulus-tap asynchrony (e.g., Aschersleben and Prinz, 1997), indicating that theorized action planning and outcome monitoring would interact with each other. Indeed, the naturalistic delay that exists between action execution and sensory outcome is thus considered to make people assign the agency, or ownership of the movement to the outcome, and learn and calibrate the prediction according to the feedback delay (Rohde and Ernst, 2016). Within this framework, one might expect that the two action sequences assigned to self and partner would be encoded with the designated ATL. Moreover, our task employed a time-offset between the co-performers’ actions, further differentiating temporal organization between one’s own and another’s behavior. In particular, our interlocked rhythmic pattern likely assigned a momentary leadership function alternatingly to the pianist who had the solo eighth note before a given unison note, as shown in Goebl and Palmer (2009). These temporal asymmetries may play an important role in defining how action representations are managed. Especially, because the ATL was only applied to the co-performer's sound, each performer might separately represent the other's action *with* the uniquely associated time schedule (ATL plus rhythm-offset). When the co-performer played a musically distinct part, maintaining and monitoring the two duet parts scheduled independently in the motor system might be computationally easier than assigning two different time schedules to the shared, or similar, action sequences. The latter scenario may be prone to introducing cross-talk between self and other schedules, leading to inaccurate timing information extracted and encoded from the other's action outcome. This view is actually compatible with another piece of our results, where at 20 ms ATL performers showed greater variability when their parts were dissimilar, and at 40 ms ATL performers showed greater variability when their parts were similar. Therefore, it is possible that performers have access to more independent timing representations of each other's actions when their action sequences are dissimilar, pointing to the possible interaction between *what* and *when* information in joint action representation. This could then lead to decreased stability under conditions that otherwise support relatively stable coordination (i.e., 20 ms ATL) but also ultimately lead to a counter against the influence of conditions that typically perturb coordination.

Dynamical systems theory offers a mathematical framework for understanding the processes giving rise to the coordination patterns observed in our study. Here, two interacting individuals are seen as a single, multi-component system living within a "phase space" which contains all of its potential behavioral states and how they change over time (Kelso, 1995). Various symmetries and asymmetries between interacting individuals can shape these behavioral possibilities (see Richardson et al., 2016), including interpersonal social psychological asymmetries. Interestingly, past work has demonstrated that pairs of individuals with distinctly different social competence scores actually achieve more stable coordination than those with similar scores during a rhythmic synchronization task (Schmidt et al., 1994). Relatedly, pairs of individuals arbitrarily assigned to different artistic preference groups displayed greater coordination than those assigned to the same group (Miles et al., 2011). Schmidt et al. (1994) suggest that the advantage of complementarity they observed may be based on associations between the asymmetry being controlled for (e.g., social competence) and asymmetries in typical interaction behaviors such that there is a natural complementarity of stability and adaptation supporting coordination. Alternatively, Miles et al. (2011) propose that their observations may be based on a desire to lessen perceived social distance, which could also lead to increased coordinative effort. These findings point to the significance of asymmetries in determining the interplay between agents. Our current results are consistent with these past findings in demonstrating that having dissimilar musical parts is sometimes associated with greater coordination than having similar parts. Furthermore, Richardson et al. (2016) emphasize that not all types of symmetries and asymmetries consistently influence collective behavior, and that this can depend on other aspects of the interaction context. In the current study, the change in the dynamical system capturing the musicians' interacting behavior precipitated by the introduction of different ATLs may have heightened the effect of asymmetry between co-performer parts so that at 10 and 40 ms ATL having dissimilar parts actually afforded more stable coordination than having similar parts.

We found no relationship between the self-reported perspective taking measure of empathy and any of the intrapersonal or interpersonal coordination measures. This contrasts with previous findings that individuals with higher levels of perspective taking behavior show greater adaptive behavior during rhythmic coordination (e.g., Pecenka and Keller, 2011; Washburn et al., 2019). However, those previous findings were obtained when the higher empathy was assumed to enhance the action simulation in in-person, simultaneous coordination. Thus, the discrepancy here could be explained by the above proposal with respect to possible temporal information representation required for encoding and monitoring the two complementary actions. More importantly, our results did reveal an interesting, novel pattern of potential associations between the within-pair, person-related asymmetry in locus of control scores and average asynchrony lead that appears to change as a function of the ATL. Specifically, the positive association observed at 10 ms ATL shifted to no association at 20 ms ATL, and a trending negative association at 40 ms ATL. The co-performer with a more internal locus of control may therefore have been more anticipatory than the co-performer with a more external locus of control at 10 ms ATL, but less anticipatory, and potentially more reactive, at 40 ms ATL. Because the tempo drift we observed was accelerating at 10 ms ATL and decelerating at 40 ms ATL, such individuals with the more internal locus of control may have driven the tempo drift exhibited by piano duet pairs.

With respect to the effects of co-performer differences on temporal asynchrony during coordination, Loehr and Palmer (2011) demonstrated that asynchronies between pianists were smaller when the co-performers' individual preferred performance rates were more similar. While we did not measure individual pianists' preferred tempo, the within-pair difference in locus of control may function in a similar manner. Notably, our findings actually indicate that within-pair difference in locus of control has a greater influence on interpersonal interaction under conditions where maintaining stable coordination is generally more difficult. Specifically, while the within-pair difference in co-performer locus of control did not impact interaction at 20 ms ATL, with the challenges to stable coordination present at 10 ms ATL and 40 ms ATL this difference did have an effect. Interestingly, our results also suggest that the individual in a pair with the more external locus of control may actually be more resilient to the coordination challenges posed by certain ATLs. This contrasts with previous findings indicating that individuals with an internal locus of control are less adaptive to the behavior of a co-performer during coordination (e.g., Fairhurst et al., 2013).

Studies with experimentally controlled musical scores allow us to observe the effects of differences in the musical structure between co-performer parts. However, in typical performance contexts such relationships between performers are likely to be dynamic, with performers exchanging who has, for example, the higher note ratio and exhibiting associated changes in temporal coordination patterns over the course of a single piece (Bishop and Goebl, 2020). Moreover, although there is evidence that performers approach etude or exercise-like material similarly to more naturalistic musical material (Brooks, 1995), it is possible that repetitive compositions lead to reduced attention or engagement with expectancy-related processes. It is therefore important that future work on NMP employ more complex musical materials as well. The incorporation of visual sensorimotor interaction within studies of NMP is also a key consideration for further study. Interestingly, Iorwerth and Knox (2019) recently illustrated that video was rarely attended by performers during NMP, despite their self-report of its importance for successful performance. In contrast, work by Hilt et al. (2020) indicates that a range of different movement kinematics sources related to ensemble performance (e.g., bow movement and head movement) each affect either inter- or intra-group coordination. Altogether, multimodal, audiovisual interaction may significantly affect the coordination of co-performers during NMP.

Notably, our evaluation of associations between individual performer personality traits and coordination during musical performance in the context of ATL is preliminary; our study was not specifically designed to include individuals with a wide range of perspective taking or locus of control characteristics or to create pairs of performers based on similarities or differences. A better understanding of how co-performer differences influence coordination, as well as other social factors related to joint action such as likeability, in contexts that involve time-shifted and asymmetric feedback such as NMP will require further targeted investigation. We also acknowledge that while our study demonstrates differences in objective measures of performance behavior in relation to ATL, we do not know what performers consciously perceived of the ATLs. Especially for the 10 ms ATL, the additional effect on top of the base apparatus latency may be not perceivable. Even with similar ensemble performance behavior there may be differences in consciousness about the effect of a delay (e.g., at the low and high ends of a range of ATLs associated with a consistent coordination regime). Other aspects of subjective performer experience in the context of NMP will also benefit from additional study. Existing work has revealed that musicians describe the physical separation associated with NMP as challenging to communication and leading to musical issues (Iorwerth and Knox, 2019). To build on this understanding, it would, for example, be valuable to investigate performer enjoyment related to ensemble coordination in the context of ATLs. Such work will be advantageously informed by the methodologies of groups like Glowinski et al. (2015), who used a combination of motion capture and semi-structured performer interviews within an Immersive Virtual Environment to identify performance strategies associated with differences in audience engagement. Relatedly, it will also be informative to gather information about the subjective experiences of audience members in relation to ATL during NMP.

Networked music performance provides an intriguing space for exploring novel composition techniques and the creation of experimental music. In an increasingly global society as well as the current COVID-19 pandemic it also provides a practical solution for musical interaction across separate geographical locations. With the need for quarantining and social distancing following the onset of the COVID-19 pandemic in early 2020 music students, educators, and professionals have sought opportunities to continue learning, rehearsing, and performing together. This has driven rapid and transformative improvements in the primary platform for NMP, a multi-machine technology called JackTrip which supports bidirectional flows of uncompressed audio over the internet at the lowest possible latency (JackTrip, 2021). Advancements have focused on ease of use and scaling across worldwide cloud infrastructure to support a range of activities, including rehearsal for orchestral-sized ensembles. This ongoing development led by developers and musical practitioners in conjunction with the recently established JackTrip Foundation constitutes a significant contribution to the potential utility of NMP across contexts (JackTrip Foundation, 2021). Other researchers have also proposed a global metronome for facilitating NMP (Oda and Fiebrink, 2016; Hupke et al., 2019a,b). Most recently this has included the presentation of an adaptive metronome capable of supporting increased synchronicity at higher delays and reducing tempo drift *via* a low-latency solution for when high-end hardware is not available (Battello et al., 2020). Collectively, these efforts are generating technologies that will increase the quality and accessibility of cyber-mediated musical interaction far beyond the needs of the current pandemic. The existence of functional, remote musical education, for example, would greatly increase the availability and frequency of music education activities worldwide.

Our study directly informs how musical interaction in NMP could be designed with the presence of the ATL. Notably, our findings indicate that ATLs around 20 ms are most likely to support stable coordination. Internet connections supporting ATLs of 20 ms or lower will therefore be more appropriate for the majority of NMP applications than those with longer latencies. Interestingly, we also observed that task-related asymmetries, such as dissimilarity between musical parts, may increase the coordinative stability between co-performers. Accordingly, individuals leading music education activities *via* NMP might aim to prioritize exercises involving complementary, asymmetric musical tasks, especially in cases where ATL is variable or cannot be kept to a minimum around 20 ms. Further experimental research into the effects of ATL on interpersonal coordination during musical performance, including the ongoing development of computational models capturing multi-agent coordination in the context of informational delays (e.g., Rottondi et al., 2016; Demos et al., 2019; Heggli et al., 2019; Román et al., 2019; Shahal et al., 2020), is invaluable to the continued improvement of NMP. Performers have also indicated that physical separation alone poses challenges to music ensemble performance independent of those introduced by ATLs, noting that the associated hindrances to tuning, blending, and taking musical risks can all inhibit creativity (Iorwerth and Knox, 2019). The significant and transformative implication for cyber-interaction is that together researchers, developers, and performers have the opportunity to: (1) understand the effects of physical separation and perceptual-motor delays on complex musical interaction, and (2) use this information to shape the evolution of technologies for robust, versatile, and rewarding NMP.

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 Stanford University Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MW created the experimental apparatus and MW and AW verified its setup. AW and MW collected the data. AW and TF analyzed the data and wrote the manuscript. All authors designed the study and edited the manuscript.

FUNDING

This research was supported by the National Science Foundation (1606904). The content was solely the responsibility of the authors and does not necessarily represent the official views of the National Science Foundation. The authors had no patents pending or financial conflicts to disclose.

REFERENCES

Aschersleben, G., and Prinz, W. (1997). Delayed auditory feedback in synchronization. *J. Motor Behav.* 29, 35–46. doi: 10.1080/00222899709603468

Badino, L., D'Ausilio, A., Glowinski, D., Camurri, A., and Fadiga, L. (2014). Sensorimotor communication in professional quartets. *Neuropsychologia* 55, 98–104. doi: 10.1016/j.neuropsychologia.2013.11.012

Bartlette, C., Headlam, D., Bocko, M., and Velikic, G. (2006). Effect of network latency on interactive musical performance. *Music Percept.* 24, 49–62. doi: 10.1525/mp.2006.24.1.49

Battello, R., Comanducci, L., Antonacci, F., Sarti, A., Delle Monache, S., Cospito, G., et al. (2020). "An adaptive metronome technique for mitigating the impact of latency in networked music performances," in *2020 27th Conference of Open Innovations Association (FRUCT)*, (U.S.: IEEE), 10–17.

Bishop, L., and Goebl, W. (2020). Negotiating a shared interpretation during piano duo performance. *Music Sci.* 3:2059204319896152.

Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge: Bradford Books.

Brooks, R. W. (1995). Mental practice and the musician: a practical approach to practice. *Update Appl. Res. Music Educ.* 13, 4–8. doi: 10.1177/875512339501300202

Cáceres, J. P., and Chafe, C. (2010). JackTrip: under the hood of an engine for network audio. *J. New Music Res.* 39, 183–187. doi: 10.1080/09298215.2010.481361

Cáceres, J. P., Hamilton, R., Iyer, D., Chafe, C., and Wang, G. (2008). "To the edge with China: explorations in network performance," in *ARTECH 2008: Proceedings of the 4th International Conference on Digital Arts*, (US: ARTECH), 61–66.

Carôt, A., Werner, C., and Fischinger, T. (2009). "Towards a comprehensive cognitive analysis of delay-influenced rhythmical interaction," in *Proceedings of the International Computer Music Conference (ICMC 2009)*, (San Francisco: International Computer Music Association).

Chafe, C. (2011). "Living with net lag," in *Audio Engineering Society Conference: 43rd International Conference: Audio for Wirelessly Networked Personal Devices*, (New York: Audio Engineering Society).

Chafe, C., Caceres, J. P., and Gurevich, M. (2010). Effect of temporal separation on synchronization in rhythmic performance. *Perception* 39, 982–992. doi: 10.1068/p6465

Chang, A., Livingstone, S. R., Bosnyak, D. J., and Trainor, L. J. (2017). Body sway reflects leadership in joint music performance. *Proc. Natl. Acad. Sci. U. S. A.* 114, E4134–E4141.

Chauvigné, L. A., and Brown, S. (2018). Role-specific brain activations in leaders and followers during joint action. *Front. Hum. Neurosci.* 12:401. doi: 10.3389/fnhum.2018.00401

Creel, S. C., Newport, E. L., and Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *J. Exp. Psychol. Learn. Memory Cogn.* 30, 1119–1130. doi: 10.1037/0278-7393.30.5.1119

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog Select. Documents Psychol.* 10:85.

Delle Monache, S., Comanducci, L., Buccoli, M., Zanoni, M., Sarti, A., Pietrocola, E., et al. (2019). A presence-and performance-driven framework to investigate interactive networked music learning scenarios. *Wirel. Commun. Mob. Comput.* 2019:4593853.

Demos, A. P., Layeghi, H., Wanderley, M. M., and Palmer, C. (2019). Staying together: a bidirectional delay–coupled approach to joint action. *Cogn. Sci.* 43:e12766.

Dowling, W. J. (1973). The perception of interleaved melodies. *Cogn. Psychol.* 5, 322–337.

Fairbanks, G., and Guttman, N. (1958). Effects of delayed feedback upon articulation. *J. Speech Hear. Res.* 1, 12–22. doi: 10.1044/jshr.0101.12

Fairhurst, M. T., Janata, P., and Keller, P. E. (2013). Being and feeling in sync with an adaptive virtual partner: brain mechanisms underlying dynamic cooperativity. *Cereb. Cortex* 23, 2592–2600. doi: 10.1093/cercor/bhs243

Fairhurst, M. T., Janata, P., and Keller, P. E. (2014). Leading the follower: an fMRI investigation of dynamic cooperativity and leader–follower strategies in synchronization with an adaptive virtual partner. *Neuroimage* 84, 688–697. doi: 10.1016/j.neuroimage.2013.09.027

Farner, S., Solvang, A., Sæbo, A., and Svensson, U. P. (2009). Ensemble hand-clapping experiments under the influence of delay and various acoustic environments. *J. Audio Eng. Soc.* 57, 1028–1041.

Finney, S. A. (1997). Auditory feedback and musical keyboard performance. *Music Percept.* 15, 153–174. doi: 10.2307/40285747

Finney, S. A., and Warren, W. H. (2002). Delayed auditory feedback and rhythmic tapping: evidence for a critical interval shift. *Percept. Psychophys.* 64, 896–908. doi: 10.3758/BF03196794

Fischer, M. H. (2012). A hierarchical view of grounded, embodied, and situated numerical cognition. *Cogn. Process.* 13, 161–164. doi: 10.1007/s10339-012-0477-5

Gates, A., Bradshaw, J. L., and Nettleton, N. C. (1974). Effect of different delayed auditory feedback intervals on a music performance task. *Percept. Psychophys.* 15, 21–25. doi: 10.3758/BF03205822

Glowinski, D., Baron, N., Shirole, K., Coll, S. Y., Chaabi, L., Ott, T., et al. (2015). Evaluating music performance and context-sensitivity with Immersive Virtual Environments. *EAI Endorsed Trans. Creative Technol.* 2:e3. doi: 10.4108/ct.2.2.e3

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

Halpern, A. R. (1984). Perception of structure in novel music. *Memory & Cognition*, 12, 163–170.

Heggli, O. A., Cabral, J., Konvalinka, I., Vuust, P., and Kringelbach, M. L. (2019). A Kuramoto model of self-other integration across interpersonal synchronization strategies. *PLoS Comput. Biol.* 15:e1007422. doi: 10.1371/journal.pcbi.1007422

Hilt, P. M., Badino, L., D'Ausilio, A., Volpe, G., Tokay, S., Fadiga, L., et al. (2020). Author correction: multi-layer adaptation of group coordination in musical ensembles. *Sci. Rep.* 10:597. doi: 10.1038/s41598-019-55965-3

Howell, P., Powell, D. J., and Khan, I. (1983). Amplitude contour of the delayed signal and interference in delayed auditory feedback tasks. *J. Exp. Psychol. Hum. Percept. Perform.* 9, 772–784. doi: 10.1037/0096-1523.9.5.772

Hupke, R., Beyer, L., Nophut, M., Preihs, S., and Peissig, J. (2019a). "A rhythmic synchronization service for music performances over distributed networks," in *Fortschritte der Akustik: DAGA, Jahrestagung für Akustik*, (Washington: PMSE).

Hupke, R., Beyer, L., Nophut, M., Preihs, S., and Peissig, J. (2019b). "Effect of a global metronome on ensemble accuracy in networked music performance," in *Audio Engineering Society Convention 147*, (New York: Audio Engineering Society).

Iorwerth, M., and Knox, D. (2019). Playing together, apart: musicians' experiences of physical separation in a classical recording session. *Music Percept. Interdiscip. J.* 36, 289–299. doi: 10.1525/mp.2019.36.3.289

Jack, R. H., Mehrabi, A., Stockman, T., and McPherson, A. (2018). Action-sound latency and the perceived quality of digital musical instruments: comparing professional percussionists and amateur musicians. *Music Percept. Interdiscip. J.* 36, 109–128. doi: 10.1525/mp.2018.36.1.109

JackTrip (2021). *Documentation*. Available Online at: https://ccrma.stanford.edu/groups/soundwire/software/jacktrip/ [accessed June 17, 2021].

JackTrip Foundation (2021). *Activities, Directors, and Affiliates*. Available Online at: https://www.jacktrip.org/foundation.html [accessed June 17, 2021].

Kelso, J. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge: MIT press.

Konvalinka, I., Bauer, M., Stahlhut, C., Hansen, L. K., Roepstorff, A., and Frith, C. D. (2014). Frontal alpha oscillations distinguish leaders from followers: multivariate decoding of mutually interacting brains. *Neuroimage* 94, 79–88. doi: 10.1016/j.neuroimage.2014.03.003

Landström, S., Bergström, J., Westerberg, E., and Hammarwall, D. (2016). NB-IoT: a sustainable technology for connecting billions of devices. *Ericsson Technol. Rev.* 4, 2–11.

Levenson, H. (1981). Differentiating among internality, powerful others, and chance. *Res. Locus Control Construct* 1, 15–63. doi: 10.1016/B978-0-12-443201-7.50006-3

Lindenberger, U., Li, S. C., Gruber, W., and Müller, V. (2009). Brains swinging in concert: cortical phase synchronization while playing guitar. *BMC Neurosci.* 10:22. doi: 10.1186/1471-2202-10-22

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

Meals, C. D., Morrison, S. J., and Confredo, D. A. (2019). The effects of temporal action-sound congruence on evaluations of conductor quality. *Music Sci.* 2:2059204319891968. doi: 10.1177/2059204319891968

Miles, L. K., Lumsden, J., Richardson, M. J., and Macrae, C. N. (2011). Do birds of a feather move together? Group membership and behavioral synchrony. *Exp. Brain Res.* 211, 495–503. doi: 10.1007/s00221-011-2641-z

Novembre, G., Ticini, L. F., Schütz-Bosbach, S., and Keller, P. E. (2012). Distinguishing self and other in joint action. Evidence from a musical paradigm. *Cereb. Cortex* 22, 2894–2903. doi: 10.1093/cercor/bhr364

Novembre, G., Ticini, L. F., Schütz-Bosbach, S., and Keller, P. E. (2013). Motor simulation and the coordination of self and other in real-time joint action. *Soc. Cogn. Affect. Neurosci.* 9, 1062–1068. doi: 10.1093/scan/nst086

Novembre, G., Ticini, L. F., Schütz-Bosbach, S., and Keller, P. E. (2014). Motor simulation and the coordination of self and other in real-time joint action. *Social Cogn. Affect. Neurosci.* 9, 1062–1068. doi: 10.1093/scan/nst086

O'Conaill, B., Whittaker, S., and Wilbur, S. (1993). Conversations over video conferences: an evaluation of the spoken aspects of video-mediated communication. *Hum. Comput. Interact.* 8, 389–428. doi: 10.1207/s15327051hci0804_4

Oda, R., and Fiebrink, R. (2016). "The global metronome: absolute tempo sync for networked musical performance," in *Proceedings of the International Conference on New Interfaces Musical Expression*, (Brisbane: NIME)26–31.

Pecenka, N., and Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Exp. Brain Res.* 211, 505–515. doi: 10.1007/s00221-011-2616-0

Pfordresher, P., and Palmer, C. (2002). Effects of delayed auditory feedback on timing of music performance. *Psychol. Res.* 66, 71–79. doi: 10.1007/s004260100075

Richardson, M. J., Washburn, A., Kallen, R. W., and Harrison, S. J. (2016). "Symmetry and the behavioral dynamics of social coordination," in *Interpersonal Coordination and Performance in Social Systems*, eds P. Passos and K. Davis (Abingdon: Routledge), 65–81.

Robinson, G. M. (1972). The delayed auditory feedback effect is a function of speech rate. *J. Exp. Psychol.* 95, 1–5. doi: 10.1037/h0033268

Rohde, M., and Ernst, M. O. (2016). Time, agency, and sensory feedback delays during action. *Curr. Opin. Behav. Sci.* 8, 193–199. doi: 10.1016/j.cobeha.2016.02.029

Román, I. R., Washburn, A., Large, E. W., Chafe, C., and Fujioka, T. (2019). Delayed feedback embedded in perception-action coordination cycles results in anticipation behavior during synchronized rhythmic action: a dynamical systems approach. *PLoS Comput. Biol.* 15:e1007371. doi: 10.1371/journal.pcbi.1007371

Rottondi, C., Chafe, C., Allocchio, C., and Sarti, A. (2016). An overview on networked music performance technologies. *IEEE Access* 4, 8823–8843. doi: 10.1109/ACCESS.2016.2628440

Rottondi, C. E. M., Buccoli, M., Zanoni, M., Garao, D. G., Verticale, G., and Sarti, A. (2015). Feature-based analysis of the effects of packet delay on networked musical interactions. *J. Audio Eng. Soc.* 1, 864–875. doi: 10.17743/jaes.2015.0074

Sänger, J., Müller, V., and Lindenberger, U. (2013). Directionality in hyperbrain networks discriminates between leaders and followers in guitar duets. *Front. Hum. Neurosci.* 7:234. doi: 10.3389/fnhum.2013.00234

Schmidt, R. C., Christianson, N., Carello, C., and Baron, R. (1994). Effects of social and physical variables on between-person visual coordination. *Ecol. Psychol.* 6, 159–183. doi: 10.1207/s15326969eco0603_1

Sebanz, N., Knoblich, G., and Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition* 88, B11–B21. doi: 10.1016/S0010-0277(03)00043-X

Shahal, S., Wurzberg, A., Sibony, I., Duadi, H., Shniderman, E., Weymouth, D., et al. (2020). Synchronization of complex human networks. *Nat. Commun.* 11, 1–10. doi: 10.1038/s41467-020-17540-7

Shaki, S., and Fischer, M. H. (2018). Deconstructing spatial-numerical associations. *Cognition* 175, 109–113. doi: 10.1016/j.cognition.2018.02.022

Timmers, R., Endo, S., Bradbury, A., and Wing, A. M. (2014). Synchronization and leadership in string quartet performance: a case study of auditory and visual cues. *Front. Psychol.* 5:645. doi: 10.3389/fpsyg.2014.00645

Tsai, C. C., Kuo, W. J., Jing, J. T., Hung, D. L., and Tzeng, O. J. L. (2006). A common coding framework in self–other interaction: evidence from joint action task. *Exp. Brain Res.* 175, 353–362. doi: 10.1007/s00221-006-0557-9

Vanzella, P., Balardin, J. B., Furucho, R. A., Zimeo Morais, G. A., Braun Janzen, T., Sammler, D., et al. (2019). fNIRS responses in professional violinists while playing duets: evidence for distinct leader and follower roles at the brain level. *Front. Psychol.* 10:164. doi: 10.3389/fpsyg.2019.00164

Vesper, C., Van Der Wel, R. P., Knoblich, G., and Sebanz, N. (2011). Making oneself predictable: reduced temporal variability facilitates joint action coordination. *Exp. Brain Res.* 211, 517–530. doi: 10.1007/s00221-011-2706-z

Volpe, G., D'Ausilio, A., Badino, L., Camurri, A., and Fadiga, L. (2016). Measuring social interaction in music ensembles. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20150377. doi: 10.1098/rstb.2015.0377

Washburn, A., Román, I., Huberth, M., Gang, N., Dauer, T., Reid, W., et al. (2019). Musical role asymmetries in piano duet performance influence alpha-band neural oscillation and Behavioral synchronization. *Front. Neurosci.* 13:1088. doi: 10.3389/fnins.2019.01088

Wing, A. M., Endo, S., Bradbury, A., and Vorberg, D. (2014). Optimal feedback correction in string quartet synchronization. *J. R. Soc. Interface* 11:20131125. doi: 10.1098/rsif.2013.1125

Wright, M., Cassidy, R. J., and Zbyszyński, M. F. (2004). "Audio and gesture latency measurements on Linux and OSX," in *Proceedings of the International Computer Music Conference*, (Miami: International Computer Music Association), 423–429.

Zamm, A., Wellman, C., and Palmer, C. (2016). Endogenous rhythms influence interpersonal synchrony. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 611–6. doi: 10.1037/xhp0000201

Conflict of Interest: CC was employed by the company JackTrip Labs, 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.

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 | Frontiers in Neuroscience

PERSPECTIVE
published: 22 June 2022
doi: 10.3389/fnins.2022.793163

Atonal Music as a Model for Investigating Exploratory Behavior

Iris Mencke[1], Diana Omigie[2], David Ricardo Quiroga-Martinez[3] and Elvira Brattico[3,4]*

[1] Department of Music, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany, [2] Department of Psychology, Goldsmiths, University of London, London, United Kingdom, [3] Department of Clinical Medicine, Center for Music in the Brain, Aarhus University and Royal Academy of Music, Aarhus, Denmark, [4] Department of Education, Psychology and Communication, University of Bari Aldo Moro, Bari, Italy

Atonal music is often characterized by low predictability stemming from the absence of tonal or metrical hierarchies. In contrast, Western tonal music exhibits intrinsic predictability due to its hierarchical structure and therefore, offers a directly accessible predictive model to the listener. In consequence, a specific challenge of atonal music is that listeners must generate a variety of new predictive models. Listeners must not only refrain from applying available tonal models to the heard music, but they must also search for statistical regularities and build new rules that may be related to musical properties other than pitch, such as timbre or dynamics. In this article, we propose that the generation of such new predictive models and the aesthetic experience of atonal music are characterized by internal states related to exploration. This is a behavior well characterized in behavioral neuroscience as fulfilling an innate drive to reduce uncertainty but which has received little attention in empirical music research. We support our proposal with emerging evidence that the hedonic value is associated with the recognition of patterns in low-predictability sound sequences and that atonal music elicits distinct behavioral responses in listeners. We end by outlining new research avenues that might both deepen our understanding of the aesthetic experience of atonal music in particular, and reveal core qualities of the aesthetic experience in general.

Keywords: new music, atonal music, exploratory behavior, predictive processing, aesthetic experience, neuroaesthetics, musical pleasure

OPEN ACCESS

Edited by:
Clara Eline James, University of Applied Sciences and Arts of Western Switzerland, Switzerland

Reviewed by:
Peter Cariani, Boston University, United States
Hamish Innes-Brown, Eriksholm Research Centre, Denmark

****Correspondence:***
Iris Mencke
iris.mencke@ae.mpg.de

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

Received: *11 October 2021*
Accepted: *12 May 2022*
Published: *22 June 2022*

Citation:
Mencke I, Omigie D, Quiroga-Martinez DR and Brattico E (2022) Atonal Music as a Model for Investigating Exploratory Behavior. Front. Neurosci. 16:793163. doi: 10.3389/fnins.2022.793163

INTRODUCTION

Western art music from the twentieth and the twenty-first centuries is regarded as the continuation of the earlier classic-romantic Western art music tradition and has been characterized by transformation and innovation since its origins in Europe around 1910. The continuous invention of novel compositional techniques and practices has ultimately led to the pluralism of music styles and idioms seen today (Taruskin, 2010). One common element of twentieth/twenty-first century Western art music is its aesthetic premise of creating and presenting something completely novel if not merely experimental (Maletz, 2011; Hiekel, 2016; Mencke et al., 2019). In fact, these aesthetic premises resemble the motivations for engaging with Western art music: one of our studies showed that expert listener's expectations revolve around the desire to experience both novelty and surprise (Mencke et al., 2022) and similarly, the audience of contemporary art has been found to seek for experiences in which they are presented with novelty and are able to engage with challenging and difficult to understand materials (Gross and Pitts, 2016).

Fundamental Characteristics of Atonal Music

Despite the tremendous variety of compositional creation in the twentieth and twenty-first centuries, the focus of this article relates to two specific characteristics of this period of music. First, we focus on atonal music (Griffiths, 2001) which we define as music that clearly lacks a tonal center and that suspends tonal hierarchical relations between scale degrees (Bigand and Poulin-Charronnat, 2016). Typically, all 12 tones of the chromatic scale are treated as equal, leading to an abstract and non-hierarchical tonal structure. Atonal music is used here as a style-independent term as it is a description of a certain feature of the music that appeared (and still appears) in various epochs of the twentieth and twenty-first centuries. However, an exemplary compositional style is serialism. It originated in the 12-tone technique developed by Arnold Schoenberg in the early 1920s, and was further developed in the 1950s by composers such as Karel Goeyvaerts, Karlheinz Stockhausen, Pierre Boulez, and Luigi Nono (Dibelius, 1998; Taruskin, 2010). While a large number of pieces composed in a serial manner are strictly atonal, non-serial music that is atonal has been composed throughout the twentieth and twenty-first centuries until today making Atonality a critical feature of many compositions (Kostka and Santa, 2018).

Second, we focus on music that lacks a clear metrical structure. In post-tonal music, perceived rhythms are often highly complex and varied leading to an a-metric structure (Kostka and Santa, 2018). A number of pieces composed during serialism can serve as examples, since the deterministic principles guiding pitch in 12-tone compositions was—among other features—transferred to note duration. Critically, the irregular rhythmic structure and free treatment of meter that results from this compositional technique (Grant, 2001; Kostka and Santa, 2018) ultimately prevents listeners from easily following and entraining with the heard music (London, 2004).

Indeed, both atonal and ametrical musical structures strongly lower the predictability of the music (Kramer, 1989). In particular, serial pieces are, according to Lerdahl, "cognitively opaque," by which he is referring to the distance between the composed and heard structure (Lerdahl, 1992, p. 115; Lerdahl, 2019). Thus, even though compositions were based on a certain set of principles (e.g., re-presenting a given row of tones reversed and/or turned upside down) for a listener this underlying non-hierarchical organization is barely perceivable. An exemplary series of pieces that resemble or represent both the atonal and the ametric aspects that the current article is targeting are the piano pieces I-IV by Karlheinz Stockhausen, composed in the early 1950s (Stockhausen, 1954; Wörner, 1973). These fall under the category of pointillistic compositions, a technique that has the tendency to "isolate the sounds into 'points'" (Kostka and Santa, 2018, p. 233) and that generate a "verticalize(d) [...] sense of time" (Kramer, 1989, p. 202: for more exemplary pieces see Mencke et al., 2019).

The Challenge of Atonal Music

With the brain's prioritization of auditory input containing cognitive reference points (Rosch, 1975) such as the tonic or the cadence in tonal music (Krumhansl and Cuddy, 2010), it is no wonder that atonal music is widely recognized as challenging (Utz, 2016). As early as in the dawn of experimental psychology research, Wilhelm Wundt proposed an inverted U-shaped relationship between stimulus intensity and pleasure, known as the Wundt-curve (Wundt, 1896): a stimulus is liked until a certain level of intensity, but if intensity further increases, pleasure will decay. This theory was extended by Daniel Berlyne, who proposed that each stimulus has a certain potential for physiological arousal: defined as "arousal potential" and referring to the excitement of the nervous system in response to the stimulus (Berlyne, 1971). As a stimulus' arousal potential is mostly modulated by variables such as complexity, surprise, ambiguity and novelty, atonal music has a high arousal potential that is likely going beyond the pleasure peak for most listeners (Berlyne, 1954, 1970; Marin et al., 2016; Marin, 2020). In turn, listeners with a preference for atonal music might particularly seek for this complexity and therefore potentially have an increased need to engage in cognitive tasks (as can be measured by the *Need for Cognition*—Scale; Cacioppo and Petty, 1982; Cacioppo et al., 1996) in conjunction with a particularly high degree of tolerance to ambiguity (McLain, 2009; for more hypotheses regarding the correlation between preferences for atonal music and interindividual differences see Mencke et al., 2019).

Atonal music typically also has a high degree of dissonance, a key factor underlying the challenge it presents listeners. Dissonant intervals lead to beating and roughness (Helmholtz, 1954) and are often experienced as sensorily unpleasant (Brattico, 2015). It is important to note that dissonance is a central component in perhaps all styles of music and in the context of tonal music, is an important means to generate moments of tension and release (Juslin, 2013; Brincker, 2015; Lehne and Koelsch, 2015; Brattico, 2021), both of which are experienced as pleasurable by many listeners. However, works of atonal music often are predominantly dissonant, which may be why listeners might find this music less appealing. Here, the idea that tolerance toward negative feelings or emotions increases in the context of art perception and that such feelings or emotions are used by artists to reinforce the intensity of an aesthetic experience (Menninghaus et al., 2017) may explain why listeners nevertheless choose to engage with this music. Acceptance of high levels of dissonance in atonal music parallels the fact that feelings of sadness are particularly appreciated in music (Taruffi and Koelsch, 2014; Sachs et al., 2015; Brattico et al., 2016; Eerola et al., 2018).

Roadmap

In the following, we are concerned with the internal states and cognitive processes that occur while listening to atonal music and argue that in the uncertain environment it provides, neural and cognitive mechanisms associated with exploratory behavior may become engaged. Further we suggest that a focus on this exploratory behavior, may be helpful when considering the predictive processes and hedonic values underlying reception of atonal music. We first give a brief summary of the perceptual and predictive mechanisms that have been suggested to play a key role in music processing and in music appreciation (section "Predictive Dynamics in Music"), before summarizing empirical findings consistent with our proposal that atonal music affords exploratory behavior (section "Neural and Behavioral Levels of Predictive Processing Under High Uncertainty"). In section "Building Predictions in Atonal Music," we elaborate on how predictive models might emerge as a listener engages with atonal music and argue that these specific model-building processes may be responsible for a sustained exploratory internal state in a listener. Finally, we present future research avenues addressing the role of exploration as a crucial facet of the engagement with atonal music, as well as with respect to aesthetic experience in general (section "Future Research Avenues").

PREDICTIVE DYNAMICS IN MUSIC

Musical Expectancy: Bottom-Up and Top-Down

When listening to music, individuals constantly generate expectations about future events and about how the music will evolve. The interplay between the violation and confirmation of these expectations is widely accepted to be a key underlying mechanism for music-induced pleasure in tonal music (Meyer, 1956; Blood and Zatorre, 2001; Huron, 2006). However, in the context of atonal music, where expectations and predictions are more difficult to establish, it seems relevant to have a closer look at the interplay between bottom-up and top-down expectations and how their weighting may differ in atonal music.

With regard to bottom-up expectations Gestalt principles such as pitch and temporal proximity (Deutsch, 1999), rhythmic grouping (Koelsch, 2012), and sound similarity (Bregman, 1990; Deutsch, 1999) have been proposed to be central. According to Lerdahl, grouping preference rules can additionally be based on changes in intensity and articulation (Lerdahl and Jackendoff, 1983; Clarke, 1999). In other words the grouping of notes or phrases into processable chunks can be carried out with regard to a number of musical properties. Bottom-up grouping is thought to happen in early processing stages, namely in short-term memory comprising a time span of 250 ms–8 s (Snyder, 2008). Furthermore, grouping processes in music are related to chunking which determines the memorability of sequences, whereby the better a musical phrase can be chunked, both the better it can be recalled (Lerdahl and Jackendoff, 1983; Bregman, 1990; Snyder, 2000) and the more efficiently it can be processed (Deutsch, 2013).

While, on the one hand, our expectations are strongly shaped by Gestalt-like principles, thanks to bottom-up grouping, musical expectancy is similarly modulated, on the other hand, by top-down mechanisms that are linked to statistical learning (Loui and Wessel, 2006; Pearce, 2018). Our brain is sensitive to auditory regularities in the environment, internalizing and using them to predict future events (Friston, 2009; Clark, 2013). In the case of music, this means that the predictive model of the musical style(s) that we grew up with is the one that we internalize most (Kliuchko et al., 2019) and for which we have the best predictive model. Processes underlying musical expectancy are not only modulated by cultural background but also by style-specific expertise and by piece-specific knowledge (Huron, 2006).

These bottom-up and top-down expectations stemming from both Gestalt principles and statistical learning have been shown to work in parallel and be critical to the reception of tonal music (Morgan et al., 2019). In atonal music, however, as we will argue in later sections of the article, the difficulty of the listener to apply clear top-down expectations may mean an upweighting of bottom-up grouping effects.

Predictive Coding of Music

One theory that combines bottom-up and top-down processes is the theory of predictive coding and having shown great success in accounting for how we respond to tonal music, it is relevant to consider it in the context of atonal music.

The predictive coding theory is based on the assumption that the brain, as a prediction machine, continuously tries to predict upcoming sensory input by means of generative cognitive models on higher levels. These top–down predictions encounter bottom-up sensory input and in cases where the model's prediction is incorrect, a prediction error signal occurs to update the model (Friston and Kiebel, 2009; Clark, 2013). To measure this prediction error signal, often times the mismatch-negativity (MMN) is used, which is a brain response to deviating sounds in a regular sound environment (Näätänen et al., 2007). The MMN is suggested to index violations of predictions that are set up by the statistics of an unfolding melody (Denham and Winkler, 2006) and is therefore regarded as a neural marker of prediction error in musical processing.

Critically, research shows that the brain effectively adapts to the statistics of incoming input and thereby to the level of predictability of the sensory input. In high uncertainty contexts, predictions therefore become attenuated or imprecise and the error minimization process is consequently reduced, if not nullified (Garrido et al., 2013; Sohoglu and Chait, 2016; Heilbron and Chait, 2018). Accordingly, it has been suggested that high predictive uncertainty of a stimulus is reflected in a weak predictive model (Vuust and Frith, 2008; Ross and Hansen, 2016; Koelsch et al., 2019) and that in such high-uncertainty conditions imprecise prediction errors have been suggested to be "effectively ignored" (Koelsch et al., 2019).

Recently, the predictive coding framework has been used to account for the enjoyment that humans derive from listening to music (Vuust and Kringelbach, 2010; Koelsch et al., 2019; Brattico, 2021). In particular, it was put forward that—since music plays with a listeners' expectations and predictions—music provides the opportunity to constantly resolve uncertainty (Koelsch et al., 2019). It has been suggested that the enjoyment of music stems from the interplay between levels of predictability that are typical for music, and how those unfold over time. In some moments, music allows us to generate strong predictions, and at other moments, our predictions are more uncertain, with such changes in the precision of predictions relating to a range of different musical properties such as to meter, melody, rhythm of harmonic progressions (Pearce and Wiggins, 2012; Koelsch et al., 2019).

In essence, the process of resolving uncertainty is proposed to be an appealing element when listening to any music that has some degree of regularity and crucially, resolving uncertainty is typical for states of exploration (Friston, 2010; Schwartenbeck et al., 2013). However, the extent to which a drive to resolve uncertainty is important in atonal music has yet to be fully explored.

NEURAL AND BEHAVIORAL LEVELS OF PREDICTIVE PROCESSING UNDER HIGH UNCERTAINTY

Generally, atonal music has indeed long been a fringe topic in the literature of music cognition, with Western tonal music being central to most studies. The atonal music studies that do exist make explicit listeners' everyday intuitions that atonal music is more difficult to recognize and remember than tonal music (see for instance: Cuddy et al., 1981; Dibben, 1994; Dowling et al., 1995; Schulze et al., 2012). An interesting question is what insights a predictive coding framework can offer the study of atonal music. Indeed, as predictive coding frameworks have predominantly dealt with Western tonal music, their application to atonal music would seem to present a particular challenge: this both when considering predictive processes in general and when considering the role predictive dynamics might play in music-induced pleasure.

Prediction in Atonal Music

Only a few studies have focused on the extent to which listeners make or do not make predictions when listening. In a series of experiments, Krumhansl et al. (1987) provided data suggesting that tonal predictions are being made even when individuals (from Western culture) listen to atonal sequences. After hearing a 12-tone row, trained musicians expected that the following tone would not have been heard in the preceding row and would not suggest any sort of tonal center. This is consistent with the interpretation of another study (Ockelford and Sergeant, 2012) suggesting that listeners' expectations in response to atonal music may follow an "anti-structure" and the claim that listeners of atonal music adapt to the unpredictability of the music by "expect(ing) the unexpected" (Huron, 2006, p. 331). In any case, atonal contexts have been shown to behaviorally evoke weaker expectancies than a major-minor tonal context with higher false alarm rates for recognizing atonal unexpected target notes (Vuvan et al., 2014). All of the above accord with the idea that atonal contexts provide a more significant challenge to a listener than a major-minor tonal context even though studies could provide evidence that a listener is implicitly internalizing other regularities than pitch as for instance with regard to timbre (Tillmann and McAdams, 2004). In recent work, we asked how high-uncertainty musical contexts affect brain activity, in particular the precision of automatically generated predictions (Quiroga-Martinez et al., 2019; Haumann et al., 2021). Consistent with research showing that early evoked responses are dampened in unpredictable contexts (Garrido et al., 2013; Hsu et al., 2015; Sohoglu and Chait, 2016; Southwell and Chait, 2018), we provided evidence that contextual uncertainty can attenuate a pre-attentive neural response to deviating sounds known as the mismatch-negativity (MMN) (Quiroga-Martinez et al., 2020a,b). However, in a yet more recent study, we demonstrated that contextual uncertainty may not always completely eliminate sensory sensitivity to deviating events (Mencke et al., 2021). In contrast to previous studies that largely used musical stimuli composed according to Western tonal rules, we created atonal melodies based on original 12-tone rows by Arnold Schoenberg and then measured how the brain responded to deviants using the MMN (Näätänen et al., 2007). In 20 non-musicians measured with magnetoencephalography (MEG) and other 39 non-musicians tested behaviorally, we found that, while the MMN response to four types of deviants (pitch, timbre, intensity, location) did not differ between tonal and atonal sequences, the behavioral accuracy and confidence in pitch deviance detection were nevertheless significantly lower in atonal sequences (Mencke et al., 2021).

Taken together, our results show that subjective ratings, which reflect processing stages in which conscious awareness is engaged, are strongly affected by the atonal structure of the stimuli even when earlier sensory processing stages addressed by the MMN may remain relatively unaffected. We regard this as evidence of a dissociation between sensory and cognitive musical expectations, whereby the latter may be most affected in the context of high uncertainty stimuli (see also: Neuloh and Curio, 2004). An interesting possibility is that the lack of accuracy and confidence that accompanies listening to atonal music enhances the adoption of searching or exploratory listening behaviors in pre-disposed listeners.

The Phenomenal Level: Listening Experts and the Experiential Dimension

Previous work shows that listeners' predictions are weaker for tonal than atonal music (Vuvan et al., 2014). Our recent data further demonstrate that processes in which conscious awareness is involved may be particularly affected by the lack of a tonal hierarchy (Mencke et al., 2021). However, relatively little is known about the phenomenal experience of atonal music. Specifically, characterizations of the nature of aesthetic experiences of atonal music, the key affective dimensions underlying engagement with this music, and the sources of enjoyment reported when listening to it, all remain largely absent.

To fill this gap, a series of interviews with experts specialized in atonal music and, by way of comparison, listening experts from the field of classic-romantic music, were conducted by the first author (Mencke et al., 2022). The aim of the study was to investigate a variety of experiential dimensions underlying an aesthetic experience with atonal music as well as to explore hedonic values, appreciation and pleasurable experiences with this music. Sixteen interviews were conducted with 8 experts in each group and the interview guide (for both groups) comprised questions about several aspects of a listening experience with the corresponding style of expertise. Physiological, cognitive and affective dimensions were covered. After the transcription of audio-recorded interviews, the textual material was analyzed qualitatively both following a deductive and an inductive step (Mayring, 2014).

The analysis revealed striking differences between the expert groups regarding how they described engagement as well as pleasurable experiences with music from their style of expertise. In the following we elaborate on a few findings of the analysis of the reports of the atonal music group that were most frequently mentioned.

First, the notion of exploration was prominent in many of the resulting themes of the qualitative analysis. Participants from the atonal music group repeatedly reported on their adoption of an exploratory attitude and on the fact that they enjoyed the active exploration of a piece of music. With regard to the latter, they emphasized their appreciation of the opportunity to continuously seek new ways of engaging with a piece of music by "probing it through listening."

Second, the results showed that pattern recognition, i.e., the perception of a memorable musical motive or phrase, may be a large source of pleasure when listening to atonal music. Specifically, when probed regarding experiences of beauty and pleasure during listening, participants reported the joy experienced when discovering a certain pattern in the music ("joy of discovery") and the high valuation of such moments of perceptual insight. One participant called this insight a "listening guide," something that leads one through a piece of atonal music. According to the participant's descriptions, such patterns were, for instance, related to a certain rhythmic structure that re-occurred from time to time. Thus, pleasurable moments were shown to emerge from recognition of an underlying pattern or structure in the music. Another finding supporting this notion are reports that increasing coherence as well as the confirmation of one's own expectations was experienced as highly positive.

Third, participants from the atonal group emphasized their adoption of an open stance and of their emergent feelings of curiosity. Here it is important to note that, while it has frequently been suggested that the curiosity and openness at a trait level correlate with a preference for (particularly complex) art (Feist and Brady, 2004; Silvia, 2008; Nusbaum and Silvia, 2011; Omigie, 2015), the relevance of state curiosity and openness during engagement with music has received only very little attention in empirical research. In one study employing a continuous rating methodology, it was shown that the perception of change while listening to music can lead to increases in feelings of curiosity as to how the music will unfold (Omigie and Ricci, 2021). Interestingly, in another recent study (Omigie and Ricci, 2022) a difference in the way state curiosity emerges in high- vs. low-uncertainty contexts was demonstrated, whereby while in low uncertainty contexts, high information content notes tended to induce curiosity, in high uncertainty contexts, both low and high information content events were able to drive curiosity. While that study did not focus on listeners that are experts in atonal music, it is interesting to consider how it aligns with the idea that the feeling of discovering a pattern (encountering a predictable event in a high entropy context) can lead especially engaged (atonal music) listeners to feel greater curiosity and further inclinations to explore the heard music as it unfolds.

Taken together, recent empirical data suggest that the complexity in atonal music may make listeners desire both coherence and a decrease in complexity, a phenomenon that is clearly in line with Berlyne's arousal theory proposing an inverted U-shaped relationship between hedonic value and the complexity of a stimulus (Wundt, 1896; Berlyne, 1970, 1971) and is further supported by other neuroscientific and behavioral studies in the context of empirical music research (Cheung et al., 2019; Gold et al., 2019). Critically, contrary to what was reported by the atonal music experts in the interview study (Mencke et al., 2022), a key role for pattern discovery was at no point mentioned by the classical music group. Rather, these listeners instead reported enjoying the clear structure of classical music, an interesting harmony or rhythm, and the "thrill of an unfamiliar interpretation" of a familiar piece (reports on the entire data see Mencke et al., 2022).

In sum, we argue that the following interrelated processes may be critical to the aesthetic experience of atonal music: The exploratory stance, by promoting the identification of sensory and perceptual features conveying coherence, may allow state curiosity to emerge. In turn, state curiosity by encouraging further engagement increases the opportunity for moments of structural insights to emerge (Brattico, 2015). This oscillation between exploring and moments of structural insights may characterize the experience of atonal music.

BUILDING PREDICTIONS IN ATONAL MUSIC

Having considered the evidence for an exploratory stance during atonal music listening, the current section elaborates on the mechanisms underlying the interplay between this stance and momentary phases of structural insight and on how predictive processes may give rise to positive hedonic values during online processing of atonal music.

Grouping and Gestalt Processes

As described above, memorable patterns can emerge in atonal music, for instance, when a repetitive structure arises in a piece based on pitch, rhythm, timbre, pitch, or loudness similarities. Critically, the recognition of memorable patterns, which constitutes moments of insight or a "listening guide," may indicate the cognitive process by which a Gestalt or a grouping—and therefore a momentary predictive model—has been developed.

Indeed, it is likely that model-building processes in the context of atonal music are strongly related to grouping principles and Gestalt heuristics (Bregman, 1990; Deutsch, 1999). That low-level predictive models are indeed present in an atonal context (Mencke et al., 2021) supports the assumption that low-level musical features may be used to produce those moments of structural insight, even when they do not reach conscious awareness. A listener of an atonal music work—more than a listener of tonal music—may have to rely on such bottom-up and implicit perceptual principles related to pitch proximity, rhythmic grouping or sound similarity (Bregman, 1990; Deutsch, 1999; Koelsch, 2012), as well as neural processes such as stimulus specific adaptation and forward masking. Sound similarity, as an example, would provide the listener an opportunity to perceive a certain instrumental group in an ensemble piece as a single Gestalt or, in terms of auditory scene analysis, as one sound stream. Another example—for instance, considering one of the early piano pieces by Stockhausen (1954)—would be a few subsequent notes that are played in the high register while being preceded and succeeded by melodies in a very low register: here a chunk or a Gestalt would be perceived based on pitch proximity. When and how Gestalts or chunks emerge and are recognized is modulated by a listener's musical and cultural background, their style-specific expertise, their piece-specific knowledge, and by cognitive abilities such as working memory capacity (Snyder, 2000; Tillmann and McAdams, 2004; Huron, 2006). By influencing the ways in which Gestalts, and thereby low-level and sensory predictions, are generated, all of these may be expected to modulate the experience of atonal music.

In sum, even though Gestalt elements are typically avoided in atonal music, growing evidence suggests that Gestalt-related processes of pattern identification and recognition should be taken into account when considering how atonal music is received. The lack of clear, familiar Gestalts in atonal music, and the complex environment that it presents to a listener, results in attention being enhanced in those moments in the music when regularity increases (Jones, 2019) or when sound events are particularly salient.

Structures of Saliency

Gestalt-related processes may be linked to so-called salient events in the context of listening to atonal music (Lerdahl, 1989). Lerdahl argued that listeners of atonal music "grab on to what they can: relative salience becomes structurally important" (Lerdahl, 1989, p. 84) and indeed, the relative salience of musical events (indexing structural importance) has been shown to be more relevant for listeners of atonal music than for those of tonal music (Deliège and Mélen, 1997; Dibben, 1999).

It has been suggested that such salient musical features function as cues, serving as "memory triggers" in situations in which they are repeated (Daynes, 2011). Interestingly, listeners have also been shown to remember and localize certain excerpts in atonal pieces if they can relate them to a specific cue (Deliège, 1989). Thus, salient cues that elicit an increase in attention and awareness could be the basis for building a memory trace of an atonal piece thanks to the opportunity they provide for the listener to form a perceptual chunk (Jones, 2015).

Salience can be generated in different ways (for instance by rhythmic, timbral or dynamic means). These structures of saliency could form so-called "event hierarchies" (Bharucha, 1984; Deutsch, 1984) and may conspicuously vary throughout a single work of music so that only provisional musical hierarchies can be generated (Imberty, 1993; Ordoñana and Laucirica, 2017). Such momentary hierarchies may only be applicable to certain phrases or parts within a piece and are therefore "extremely fluid for the hearer" (Imberty, 1993, p. 331) inasmuch as they do not have a fixed or a definite structure. Only temporarily valid predictive models can be built: one chunk that was built in a certain phrase of a musical work and that provided a momentary stability and a momentary predictive model might not be applicable in the next phrase of a work.

We therefore suspect that the generation of predictive models in the context of listening to atonal music is a highly transitory and fluid process, characterized by only temporary stability and temporary moments of increased predictability. It requires a high degree of adaptability of the listener and thus prompts the continuous exploratory behavior that is typical for an aesthetic experience with atonal music.

Exploration in Atonal Music

Taking into account the reports from dedicated and professional atonal music listening experts about the adoption of an exploratory state even after many years of exposure (Mencke et al., 2022), and considering the deliberate goal of avoiding predictability, particularly in serialist compositions (Stockhausen, 1963; Boulez, 1972; Kramer, 1989; Lerdahl, 1992; Hiekel, 2016), we suggest that atonal music is an artistic language that, more than other musical styles, affords its listeners an exploratory attitude.

Exploratory behavior is essential in environments that are novel, have surprising elements, and are complex (van Lieshout et al., 2020). Humans intrinsically seek knowledge (Berlyne, 1954; Perlovsky, 2010), have a drive for curiosity (Jepma et al., 2012), and actively engage with novel environments (Kidd and Hayden, 2015). Exploratory behavior can be seen as resulting from a desire to reduce uncertainty, a mechanism that all biological agents share (Friston, 2010; Schwartenbeck et al., 2013). Some studies and theoretical proposals suggest that uncertainty reduction is linked to positive affect and argue that it allows an agent to either confirm or update their existing predictive models (Van de Cruys, 2017; Koelsch et al., 2019; Kraus, 2020). In the case of atonal music, this may be particularly relevant since increased regularity in these uncertain environments are reported to be particularly pleasant (Mencke et al., 2022). A recent study using musical stimuli has shown that this model update is linked to activity in the nucleus accumbens (Gold et al., 2019), a central part of the dopaminergic mesolimbic reward pathway (Koelsch, 2014) and known to index pleasurable emotional peak experiences (Salimpoor et al., 2011). The hedonic value underlying atonal music might be closely linked to the positive affect stemming from a momentary reduction of uncertainty.

While exploration can be defined as "learning about the properties of an uncertain environment" (Gazzaniga et al., 2010, p. 1,065), a complementary behavior termed "exploitation" refers to a state in which an individual benefits from a familiar environment in which they know where rewards can be obtained (Kidd and Hayden, 2015). Arguably, Western tonality, with its inherent tonal and metrical hierarchy, offers a (Western) listener the opportunity for immediate exploitation (Meyer, 1956; Huron, 2006; Vuust and Frith, 2008; Rohrmeier and Koelsch, 2012;

Salimpoor et al., 2015; Koelsch et al., 2019). In contrast, for atonal music, which lacks such a fundamental predictability there is no structure to be gleaned —especially when we first listen—, and accordingly no immediate exploitation to be afforded.

Rather, we argue, atonal music, by being minimally predictable, affords a mode of exploration that allows brief pleasurable moments of insight to emerge but these insights may disappear as quickly as they emerge. This exploratory state is likely to be highly relevant for a number of different sorts of engagements with art, from abstract painting and contemporary dance to modernist poetry (Saklofske, 1975; Cupchik and Gebotys, 1990). It might additionally be relevant for many other genres from twentieth to twenty-first century art music that present listeners with unconventional musical structures. Accordingly, the minority of individuals who are willing to engage in atonal music are likely also willing to engage with these kinds of artistic languages (Mencke et al., 2019). Finally, an exploratory state might also be adopted if a listener engages with music from an unfamiliar culture as for instance when Western listeners encounter music based on the pentatonic scale.

FUTURE RESEARCH AVENUES

The constant generation of new predictive models facilitated by atonal music may provide unique insights, not only with respect to how a psychological state of listening to atonal music can be characterized, but also with respect to aesthetic experiences more generally.

A first issue pertains to the role of pattern discovery in music and other cross-modal aesthetic domains and its relation to the hedonic value. The cognitive mechanism related to the successful recognition of a musical pattern could be regarded as the auditory analog of the "Aesthetic Aha" in the visual (art) domain, which refers to a pleasurable moment stemming from pattern recognition (Muth and Carbon, 2013; Graf and Landwehr, 2017; Muth et al., 2018). It has been argued that the pleasurable effect originates in the sudden increase of processing fluency (Topolinski and Reber, 2010), a cognitive process that has been associated with liking (Reber et al., 2004). This effect is corroborated by studies showing that moments of insight evoke intense positive feelings (Shen et al., 2016; Webb et al., 2018). With regard to problem solving or verbal comprehension, the cognitive process of insight is regarded as an unexpected solution for a problem (Subramaniam et al., 2009) and a moment of sudden comprehension (Bowden et al., 2005; Kounios and Beeman, 2009). In one study investigating verbal problem-solving, insight moments were shown to be reflected in an increase in synchronous gamma-band oscillatory brain activity in the right anterior superior temporal gyrus, that was preceded by an increase in alpha-band activity at the right occipital cortex (Jung-Beeman et al., 2004). The authors suggested that these electrophysiological signatures point to the transition from a pre-attentive to an attentive state and reflect the "conscious availability of a solution" (Jung-Beeman et al., 2004, p. 506). In a more recent study that used auditory sequences, it was shown that the shift from a random to a regular sequence was accompanied by a sustained increase in the amplitude of neural signals (Barascud et al., 2016; Sohoglu and Chait, 2016; Southwell et al., 2017).

For this reason, it would be interesting to study whether such moments of insight similarly emerge in the context of music listening, particularly in the context of high-uncertainty music such as atonal music. Here, such moments might potentially emerge when an auditory object is perceived (Winkler et al., 2009), which could lead to an increase in conscious auditory perception or auditory awareness (Gutschalk et al., 2008; Dykstra et al., 2017). In order to study this, a series of musical stimuli could be created in which the complexity is gradually modified, either with regard to pitch or meter. While simpler stimuli would offer many ways to form chunks, the more complex stimuli would hamper the formation of chunks or auditory objects. Using magneto- or electroencephalography (MEG/EEG), brain responses could be analyzed with the aim to see whether neural activity evolves in comparable patterns such as those found for verbal moments of insight. Complementary data collection of ratings addressing (a range of) questions related to liking or preferences would shed light on any relationships to hedonic values.

Second, the predominance of exploration afforded by atonal music opens up the opportunity to study effects that stem from an aesthetic attitude—a state of mind that, according to philosophical aesthetics, "is entered into, voluntarily and consciously, by an individual, making that individual receptive to having an aesthetic experience" (Fenner, 1998, p. 1954) and thereby involves a certain intentional stance of the listener (Fenner, 1996, 1998; Kemp, 1999; Levinson, 2009; Brattico and Pearce, 2013; Hodges, 2016). Some scholars conceptualize this stance as an attentional focus on the formal and perceptual properties of an object in the context of an aesthetic experience (Levinson, 2009; Juslin and Isaksson, 2014). This is in line with the conceptualization of an aesthetic experience as a person's phenomenal state while consciously immersing and interacting with the music and in which the attention is directed to the music's perceptual and formal properties as well as its cognitive and affective interpretations (Brattico and Pearce, 2013; Wald-Fuhrmann et al., 2021). It has been proposed that an aesthetic attitude entails a focus on "sensory impressions, based on low-level features of the music" (Juslin, 2013, p. 248), i.e., features of the music that potentially serve as important cues in order to build Gestalts. Reybrouck (2015) proposes that art contexts force an individual to explore the content of an artwork, which may, in turn, cause this focus on sensory or perceptual properties. One conception and operationalization of the aesthetic attitude might therefore be that it evokes a particular attentional focus on sensory and perceptual properties of the music.

The empirical findings summarized in chapter three indeed support the idea that a focus on perceptual properties is afforded by atonal music, and thereby corroborates the proposal that "some pieces of music will "invite" an aesthetic attitude to a greater extent than other pieces (because of certain formal features)" (Juslin, 2013, p. 247). Atonal music might therefore be helpful when studying behavioral and neural correlates of an aesthetic attitude. In an experiment, one could utilize different musical stimuli with varying degrees of predictability and complexity (for instance, classical music vs. jazz music vs. atonal music). In order to investigate whether the attentional focus shifts as a function of complexity, participants' responses to questions relating to a number of different low-, medium-, and high-level features of the music could be collected. Based on what presented above, we would predict a positive correlation between degree of complexity and attentional focus on low-level properties of the music. By contrast, in music that provides a certain structure and clear anchor points, such as a tonal hierarchy, the attentional focus on sensory properties would remain in the background, i.e., such basic properties would remain subconscious.

Here it is worth noting an interesting alternative strategy for dealing with uncertainty that was described by the interviewed participants: namely, the adoption of an attitude in which they tried to avoid an analytic and structured listening mode and instead switched to a more free, open-ended listening mode. One respondent said: "If you go in without any expectation of understanding and just let yourself be affected, then you end up understanding more than if you had gone in already expecting to acquire knowledge." Thus, the acceptance or awareness of not being able to fully predict and ultimately to exploit the music for an open-ended exploratory listening experience might be another, complementary strategy of how to deal with the perceptual challenge present in this music. This might lead to the generation of subjective meaning that goes beyond the positive experience of successfully generating or refining predictive models. Thus, how this particular aspect of "guidedness" can be conceptualized, what it underlies and which effects it has, and ultimately how this mode of open-ended exploration interacts with a mode of active exploration, represent another novel potential research avenue.

CONCLUSION

In this article we aimed to provide a novel perspective on how an aesthetic experience with atonal music could be characterized. With reference to neural and behavioral evidence, we emphasized that an exploratory state is important and often adopted given that it facilitates the discovery of novel reference points. Importantly, this state becomes particularly crucial in relation to the poor prospects of exploitation. Future research that aims to study the exploratory state in the context of an aesthetic experience should therefore include atonal music in their research paradigms.

AUTHOR CONTRIBUTIONS

IM generated the idea and hypothesis for this article and wrote the first draft of the manuscript. All authors equally contributed to the further development and refinement of the argumentation and to writing and revising the article.

FUNDING

The Center for Music in the Brain was funded by the Danish National Research Foundation (DNRF 117).

ACKNOWLEDGMENTS

We thank the reviewers for highly valuable comments on earlier versions of the manuscript.

REFERENCES

Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., and Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proc. Natl. Acad. Sci. U.S.A.* 113, E616–E625. doi: 10.1073/pnas.1508523113

Berlyne, D. E. (1954). A theory of human curiosity. *Br. J. Psychol.* 45:180.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Percept. Psychophys.* 8, 279–286. doi: 10.3758/BF03212593

Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. New York, NY: Appleton-Century-Crofts.

Bharucha, J. J. (1984). Event hierarchies, tonal hierarchies and assimilation: a reply to deutsch and dowling. *J. Exp. Psychol. Gen.* 113, 421–425.

Bigand, E., and Poulin-Charronnat, B. (2016). "Tonal cognition," in *Oxford Handbook of Music Psychology Ian Cross*, 2nd Edn, eds M. Thaut and S. Hallam (Oxford: OUP), 1–19.

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

Boulez, P. (1972). *Werkstatt-Texte*. Berlin: Propyläen.

Bowden, E. M., Jung-Beeman, M., Fleck, J., and Kounios, J. (2005). New approaches to demystifying insight. *Trends Cogn. Sci.* 9, 322–328. doi: 10.1016/j.tics.2005.05.012

Brattico, E. (2015). "From pleasure to liking and back: bottom-up and top-down neural routes to the aesthetic enjoyment," 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), 303–318.

Brattico, E. (2021). "The empirical aesthetics of music," in *The Oxford Handbook of Empirical Aesthetics*, eds M. Nadal and O. Vartanian (Oxford: Oxford University Press), 1–38.

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

Brattico, E., Bogert, B., Alluri, V., Tervaniemi, M., Eerola, T., and Jacobsen, T. (2016). It's sad but i like it: The neural dissociation between musical emotions and liking in experts and laypersons. *Front. Hum. Neurosci.* 9:676. doi: 10.3389/fnhum.2015.00676

Bregman, A. (1990). *Auditory Scene Analysis: the Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Brincker, M. (2015). "The aesthetic stance – on the conditions and consequences of becoming a beholder," in *Aesthetics and the Embodied Mind: Beyond Art Theory and the Cartesian Mind- Body Dichotomy Contributions to Phenomenology*, ed. A. Scarinzi (Berlin: Springer), 117–138. doi: 10.1007/978-94-017-9379-7_8

Cacioppo, J. T., and Petty, R. E. (1982). The need for cognition. *J. Pers. Soc. Psychol.* 42, 116–131. doi: 10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., Blair, W., and Jarvis, G. (1996). Dispositional differences in cognitive motivation: the life and times of individuals varying in need for cognition. *Psychol. Bull.* 119, 197–253.

Cheung, V. K. M., Harrison, P. M. C., Meyer, L., Pearce, M. T., Haynes, J.-D., and Koelsch, S. (2019). Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. *Curr. Biol.* 29, 4084–4092.e4. doi: 10.1016/j.cub.2019.09.067

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–253. doi: 10.1017/S0140525X12000477

Clarke, E. F. (1999). "Rhythm and timing in music," in *The Psychology Of Music*, ed. D. Deutsch (Cambridge, MA: Academic Press), 473–500.

Cuddy, L. L., Cohen, A. J., and Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 869–883. doi: 10.1037//0096-1523.7.4.869

Cupchik, G. C., and Gebotys, R. J. (1990). Interest and pleasure as dimensions of aesthetic response. *Empir. Stud. Arts* 8, 1–14.

Daynes, H. (2011). Listeners' perceptual and emotional responses to tonal and atonal music. *Psychol. Music* 39, 468–502. doi: 10.1177/0305735610378182

Deliège, I. (1989). A perceptual approach to contemporary musical forms. *Contemp. Music Rev.* 4, 213–230. doi: 10.1080/07494468900640301

Deliège, I., and Mélen, M. (1997). "Cue abstraction in the representation of musical form," in *Perception And Cognition Of Music*, eds I. Deliege and J. A. Sloboda (Abingdon: Taylor and Francis), 359–382. doi: 10.4324/9780203344262-28

Denham, S. L., and Winkler, I. (2006). The role of predictive models in the formation of auditory streams. *J. Physiol. Paris* 100, 154–170. doi: 10.1016/j.jphysparis.2006.09.012

Deutsch, D. (1984). Two issues concerning tonal hierarchies: comment on castellano, bharucha, and krumhansl. *J. Exp. Psychol. Gen.* 113, 413–416. doi: 10.1037//0096-3445.113.3.413

Deutsch, D. (1999). *The Psychology Of Music*. Cambridge, MA: Academic Press.

Deutsch, D. (2013). "Grouping mechanisms in music," in *The Psychology of Music*, ed. D. Deutsch (Cambridge, MA: Elsevier Academic Press), 183–248. doi: 10.1016/B978-0-12-381460-9.00006-7

Dibben, N. (1994). The cognitive reality of hierarchic structure in tonal and atonal music. *Music Percept.* 12, 1–25.

Dibben, N. (1999). The perception of structural stability in atonal music: the influence of salience, stability, horizontal motion, pitch commonality, and dissonance. *Perception* 16, 265–294. doi: 10.2307/40285794

Dibelius, U. (1998). *Moderne Musik Nach 1945*. München: Piper.

Dowling, W. J., Kwak, S., and Andrews, M. W. (1995). The time course of recognition of novel melodies. *Percept. Psychophys.* 57, 136–149. doi: 10.3758/bf03206500

Dykstra, A. R., Cariani, P. A., and Gutschalk, A. (2017). A roadmap for the study of conscious audition and its neural basis. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160103. doi: 10.1098/rstb.2016.0103

Eerola, T., Vuoskoski, J. K., Peltola, H. R., Putkinen, V., and Schäfer, K. (2018). An integrative review of the enjoyment of sadness associated with music. *Phys. Life Rev.* 25, 100–121. doi: 10.1016/j.plrev.2017.11.016

Feist, G. J., and Brady, T. R. (2004). Openness to experience, non-conformity, and the preference for abstract art. *Empir. Stud. Arts* 22, 77–89. doi: 10.2190/Y7CA-TBY6-V7LR-76GK

Fenner, D. E. W. (1996). *The Aesthetic Attitude*. Atlantic Highlands, NJ: Humanities Press.

Fenner, D. E. W. (1998). "Aesthetic Attitude," in *Encyclopedia Of Aesthetics*, ed. M. Kelly (New York, NY: Oxford University Press).

Friston, K. J. (2009). The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* 13, 293–301. doi: 10.1016/j.tics.2009.04.005

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K. J., and Kiebel, S. (2009). Cortical circuits for perceptual inference. *Neural Netw.* 22, 1093–1104. doi: 10.1016/j.neunet.2009.07.023

Garrido, M. I., Sahani, M., and Dolan, R. J. (2013). Outlier responses reflect sensitivity to statistical structure in the human brain. *PLoS Comput. Biol.* 9:e1002999. doi: 10.1371/journal.pcbi.1002999

Gazzaniga, M. S., Ivry, R. B., and Mangun, G. R. (2010). *Cognitive Neuroscience; the Biology of the Mind*, 4th Edn. New York, NY: W.W. Norton.

Gold, B. P., Pearce, M. T., Mas-Herrero, E., Dagher, A., Zatorre, R. J., and Zatorre, R. J. (2019). Predictability and uncertainty in the pleasure of music: a reward for learning? *J. Neurosci.* 39, 9397–9409. doi: 10.1523/JNEUROSCI.0428-19.2019

Graf, L. K. M., and Landwehr, J. R. (2017). Aesthetic pleasure versus aesthetic interest: the two routes to aesthetic liking. *Front. Psychol.* 8:15. doi: 10.3389/fpsyg.2017.00015

Grant, M. J. (2001). *Serial Music, Serial Aesthetics: Compositional Theory in Post-War Europe*. Cambridge: Cambridge University Press.

Griffiths, P. (2001). *Serialism. Grove Music Online* (Oxford: Oxford University Press). 1–16.

Gross, J., and Pitts, S. (2016). Audiences for the contemporary arts: Exploring varieties of participation across art forms in Birmingham, UK. *Participations* 13, 4–23.

Gutschalk, A., Micheyl, C., and Oxenham, A. J. (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol.* 6:1156–1165. doi: 10.1371/journal.pbio.0060138

Haumann, N. T., Lumaca, M., Kliuchko, M., Santacruz, J. L., Vuust, P., and Brattico, E. (2021). Extracting human cortical responses to sound onsets and acoustic feature changes in real music, and their relation to event rate. *Brain Res.* 1754, 147248. doi: 10.1016/j.brainres.2020.147248

Heilbron, M., and Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex? *Neuroscience* 389, 54–73. doi: 10.1016/j.neuroscience.2017.07.061

Helmholtz, H. L. F. V. (1954). *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, ed. A. J. Elli (New York, NY: Dover Publications). (original work published in 1863).

Hiekel, P. (2016). "Neue musik," in *Lexikon Neue Musik*, eds P. Hiekel and C. Utz (Stuttgart:), 434–444.

Hodges, D. A. (2016). "The Neuroaesthetics of Music," in *Oxford Handbook of Music Psychology*, eds S. Hallam, I. Cross, and M. Thaut (Oxford: Oxford University Press), doi: 10.1093/oxfordhb/9780198722946.013.20

Hsu, Y., Le Bars, S., Hämäläinen, J. A., and Waszak, F. (2015). Distinctive representation of mispredicted and unpredicted prediction errors in human electroencephalography. *J. Neurosci.* 35, 14653–14660. doi: 10.1523/JNEUROSCI.2204-15.2015

Huron, D. (2006). *Sweet Anticipation: Music And The Psychology Of Expectation*. Cambridge, MA: MIT Press.

Imberty, M. (1993). How do we perceive atonal music? Suggestions for a theoretical approach. *Contemp. Music Rev.* 9, 325–337.

Jepma, M., Verdonschot, R. G., van Steenbergen, H., Rombouts, S. A. R. B., and Nieuwenhuis, S. (2012). Neural mechanisms underlying the induction and relief of perceptual curiosity. *Front. Behav. Neurosci.* 6:5. doi: 10.3389/fnbeh.2012.00005

Jones, M. R. (2015). "Musical time," in *Oxford Handbook Music Psychology*, eds S. Hallam, I. Cross, and M. Thaut (Oxford: Oxford University).

Jones, M. R. (2019). *Time Will Tell – A Theory of Dynamic Attending*. New York, NY: Oxford University Press.

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., et al. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biol.* 2:500–510. doi: 10.1371/journal.pbio.0020097

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 Isaksson, S. (2014). Subjective criteria for choice and aesthetic judgment of music: a comparison of psychology and music students. *Res. Stud. Music Educ.* 36, 179–198. doi: 10.1177/1321103X14540259

Kemp, G. (1999). The aesthetic attitude. *Br. J. Aesthet.* 39, 392–399.

Kidd, C., and Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron* 88, 449–460. doi: 10.1016/j.neuron.2015.09.010

Kliuchko, M., Brattico, E., Gold, B. P., Tervaniemi, M., Bogert, B., Toiviainen, P., et al. (2019). Fractionating auditory priors: a neural dissociation between active and passive experience of musical sounds. *PLoS One* 14:e0216499. doi: 10.1371/JOURNAL.PONE.0216499

Koelsch, S. (2012). *Brain and Music*. West Sussex: Wiley-Blackwell.

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

Koelsch, S., Vuust, P., and Friston, K. J. (2019). Predictive processes and the peculiar case of music. *Trends Cogn. Sci.* 23, 63–77. doi: 10.1016/J.TICS.2018.10.006

Kostka, S., and Santa, M. (2018). *Materials and Techniques of Post-Tonal Music*, 5th Edn. New York, NY: Routledge.

Kounios, J., and Beeman, M. (2009). The aha! moment: the cognitive neuroscience of insight. *Curr. Dir. Psychol. Sci.* 18, 210–216. doi: 10.1111/j.1467-8721.2009.01638.x

Kramer, J. D. (1989). *The Time of Music*. New York, NY: Schirmer Books.
Kraus, N. (2020). The joyful reduction of uncertainty: music perception as a window to predictive neuronal processing. *J. Neurosci.* 40, 2790–2792. doi: 10.1523/JNEUROSCI.0072-20.2020
Krumhansl, C. L., and Cuddy, L. L. (2010). "A theory of tonal hierarchies in music," in *Music Perception, Springer Handbook of Auditory Research*, eds M. R. Jones, R. R. Fay, and A. N. Popper (Berlin: Springer), 51–87. doi: 10.1007/978-1-4419-6114-3_3
Krumhansl, C. L., Sandell, G. J., and Sergeant, D. C. (1987). The perception of tone hierarchies and mirror forms in twelve-tone serial music. *Music Percept.* 5, 31–78.
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
Lerdahl, F. (1989). Atonal prolongational structure. *Contemp. Music Rev.* 4, 65–87. doi: 10.1080/07494468900640211
Lerdahl, F. (1992). Cognitive constraints on compositional systems. *Contemp. Music Rev.* 6, 97–121. doi: 10.1080/07494469200640161
Lerdahl, F. (2019). *Composition And Cognition: Reflections on Contemporary Music and the Musical Mind*. Berkeley, CA: University of California Press.
Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory Of Tonal Music*. Cambridge, MA: MIT Press.
Levinson, J. (2009). The aesthetic appreciation of music. *Br. J. Aesthet.* 49, 415–425. doi: 10.1093/aesthj/ayp043
London, J. (2004). *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford: Oxford University Press.
Loui, P., and Wessel, D. (2006). "Acquiring new musical grammars: a statistical learning approach," in *Proceedings of the 9th International Conference on Music Perception and Cognition*, eds M. Baroni, A. R. Addessi, R. Caterina, and M. Costa (Bologna: ICMPC-ESCOM), 1009–1017.
Maletz, H. (2011). *Leidenschaft? Neue Musik. Über Klänge, Laute, Zeichen Bis Zu Jazz und Pop*. Münster: Lit Verlag.
Marin, M. (2020). "The role of collative variables in aesthetic experiences," in *The Oxford Handbook of Empirical Aesthetics*, eds M. Nadal and O. Vartanian (Oxford: Oxford University Press).
Marin, M. M., Lampatz, A., Wandl, M., and Leder, H. (2016). Berlyne revisited: evidence for the multifaceted nature of hedonic tone in the appreciation of paintings and music. *Front. Hum. Neurosci.* 10:536. doi: 10.3389/fnhum.2016.00536
Mayring, P. (2014). *Qualitative Content Analysis Theoretical Foundation, Basic Procedures and Software Solution*. Klagenfurt: Beltz Verlag.
McLain, D. L. (2009). Evidence of the properties of an ambiguity tolerance measure: the multiple stimulus types ambiguity tolerance scale-ii (mstat-ii). *Psychol. Rep.* 105, 975–988. doi: 10.2466/PR0.105.3.975-988
Mencke, I., Omigie, D., Wald-Fuhrmann, M., and Brattico, E. (2019). Atonal music: can uncertainty lead to pleasure? *Front. Neurosci.* 12:979. doi: 10.3389/FNINS.2018.00979
Mencke, I., Quiroga-Martinez, D. R., Omigie, D., Schwarzacher, F., Haumann, N. T., Michalareas, G., et al. (2021). Prediction under uncertainty: dissociating sensory from cognitive expectations in highly uncertain musical contexts. *Brain Res.* 1773:147664. doi: 10.1016/j.brainres.2021.147664
Mencke, I., Seibert, C., Brattico, E., and Wald-Fuhrmann, M. (2022). Comparing the aesthetic experience of classic-romantic and contemporary classical music: An interview study. *Psychol. Music* doi: 10.1177/03057356221091312
Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Jacobsen, T., and Koelsch, S. (2017). The distancing-embracing model of the enjoyment of negative emotions in art reception. *Behav. Brain Sci.* 40, 1–63. doi: 10.1017/S0140525X17000309
Meyer, L. B. (1956). *Emotion And Meaning In Music*. Chicago, IL: University of Chicago Press.
Morgan, E., Fogel, A., Nair, A., and Patel, A. D. (2019). Statistical learning and Gestalt-like principles predict melodic expectations. *Cognition* 189, 23–34. doi: 10.1016/j.cognition.2018.12.015
Muth, C., and Carbon, C. C. (2013). The aesthetic aha: on the pleasure of having insights into gestalt. *Acta Psychol. (Amst)* 144, 25–30. doi: 10.1016/j.actpsy.2013.05.001
Muth, C., Hesslinger, V. M., and Carbon, C. C. (2018). Variants of semantic instability (SeIns) in the arts: a classification study based on experiential reports. *Psychol. Aesthetics Creat. Arts* 12, 11–23. doi: 10.1037/aca0000113
Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026
Neuloh, G., and Curio, G. (2004). Does familiarity facilitate the cortical processing of music sounds? *Neuroreport* 15, 2471–2475. doi: 10.1097/00001756-200411150-00008
Nusbaum, E. C., and Silvia, P. J. (2011). Shivers and timbres: personality and the experience of chills from music. *Soc. Psychol. Personal. Sci.* 2, 199–204. doi: 10.1177/1948550610386810
Ockelford, A., and Sergeant, D. (2012). Musical expectancy in atonal contexts: musicians' perception of "antistructure.". *Psychol. Music* 41, 139–174. doi: 10.1177/0305735612442582
Omigie, D. (2015). Dopamine and epistemic curiosity in music listening Dopamine and epistemic curiosity in music listening. *Cogn. Neurosci.* 6, 222–224. doi: 10.1080/17588928.2015.1051013
Omigie, D., and Ricci, J. (2022). Accounting for expressions of curiosity and enjoyment during music listening. *Psychol. Aesthet. Creat. Arts* 1–17. doi: 10.1037/aca0000461 [Epub ahead of print].
Omigie, D., and Ricci, J. (2021). Curiosity emerging from the perception of change in music. *Empir. Stud. Arts* 40, 296–316. doi: 10.1177/02762374211059460
Ordoñana, J. A., and Laucirica, A. (2017). Structural segmentation of toru takemitsu's piece, itinerant, by advanced level music graduate students. *Iperception* 8, 1–17. doi: 10.1177/2041669517705387
Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. *Ann. N. Y. Acad. Sci.* 1423, 378–395. doi: 10.1111/nyas.13654
Pearce, M. T., and Wiggins, G. A. (2012). Auditory expectation: the information dynamics of music perception and cognition. *Top. Cogn. Sci.* 4, 625–652. doi: 10.1111/j.1756-8765.2012.01214.x
Perlovsky, L. I. (2010). Musical emotions: functions, origins, evolution. *Phys. Life Rev.* 7, 2–27. doi: 10.1016/j.plrev.2009.11.001
Quiroga-Martinez, D. R., Hansen, N. C., Højlund, A., Pearce, M. T., Brattico, E., and Vuust, P. (2019). Reduced prediction error responses in high-as compared to low-uncertainty musical contexts. *Cortex* 120, 181–200. doi: 10.1016/j.cortex.2019.06.010
Quiroga-Martinez, D. R., Hansen, N. C., Højlund, A., Pearce, M. T., Brattico, E., and Vuust, P. (2020a). Musical prediction error responses similarly reduced by predictive uncertainty in musicians and non-musicians. *Eur. J. Neurosci.* 51, 2250–2269. doi: 10.1111/ejn.14667
Quiroga-Martinez, D. R., Hansen, N. C., Højlund, A., Pearce, M. T., Brattico, E., and Vuust, P. (2020b). Decomposing neural responses to melodic surprise in musicians and non-musicians: evidence for a hierarchy of predictions in the auditory system. *Neuroimage* 215:116816. doi: 10.1016/j.neuroimage.2020.116816
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
Reybrouck, M. (2015). Music as environment: an ecological and biosemiotic approach. *Behav. Sci. (Basel)* 5, 1–26. doi: 10.3390/bs5010001
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
Rosch, E. (1975). Cognitive reference points. *Cogn. Psychol.* 7, 532–547.
Ross, S., and Hansen, N. C. (2016). Dissociating prediction failure: considerations from music perception. *J. Neurosci.* 36, 3103–3105. doi: 10.1523/JNEUROSCI.0053-16.2016
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
Saklofske, D. H. (1975). Aesthetic complexity and exploratory behavior. *Percept. Mot. Skills* 41, 363–368. doi: 10.2466/pms.1975.41.2.363
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., Zald, D. H., Zatorre, R. J., Dagher, A., and McIntosh, 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

Schulze, K., Dowling, W. J., and Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Percept.* 61, 137–170. doi: 10.1525/rep.2008.104.1.92.This

Schwartenbeck, P., FitzGerald, T., Dolan, R. J., and Friston, K. J. (2013). Exploration, novelty, surprise, and free energy minimization. *Front. Psychol.* 4:710. doi: 10.3389/fpsyg.2013.00710

Shen, W., Yuan, Y., Liu, C., and Luo, J. (2016). In search of the "Aha!" experience: elucidating the emotionality of insight problem-solving. *Br. J. Psychol.* 107, 281–298. doi: 10.1111/bjop.12142

Silvia, P. J. (2008). Interest – the curious emotion. *Curr. Dir. Psychol. Sci.* 17, 57–60. doi: 10.1111/j.1467-8721.2008.00548.x

Snyder, B. (2000). *Music and Memory: An Introduction*. Cambridge, MA: MIT Press.

Snyder, B. (2008). "Memory for music," in *Oxford Handbook Music Psychology*, eds S. Hallam, I. Cross, and M. Thaut (Oxford: Oxfords University), 1–17. doi: 10.1093/oxfordhb/9780199298457.013.0010

Sohoglu, E., and Chait, M. (2016). Detecting and representing predictable structure during auditory scene analysis. *Elife* 5:e19113. doi: 10.7554/eLife.19113

Southwell, R., and Chait, M. (2018). Enhanced deviant responses in patterned relative to random sound sequences. *Cortex* 109, 92–103. doi: 10.1016/j.cortex.2018.08.032

Southwell, R., Baumann, A., Gal, C., Barascud, N., Friston, K. J., and Chait, M. (2017). Is predictability salient? A study of attentional capture by auditory patterns. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160105. doi: 10.1098/rstb.2016.0105

Stockhausen, K. (1954). *Klavierstücke I-IV*. London: Universal Edition.

Stockhausen, K. (1963). *Texte zur Elektronischen und Instrumentalen Musik, Aufsätze 1952-1962*, Vol. 1. Köln: DuMont.

Subramaniam, K., Kounios, J., Parrish, T. B., and Jung-Beeman, M. (2009). A brain mechanism for facilitation of insight by positive affect. *J. Cogn. Neurosci.* 21, 415–432. doi: 10.1162/jocn.2009.21057

Taruffi, L., and Koelsch, S. (2014). The paradox of music-evoked sadness: an online survey. *PLoS One* 9: e110490. doi: 10.1371/journal.pone.0110490

Taruskin, R. (2010). *Music In The Early Twentieth Century*. New York, NY: Oxford University Press.

Tillmann, B., and McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1131–1142. doi: 10.1037/0278-7393.30.5.1131

Topolinski, S., and Reber, R. (2010). Gaining insight into the "Aha" experience. *Curr. Dir. Psychol. Sci.* 19, 402–405. doi: 10.1177/0963721410388803

Utz, C. (2016). "Wahrnehmung," in *Lexikon Neue Musik*, eds P. Hiekel and C. Utz (Kassel: Baerenreiter), 600–609.

Van de Cruys, S. (2017). "Affective value in the predictive mind," in *Philosophy and Predictive Processing*, eds T. Metzinger and W. Wiese (Frankfurt am Main: MIND Group), 1–21. doi: 10.15502/9783958573253

van Lieshout, L. L., de Lange, F. P., and Cools, R. (2020). Why so curious? Quantifying mechanisms of information seeking. *Curr. Opin. Behav. Sci.* 35, 112–117. doi: 10.1016/j.cobeha.2020.08.005

Vuust, P., and Frith, C. (2008). Anticipation is the key to understanding music and the effects of music on emotion. *Behav. Brain Sci.* 31, 599–600. doi: 10.1017/S0140525X08005542

Vuust, P., and Kringelbach, M. L. (2010). The pleasure of making sense of music. *Interdiscip. Sci. Rev.* 35, 166–182. doi: 10.1179/030801810X12723585301192

Vuvan, D. T., Podolak, O. M., and Schmuckler, M. A. (2014). Memory for musical tones: the impact of tonality and the creation of false memories. *Front. Psychol.* 5:582. doi: 10.3389/fpsyg.2014.00582

Wald-Fuhrmann, M., Egermann, H., O'Neill, K., Czepiel, A., Weining, C., Meier, D., et al. (2021). Music listening in classical concerts: theory, literature review, and research program. *Front. Psychol.* 12:638783. doi: 10.3389/fpsyg.2021.638783

Webb, M. E., Little, D. R., and Cropper, S. J. (2018). Once more with feeling: Normative data for the aha experience in insight and noninsight problems. *Behav. Res. Methods* 50, 2035–2056. doi: 10.3758/s13428-017-0972-9

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

Wörner, K. H. (1973). *Stockhausen – Life and Work*. Berkeley, CA: University of California Press.

Wundt, W. (1896). *Grundriss Der Psychologie*. Leipzig: Wilhelm Engelmann.

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: 28 May 2020
doi: 10.3389/fpsyg.2020.01001

What Determines the Perception of Segmentation in Contemporary Music?

Michelle Phillips[1], Andrew J. Stewart[2], J. Matthew Wilcoxson[3], Luke A. Jones[2], Emily Howard[1], Pip Willcox[3], Marcus du Sautoy[4] and David De Roure[3,5]*

[1] The Royal Northern College of Music, Manchester, United Kingdom, [2] Division of Neuroscience and Experimental Psychology, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, The University of Manchester, Manchester, United Kingdom, [3] Oxford e-Research Centre, Mathematical, Physical and Life Sciences Division, University of Oxford, Oxford, United Kingdom, [4] Mathematical Institute, University of Oxford, Oxford, United Kingdom, [5] The Alan Turing Institute, London, United Kingdom

OPEN ACCESS

Edited by:
Michiel M. Spapé,
University of Helsinki, Finland

Reviewed by:
Kentaro Katahira,
Nagoya University, Japan
Justin London,
Carleton College, United States

****Correspondence:***
Michelle Phillips
Michelle.Phillips@rncm.ac.uk;
michelle.phillips@btinternet.com

Specialty section:
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

Received: *22 November 2019*
Accepted: *22 April 2020*
Published: *28 May 2020*

Citation:
Phillips M, Stewart AJ, Wilcoxson JM, Jones LA, Howard E, Willcox P, du Sautoy M and De Roure D (2020) What Determines the Perception of Segmentation in Contemporary Music? Front. Psychol. 11:1001. doi: 10.3389/fpsyg.2020.01001

Background: This study concerns the perception of musical segmentation during listening to live contemporary classical music. Little is known about how listeners form judgments of musical segments, particularly when typical section markers, such as cadences and fermatas, are absent [e.g., Sears et al. (2014)] or when the music is non-tonal (e.g., in much contemporary classical music).

Aims: The current study aimed to examine the listeners' segmentation decisions in a piece of contemporary music, Ligeti's "Fanfares"?

Methods: Data were gathered using a smartphone application [Practice & Research in Science & Music (PRiSM) Perception App] designed for this study by the Royal Northern College of Music (RNCM) Centre for PRiSM and the Oxford e-Research Centre. A total of 259 audience participants were asked to "tap" when they felt that a section had ended. Subjective responses were captured, as well as contextual data about the participants.

Results: The audience members demonstrated high levels of agreement regarding segmentation, mostly at places in the music involving breaks in the musical texture (one piano hand resting), changes in dynamic (volume), and changes in register/pitch. A sense of familiarity with contemporary repertoire did seem to influence the responses—the participants who self-reported being familiar with contemporary music used a wider range of cues to make their segmentation decisions. The self-report data analysis suggested that the listeners were not always aware of how they made decisions regarding segmentation. The factors which may influence their judgment of musical segmentation are, to some extent, similar to those identified by music analysis (Steinitz, 2011) but different in other ways. The effect of musical training was found to be quite small.

Conclusion: Whether musically trained and/or familiar with contemporary classical music or not, the listeners demonstrate commonalities in segmentation, which they are not always aware of. This study has implications for contemporary composers, performers, and audiences and how they engage with new music in particular.

Keywords: music perception, large-scale form, segmentation, contemporary music, Ligeti, live music

INTRODUCTION

When we listen to a piece of music, we perceive the information as consisting of separate units or events (Drake and Bertrand, 2001). Such a process may be referred to as grouping (Lerdahl and Jackendoff, 1983), chunking (Miller, 1956), or segmentation (Cambouropoulos, 2006). How listeners infer such chunks of information or segment the auditory stream may be influenced by multiple factors. These may include individual differences, such as musical expertise, and/or features in the music itself, for example, when the music changes significantly in some way (for example, in modality or tempo). Investigating how musical information is parsed in real time, i.e., what determines how one event is deemed to have ended and another begun, may grant valuable insights into the ways in which expertise may shape the experience of music and how some aspects in the musical surface or structure may be more perceptually salient than others.

Most popular studies and models of the perception of musical form (e.g., Meyer, 1956; Lerdahl and Jackendoff, 1983) are based on the perception of tonal music. Contemporary (or "new") music is rarely tonal and hence does not include related features such as diatonic harmony, harmonic tension and release, harmonic closure (e.g., at the cadence), or melody as conceived in the conventional 19th-century sense. Tonal music is often constructed as a series of organized events existing in a hierarchy. However, this may not be the case for non-tonal music (Dibben, 1994).

Such models of music listening, although perhaps addressing tonal music most often, may have relevance for the perception of contemporary music. Temperley's (2007) theories of music and probability not only examine tonal hierarchies in their discussions of patterns but also discuss how music perception relies on familiarity with musical structures (relevant for contemporary music) and the probabilities regarding which events follow which others. Other notable music scholars have theorized about musical form (Rothstein, 1991; Caplin, 2001; Marsden, 2005). However, such works of theory and analysis, while often discussing notions of perception and building on the work of Schenker (1979) and Lerdahl and Jackendoff (1983), rarely include non-tonal music as their focus and do not examine perception empirically. Notable scholars have taken non-tonal music as their focus, for example, Hanninen's (2003) concept of "recontextualization," particularly with regards to her work on Feldman's music, Straus's (2016) theories of pitch-class sets, and Chew's (2001) work with mathematical models of perception of tonal music, and much of this work includes reference to how a listener may parse (or segment) musical information. However, again this literature takes a theoretical stance and does not seek to examine the perception of non-tonal music empirically.

A common assumption is that "atonal music is generally perceived as opaque and difficult to understand because it rarely makes aural sense to the uninitiated listener" (Spies, 2005). Such a stance raises the question of whether tonal and non-tonal music are perceived differently from one another, and if so, how? Also, to what extent might non-tonal music be more difficult to understand, and does this depend at all on expertise? Research regarding the perception of musical form and factors which may influence this (such as a listener's level of musical training) has only recently begun to interrogate ways in which contemporary classical music is processed and stored (e.g., Schulze et al., 2012), and there are many unanswered questions in this field despite the obvious relevance to modern-day listeners, composers, and performers.

Over the last 30 years, multiple studies have explored the perception of musical form in tonal music. Most have suggested that local cues (small scale and short duration) may take priority over global relationships (large scale and overall formal structure) in the perception of musical information (Deliège et al., 1996; Tillmann and Bigand, 2004; Rolison and Edworthy, 2012). Moreover, studies which have examined the impact of rearranging segments of music on the listener have found that the participants may not be able to differentiate between the original and the rearranged pieces of music (Tillmann and Bigand, 1996) and that the listeners may not prefer an original over an arrangement (Eitan and Granot, 2008). Two studies by Granot and Jacoby (2011, 2012) suggested that the participants may demonstrate sensitivity to some aspects of musical form (e.g., the ABA structure, the placement of the development section, and features at the beginning and end of the work) but not to others (e.g., overall harmonic structure). In summary, this research on the perception of large-scale form suggests that global relationships in the structure of music may not be available to perception, or if they are, such perception may not be as prominent as the recognition of smaller-scale events.

The question of how such local events are parsed, or segmented, when listening to large-scale form has only recently begun to be explored. Empirical work has suggested that, in tonal music, cadences (Tillmann et al., 1998; Tillmann and Bigand, 2004) and long notes and rests (Bruderer and McKinney, 2008) influence the judgment of segmentation during music listening. Studies of non-tonal music are fewer. Clarke and Krumhansl (1990) undertook five experiments which investigated various ways in which trained and untrained listeners perceive musical form. Experiments 1–3 used Stockhausen's *Klavierstueck IX* and experiments 4–6 used Mozart's *Fantasie in C minor*, K. 475, as their stimulus set. The authors state that, overall, the results of these experiments suggest that the attempts to model the perception of large-scale form and predict the perception of segments (in this case, using the model proposed by Lerdahl and Jackendoff, 1983) may be successful such that the listeners largely agree on the segment boundaries (for both the works by Mozart and by Stockhausen) and that the listeners are largely accurate in judging the location of segments. These results were relevant when listening to both tonal and non-tonal stimuli, which the authors attribute to "the two pieces share[ing] some high-level property on which listeners focus" (p. 249). This study laid important groundwork for questions of perception of musical structure (including segmentation) in tonal and non-tonal music. However, the study also invited a more thorough investigation of the perception of contemporary music (the structure of Stockhausen's *Klavierstueck IX* may not be representative of non-tonal music more generally or shares attributes with later contemporary music), with a larger participant pool representative of a broader demographic listening in an ecologically valid environment. The question of how segments are perceived in a live performance of contemporary non-tonal music is central to the current study.

Clarke and Krumhansl's (1990) study used musically trained participants. There is varying opinion regarding whether listeners with musical training perceive musical structure differently from those with no training. Deliège et al. (1996) found differences, and Eitan and Granot (2008) found some evidence that musical training may lead to the preference of a hybrid version over an original. Phillips and Cross (2011) found that musically trained listeners, when asked to judge duration in retrospect, judged the length of an extract of tonal music to be shorter than those with no musical training. Ockelford and Sergeant (2012) suggested that musicians may process non-tonal structure differently due to their training; they asked 14 musicians to listen to tone-rows and rate the extent to which notes "fit" with the other notes (probe tone paradigm) and found that musically trained listeners may impose tonal frameworks on the perception of non-tonal music (however, the study does not include a non-musician group for comparison). In contrast, Tillmann and Bigand (2004) found that the participants prioritized local over global features in perception, regardless of musical training. Bigand (2002) conducted a series of experiments which explored the perception of melody and harmony and, finding no difference in perception between trained and untrained listeners, concluded that mere exposure to music in everyday life results in everyone being an "experienced listener" (p. 304). A review of studies which examined various ways in which music is processed (perception of tension and relaxation, link between theme and variations, expectation generation, locating local features in global structure, and emotional response) by Bigand and Poulin-Charronnat (2006) also found that music perception depends on exposure to music rather than on formal training. Hence, studies do not convincingly support claims that musical training may alter the perception of musical structure.

The exposure effect (Zajonc, 2001) has been demonstrated to be important in music listening. For example, listeners exposed to a new musical scale for 25–30 min may show "extensive learning as characterized by recognition, generalization, and sensitivity to the event frequencies in their given grammar, as well as increased preference for repeated melodies in the new musical system" (Loui et al., 2010, p. 377). The proposition in this study, that "knowledge of musical structure is implicitly acquired from passive exposure to acoustical and statistical properties of musical sounds in the environment" (p. 386), is important when considering the perception of contemporary music; how this is perceived may depend on the extent to which a listener has been exposed to a similar musical system, or grammar, previously. However, it may not be the case that a grammar may be assumed for contemporary classical music, and there is as yet no convincing evidence of such. It is therefore also necessary to consider empirical studies of learned familiarity with contemporary music. For example, Western listeners may acquire knowledge of the structure of non-Western music through exposure (Stevens et al., 2013), and both musicians and non-musicians may acquire knowledge of sequences of melodies through exposure to a new musical grammar (Rohrmeier et al., 2011). On the other hand, studies have suggested that non-tonal music may be more difficult to store in working memory than tonal music for both musicians and non-musicians (Krumhansl, 1979). Schulze et al. (2012) asked musician and non-musician listeners to indicate whether tonal and non-tonal sequences were the same or different from the ones previously heard and found that both musicians and non-musicians performed better for tonal sequences than for non-tonal ones. These studies suggest that not only may working memory be better for tonal than non-tonal music but also that this may be the case regardless of musical training. It is not clear whether these two theories—that non-tonal music may be processed differently or that processing depends on exposure—are competing or two sides of the same coin (i.e., that processing depends on the *extent* of the exposure). Given that it is also not clear whether those who listen to contemporary classical music might acquire a set of expectations based on a learned grammar, this seems to be an important research question to pose in the current study.

Perception of structure in music relies on the way in which information is grouped and stored in memory or segmented. Moreover, the way in which musical information is organized and structured also plays a significant role in how that information is processed and remembered (e.g., Lerdahl, 1992 and notions of tonal pitch space, and probe tone and tonal hierarchy experiments such as Krumhansl and Kessler, 1982). Experimental work has recently begun to explore the sense of segmentation in contemporary music. Hartmann et al. (2017) asked 18 musicians and 18 non-musicians to listen to six 2-minute stimuli of unfamiliar music and to note segment boundaries (which they define for the participants as "instants of significant musical change," p. 6) by pressing a computer keyboard space bar. In a second experiment, the same 18 musicians were invited to listen and mark instants of significant change once again and then to revisit the places they had marked and move these as they wished after the second occasion of hearing. They then rated the strength of each change that they had marked. The study indicated that musical training may play a role in the perception of segmentation.

The question of what constitutes a "significant change" in music perception is an important one. Models of segmentation have highlighted relevant factors including emotion (Aljanaki et al., 2015), rhythm, timbre, and harmony (Jensen, 2007), and novelty (Foote, 2000). However, such modeling exercises have largely sought to represent experiences of tonal music (for example, Foote, 2000 discusses a change of tonality as representing a significant change). Most of these musical factors are relevant to contemporary music listening, but few studies have sought to explore how these may lead to the perception of change in non-tonal music.

The participants in all studies discussed so far listened to stimuli in a laboratory or office environment. Only recently have studies begun to collect data in a live concert hall setting, which could be considered a more ecologically valid setting. As Broughton et al. (2008) note: "There is a need to understand how audience members respond to a concert as the music unfolds in time" (p. 366). A live performance may be considered more immersive than listening to a recording, involving auditory and visual information and a close focus on the performance as the main object on which attention is focused. Indeed the results of the study by Broughton et al. (2008) do suggest that audience response is partly based on the bodily movement of the performer (in this case, a solo marimba player). During a concert hall performance, the listeners have usually elected to listen attentively to music. Sound quality is also likely to differ between a live concert half performance and a recording, with the latter preserving more of the detail of the acoustic sound (without recording, editing, and altering this and without the sound being passed through headphones or speakers or listened to in an otherwise noisy environment). It could thus be deduced that a live music listening experience may result in more attention being paid to the piece as it unfolds in time and to the chunks (or segments) of information in the auditory stream. Experiments conducted in a live concert hall environment may also further the understanding of what it means to be a performer or a member of the audience in a contemporary music performance in the modern day. A performance could be considered a co-creation of, in this case, the composer's work, a pianist, and the audience as a collective (rather than a set of isolated individuals), in which relevant experiences include other audients' reactions and the pianist's unique interpretation and realization. An experiment in a live concert hall setting is therefore communal, and the liveness of the assessment is an essential part of it. The findings of such a study may aid the knowledge of how performers and venues may adapt programs to audience demographics:

> The application of experimental techniques to the study of music performance in its own environment builds new pathways for performing musicians, teachers, researchers, and those involved in the presentation of music performance to better understand the behavior and development of audience members. Such research in the future will be of great interest to the aforementioned groups and impact upon the creation, presentation, and programming of live concert music. (Broughton et al., 2008, p. 369)

Egermann et al. (2013) asked 50 concert audience members to provide subjective responses to a live flute performance. In line with theories relating to the exposure effect, they found that the listeners' expectations regarding musical structure were a strong predictor of emotional response. This study provides support for notions such as the "experienced listener" discussed above, i.e., that familiarity with, or exposure to, music may influence the perception of segmentation and structure.

Results from relevant literature such as those studies discussed here suggest that there is a need for investigation of the experience of musical form as it unfolds in a live concert hall setting, where much of the real audience listening experience may be preserved (note that this method also involves limitations in the current study, which are discussed in the relevant section later). Furthermore, such investigation needs to advance the understanding of music created today rather than of tonal music composed in previous centuries. Studies should aim for data gathered from larger audiences, in excess of 50 participants, which is the maximum in existing studies. Finally, data on environmentally valid listening experiences should be gathered *via* means which are familiar to listeners and as unobtrusive as possible on the listening experience. This study seeks to address these gaps in knowledge in this field.

THE CURRENT STUDY

The aims of this study were to examine how audience members perceive segmentation in a live performance of contemporary classical music and whether perception varies according to musical training and/or familiarity with contemporary music in general. The study sought to address the gaps in existing empirical work discussed above by employing a smartphone app designed for the study, which allowed live concert hall audience participants to tap their devices when they considered a segment to have ended.

Research Questions

1. Is there evidence of agreement among audience members about segment boundaries in a piece of contemporary music?
2. What musical and sonic features, if any, occur where there is agreement on segment boundaries?
3. Does familiarity with contemporary music influence the decisions relating to segmentation?
4. What musical and sonic features do listeners self-report to have guided their decisions regarding segmentation and to what extent do these match the empirical tapping data collected?
5. Does musical training influence the decisions of segmentation?

MATERIALS AND METHODS

Ethical Considerations

The study was granted ethical approval by the Royal Northern College of Music (RNCM) Ethics Committee (REC 131, approved 13 September 2017). The first screen of the app contained information relating to ethical consent and details of how the data would be processed, stored, and used.

Participants

The 259 participants were audience members who attended an evening concert event in October 2017 as part of the Manchester Science Festival (MSF), which is run annually by the Science and Industry Museum in Manchester. The event was advertised in MSF publicity in print and online and in the event brochure of the RNCM, where the event took place. Those who purchased tickets for the event were subsequently contacted by email and invited to download the Practice and Research in Science and Music (PRiSM) Perception App (free to download *via* Apple and Android application online stores) in advance of their attendance at the event. Participation in the study was optional, and the audience members could attend the performance without taking part in the study. The 259 audience members who did opt to take part represented 43.2% of the 600 total attendees at the event. The performance was part of the launch event for the PRiSM research center titled "The Music of Proof: What Does Maths Sound Like?" Audience members included musicians from the RNCM and members (both adults and children) of the general public with an interest in music, maths, or science communication (as the event was part of a science festival).

Expert musicians were those that reported having had 10 or more years of musical training (52 participants in total). Those familiar with contemporary classical music were considered to be the participants who gave a response of 5, 6, or 7 to the question "As a listener, how familiar are you with twentieth-century music?" and those unfamiliar were the participants who responded with 1, 2, or 3. Although there was some overlap in these groups (those with musical training and those familiar with contemporary classical music), the overlap was only partial (46.67% of the participants who responded with 5–7 on a seven-point Likert scale to the question regarding their familiarity with contemporary classical music were also considered to be musicians for the purposes of this study, i.e., they had 10 or more years of musical training).

Apparatus

The audience responses were captured on the participants' own smartphones or tablets *via* the PRiSM Perception App, which was designed for this study. The source code used for the mobile app is registered under DOI "doi: 10.5281/zenodo. 2542790". Each device was used by one person and data were captured in real time.

After an "about this app" screen (described above and which included the ethics statement), the app consisted of three pages on which the participants were asked to enter data:

1. "Your profile" (date of birth, musical and mathematical training and experience, education, and how often the participants listen to music).
2. "Performances" (including the button to tap in response to the live performance).
3. "Questions" (questions relating to the experience of the stimulus).

Stimulus

The participants heard the solo piano piece "Fanfares" from *Etudes* (Book 1, 1985) by Ligeti. The piece remains at a consistent tempo throughout, and while one piano hand plays quavers (seven quavers per bar), the other plays a motif often referred to as "horn fifths" or a variation on this motif. The material switches between the piano hands and changes in register and dynamic (volume) throughout the piece (although these changes are less overt in the second half of the piece). Although occasionally one piano rests for a number of quaver beats, there is never a break in both piano hands until the final bars of the piece. Thus, the work could be seen as being made up of 8 to 10 bar "phrases," but the motion is never interrupted completely, and those segment boundaries commonly found in tonal music (cadences, fermatas) are not present. Steinitz's (1996, 2011) expert analysis of the work focuses on the first 45 bars and outlines how the motif changes hands every 8 to 10 bars, with the accompanying ostinato being played 208 times in the duration of the piece.

This contemporary study for solo piano was performed in a large concert hall in front of 600 audience members, of which 259 opted to take part in the current study. The work was selected for the experiment for the following reasons:

1. As a composer, Ligeti is firmly established as representative of contemporary composition.
2. Ligeti's *Etudes* is considered an important part of contemporary music repertoire and is regularly performed.
3. "Fanfares" includes features common to contemporary music (multiple series of notes which repeat, including an ostinato pattern, and a recurring motif commonly referred to as "horn fifths").
4. "Fanfares" lacks many of those features commonly found at points considered to mark section boundaries in tonal music (silences, harmonic closure, and changes of meter or tempo).

The live performance of this work lasted for 3 minutes and 26 seconds (i.e., 206 s).

Procedure

The participants could opt to fill in the "About You" page of the smartphone app before they arrived at the performance or afterward (it was compulsory to complete this page on the app in order to proceed to the experiment page). Immediately before the performance began, the participants were advised to take out their devices and open the "Performances" page on the PRiSM Perception App, which asked them to wait for the instruction to begin the participation in the study. A countdown was then given to the participants by a member of the PRiSM research team, at the end of which they were instructed to tap their screen to synchronize the devices to the timings of the performance. The performance then began, and the participants tapped the green button displayed on the app whenever they felt that a section had ended. The participants could also tap a button to mark the previous tap as an "error."

Following the experiment, the participants were asked to respond to a final series of questions concerning their experience of the piece (on the "Questions" page of the app), including their enjoyment of and familiarity with the piece, familiarity with contemporary music as a genre in general, and how they made decisions regarding section boundaries. A full list of the questions can be found in **Appendix 1**.

Analysis

The following steps were used to analyze the data:

1. A chi-square goodness-of-fit analysis was performed to examine whether the total taps were equally spread across the performance or whether they were not (research question 1).
2. The taps were divided into 2-s windows, i.e., the taps were grouped into clusters of 2 s each throughout the duration of the piece. 2 s was chosen as the relevant size of the groups as each bar in the performance lasts approximately 2 s, and the maximum size of one of the "breaks" (where one piano hand rests for a number of quaver beats) is 2 s. Therefore, 2 s is the upper limit for where we would expect audiences to identify a new segment. Additionally, comparable empirical studies have used windows [termed by Broughton et al. (2008) as a "lag," which they defined as 1.5 s for the purpose of their study] varying between 0.5 s (Geringer, 1995) and 3.25 s (Krumhansl, 1996). 2 s is therefore within the normal range established in the field when examining how long after an event a participant can be considered to have responded to that event.
3. The top five 2-s windows where the participants tapped were examined to explore which musical features occurred at these points (research question 2).
4. The top 10 2-s windows where the participants tapped were examined for those classified as familiar and not familiar with contemporary music (research question 3).
5. The participants' self-reports of how they decided where to tap (part of the post-performance "Questions" page of the app) were analyzed to determine what the most common factors were in guiding the participants' decisions regarding segmentation (research question 4).
6. The results of step 5 above were used, along with relevant research, to construct a new prediction of where the participants may tap.
7. This new prediction was compared to the actual groups of taps, including an investigation of the relevance of this new prediction to musicians *vs.* non-musicians (research question 5).

RESULTS

The 259 participants tapped an average of 8.16 times (standard deviation: 4.43) during the 3.78-min performance (minimum taps: one, maximum taps: 21). The taps which the participants marked as "errors" were removed prior to analysis. **Figure 1** shows all the taps by all the participants during the performance of the work, split into 2-s windows.

A chi-square test of goodness-of-fit was performed to determine whether the taps were spread evenly over the duration of the performance. The analysis suggested that the taps were not equally distributed across the total duration of the performance, $\chi^2(36) = 543.81$, $p < 0.0005$.

The five most common 2-s windows where the participants tapped occurred at the following points in the musical work (letters A–E below refer to the labels in **Figure 1**):

- A—bar 63, break of eight quavers in one hand, followed by a change in register and dynamics (*ppp* to *ff*) in the right hand.
- B—bar 74, break of 12 quavers in one hand, followed by a significant change in register in the left hand.
- C—bar 88, break of 10 quavers in one hand, followed by a change in register in the left hand, and dynamics in the left (*pp* to *ff*) and the right hand (*pppp* to *mp*).

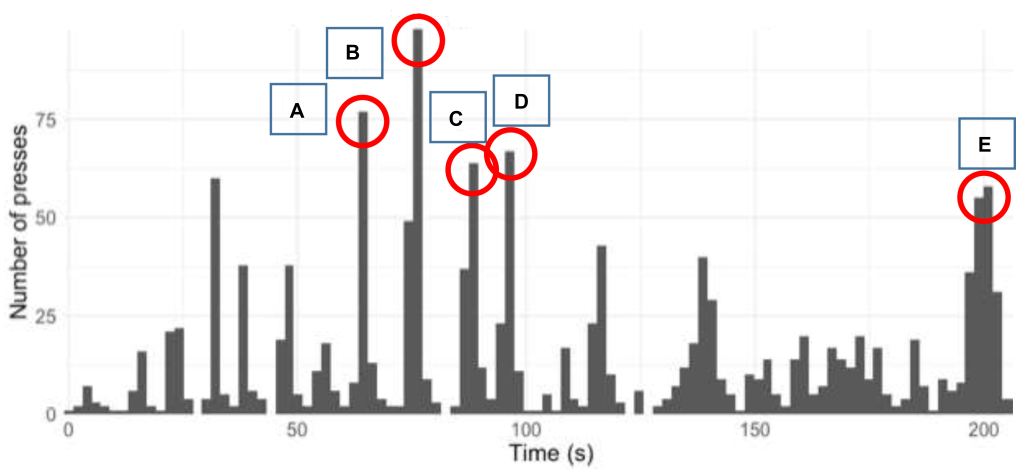

FIGURE 1 | All taps ("presses") by all participants during the performance of the work, split into 2-s windows (capital letters A–E refer to the aspects of this analysis which will be referred to in the outline of the results of the study).

- D—bar 96, break of 10 quavers in one hand, followed by a change in register in the left hand, and dynamics in the left (*mp* to *pppp*) and the right hand (*mp* to *pp*).
- E—bar 210, penultimate note (semibreve) of the piece.

From these five points, the following observations may be made: the factors which governed the decisions in relation to segmentation for all participants include: a break (quaver rests) in one piano hand, a change in register (or pitch), and a change in dynamic (or volume/listening level). The listeners also marked a segment break at the end of the piece when the consistent motion of the notes is replaced by one long note.

If points A–E above are reordered with the point which received the highest number of taps first, the following list results:

1. B—bar 74, break of 12 quavers in one hand, followed by a significant change in register in the left hand.
2. D—bar 96, break of 10 quavers in one hand, followed by a change in register in the left hand, and dynamics in the left (*mp* to *pppp*) and the right hand (*mp* to *pp*).
3. A—bar 63, break of eight quavers in one hand, followed by a change in register and dynamics (*ppp* to *ff*) in the right hand.
4. E—bar 210, penultimate note (semibreve) of the piece.
5. C—bar 88, break of 10 quavers in one hand, followed by a change in register in the left hand, and dynamics in the left (*pp* to *ff*) and the right hand (*pppp* to *mp*).

This reordering could suggest that a break (quaver rests) in one piano hand is the most common factor in leading the listeners to perceive segmentation.

Figures 2, **3** give the equivalent visualization in **Figure 1** of taps in 2-s windows for those who self-reported as being familiar (**Figure 2**) or not familiar (**Figure 3**) with contemporary music.

As the participants filled in this page of questions (which included familiarity ratings) after the performance, this gave the participants the option of not completing these questions. A total of 85 out of the 259 participants who provided tapping responses completed these questions. The method of analysis of these familiar and not familiar groups was informed by this fact, i.e., not all participants whose data features in **Figure 3** responded to this part of the smartphone app.

The top 10 points where all the participants, familiar and not familiar with contemporary music, tapped are outlined in **Table 1**.

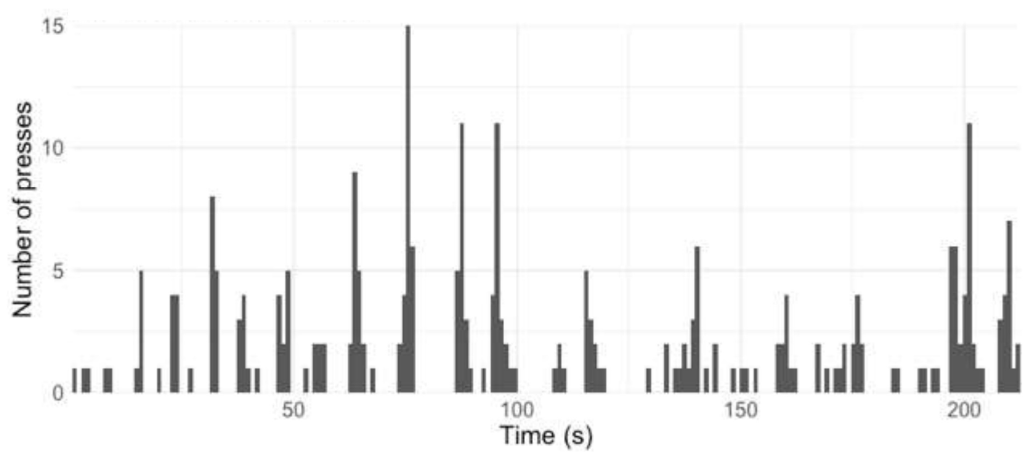

FIGURE 2 | All taps ("presses") by all participants familiar with contemporary music, split into 2-s windows.

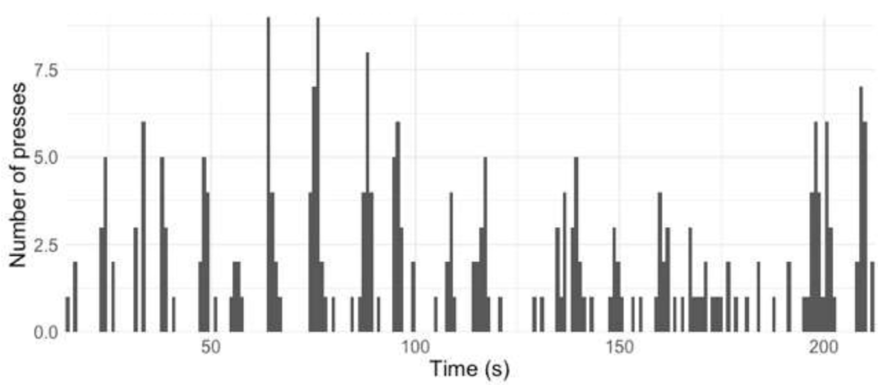

FIGURE 3 | All taps ("presses") by all participants not familiar with contemporary music, split into 2-s windows.

From this table, it can be seen that the familiar and the not-familiar groups both tapped at the main points at 76 s ("1" in the notes column in **Table 1**), 88 s ("2"), and 64 s ("3"). There is hence some agreement between these two groups on points of segmentation.

However, the remaining four points in **Table 1** differ between the groups. First, the familiar group features six times in **Table 1**, compared with four times in the not-familiar group. This may suggest that those not familiar with contemporary music were in more agreement regarding segmentation with others than those in the familiar group. The remaining four points occur at the following points in the music (the numbers below refer to the "notes" column in **Table 1**):

4. In this bar, the music remains in the *pp* (right hand) and *pppp* (left hand) dynamic range (as for the eight bars before this), but the right hand changes register by moving upwards by c. three octaves.
5. This is the final bar of the music (two semibreves tied together).
6. In this bar, the music changes in dynamic (*mp* to *pp* in the left hand and *pp* to *f* in the right hand) and the two piano hands swap roles in terms of the musical information that they play—the right hand now plays the "melody" ("horn fifths") material, while the left hand plays the quaver ostinato.
7. This is 1 s before the window at 76 s and is therefore most likely to be a marker of earlier responses to the same event.

Points 4 and 6 above suggest that there may be musical and sonic features which lead the listeners to mark a point of segmentation (change of register and change of musical material played by each piano hand) but that these may be detected more by the listeners who are familiar with contemporary classical music than those who are not.

In the post-performance "Questions" page of the app, the participants were asked to state in a blank text box the basis on which they made their decisions of segmentation. The responses were categorized into overarching themes, and the results.

TABLE 1 | Top 10 points (in seconds) where all participants, familiar and not familiar with contemporary music, split into 2-s windows, tapped (the "Notes" column refers to the comments in the results section).

Condition (familiar *vs*. not familiar)	Time point in the duration of the piece (s)	Number of taps in this 2-s window	Notes
Familiar	32	8	6
Not familiar	64	9	3
Familiar	64	9	3
Not familiar	75	7	7
Familiar	76	15	1
Not familiar	76	9	1
Familiar	88	11	2
Not familiar	88	8	2
Familiar	96	11	4
Familiar	201	11	5

The participants reported that the main factors used to make their decisions about where a segment ended were speed, register/pitch (a change in pitch or register), and a change in melody, theme, phrase, or motif. The work did not change tempo throughout but did include changes in various places in relation to pitch and melody. Any perception of a change of speed could not have resulted from changes in tempo in the music but may have been caused by changes in other musical characteristics.

Based on the self-reported data, it is possible to form new predictions regarding where the participants in the study may have tapped. Such predictions can also be informed by existing formal music analysis of the piece. At this point, it is important to acknowledge that all of these features, which may result in the listeners choosing to mark a segment boundary, may be thought of as distinct phenomena. On the one hand, the features may be the musical material itself, such as the melodic material, the notes themselves, and any moments where one hand of the piano has a break (rests). The second category of features could be termed the "sonic" features of the piece. Such sonic features include those aspects which may be used to determine auditory streams (in line with Bregman's, 1990 theories of "auditory scene analysis"). These include changes in the quality of the sound, such as its pitch height (register) and listening level (or volume/dynamic).

The following segmentation analysis is based on the main musical and sonic characteristics which listeners self-reported to have guided their judgments of segmentation (see **Table 2**, which shows the main musical characteristics—other than "speed," as the piece did not change tempo during the performance—as "melody," "register/pitch," and "dynamic"), plus segment boundaries identified by music analysis (note that the first five segments in **Table 2** are also identified as the first five segments in the analysis of the work by Steinitz, 2011).[1]

As can be seen from the discussion above, the top five segments in which the participants tapped most commonly (at the beginning of the 2-s window) are 9, 15, 12, 13, and 16. The musical and the sonic features of these include a change in dynamics (or volume/listening level), a change in the kind of musical material which makes up the "melody" (i.e., a different motif to the "horn fifths"), a break in the musical material in one piano hand (rests), and a change in register/pitch. Given the balance of these features in these top five segments, the importance of each of these features in the perception of segmentation seems to be in the following order of priority:

1. Melody + dynamics.
2. Break in the musical material (one piano hand has rests).
3. Register/pitch.

[1]It should be noted that, in the live performance which formed the basis for this study, the pianist skipped two bars (c. 2 s) of the original musical notation. The mistake was not audible, as verified during discussions with expert audience members (including Professor Richard Steinitz, author of Steinitz, 2011) but was detected during the analysis. This two-bar skip did not happen across any of the segment boundaries in **Table 2** or at any of the points at which the audience marked on the main segment boundaries in the piece.

A change in dynamics occurs at three of these top five segments and is the most common feature out of this list of three.

Segments 9, 15, 12, 13, and 16 (marked with an asterisk in **Table 2**) are the five segments identified in **Figure 1** as being in the top five 2-s windows in which the participants tapped. The participants self-report that the main factors influencing their segmentation decisions are speed, register/pitch, melody, and dynamic. However, **Table 2** suggests that these self-reports of which factors influence the decisions of segmentation are only partially accurate. The participants mainly appear to use dynamic, melody, a break in the musical texture (rests in one piano hand), and register/pitch to guide their decisions rather than changes in tempo.

These top five segments in **Table 2** (9, 15, 12, 13, and 16) occur at bars 46, 114, 74, 88, and 116, respectively. This means that, when predictions of perceptual segmentation are formed using music analysis and self-report data regarding how the listeners make their decisions, these are the top five places where the predictions match the main points at which people tapped. However, the top five points at which all the participants tapped as per **Figure 1**, which shows the top taps when data are split into 2-s bins (rather than according to prediction of segmentation), occur at bars 63, 74, 88, 96, and 210. In summary, **Figure 1** and **Table 2** both have bars 74 and 88 in common, but they differ on their other top five points. This suggests that music analysis and self-report data alone do not account for how segmentation is perceived.

Logistic regression models were built to determine whether tapping (or not tapping) at a segment boundary was predicted by the degree of musical training that the participants had. We found that music training did predict whether people tapped at the segment 13 boundary (beta = 0.13308, $z = 2.103$, $p = 0.0355$) and at the segment 15 boundary (beta = 0.12626, $z = 2.373$, $p = 0.0177$), but not at any other boundary (all other ps > 0.05). As the level of musical training increased, the participants were more likely to tap at these segments. As can be seen in **Table 2**, segment 13 involves a change in dynamics, and segment 15 involves a break (one piano hand rests). This may be interpreted as suggesting that musicians and non-musicians rely differently on these cues for their judgment of segmentation, with musical training resulting in listeners being more likely to tap in response to some musical and sonic features. However, the effect of musical training is quite small as it only affected two of the segments, and there is no clear evidence that there are distinct features, or cues, at these points in the music. This should not, therefore, be considered a significant result.

All data and the script for the models discussed above are freely available at https://github.com/ajstewartlang/RNCM_Ligeti.

TABLE 2 | Prediction of segmentation in the piece based on participant self-reports of what musical features lead to segmentation decisions.

Segment number	Bars	Reason for production of segmentation identification (musical feature)	Number of taps at the beginning of this segment/end of the previous segment (see Figure 4)
1	1–8	Register/pitch	N/A
2	8–9	Break	6
3	10–17	Register/pitch	16
4	18	Break	21
5	18–26	Register/pitch + dynamic	17
6	26–27	Break	8
7	28–36	Register/pitch + dynamic	5
8	37–45	Register/pitch + dynamic	6
9*	46–53	Melody + dynamics	38
10	54–62	Register/pitch	18
11	63–73	Register/pitch + dynamic	13
12*	74–87	Register/pitch + dynamic	30
13*	88–95	Dynamics	30
14	96–113	Register/pitch + dynamic	26
15*	114–115	Break	37
16*	116–122	Register/pitch	29
17	123–129	Register/pitch	6
18	130–141	Register/pitch	3
19	142–170	Dynamic	9
20	170–177	Dynamic	18
21	177–end	Dynamic	5

DISCUSSION

The research questions will be addressed in turn:

Is There Evidence of Agreement Among Audience Members About Segment Boundaries in a Piece of Contemporary Music?

The results outlined above suggest that the taps were not evenly distributed across the piece of music (as can be seen clearly in **Figure 1**). We would have expected to see a more even distribution had taps been random. This suggests that there are commonalities in how segmentation is perceived by the listeners to this piece of music.

What Musical and Sonic Features, If Any, Occur Where There is Agreement on Segment Boundaries?

The musical characteristics which influence such decisions include a break (quaver rests) in one piano hand, register (pitch), dynamic (volume), and iteration or breakdown of an established pattern (i.e., the end of the piece, when the continuous ostinato and melody pattern ceases). Of these, a break in the continuous quaver pattern in one piano hand seems to be the most common feature which leads to perceptual segmentation.

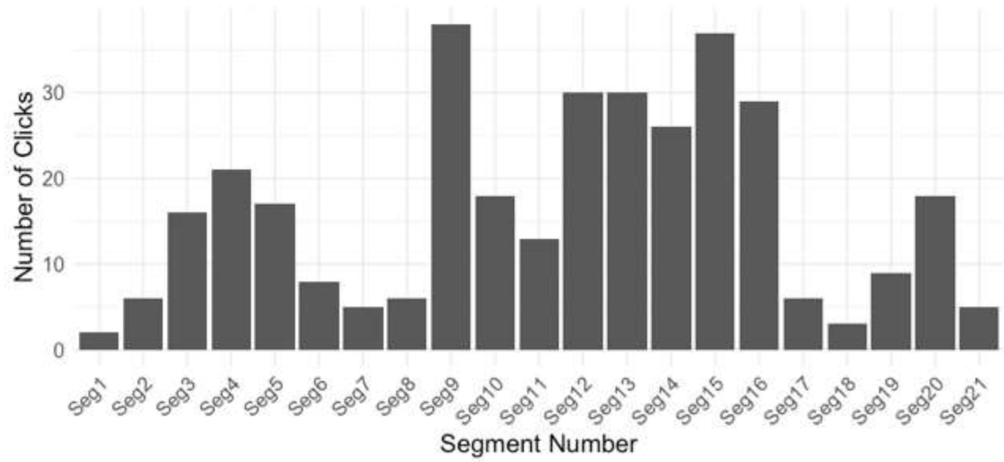

FIGURE 4 | Number of taps ("clicks") at each segment boundary outlined in the final column of **Table 2**.

Does Familiarity With Contemporary Music Influence Decisions Relating to Segmentation?

The results suggest that familiarity with contemporary classical music in general (not only with this specific piece of music) may influence the segmentation decisions when listening to this piece. More of the top 10 groups of taps, for those who self-reported being familiar with contemporary music or not, were found in listeners who were familiar. This suggests that these listeners pick up on a wider range of cues to make their decisions of segmentation. The features which resulted in the listeners marking a point of segmentation, such as change of register and change of musical material played by each piano hand, seem to be detected more by listeners familiar with contemporary classical music than those who were not.

What Musical and Sonic Features Do Listeners Self-Report to Have Guided Their Decisions Regarding Segmentation, and to What Extent Do These Match the Empirical Tapping Data Collected?

The participants self-reported most commonly that speed influenced their decisions regarding segmentation. As this piece of music remained at one constant tempo throughout, this finding suggests that the listeners are not fully aware of how decisions regarding segmentation are made. The data analysis also suggests that the top five points at which the listeners perceived segmentation are different, depending on whether the data are analyzed in 2-s windows (**Figure 1**) or compared with a model of perceptual segmentation which was formed using self-report data and music analysis (**Table 2**).

Taken together, these findings indicate that the factors guiding the perception of a piece of music as it is realized in time may not be predicted by the listeners' own accounts of perception or by an examination of the musical score. These data provide further support to the notion that the structures which may be relevant during an analysis of a musical score may not be the same as those relevant in the perception of the music as it unfolds in time (relevant debates in this area have concerned, for example, music and the golden section; see e.g., Cook, 1987; Phillips, 2019). Also, regardless of the level of musical training or familiarity with contemporary classical music, the listeners may not be able to entirely predict how a piece of music may be perceived. This has relevance for listeners, composers, and performers today.

One possible reason that the participants self-reported speed as playing a role in their decisions may be that speed was a proxy for a different musical characteristic. For example, perhaps *note density* (number of onsets in a given temporal window) was interpreted as speed. This interpretation proposes that a point where there are fewer notes in the bar, for example, where one hand of the piano does not play, could sound like a change in tempo. The results discussed in relation to research question 1 suggest that the listeners commonly marked a segment when one piano hand was resting and hence where there were fewer notes in the piece at the point. Perhaps note density is key in understanding the sense of segmentation during live music listening.

Does Musical Training Influence Decisions of Segmentation?

Although the multiple logistic regression models used to analyze these data suggested that musicians were slightly more likely to tap at certain segments than non-musicians, this effect was found to be quite small. We therefore consider that there is no clear evidence for an effect of musical training on perceptual segmentation.

Limitations

Data collection took place in a live concert hall environment, using a smartphone application which could be freely downloaded onto the participants' personal devices. This allowed for high-resolution and non-obtrusive data capture, which may be conducted in a silent, low-light environment. Data collection by smartphone application also allowed for the participants to contribute responses both within (real-time responses to musical structure) and outside (demographic data and subjective responses to music) the concert hall environment. However, this element of the study design also resulted in some limitations to the study. Firstly, peripheral or direct vision of the other participants' activity on the app could have altered behavior. It may have been the case that the listeners tapped due to seeing a neighbor tap, possibly demonstrating a desire to conform (Asch, 1951, replicated by Brandstetter et al., 2014). The results of the study by Brandstetter et al. (2014) suggest that the participants are even more likely to conform if a situation contains a high degree of uncertainty. The current experiment could be considered to include uncertainty, first, as it is not common that listeners are asked to respond to a concert hall performance in real time and, second, as conventional segment boundaries in classical music (silence, cadences, and pauses) were not available in this piece of music, and the decisions regarding segmentation may therefore have been difficult. However, not all audience members who attended the event took part in the experiment; 259 of the 600 audience members chose to download the app and take part in the study, i.e., 43%. The participants were not asked for their seat numbers and no data were collected on their physical clustering or dispersal. Those taking part in the study were not allocated particular seats apart from the non-participants. Therefore, it was not the case that all the participants were sitting next to someone else who was also participating. It is therefore unlikely that conformity could account for the main findings of this study.

A further limitation of this study was that, although the first research questions asked "is there evidence of agreement among audience members about segment boundaries in a piece of contemporary music?," only one piece of contemporary music was used for the experiment. The results cannot be considered as generalizable to all contemporary classical music. However, the results may be relevant for considerations of perception of segmentation in music which does not contain any of those segment boundaries that have been shown to influence the perception in tonal music, and the method, app, and data analysis could be reproduced for comparative studies of future live performances.

The performance lasted for 206 s. This brings with it another limitation, which is that the participants' responses may have changed over the duration of the listening period (e.g., they may have paid more attention to the performance at the beginning and therefore tapped more or less than in the middle). However, the taps at each of the segments outlined in **Table 2** suggest that it is not the case that the listeners tapped less frequently in the second half of the piece than in the first half; there are 100 taps in the first 10 segments in this table and 115 taps in the segments in the second half. However, the cues to which the listeners responded may indeed have changed over the course of the work. Although **Table 2** does not suggest this, changing responses over 206 s could have played a role in the results. Future studies could create multiple versions of the stimulus in which the segments are rearranged (similar to that of Eitan and Granot, 2008), and the listeners are asked to respond to segmentation in these new versions.

There are multiple different ways in which these data could have been analyzed. The analysis above addresses the research questions and leads to valuable insights which will hopefully pave the way for future studies. One aspect which could be examined in such studies includes the size of the temporal windows into which the tapping data were split. The current study used windows of 2 s due to the musical information in the piece, the rate at which events occurred, and existing standards in other relevant empirical work. However, future studies could consider different temporal windows.

CONCLUSION

Overall, this study suggests that listeners do agree on segment boundaries when listening to this piece of contemporary classical music in a live concert hall setting as evidenced in their responses to segmentation in real time using a bespoke smartphone app. The segments which were identified by the participants seem to have occurred where one hand in the piano part had a break of 10–13 quavers (N.B.: some audience members could see the pianist's hands and the piano keyboard, but only from a distance), where the music changed significantly in register/pitch, dynamic (volume), or where there was a change in the melody or motif (e.g., at the end of the piece). The musical factors which influenced segmentation were partly evidenced in the participants' self-reports of how they made decisions, although speed was the most commonly reported reason for identifying change, and the piece was at a consistent tempo throughout. The perception of a change of speed may actually be a response to a change in note density. A break in one piano hand also leads to the perception of segmentation, which is a factor that could be linked to previous empirical studies of tonal music, which have also found that a rest, or break, provides a sense of a segment having ended. The marking of a segment boundary at a point where one piano hand rests could also be linked to the idea that the listeners perceive segmentation when there is a change in an aspect of the texture or timbre of the piece in general.

The results suggest that familiarity with contemporary classical music as a genre may influence the perception of segmentation. Those familiar with this genre seemed to use a wider range of cues to inform their responses (taps), including a change of register and a change of musical material played by each piano hand. These findings could be interpreted as supportive of the notion of the "experienced listener"—general music listening experience may play as significant a role in music perception as formal music training, or perhaps an even greater role.

The results of this study shed valuable light on how the listeners with varying levels of experiences of listening to contemporary classical music may parse musical information in real time in an ecologically valid setting. Such findings may be of interest to composers and performers who could make decisions regarding how they use musical material to suggest a particular structure to a listener (e.g., a performer could alter the dynamic level in different segments). However, given the results of the self-report data above, it is unlikely to be the case that composers or performers could predict all aspects of how segmentation may be perceived.

A further implication of these findings relates to concert hall programming and audience engagement. The perception of contemporary classical music may change as a listener becomes more familiar with this repertoire. Access to contemporary music may appeal to new audiences if these audiences can be gradually exposed to the musical and the sonic features of contemporary classical music, and it may not be the case that marketing should be targeted preferably at those potential audience members with formal musical training.

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 RNCM Ethics Committee, Royal Northern College of Music. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MP contributed to experimental design, literature review, data collection, data analysis, writing, review, and submission of the manuscript. JW contributed to experimental design and building of mobile phone software application. EH contributed to experimental design, data collection, and review of the final manuscript. MS contributed to experimental design, data collection, and review of the final manuscript. PW contributed to experimental design and review of the final manuscript. LJ contributed to data analysis and review of the final manuscript. AS contributed to data analysis and review of the final manuscript. DD contributed to experimental design, building of mobile phone software, application data collection, data analysis, advice on, and review of the final manuscript.

FUNDING

This research was supported by the Royal Northern College of Music. Development of the PRiSM perception application was funded by "Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption," awarded by the Engineering and Physical Sciences Research Council (EP/L019981/1). The RNCM Centre for Practice & Research in Science & Music (PRiSM), is funded by the Research England fund Expanding Excellence in England (E3).

ACKNOWLEDGMENTS

The authors note their gratitude to the two reviewers whose generous and rigorous comments on an earlier draft of this paper enabled them to improve the clarity of its methods of analysis and broaden its frames of reference. Many thanks to the Museum of Science and Industry (Manchester) and the Manchester Science Festival team. Thanks to Professor Richard Steinitz for the invaluable expertise.

REFERENCES

Aljanaki, A., Wiering, F., and Veltkamp, R. C. (2015). MediaEval 2015: a segmentation-based approach to continuous emotion tracking. *Paper Presented at the MediaEval* (Utrecht: Utrecht University).

Asch, S. E. (1951). "Effects of group pressure upon the modification and distortion of judgments," in *Groups, Leadership And Men; Research In Human Relations*, ed. H. Guetzkow (New York, NY: Carnegie Press), 177–190.

Bigand, E. (2002). More about the musical expertise of musically untrained listeners. *Paper Presented at the Conference on Neurosciences and Music: Mutual Interactions and Implications on Developmental Functions*, Venice. doi: 10.1196/annals.1284.041

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

Brandstetter, J., Rácz, P., Beckner, C., Sandoval, E. B., Hay, J., and Bartneck, C. (2014). "A peer pressure experiment: recreation of the Asch conformity experiment with robots," in *Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Chicago, IL.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization Of Sound*. Cambridge, MA: Cambridge University Press.

Broughton, M., Stevens, C., and Schubert, E. (2008). "Continuous self-report of engagement to live solo marimba performance," in *Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC10)*, Sapporo.

Bruderer, M. J., and McKinney, M. F. (2008). Perceptual evaluation of models for music segmentation. *Paper Presented at the Proceedings of the 4th Conference on Interdisciplinary Musicology*, Thessaloniki.

Cambouropoulos, E. (2006). Musical parallelism and melodic segmentation: a computational approach. *Music Percept.* 23, 249–268.

Caplin, W. E. (2001). *Classical form: A Theory Of Formal Functions For The Instrumental Music Of Haydn, Mozart, And Beethoven*. Oxford: Oxford University Press.

Chew, E. (2001). "Modeling tonality: Applications to music cognition," in *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*, New York, NY.

Clarke, E. F., and Krumhansl, C. L. (1990). Perceiving musical time. *Music Percept.* 7, 213–252.

Cook, N. (1987). Musical form and the listener. *J. Aesthet. Art Critic.* 46, 23–29.

Deliège, I., Melen, M., Stammers, D., and Cross, I. (1996). Musical schemata in real-time listening to a piece of music. *Music Percept.* 14, 117–159.

Dibben, N. (1994). The cognitive reality of hierarchic structure in tonal and atonal music. *Music Percept.* 12, 1–25.

Drake, C., and Bertrand, D. (2001). The quest for universals in temporal processing in music. *Psychol. Sci.* 13, 71–74. doi: 10.1111/j.1749-6632.2001.tb05722.x

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

Eitan, Z., and Granot, R. Y. (2008). Growing oranges on mozart's apple tree: "inner form" and aesthetic judgment. *Music Percept.* 25, 397–418.

Foote, J. (2000). Automatic audio segmentation using a measure of audio novelty. *Paper Presented at the 2000 IEEE International Conference on Multimedia and Expo. ICME2000*, London.

Geringer, J. M. (1995). Continuous loudness judgments of dynamics in recorded music excerpts. *J. Res. Music Educ.* 43, 22–35.

Granot, R. Y., and Jacoby, N. (2011). Musically puzzling I: sensitivity to overall structure in the sonata form? *Mus. Sci.* 15, 365–386.

Granot, R. Y., and Jacoby, N. (2012). Musically puzzling II: sensitivity to overall structure in a Haydn E-minor sonata. *Mus. Sci.* 16, 67–80.

Hanninen, D. A. (2003). A theory of recontextualization in music: analyzing phenomenal transformations of repetition. *Music Theor. Spect.* 25, 59–97.

Hartmann, M., Lartillot, O., and Toiviainen, P. (2017). Interaction features for prediction of perceptual segmentation: effects of musicianship and experimental task. *J. New Music Res.* 46, 156–174.

Jensen, K. (2007). Multiple scale music segmentation using rhythm, timbre, and harmony. *EURASIP J. Appl. Signal Proc.* 2007:159.

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cogn. Psychol.* 11, 346–374.

Krumhansl, C. L. (1996). A perceptual analysis of mozart's piano sonata K. 282: segmentation, tension, and musical ideas. *Music Percept. Interdiscipl. J.* 13, 401–432.

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.

Lerdahl, F. (1992). Cognitive constraints on compositional systems. *Contemp. Music Rev.* 6, 97–121.

Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. London: MIT Press.

Loui, P., Wessel, D. L., and Kam, C. L. H. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Percept.* 27, 377–388. doi: 10.1525/mp.2010.27.5.377

Marsden, A. (2005). Generative structural representation of tonal music. *J. New Music Res.* 34, 409–428.

Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago: Chicago University Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97. doi: 10.1037/0033-295x.101.2.343

Ockelford, A., and Sergeant, D. (2012). Musical expectancy in atonal contexts: musicians' perception of "antistructure". *Psychol. Music* 41, 139–174.

Phillips, M., and Cross, I. (2011). "About musical time–effect of age, enjoyment, and practical musical experience on retrospective estimate of elapsed duration during music listening," in *Proceedings Of The Multidisciplinary Aspects Of Time And Time Perception*, Athens.

Phillips, M. E. (2019). Rethinking the role of the golden section in music and music scholarship. *Creativ. Res. J.* 31, 419–427.

Rohrmeier, M., Rebuschat, P., and Cross, I. (2011). Incidental and online learning of melodic structure. *Conscious. Cogn.* 20, 214–222. doi: 10.1016/j.concog.2010.07.004

Rolison, J. J., and Edworthy, J. (2012). The role of formal structure in liking for popular music. *Music Percept. Interdiscipl. J.* 29, 269–284.

Rothstein, W. (1991). On implied tones. *Music Analy.* 10, 289–328.

Schenker, H. (1979). *Free Composition (der Freie Satz): Heinrich Schenker*. London: Longman.

Schulze, K., Dowling, W. J., and Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Percept.* 29, 255–267.

Sears, D., Caplin, W. E., and McAdams, S. (2014). Perceiving the classical cadence. *Music Percept.* 31, 397–417.

Spies, B. (2005). Facilitating access to atonal music: ligeti's second string quartet. *J. Musical Arts Africa* 2, 55–69.

Steinitz, R. (1996). The dynamics of disorder. *Musical Times* 137, 7–14.

Steinitz, R. (2011). *Gyorgy Ligeti*. London: Faber & Faber.

Stevens, C. J., Tardieu, J., Dunbar-Hall, P., Best, C. T., and Tillmann, B. (2013). Expectations in culturally unfamiliar music: influences of proximal and distal cues and timbral characteristics. *Front. Psychol.* 4:789. doi: 10.3389/fpsyg.2013.00789

Straus, J. N. (2016). *Introduction to Post-Tonal Theory*. New York, NY: WW Norton & Company.

Temperley, D. (2007). *Music and Probability*. Cambridge, MA: MIT Press.

Tillmann, B., and Bigand, E. (1996). Does formal musical structure affect perception of musical expressiveness? *Psychol. Music* 24, 3–17.

Tillmann, B., and Bigand, E. (2004). The relative importance of local and global structures in music perception. *J. Aesthet. Art Critic.* 62, 211–222.

Tillmann, B., Bigand, E., and Madurell, F. (1998). Local versus global processing of harmonic cadences in the solution of musical puzzles. *Psychol. Res.* 61, 15–17.

Zajonc, R. B. (2001). Mere exposure: a gateway to the subliminal. *Curr. Direct. Psychol. Sci.* 10, 224–228.

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.

APPENDIX 1—SURVEY QUESTIONS

Q1. What is your date of birth (YYYY-MM-DD)?

Q2. How many years of formal music training have you had (including A-level and any instrumental, vocal, or composition lessons)? [options: 0 1 2 3 4 5 6 7 8 9 10+]

Q3. Do you currently play a musical instrument, sing, or compose, and if so for how long? (Please select zero if you do not have any) [options: 0 1 2 3 4 5 6 7 8 9 10+]

Q4. How many years of formal mathematics training have you had (including A-level and any further study of mathematics)? [options: 0 1 2 3 4 5 6 7 8 9 10+]

Q5. Do you currently work in a field which requires mathematical skills, and if so for how long have you worked in this area? (Please select zero if you do not work with mathematics) [options: 0 1 2 3 4 5 6 7 8 9 10+]

Q6. What is your highest level of formal qualification? [options: GCSEs, A-levels, Bachelor's Degree, Master's Degree, Ph.D., other (please state in the text box)]

Q7. How often do you listen to music (of any style)? [options: never, occasionally, sometimes, most days, every day]

Q8. How long did you think that the piece of music lasted? [minutes and a seconds box, with the options that either can be left blank but not both]

Q9. Please describe the piece in three words [followed by a text box]

Q10. How did you decide when a section had ended [followed by a text box]?

Q11. How much did you enjoy the piece of music? (from 1 = I did not enjoy it at all to 7 = I enjoyed it a lot) [options 1–7]

Q12. As a listener, how familiar are you with twentieth-century classical music? (from 1 = I am not familiar with it at all to 7 = I am very familiar with it) [options 1–7]

Q13. How often do you listen to twentieth-century classical music? (from 1 = never listen to 7 = listen every day) [options 1–7]

Q14. How familiar are you with the piece performed? (from 1 = I have never heard of it to 7 = I have heard it many times) [options 1–7]

Q15. Does participation in a scientific experiment such as this increase or decrease your enjoyment of a performance? (from 1 = it significantly decreases my enjoyment to 7 = it significantly increases my enjoyment) [options 1–7]

Q14. What motivated you to come to tonight's event (select all that apply)? [options: I wanted to learn more about music and maths working together, I am a regular attendee of RNCM events, I wanted to take part in a scientific study, a friend/family member asked me to come along]

Q16. Please use this box for any other comments you wish to make [followed by a text box]

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