

The background of the entire page is a complex network of overlapping circles in various colors (blue, green, yellow, orange, red, purple) connected by thin, multi-colored lines, creating a web-like or neural network pattern. A solid orange horizontal band is positioned at the top, containing the title text.

# MUSIC AND THE FUNCTIONS OF THE BRAIN: AROUSAL, EMOTIONS, AND PLEASURE

**EDITED BY :** Mark Reybrouck, Tuomas Eerola and Piotr Podlipniak  
**PUBLISHED IN :** Frontiers in Psychology and Frontiers in Neuroscience



# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2018 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-452-5

DOI 10.3389/978-2-88945-452-5

## 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: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# MUSIC AND THE FUNCTIONS OF THE BRAIN: AROUSAL, EMOTIONS, AND PLEASURE

Topic Editors:

**Mark Reybrouck**, University of Leuven, University of Ghent, Belgium

**Tuomas Eerola**, Durham University, United Kingdom

**Piotr Podlipniak**, Adam Mickiewicz University in Poznań, Poland



Visualisation of 100 emotion words in the context of music (unpublished data and visualisation by Tuomas Eerola).

Music impinges upon the body and the brain. As such, it has significant inductive power which relies both on innate dispositions and acquired mechanisms and competencies. The processes are partly autonomous and partly deliberate, and interrelations between several levels of processing are becoming clearer with accumulating new evidence. For instance, recent developments in neuroimaging techniques, have broadened the field by encompassing the study of cortical and subcortical processing of the music. The domain of musical emotions is a typical example with a major focus on the pleasure that can be derived from listening to music. Pleasure, however, is not the only emotion to be induced and the mechanisms behind its elicitation are far from understood. There are also mechanisms related to arousal and activation that are both less differentiated and at the same time more complex than the assumed mechanisms that trigger basic emotions. It is imperative, therefore, to investigate what pleasurable and mood-modifying effects music can have on human beings in real-time listening situations. This e-book is an attempt to answer these questions. Revolving around the specificity of music experience in terms of perception,

emotional reactions, and aesthetic assessment, it presents new hypotheses, theoretical claims as well as new empirical data which contribute to a better understanding of the functions of the brain as related to musical experience.

**Citation:** Reybrouck, M., Eerola, T., Podlipniak, P., eds. (2018). Music and the Functions of the Brain: Arousal, Emotions, and Pleasure. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-452-5

# Table of Contents

- 05 Editorial: Music and the Functions of the Brain: Arousal, Emotions, and Pleasure**  
Mark Reybrouck, Tuomas Eerola and Piotr Podlipniak

## **Part I: Hypotheses and Theory**

- 07 Music and Its Inductive Power: A Psychobiological and Evolutionary Approach to Musical Emotions**  
Mark Reybrouck and Tuomas Eerola
- 21 Global Sensory Qualities and Aesthetic Experience in Music**  
Pauli Brattico, Elvira Brattico and Peter Vuust
- 34 Constituents of Music and Visual-Art Related Pleasure – A Critical Integrative Literature Review**  
Marianne Tiihonen, Elvira Brattico, Johanna Maksimainen, Jan Wikgren and Suvi Saarikallio
- 46 Arousal Rules: An Empirical Investigation into the Aesthetic Experience of Cross-Modal Perception with Emotional Visual Music**  
Irene Eunyoung Lee, Charles-Francois V. Latchoumane and Jaeseung Jeong

## **Part II: Empirical Studies**

- 77 Pitch Syntax Violations Are Linked to Greater Skin Conductance Changes, Relative to Timbral Violations – The Predictive Role of the Reward System in Perspective of Cortico-subcortical Loops**  
Edward J. Gorzelańczyk, Piotr Podlipniak, Piotr Walecki, Maciej Karpiński and Emilia Tarnowska
- 88 The Pleasure Evoked by Sad Music Is Mediated by Feelings of Being Moved**  
Jonna K. Vuoskoski and Tuomas Eerola
- 99 A Neurodynamic Perspective on Musical Enjoyment: The Role of Emotional Granularity**  
Nathaniel F. Barrett and Jay Schulkin
- 105 Musicians Are Better than Non-musicians in Frequency Change Detection: Behavioral and Electrophysiological Evidence**  
Chun Liang, Brian Earl, Ivy Thompson, Kayla Whitaker, Steven Cahn, Jing Xiang, Qian-Jie Fu and Fawen Zhang
- 119 High-Resolution Audio with Inaudible High-Frequency Components Induces a Relaxed Attentional State without Conscious Awareness**  
Ryuma Kuribayashi and Hiroshi Nittono

**131 *Emotional Responses to Music: Shifts in Frontal Brain Asymmetry Mark Periods of Musical Change***

Hussain-Abdulah Arjmand, Jesper Hohagen, Bryan Paton and Nikki S. Rickard

**Part III: Clinical Applications**

**144 *Reviewing the Effectiveness of Music Interventions in Treating Depression***

Daniel Leubner and Thilo Hinterberger



# Editorial: Music and the Functions of the Brain: Arousal, Emotions, and Pleasure

Mark Reybrouck<sup>1\*</sup>, Tuomas Eerola<sup>2</sup> and Piotr Podlipniak<sup>3</sup>

<sup>1</sup> Musicology Research Unit, KU Leuven, Leuven, Belgium, <sup>2</sup> Department of Music, Durham University, Durham, United Kingdom, <sup>3</sup> Institute of Musicology, Adam Mickiewicz University in Poznan, Poznan, Poland

**Keywords:** music, functions of the brain, arousal, emotions, pleasure-pain principle

## Editorial on the Research Topic

### Music and the Functions of the Brain: Arousal, Emotions, and Pleasure

Music impinges upon the body and the brain and has inductive power, relying on both innate dispositions and acquired mechanisms for coping with the sounds. This process is partly autonomous and partly deliberate, but multiple interrelations between several levels of processing can be shown. There is, further, a tradition in neuroscience that divides the organization of the brain into lower and higher functions. The latter have received a lot of attention in music and brain studies during the last decades. Recent developments in neuroimaging techniques, however, have broadened the field by encompassing the study of both cortical and subcortical processing of the sounds. Much is still to be investigated but some major observations seem already to emerge. The domain of music and emotions is a typical example with a major focus on the pleasure that can be derived from listening to music. Pleasure, however, is not the only emotion that music can induce and the mechanisms behind its elicitation are far from understood. There are also mechanisms related to arousal and activation that are both less differentiated and at the same time more complex than the assumed mechanisms triggering basic emotions. It is tempting, therefore, to bring together contributions from neuroscience studies with a view to cover the possible range of answers to the question what pleasurable or mood-modifying effects music can have on human beings in real-time listening situations.

These questions were the starting point for a special research topic about music and the functions of the brain that was launched simultaneously in *Frontiers in Psychology* and *Neuroscience*. Scientists working on music from separate disciplines such as neuroscience, musicology, comparative musicology, ethology, biology, psychology, evolutionary psychology, and psychoacoustics were invited to submit original empirical research, fresh hypothesis and theory articles, and perspective and opinion pieces reflecting on this topic. Articles of interest could include research themes such as arousal, emotion and affect, musical emotions as core emotions, biological foundation of aesthetic experiences, music-related pleasure and reward centers in the brain, physiological reactions to music, automatically triggered affective reactions to sound and music, emotion and cognition, evolutionary sources of musical sensitivity, affective neuroscience, neuro-affective foundations of musical appreciation, cognition and affect, emotional and motor induction in music, brain stem reflexes to sound and music, activity changes in core emotion networks triggered by music, and potential clinical and medical-therapeutic applications and implications of this knowledge. The response to the call for papers yielded a wealth of proposals with 11 accepted papers by 43 contributing authors. Most of them originate from a neuroscientific orientation with only some contributions from the comparative and ethological approach. The common feature between all contributions was rigorous application of methods and inferences

## OPEN ACCESS

### Edited by:

Isabelle Peretz,  
Université de Montréal, Canada

### \*Correspondence:

Mark Reybrouck  
Mark.Reybrouck@kuleuven.be

### Specialty section:

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

**Received:** 30 December 2017

**Accepted:** 24 January 2018

**Published:** 09 February 2018

### Citation:

Reybrouck M, Eerola T and  
Podlipniak P (2018) Editorial: Music  
and the Functions of the Brain:  
Arousal, Emotions, and Pleasure.  
*Front. Psychol.* 9:113.  
doi: 10.3389/fpsyg.2018.00113

made with empirical data. As a whole, the topic seems to be timely, being exemplary of the increased interest toward music and emotion. There are, however, discrepancies in theories, observations, and approaches, as exemplified in the individual contributions.

This e-book is the outcome of this research topic. The bulk of contributions revolves around the specificity of music experience in terms of perception, emotional reactions, and aesthetic assessment. Since these constituents are also part of the experience of visual art, it seemed to be a fruitful strategy to analyze similarities and differences between these two modalities. As a whole, these studies present new data as well as new hypotheses and theoretical claims which can contribute to a better understanding of the functions of the brain as related to musical experience.

The contributions can be divided roughly in theoretical papers, empirical papers, and one applied paper. As to the theoretical papers, the contribution by Reybrouck and Eerola emphasizes the roots of the emotion induction and expression and provides a synthesis of a hierarchical framework of emotions spanning core affects, basic emotions and aesthetic emotions. Brattico et al. introduce a promising new framework to implement the statistical analysis of global sensory properties used in visual art into neuroaesthetical research of music. Other papers are also illustrative of a positive trend that there is more emphasis to attempt to come up with theories that would cover other domains than music alone. Tiihonen et al. concentrate on the conceptualization of pleasure elicited by music and visual-art in empirical studies. They provide a theoretical synthesis and demonstrate that pleasure is often an ill-defined term which is used differently in research on music and visual arts. Lee et al. compare the aesthetic experiences induced by music and visual stimuli by focusing on the cross-modal perception in the aesthetic experience of emotional visual music. They emphasize the differences in conveying emotional meaning between auditory and visual channel. The empirical papers, on the other hand, make up the bulk of

the contributions. Gorzelańczyk et al. suggest that subcortical structures are involved in processing the syntax of music. Vuoskoski and Eerola promote the view that there are important individual differences in how people experience paradoxical emotions such as pleasure experienced during listening to sad music. Barrett and Schulkin argue on similar lines but stress also the role of granularity in processing emotional contents. Liang et al. highlight how the factors related to musical expertise provide easily measurable differences in pitch discrimination. Kuribayashi and Nittono suggest that there are even more subtle reactions to audio, even to frequencies which are not assumed to carry much relevant information. And Arjmand et al. provide evidence that central markers of emotion, such as frontal asymmetry, are sensitive to high-level musical cues associated with positive affect. The contribution by Leubner and Hinterberger, finally, is the only example of an applied paper, and reviews the extant studies whether music intervention could significantly influence the emotional state of people with depression.

To our pleasure, several of the articles in this e-book have already been accessed thousands of times, indicating a genuine value of the novel angles, ideas, and findings offered within the contributions.

## AUTHOR CONTRIBUTIONS

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

**Conflict of Interest Statement:** 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.

*Copyright © 2018 Reybrouck, Eerola and Podlipniak. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Music and Its Inductive Power: A Psychobiological and Evolutionary Approach to Musical Emotions

Mark Reybrouck<sup>1\*</sup> and Tuomas Eerola<sup>2</sup>

<sup>1</sup> Faculty of Arts, Musicology Research Group, KU Leuven – University of Leuven, Leuven, Belgium, <sup>2</sup> Department of Music, Durham University, Durham, UK

## OPEN ACCESS

### Edited by:

Sonja A. Kotz,  
Maastricht University, Netherlands  
and Max Planck Institute for Human  
Cognitive and Brain Sciences,  
Germany

### Reviewed by:

Mireille Besson,  
Institut de Neurosciences Cognitives  
de la Méditerranée (CNRS), France  
Psyche Loui,  
Wesleyan University, USA

### \*Correspondence:

Mark Reybrouck  
Mark.Reybrouck@kuleuven.be

### Specialty section:

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

**Received:** 22 November 2016

**Accepted:** 16 March 2017

**Published:** 04 April 2017

### Citation:

Reybrouck M and Eerola T (2017)  
Music and Its Inductive Power:  
A Psychobiological and Evolutionary  
Approach to Musical Emotions.  
Front. Psychol. 8:494.  
doi: 10.3389/fpsyg.2017.00494

The aim of this contribution is to broaden the concept of musical meaning from an abstract and emotionally neutral cognitive representation to an emotion-integrating description that is related to the evolutionary approach to music. Starting from the dispositional machinery for dealing with music as a temporal and sounding phenomenon, musical emotions are considered as adaptive responses to be aroused in human beings as the product of neural structures that are specialized for their processing. A theoretical and empirical background is provided in order to bring together the findings of music and emotion studies and the evolutionary approach to musical meaning. The theoretical grounding elaborates on the transition from referential to affective semantics, the distinction between expression and induction of emotions, and the tension between discrete-digital and analog-continuous processing of the sounds. The empirical background provides evidence from several findings such as infant-directed speech, referential emotive vocalizations and separation calls in lower mammals, the distinction between the acoustic and vehicle mode of sound perception, and the bodily and physiological reactions to the sounds. It is argued, finally, that early affective processing reflects the way emotions make our bodies feel, which in turn reflects on the emotions expressed and decoded. As such there is a dynamic tension between nature and nurture, which is reflected in the nature-nurture-nature cycle of musical sense-making.

**Keywords:** induction, emotions, music and evolution, psychobiology, affective semantics, musical sense-making, adaptation

## INTRODUCTION

Music is a powerful tool for emotion induction and mood modulation by triggering ancient evolutionary systems in the human body. The study of the emotional domain, however, is complicated, especially with regard to music (Trainor and Schmidt, 2003; Juslin and Laukka, 2004; Scherer, 2004; Juslin and Västfjäll, 2008; Juslin and Sloboda, 2010; Coutinho and Cangelosi, 2011), due mainly to a lack of descriptive vocabulary and an encompassing theoretical framework. According to Sander, emotion can be defined as “an event-focused, two-step, fast process consisting of (1) relevance-based emotion elicitation mechanisms that (2) shape a multiple emotional response (i.e., action tendency, autonomic reaction, expression, and feeling” (Sander, 2013, p. 23). More in general, there is some consensus that emotion should be viewed as a compound of action tendency, bodily responses, and emotional experience with cognition being considered as part

of the experience component (Scherer, 1993). Emotion, in this view, is a multicomponent entity consisting of subjective experience or feeling, neurophysiological response patterns in the central and autonomous nervous system, and motor expression in face, voice and gestures (see Johnstone and Scherer, 2000 for an overview). These components—often referred to as the *emotional reaction triad*—embrace the evaluation or appraisal of an antecedent event and the action tendencies generated by the emotion. As such, emotion can be considered as a phylogenetically evolved, adaptive mechanism that facilitates the attempt of an organism to cope with important events that affect its well-being (Scherer, 1993). In this view, changes in one of the components are integrated in order to mobilize all resources of an organism and all the systems are coupled to maximize the chances to cope with a challenging environment.

Emotions—and music-induced emotions in particular, —are thus difficult to study adequately and this holds true also for the idiosyncrasies of individual sense-making in music listening. Four major areas, however, have significantly advanced the field: (i) the development of new research methods (continuous, real-time and direct recording of physiological correlates of emotions), (ii) advanced techniques and methods of neuroscience (including fMRI, PET, EEG, EMG and TMS), (iii) theoretical advances such as the distinction between felt and perceived emotions and acknowledgment of various induction mechanisms, and (iv) the adoption of evolutionary accounts. The development of new research methods, in particular, has changed dramatically the field, with seminal contributions from neuropsychology, neurobiology, psychobiology and affective neuroscience. There is, however, still need of a conceptual and theoretical framework that brings all findings together in a coherent way.

In order to address this issue, we organize our review of the field on three broad theoretical frameworks that are indispensable for the topic, namely an evolutionary, embodied and reflective one (see **Figure 1**). Within these frameworks, we focus on the levels and emphasis of the processes involved and connect the types of emotion conceptualizations involved to these frameworks. For instance, the levels of processes are typically divided into low-level and high-level processes, the emphasis of the emotion ranges from recognition to experience of emotion, and the types of emotions involved in these frameworks are usually tightly linked to the levels and emphases. Emotion recognition, e.g., is typically associated with utilitarian emotions, whereas higher level and cognitively mediated reflective emotions that are largely the product of emotion experience might be better conceptualized by aesthetic emotions. The embodied framework does break these dichotomies of high and low and recognition and experience in postulating processes that are flexible, fluid and driven through modality-specific systems that emphasize the interaction between the events offered by the environment, the sensory processes and the acquired competencies for reacting to them in an appropriate fashion.

In what follows, we will start from an evolutionary approach to musical emotions—defining them to some extent as adaptations—, looking thereafter toward the contributions from affective semantics and the embodied framework for explaining

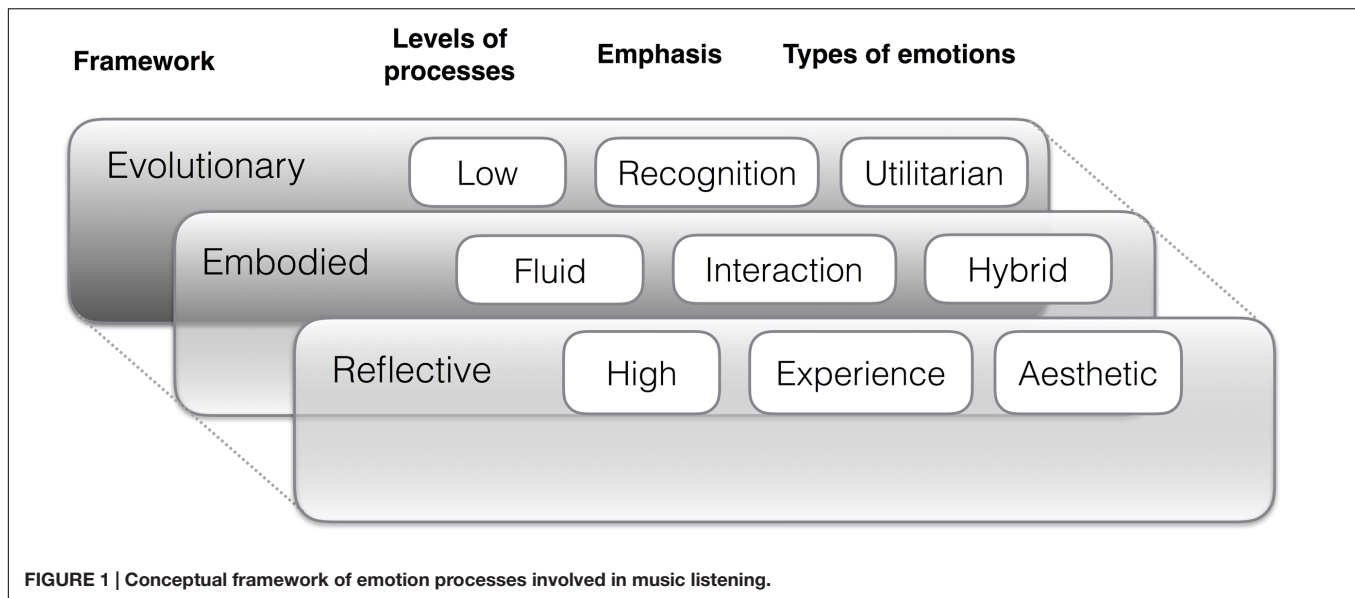
musical emotions from a neuroscientific perspective. We then move onto some psychobiological claims to end with addressing the issue of modulation of emotions by aesthetic experience. In doing so we will look at some conceptual challenges associated with emotions before moving onto emotional meanings in music with the aim to connect experience and meaning-making in the context of emotions to the functions of emotions within an evolutionary perspective. The latter, finally, will be challenged to some extent.

## EVOLUTIONARY CLAIMS: EMOTIONS AS ADAPTATIONS

The neurosciences of music have received a lot of attention in recent research. The *neuroaesthetics* of music, however, remains still somewhat undeveloped as most of the experiments that have been conducted aimed at studying the neural effects on perceptual and cognitive skills rather than on aesthetic or affective judgments (Brattico and Pearce, 2013). Psychology and neuroscience, up to now, have been preoccupied mostly with the *cortico-cognitive systems* of the human brains rather than with *subcortical-affective* ones. Affective consciousness, as a matter of fact, needs to be distinguished from more cognitive forms which generate propositional thoughts about the world. These evolutionary younger cognitive functions add an enormous richness to human emotional life but they neglect the fact that the “energetic” engines for affect are concentrated sub-neocortically. Without these ancestral emotional systems of our brains, music would probably become a less meaningful and desired experience (Panksepp and Bernatzky, 2002; Panksepp, 2005).

In order to motivate these claims, there is need of bottom-up evolutionary, and mainly adaptationist proposals in search of the origins of aesthetic experiences of music, starting from the identification of universal musical features that are observable in all cultures of the world (Brattico et al., 2009–2010). The exquisite sensitivity of our species to emotional sounds, e.g., may function as an example of the survival advantage conferred to operate within small groups and social situations where reading another person's emotional state is of vital importance. This is akin to privileged processing of human faces, which is another highly significant social signal that has been a candidate for evolutionary selection. Processing affective sounds, further, is assumed to be a crucial element for the affective-emotional appreciation of music, which, in this view, can arouse basic emotional circuits at low hierarchical levels of auditory input (Panksepp and Bernatzky, 2002).

Music has been considered from an *evolutionary* perspective in several lines of research, ranging from theoretical discussions (see Brattico et al., 2009–2010; Cross, 2009–2010; Lehman et al., 2009–2010; Livingstone and Thompson, 2009–2010; Honing et al., 2015), to biological (Peretz et al., 2015) and cross-cultural (Trehub et al., 2015), and cross-species evidence (Merchant et al., 2015). Although these various accounts have not fully unpacked the functional role of emotions in the origins of music, certain agreed positions have emerged. For instance, music is conceived as a universal phenomenon with *adaptive*



power (Wallin et al., 2000; Huron, 2003; Justus and Hutsler, 2005; McDermott and Hauser, 2005; Dissanayake, 2008; Cross, 2009–2010). Neuroscientists as LeDoux (1996) and Damasio (1999) have argued that emotions did not evolve as conscious feelings but as adaptive bodily responses that are controlled by the brain. LeDoux (1989, 1996), moreover, has proposed two separate neural pathways that mediate between sensory stimuli and affective responses: a *low road* and a *high road*. The “low road” is the subcortical pathway that transmits emotional stimuli directly to the amygdala—a brain structure that regulates behavioral, autonomic and endocrine responses—by way of connections to the brain stem and motor centers. It bypasses higher cortical areas which may be involved in cognition and consciousness and triggers emotional responses (particularly fear responses) without cognitive mediation. As such, it involves *reactive activity* that is pre-attentive, very fast and automatic, with the “startle response” as the most typical example (Witvliet and Vrana, 1996; Błaszczuk, 2003). Such “primitive” processing has considerable adaptive value for an organism in providing levels of elementary forms of decision making which rely on sets of neural circuits which do the deciding (Damasio, 1994; Lavender and Hommel, 2007). It embraces mainly physiological constants, such as the induction or modification of arousal as well as bodily reactions with a whole range of autonomic reactions. The “high road,” on the contrary, passes through the amygdala to the higher cortical areas. It allows for much more fine-grained processing of stimuli but operates more slowly.

Primitive processing is to be found also in the processing of emotions, which, at their most elementary level, may behave as reflexes in their operation. Occurring with rapid onset, through automatic appraisal and with involuntary changes in physiological and behavioral responses (Peretz, 2001), this level is analogous to the functioning of innate *affect programs* (Griffiths, 1997), which can be assigned to an inherited subcortical structure that can instruct and control a variety of muscles and glands to

respond with unique patterns of activity that are characteristic of a given affect (Tomkins, 1963). Defined in this way, affect programs related to music should be connected to rapid, automatic responses caused by sudden loud sounds (brain stem reflex in the BRECVEMA model, see below). However, a broader interpretation of affect programs as being embodied and embedded in body states and their simulations would put the majority of the emotions into this elementary level (Niedenthal, 2007). In our view, such a broadened embodied view may be a more fruitful way of mapping out the links between the stimuli and emotions than the rather narrow definition of affect programs.

Musically induced emotions, considered at their lowest level, can be conceived partly as *reactive behavior* that points into the direction of automatic processing, involving a lot of biological regulation that engages evolutionary older and less developed structures of the brain. They may have originated as adaptive responses to acoustic input from threatening and non-threatening sounds (Balkwill and Thompson, 1999) which can be considered as quasi-universal reactions to auditory stimuli in general and by extension also to sounding music. Dealing with music, in this view, is to be subsumed under the broader category of “coping with the sounds” (Reybrouck, 2001, 2005). It means also that the notion of musicality, seen exclusively as an evolved trait that is specifically shaped by natural selection, has been questioned to some extent, in the sense that the role of learning and culture have been proposed as possible alternatives (Justus and Hutsler, 2005).

From an *evolutionary* perspective, music has often been viewed as a by-product of natural selection in other cognitive domains, such as, e.g., language, auditory scene analysis, habitat selection, emotion, and motor control (Pinker, 1997; see also Hauser and McDermott, 2003). Music, then, should be merely *exaptive*, which means that is only an evolutionary by-product of the emergence of other capacities that have direct adaptive value.

As such, it should have no role in the survival as a species but should have been derived from an optimal instinctive sensitivity for certain sound patterns, which may have arisen because it proved adaptive for survival (Barrow, 1995). Music, in this view, should have exploited parasitically a capacity that was originally functional in primitive human communication [still evident in speech, note the similarity of affective cues in speech and music (Juslin and Laukka, 2003)] but that fell into disuse with the emergence of finer shades of differentiation in sound pattern that emerged with the emergence of music (Sperber, 1996). As such, processes other than direct adaptation, such as cultural *transmission* and *exaptation*, seem suited to complement the study of biological and evolutionary bases of dealing with music (Tooby and Cosmides, 1992; Justus and Hutsler, 2005, see also below).

A purely *adaptationist* point of view has thus been challenged with regard to music. In a rather narrow description, the notion of adaptation revolves around the concepts of *innate constraint* and *domain specificity*, calling forth also the *modularity approach* to cognition (Fodor, 1983, 1985), which states that some aspects of cognition are performed by mental modules or mechanisms that are specific to the processing of only one kind of information. They are largely innate, fast and unaffected by the content of other representations, and are implemented by specific localizable brain regions. Taken together, such qualities can be referred to as “domain specificity,” “innate constraints,” “information encapsulation” and “brain localization” (see Justus and Hutsler, 2005).

Several attempts have been made to apply the modular approach to the domain of music. It has been shown, e.g., that the representation of pitch in terms of a tonal system can be considered as a module with specialized regions of the cortex (Peretz and Coltheart, 2003). Much of music processing occurs also implicitly and automatically, suggesting some kind of information encapsulation. It can be questioned, however, whether the relevant cortical areas are really domain-specific for music. The concept of modularity, moreover, has been criticized, as different facets of modularity are dissociable with the introduction of the concept of *distributivity* as a possible alternative (Dick et al., 2001). One way in which this dissociation works is the discovery of emergent modules in the sense that predictable regions of the cortex may become informationally encapsulated and/or domain specific, without the outcome having been planned by the genome (Karmiloff-Smith, 1992). The debate concerning the innateness of music processing, however, is not conclusive. A lot of research still has to be done to address the ways in which a domain is innately constrained (Justus and Hutsler, 2005). Most of the efforts, up to now, have concentrated on *perception* and *cognition*, with the importance of octave equivalence and other simple pitch ratios, the categorization of discrete tone categories within the octave, the role of melodic contour, tonal hierarchies and principles of grouping and meter as possible candidate constraints. Music, however, is not merely a cognitive domain but calls forth experiential claims as well, with many connections with the psychobiology and neurophysiology of affection and emotions. *Affective neuroscience* has already extended current knowledge of the emotional brain to some

extent (Davidson and Sutton, 1995; Panksepp, 1998; Sander, 2013), but a lot of work still has to be done.

Dealing with musically induced emotions, further, can be approached from different scales of description: the larger evolutionary scale (*phylogeny*) and the scale of individual human development (*ontogeny*).

An abundance of empirical evidence has been gathered from developmental (newborn studies and infant-directed speech) (Trehub, 2003; Falk, 2009) and comparative research between humans and non-human animals (referential emotive vocalizations and separation calls). It has been shown, e.g., that evolution has given *emotional sound* special time-forms that arise from frequency and amplitude modulation of relatively simple acoustic patterns (Panksepp, 2009–2010). As such, there are means of sound communication in general which are partly shared among living primates and other mammals (Hauser, 1999) and which are the result of brain evolution with the appearance of separate layers that have overgrown the older functions without actually replacing them (Striedter, 2005, 2006). By using sound carriers, humans seem to be able to transmit information such as spatial location, structure of the body, sexual attractiveness, emotional states, cohesion of the group, etc. Some of it is present in all sound messages, but other kinds of information seem to be restricted to specific ways of sound expression (Karpf, 2006). The communicative accuracy of these sets of information, however, has been rarely if at all studied except for emotion states.

This is the case even more for singing, as a primitive way of music realization that was probably previous to any kind of instrumental music making (Geissmann, 2000; Mithen, 2006) and which contains different degrees of motor, emotional and cognitive elements which are universal for us as a species. Generalizing a little, there are special forms of human sound expression that allow communication with other species and reactions to sound stimuli that are similar to those of animals. On the other hand, there seems to be a set of specific sound features belonging exclusive to man—music features such as, e.g., tonality and isometry—, which are strongly connected with emotion expression but which are absent in other kinds of human sound communication (see Gorzelańczyk and Podlipniak, 2011). This is obvious in speech and music and even in some animal vocalizations. The acoustic measures of speech, e.g., can be subdivided into four categories: time-related measures (temporal sequence of different types of sound and silence as carriers of affective information), intensity-related measures (amount of energy in the speech signal), measures related to fundamental frequency ( $F_0$  base level and  $F_0$  range; relative power of fundamental frequency and the harmonics  $F_1$ ,  $F_2$ , etc.), and more complicated time-frequency-energy measures (specific patterns of resonant frequencies such as formants). Three of them are linked to the perceptual dimensions of speech rate, loudness and pitch, the fourth is related to the perceived timbre and voice quality (Johnstone and Scherer, 2000). Taken together, these measures have made it possible to measure the encoding of vocal affect, at least for some commonly studied emotions such as stress, anger, fear, sadness, joy, disgust, and boredom with most consistency in the findings for arousal. The search for emotion-specific acoustic patterns with similar arousal, however, is still

a subject of ongoing research (Banse and Scherer, 1996; Eerola et al., 2013).

## AFFECTIVE SEMANTICS AND THE EMBODIED FRAMEWORK

Music can be considered as a sounding and temporal phenomenon, with the experience of time as a critical factor for musical sense-making. Such an *experiential approach* depends on perceptual bonding and continuous processing of the sound (Reybrouck, 2014, 2015). It can be questioned, in this regard, whether the standard self-report instruments of induced emotions (Eerola and Vuoskoski, 2013) are tapping onto the experiential level or whether that experiential level is inaccessible by such methods, although it may be partially accessible by introspection and verbalization. To address this question, a distinction should be made between the recognition of emotions and the emotions as felt. The former can be considered as a “cognitive-discrete” process which is reducible to categorical assessments of the affective qualia of sounds; the latter calls forth a continuous experience which entails a conception of “music-as-felt” rather than a disembodied approach to musical meaning (Nagel et al., 2007; Schubert, 2013). Though the distinction has received already some attention, there is still need of a conceptual and theoretical framework that brings together current knowledge on perceived and induced emotions in a coherent way. Ways of handling time and experience in music and emotion research up to now have not been neglected (Jones, 1976; Jones and Boltz, 1989) with a significant number of continuous rating studies (Schubert, 2001, 2004), but the study of time has not been the real strength of this research. It can be argued, therefore, that time is not merely an empty perception of duration. It should be considered, on the contrary, as one of the contributing dimensions in the study of emotions in their dynamic form. It calls forth the role of *affective semantics*—a term coined by Molino (2000)—, which aims at describing the meaning of something not in terms of abstract and emotionally neutral cognitive representations, but in a way that is dependent mainly on the integration of emotions (Brown et al., 2004; Menon and Levitin, 2005; Panksepp, 2009–2010). Musical semantics, accordingly, is in search not only of the *lexico-semantic* but also of the *experiential* dimension of meaning, which, in turn, is related to the affective one. Affective semantics, as applied to music, should be able to recognize the emotional meanings which particular sound patterns are trying to convey. It calls forth a continuous rather than a discrete processing of the sounds in order to catch the expressive qualities that vary and change in a dynamic way. Emotional expressions, in fact, are not homogeneous over time, and many of music’s most expressive qualities relate to structural changes over time, somewhat analogous to the concept of *prosodic contours* which is found in vocal expressions (Banse and Scherer, 1996; Scherer, 2003; Belin et al., 2008; Hawk et al., 2009; Sauter et al., 2010; Lima et al., 2013).

The strongest arguments for the introduction of affective semantics in music emotion research come from the

developmental perspective (Trainor and Schmidt, 2003): caregivers around the world sing to infants in an *infant-directed* singing style—using both lullaby and playsong style—which is probably used in order to express emotional information and to regulate the infant’s state. This style—also known as *motherese*—is distinct from other types of singing and young infants are very responsive to it. Additional empirical grounding, moreover, comes from primate vocalizations, which are coined as *referential emotive vocalizations* (Frayer and Nicolay, 2000) and *separation calls* (Newman, 2007). Embracing a body of calls that serve a direct emotive response to some object or events in the environment, they exhibit a dual acoustic nature in having both a referential and emotive meaning (Briefer, 2012).

It is arguable, further, that the affective impact of music could be traced back to similar grounds, being generated by the modulation of sound with a close connection between primitive *emotional dynamics* and the essential *dynamics of music*, both of which appear to be biologically grounded as innate release mechanisms that generate instinctual emotional actions (Burkhardt, 2005; Panksepp, 2009–2010; Coutinho and Cangelosi, 2011). Along with the evolved appreciation of temporal progressions (Clynes and Walker, 1986) they can generate, relive, and communicate emotion intensity, helping to explain why some emotional cues are so easily rendered and recognized through music. This can be seen in the rare cases, where music expressing particular emotions have been exposed to listeners from distinct cultures, at least concerning basic or primary emotions, such as happy, sad, and angry (Balkwill and Thompson, 1999; Fritz et al., 2009). The case seems to be more complicated, however, with regard to secondary or aesthetic emotions such as, e.g., spirituality and longing (Laukka et al., 2013).

As such, there is more to music than the recognition of discrete elements and the way they are related to each other. As important is a description of “music-as-felt,” somewhat analogous to the distinction which has been made between the *vehicle* and the *acoustic mode* of sense-making (Frayer and Nicolay, 2000). The latter refers to particular sound patterns being able to convey emotional meanings by relying on the immediate, on-line emotive aspect of sound perception and production and deals with the emotive interpretation of musical sound patterns; the vehicle mode, on the other hand, involves referential meaning, somewhat analogous to the lexico-semantic dimension of language, with arbitrary sound patterns as vehicles to convey symbolic meaning. It refers to the off-line, referential form of sound perception and production, which is a representational mode of dealing with music that results from the influence of human linguistic capacity on music cognition and which reduces meaning to the perception of “disembodied elements” that are dealt with in a propositional way.

The online form of sound perception—the acoustic mode—is somewhat related to the Clynes’ concept of *sentic modulation* (Clynes, 1977), as a general modulatory system that is involved in conveying and perceiving the intensity of emotive expression by means of three graded spectra of tempo modulation, amplitude modulation, and register selection, somewhat analogous to the well-known rules of prosody. In addition, there is also timbre

as a separate category (Menon et al., 2002; Eerola, 2011), which represents three major dimensions of sounds, namely the temporal (attack time), spectral (spectral energy distribution) and spectro-temporal (spectral flux) (Eerola et al., 2012, p. 49). The very idea of sentic modulation has been taken up in recent studies about emotional expression that is conveyed by *non-verbal vocal expressions*. Examples are the modifications of prosody during expressive speech and non-linguistic vocalizations such as breathing sounds, crying, hums, grunts, laughter, shrieks, and sighs (Juslin and Laukka, 2003; Scherer, 2003; Thompson and Balkwill, 2006; Bryant and Barrett, 2008; Pell et al., 2009; Bryant, 2013) and non-verbal affect vocalizations (Bradley and Lang, 2000; Belin et al., 2008; Redondo et al., 2008; and Reybrouck and Podlipniak, submitted, for an overview). Starting from the observation that the body usually responds physically to an emotion, it can be claimed that physiological responses act as a trigger for appropriate actions with the motor and visceral systems acting as typical manifestations, but other modalities are possible as well. As such, the concept of sentic modulations can be related to Niedenthal's embodied approach to multimodal processing, surpassing the muscles and the viscera in order to focus on modality-specific systems in the brain perception, action and introspection that are fast, refined and flexible. They can even be reactivated without their output being observable in overt behavior with embodiment referring both to actual bodily states and simulations of the modality-specific systems in the brain (Niedenthal et al., 2005; Niedenthal, 2007).

The musical-emotional experience, further, has received much impetus from theoretical contributions and empirical research (Eerola and Vuoskoski, 2013). Impinging upon the body and its physiological correlates, it calls forth an embodied approach to musical emotions which goes beyond the standard cognitivist approach. The latter, based on appraisal, representation and rule-based or information-processing models of cognition, offers rather limited insights of what a musical-emotional experience entails (Schiavio et al., 2016; see also Scherer, 2004 for a critical discussion). Alternative embodied/enactive models of mind—such as the “4E” model of cognition (*embodied, embedded, enactive, and extended*, see Menary, 2010)—have challenged this approach by emphasizing meaning-making as an ongoing process of dynamic interactivity between an organism and its environment (Barrett, 2011; Maiese, 2011; Hutto and Myin, 2013). Relying on the basic concept of “enactivism” as a cross-disciplinary perspective on human cognition that integrates insights from phenomenology and philosophy of mind, cognitive neuroscience, theoretical biology, and developmental and social psychology (Varela et al., 1991; Thompson, 2007; Stewart et al., 2010), enactive models understand cognition as embodied and perceptually guided activity that is constituted by circular interactions between an organism and its environment. Through continuous sensorimotor loops (defined by real-time perception/action cycles), the living organism—including the music listener/performer—enacts or brings forth his/her own domain of meaning (Reybrouck, 2005; Thompson, 2005; Colombetti and Thompson, 2008) without separation between the cognitive states of the organism, its physiology, and the environment in which it is embedded. Cognition and mind,

in this view, originate in a continuous interplay between an organism and its environment as an evolving dynamic system (Hurley, 1998).

Starting from the observation that the body usually responds physically to an emotion, it can be claimed, further, that physiological responses act as a trigger for appropriate actions with the motor and visceral systems acting as typical manifestations. Other modalities, however, are possible as well, as exemplified in Niedenthal's embodied approach to multimodal processing, surpassing the muscles and the viscera in order to focus on modality-specific systems in the brain—perception, action and introspection—that are fast, refined and flexible. They can even be reactivated without their output being observable in overt behavior. Embodiment, then, is referring both to actual bodily states or simulations of the modality-specific systems in the brain (Niedenthal et al., 2005; Niedenthal, 2007).

## INDUCTION OF EMOTIONS: PSYCHOBIOLOGICAL CLAIMS

Music may be considered as something that catches us and that induces several reactions beyond conscious control. As such, it calls forth a deeper affective domain to which cognition is subservient, and which makes the brains such receptive vessels for the emotional power of music (Panksepp and Bernatzky, 2002). The auditory system, in fact, evolved phylogenetically from the *vestibular system*, which contains a substantial number of acoustically responsive fibers (Koelsch, 2014). It is sensitive to sounds and vibrations—especially those of loud sounds with low frequencies or with sudden onsets—and projects to the reticular formation and the parabrachial nucleus, which is a convergence site for vestibular, visceral and autonomic processing. As such, subcortical processing of sounds gives rise not only to auditory sensations but also to muscular and autonomic responses. It has been shown, moreover, that intense hedonic experiences of sound and pleasurable aesthetic responses to music are reflected in the listeners' autonomic and central nervous systems, as evidenced by objective measurements with polygraph, EEG, PET or fMRI (Brattico et al., 2009–2010). Though these measures do not always differentiate between specific emotions, they indicate that the *reward system* can be heavily activated by music (Blood and Zatorre, 2001; Salimpoor et al., 2015). But other brain structures can be activated as well, more particularly those brain structures that are crucially involved in emotion, such as the amygdala, the nucleus accumbens, the hypothalamus, the hippocampus, the insula, the cingulate cortex and the orbitofrontal cortex (Koelsch, 2014).

Emotional reactions to music, further, activate the same cortical, subcortical and autonomic circuits, which are considered as the essential survival circuits of biological organisms in general (Blood and Zatorre, 2001; Trainor and Schmidt, 2003; Salimpoor et al., 2015). The subcortical processing affects the body through the basic mechanisms of chemical release in the blood and the spread of neural activation. The latter, especially, invites listeners to react bodily to music with a whole bunch of autonomic reactions such as changes in heart rate, respiration rate, blood

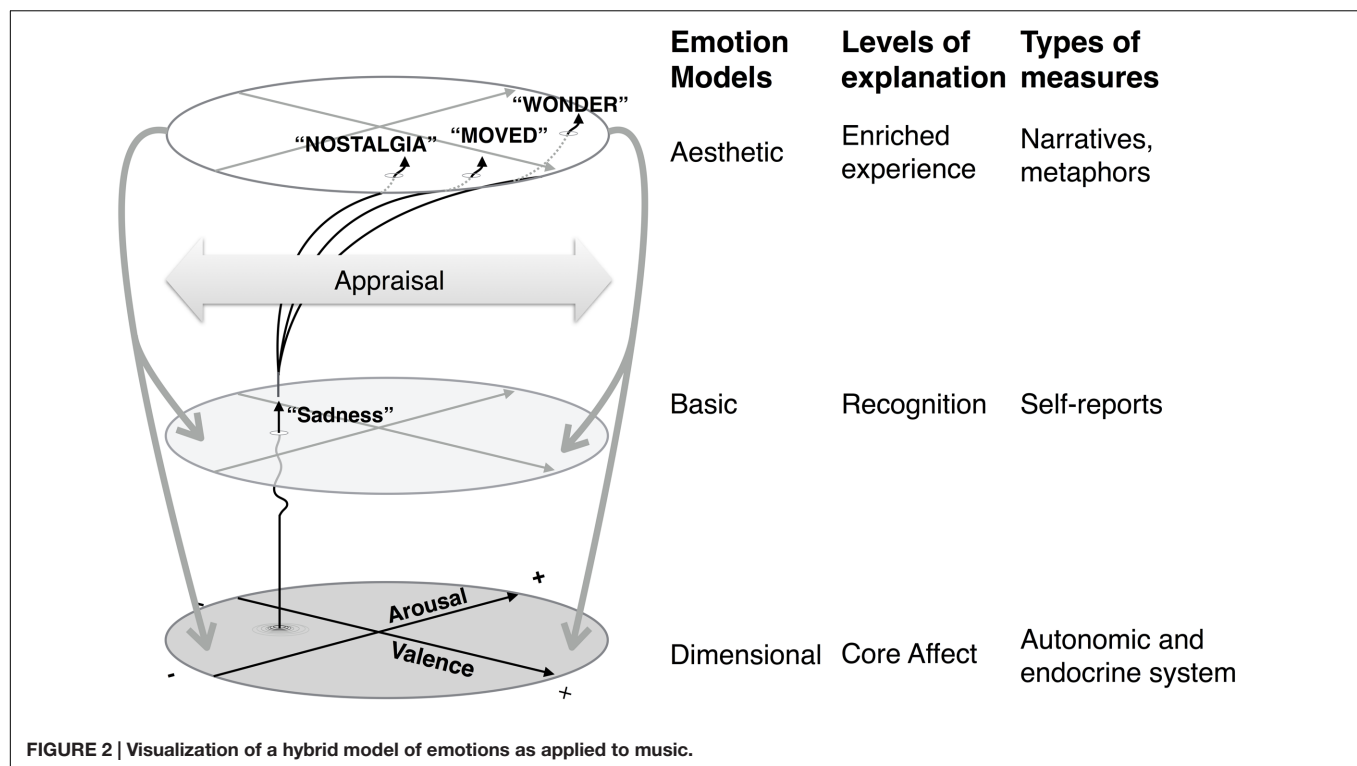
flow, skin conductance, brain activation patterns, and hormone release (oxytocin, testosterone), all driven by the phylogenetically older parts of the nervous system (Ellis and Thayer, 2010). These reactions can be considered the “physiological correlates” of listening (see Levenson, 2003, for a general review), but the question remains whether such measures provide sufficient detailed information to distinguish musically induced physiological reactions from mere physiological reactions to emotional stimuli in general (Lundqvist et al., 2009). Recent physiological studies have shown that pieces of music that express different emotions may actually produce distinct physiological reactions in listeners (see Juslin and Laukka, 2004 for a critical review). It has been shown also that performers are able to communicate at least five emotions (happiness, anger, sadness, fear, tenderness) with this proviso that this communication operates in terms of broader emotional categories than the finer distinctions which are possible within these categories (Juslin and Laukka, 2003). Precision of communication, however, is not a primary criterion by which listeners value music and reliability is often compromised for the sake of other musical characteristics. Physiological measures may thus be important, but establishing clear-cut and consistent relationships between emotions and their physiological correlates remains difficult, though some studies have received some success in the case of few basic emotions (Juslin and Laukka, 2004; Lundqvist et al., 2009).

Music thus has *inductive power*. It engenders physiological responses, which are triggered by the central nervous system and which are proportional to the way the information has been received, analyzed and interpreted through instinctive, emotional pathways that are ultimately concerned with maintaining an internal environment that ensures survival (Schneck and Berger, 2010). Such dynamically equilibrated and delicately balanced internal milieu (*homeostasis*), together with the physiological processes which maintain it, relies on finely tuned control mechanisms that keep the body operating as closely as possible to predetermined baseline physiological quantities or *reference set-points* (blood pressure, pulse rate, breathing rate, body temperature, blood sugar level, pH, fluid balance, etc.). Sensory stimulation of all kinds can change and disturb this equilibrium and invite the organism to adapt these basic reference points, mostly after persisting and continuous disturbances that act as environmental or driving forces to which the organism must adapt. There are, however, also short term immediate reactions to the music as a driving force, as evidenced from neurobiological and psychobiological research that revolves around the central axiom of *psychobiological equivalence* between percepts, experience and thought (Reybrouck, 2013). This axiom addresses the central question whether there is some lawfulness in the coordination between sounding stimuli and the responses of music listeners in general. A lot of empirical support has been collected from studies of *psychophysical dimensions* of music as well as *physiological reactions* that have shown to be their correlates (Peretz, 2001, 2006; Scherer and Zentner, 2001; Menon and Levitin, 2005; van der Zwaag et al., 2011). Psychophysical dimensions, as considered in a musical context, can be defined as any property of sound that can be perceived independently of musical experience, knowledge, or enculturation, such as, e.g.,

speed of pulse or tempo. A distinction should be made, however, between the *psychophysics* of perception and the *psychobiology* of the bodily reactions to the sounds. The psychophysics features suggest a reliable correlation between acoustic signals and their perceptual processing, with a special emphasis on the study of how individual features of music contribute to its emotional expression, embracing psychoacoustic features such as loudness, roughness and timbre (Eerola et al., 2012). The psychobiological claims, on the other hand, are still subject of ongoing research. Some of them can be subsumed under the sensations of *peak experience*, *flow* and *shivers* or *chills* (Panksepp and Bernatzky, 2002; Grewe et al., 2007; Harrison and Loui, 2014) as evidence for particularly strong emotional experiences with music (Gabrielsson and Lindström, 2003; Gabrielsson, 2010). Such intensely pleasurable experiences are straightforward to be recorded behaviorally and have the additional advantage of producing characteristic physiological markers including changes in heart rate, respiration amplitude, and skin conductance (e.g., Blood and Zatorre, 2001; Sachs et al., 2016). They are associated mainly with changes in the autonomic nervous system and with metabolic activity in the cerebral regions, such as ventral striatum, amygdala, insula, and midbrain, usually devoted to motivation, emotion, arousal, and reward (Blood and Zatorre, 2001). Their association with subcortical structures indicates also their possible association with ancestral behavioral patterns of the prehistoric individual, making them relevant for the evaluation of the evolutionary hypothesis on the origin of aesthetic experience of music (Brattico et al., 2009–2010). Such peak experiences, however, are rather rare and should not be taken as the main starting point for a generic comparative perspective on musical emotions. Some broader *vitality effects*, such as those exemplified in the relations between personal feelings and the dynamics of infant's movements and the sympathetic responses by their caregivers in a kind of mutual attunement (Stern, 1985, 1999; see also Malloch and Trevarthen, 2009), as well as the creation of tensions and expectancies may engender also some music-specific emotional reactions. The general assumption, then, is that musically evoked reactions emerge from “presemantic acoustic dynamics” that evolved in ancient times, but that still interact with the intrinsic emotional systems of our brains (Panksepp, 1995, p. 172).

## AN INTEGRATED FRAMEWORK OF MUSIC EMOTIONS AND THEIR UNDERLYING MECHANISMS

What are these presemantic acoustic dynamics? Here we should make a distinction between the *structural features* of the music which induce emotions and their *underlying mechanisms*. As to the first, musical cues such as mode, followed by tempo, register, dynamics, articulation, and timbre (Eerola et al., 2013) seem to be important, at least in Western music. Increases in perceived complexity, moreover, has been shown also to evoke arousal (Balkwill and Thompson, 1999). Being grounded in the dispositional machinery of individual music users these features may function as universal cues for the



emotional evaluation of auditory stimuli in general. Much more research, however, is needed in order to trace their underlying mechanisms. A major attempt has been made already by Juslin and Västfjäll (2008) and Liljeström et al. (2013) who present a framework that embraces eight basic mechanisms (brain stem reflexes, rhythmic entrainment, evaluative conditioning, emotional contagion, visual imagery, episodic memory, musical expectancy and aesthetic judgment—commonly referred to as BRECVEMA). In addition to these mechanisms, an *integrated framework* has been proposed also by Eerola (2017), with low-level measurable properties being capable of producing highly different higher-level conceptual interpretations (see Figure 2). Its underlying machinery is best described in dimensional terms (core affects as valence and arousal) but conscious interpretations can be superposed on them, allowing a categorical approach that relies on higher-level conceptual categories as well. As such, the model can be considered a hybrid model that builds on these existing emotion models and attempts to clarify the levels of explanations of emotions and the typical measures related to these layers of explanations. Although this is a simplification of a complex process, the purpose is to emphasize the disparate conceptual issues brought under the focus at each different level, which is a notion put forward in the past (e.g., Leventhal and Scherer, 1987). The types of measures of emotions alluded to in the model are not merely alternative instruments but profoundly different ontological stances which capture biological reductionism (all physiological responses), psychological (all behavioral responses including self-reports) and phenomenological (various experiential including narratives and metaphors) perspectives.

The dimensional perspective on emotions has fostered already a long program of research with objectless dimensions such as pleasure–displeasure (pleasure or valence) and activation–deactivation (arousal or energy). Their combination—called *core affect*—can be considered as a first primitive that is involved in most psychological events and makes them “hot” or emotional. Involving a pre-conceptual process, a neurophysiological state, core affect is accessible to consciousness as a simple non-reflective feeling, e.g., feeling good or bad, feeling lethargic or energized. Perception of the affective quality is the second primitive. It is a “cold” process which is made hot by being combined with a change in core affect (Russell, 2003, 2009).

The dimensional approach has been challenged to some extent. Eerola’s *hybrid model* (Eerola, 2017) assigns three explanatory levels of affects, starting from low level sensed emotions (core affect), proceeding over perceived or recognized emotions (basic emotions), and ending with experienced and felt emotions (high-level complex emotions). It takes as the lowest level core affect, as a neurophysiological state which is accessible to consciousness as a simple primitive non-reflective feeling (Russell and Barrett, 1999). It reflects the idea that affects arise from the core of the body and neural representations of the body state. The next higher level organizes emotions by conceiving of them in terms of discrete categories such as fear, anger, disgust, sadness, and surprise (Matsumoto and Ekman, 2009; and Sander, 2013 for a discussion of number and label of the categories). Both levels have furthered an abundance of theoretical and empirical research with a focus on the development of emotion taxonomies which all offer distinct

ways to tackle musical emotions. Both the dimensional and basic emotions model, however, seem to overlap considerably, and this holds true especially for artworks and objects in nature (Eerola and Vuoskoski, 2011) which are not always explained in terms of dimensions or discrete patterns of emotions that are involved in everyday survival (Sander, 2013). As such, there is also a level beyond core affects and the perception of basic emotions which is not reducible to mere reactions to the environment, and that encompasses complex emotions that are more contemplative, reflected and nuanced, somewhat analogous to other complex emotions such as moral, social and epistemic ones (see below).

While such a hybrid model may reconcile some of the discrepancies in the field, its main contribution is to make us aware of how the conceptual level of emotions under the focus lends itself to different mechanisms, emotion labels and useful measures. The shortcoming of the model is an impression that it offers a way to reduce complex, aesthetic emotions into simpler basic emotions and the latter into underlying core affects. Whilst some of such trajectories could be traced from the lowest to highest level (i.e., measurement of core affects via psychophysiology, recognition of the emotions expressed, and reflection of what kind of experience the whole process induces in the perceiver), it is fundamentally not a symmetrical and reversible process. One cannot reduce the experience of longing (a complex, aesthetic emotion) into recognition of combination of basic emotions nor predict the exact core affects related to such emotional experience. At best, one level may *modulate* the processes taking place in the lower levels (as depicted with the downward arrows in **Figure 2**). The extent of such top-down influence has not received sufficient attention to date, although top-down information such as extramusical information has been demonstrated to impact music-induced emotions (Vuoskoski and Eerola, 2015). However, such top-down effects on perception are well known in perceptual literature (Rahman and Sommer, 2008) and provide evidence against strictly modular framework. Despite this shortcoming, the hybrid model does organize the range of processes in a functional manner.

## EMOTIONS MODULATED BY AESTHETIC EXPERIENCE

In what preceded we have emphasized the bottom-up approach to musically induced emotions, taking as a starting point that affective experience may reflect an evolutionary primitive form of consciousness above which more complex layers of consciousness can emerge (Panksepp, 2005). Many higher neural systems are in fact involved in the various distinct aspects of experiencing and recognizing musical emotions, but a great deal of the emotional power may be generated by lower subcortical regions where basic affective states are organized (Panksepp, 1998; Damasio, 1999; Panksepp and Bernatzky, 2002). This lower level processing, however, can be modified to some extent by other variables such as

repeated encounters with the stimulus—going from mere exposure, over habituation and sensitization—, co-occurrence with other stimuli (classical and evaluative conditioning) and varying internal states such as, e.g., motivation (Moors, 2007, p. 1241).

A real aesthetic experience of music, moreover, can be defined as an experience “in which the individual immerses herself in the music, dedicating her attention to perceptual, cognitive, and affective interpretation based on the formal properties of the perceptual experience” (Brattico and Pearce, 2013, p. 49). This means that several mechanisms may be used for the processing, elicitation, and experience of emotions (Storbeck and Clore, 2007).

Musical sense-making, in this view, has to be broadened from a mere cognitive to a more encompassing approach that includes affective semantics and embodied cognition. What really counts in this regard, is the difficult relationship between *emotion* and *cognition* (Panksepp, 2009–2010). Cognition, regarded in a narrow account, is contrasted mainly with emotion and cognitive output is defined as information that is not related to emotion. It is coined “cold” as contrasted with “hot” affective information processing (Eder et al., 2007). Recent neuroanatomic studies, however, seem to increasingly challenge the idea of specialized brain structures for cognition versus emotion (Storbeck and Clore, 2007), and there is also no easy separation between cognitive and emotional components insofar as the functions of these areas are concerned (Ishizu and Zeki, 2014). Some popular ideas about cognition and emotion such as *affective independence*, *affective primacy* and *affective automaticity* have been questioned accordingly (Storbeck and Clore, 2007, pp. 1225–1226): the affective independence hypothesis states that emotion is processed independently of cognition via a subcortical low route; affective primacy claims precedence of affective and evaluative processing over semantic processing, and affective automaticity states that affective processes are triggered automatically by affectively potent stimuli commandeering attention. A more recent view, however, is the suggestion that affect modifies and regulates cognitive processing rather than being processed independently. Affect, in this view, probably does not proceed independently of cognition, nor does it precede cognition in time. (Storbeck and Clore, 2007, pp. 1225–1226).

As such, there is some kind of overlap between music-evoked complex and/or “aesthetic emotions” and so-called “everyday emotions” (Koelsch, 2014). Examples of the latter are anger, disgust, fear, enjoyment, sadness, and surprise (see Matsumoto and Ekman, 2009). They are mainly reducible to the basic emotions—also called “primary,” “discrete” or “fundamental” emotions—which have been elaborated in several taxonomies. Examples of the former are wonder, nostalgia, transcendence (see Zentner et al., 2008; Trost et al., 2012; Taruffi and Koelsch, 2014). They are typically elicited when people engage with artworks (including music) and objects or scenes in nature (Robinson, 2009; see Sander, 2013 for an overview) and can be related to “epistemic emotions” such as interest, confusion, surprise or awe (de Sousa, 2008) though the latter have not yet been the focus of much research in affective neuroscience. As explained in the hybrid model (Eerola, 2017), however,

they tend to be rare, less stable and more reliant on the various other factors related to meaning-generation in music (Vuoskoski and Eerola, 2012). Related topics, such as novelty processing, have been investigated extensively—with a key role for the function of the amygdala—as well as the role emotions, which are not directed at knowing, can have for epistemic consequences. Fear, for instance, can lead to an increase in vigilance and attention with better knowledge of the situation in order to evaluate the possibilities for escape (Sander, 2013).

The everyday/aesthetic dichotomy, further, is related also to the distinction between *utilitarian* and *aesthetic* emotions (Scherer and Zentner, 2008). The latter occur in situations that do not trigger self-interest or goal-directed action and reflect a multiplicative function of structural features of the music, listener features, performer features and contextual features leading to distinct kinds of emotion such as wonder, transcendence, entrainment, tension and awe (Zentner et al., 2008). It is possible, however, to combine aesthetic and non-aesthetic emotions when asked to describe retrospectively felt and expressed musical emotions. As such, nine factors have been described—commonly known as the *Geneva Emotional Music Scale* or GEMS (see Zentner et al., 2008), namely wonder, transcendence, tenderness, nostalgia, peacefulness, power, joy, tension and sadness. Awe, nostalgia, and enjoyment, among the aesthetic emotions, have attracted the most detailed research with aesthetic awe being crucial in distinguishing a peak aesthetic experience of music from everyday casual listening (Gabrielsson, 2010; Brattico and Pearce, 2013, p. 51), although studies that induce a range of emotions in laboratory conditions may fail to arouse the special emotions such as awe, wonder and transcendence (Vuoskoski and Eerola, 2011).

## CONCLUSION AND PERSPECTIVES: NATURE MEETS NURTURE

In this paper, we explored the evolutionary groundings of music-induced emotions. Starting from a definition of emotions as adaptive processes we tried to show that music-induced emotions reflect ancient brain functions. The inductive power of such functions, however, can be expanded or even overruled to some extent by the evolutionary younger regions of the brain. The issue whether an emotional modulation of sensory input is “top-down” and dependent upon input from “higher” areas of the brain or whether it is “bottom-up,” or both, is up to now an unresolved question (Ishizu and Zeki, 2014). Affect and cognition, in fact, have long been treated as independent domains, but current evidence seems to suggest that both are in fact highly interdependent (Storbeck and Clore, 2007). Although we may never know with certainty “the evolutionary and cultural transitions that led from our acoustic-emotional sensibilities to an appreciation of music” it may be suspected that the role of subcortical systems in the way we are affected by music has been greatly underestimated (Panksepp and Bernatzky, 2002, p. 151). Music establishes

affective resonances within the brain, and it is within an understanding of the ingrained emotional processes of the mammalian brain that the essential answers to these questions will be found, which could imply that affective sounds are related to primitive reactions with adaptive power and that somehow music capitalizes on these reactive mechanisms. In this view, early affective processing—as relevant in early infancy and prehistory—, should reflect the way the emotions make our bodies feel, which in turn reflects on the emotions expressed and decoded.

Music-induced emotions, moreover, have recently received considerable impetus from neurobiological and psychobiological research. The full mechanisms behind the proposed induction mechanisms, however, are not yet totally clear. Emotional processing holds a hybrid position: it is the place where nature meets nurture with emotive meaning relying both on pre-programmed reactivity that is based on wired-in circuitry for perceptual information pickup (*nature*) and on culturally established mechanisms for information processing and sense-making (*nurture*). It makes sense, therefore, to look for mechanisms that underlie the inductive power of the music and to relate them with evolutionary claims and a possible adaptive function of music. Especially important here is the distinction between the acoustic and the vehicle mode of listening and the related distinction between the on-line and off-line mode of listening. Much more research, however, is needed in order to investigate the relationship between music-specific or aesthetic emotions and everyday or utilitarian emotions (Scherer and Zentner, 2008; Reybrouck and Brattico, 2015). The latter are triggered by the need to adapt to specific situations that are of central significance to the individual’s interests and well-being; the former are triggered in situations that usually have no obvious material effect on the individual’s well-being. Rather than relying on categorical models of emotion by blurring the boundaries between aesthetic and utilitarian emotions we should take care to reflect also the nuanced range of emotive states, that music can induce. As such, there should be a dynamic tension between the “nature” and the “nurture” side of music processing, stressing the role of the musical experience proper. Music, in fact, is a sounding and temporal phenomenon which has inductive power. The latter involves ongoing epistemic interactions with the sounds, which rely on low-level sensory processing as well as on principles of cognitive mediation. The former, obviously, refer to the nature side, the latter to the nurture side of music processing. Cognitive processing, however, should take into account also the full richness of the sensory experience. What we argue for, therefore, is the reliance on the nature side again, which ends up, finally, in what may be called a “*nature-nurture-nature cycle*” of musical sense-making, starting with low-level processing, over cognitive mediation and revaluing the sensory experience as well (Reybrouck, 2008).

## AUTHOR CONTRIBUTIONS

The first draft of this paper was written by MR. The final elaboration was written jointly by MR and TE.

## REFERENCES

- Balkwill, L.-L., and Thompson, W. (1999). A cross-cultural investigation of the perception of emotion in music: psychophysical and cultural cues. *Music Percept.* 17, 43–64. doi: 10.2307/40285811
- Banase, R., and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70, 614–636. doi: 10.1037/0022-3514.70.3.614
- Barrett, L. (2011). *Beyond the Brain: How the Body and Environment Shape Animal and Human Minds*. Princeton, NJ: Princeton University Press.
- Barrow, J. D. (1995). *The Artful Universe*. Oxford: Clarendon Press.
- Belin, P., Fillion-Bilodeau, S., and Gosselin, F. (2008). The montreal affective voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behav. Res. Methods* 40, 531–539. doi: 10.3758/BRM.40.2.531
- Błaszczak, J. (2003). Startle response to short acoustic stimuli in rats. *Acta Neurobiol. Exp.* 63, 25–30.
- 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
- Bradley, M. M., and Lang, P. (2000). Affective reactions to acoustic stimuli. *Psychophysiology* 37, 204–215. doi: 10.1111/1469-8986.3720204
- Brattico, E., Brattico, P., and Jacobsen, T. (2009–2010). The origins of the aesthetic enjoyment of music – A review of the literature. *Musicae Sci.* 13, 15–39. doi: 10.1177/1029864909013002031
- Brattico, E., and Pearce, M. (2013). The neuroaesthetics of music. *Psychol. Aesthet. Creat. Arts* 7, 48–61. doi: 10.1037/a0031624
- Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *J. Zool.* 288, 1–20. doi: 10.1111/j.1469-7998.2012.00920.x
- Brown, S., Martinez, M., and Parsons, L. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. *Neuroreport* 15, 2033–2037. doi: 10.1097/00001756-200409150-00008
- Bryant, G. (2013). Animal signals and emotion in music: coordinating affect across groups. *Front. Psychol.* 4:990. doi: 10.3389/fpsyg.2013.00990
- Bryant, G. A., and Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *J. Cogn. Cult.* 8, 135–148. doi: 10.1163/156770908X289242
- Burkhardt, R. W. (2005). *Patterns of Behavior*. Chicago, IL: University of Chicago Press.
- Clynes, M. (1977). *Sentics: The Touch of Emotion*. London: Souvenir Press.
- Clynes, M., and Walker, J. (1986). Music as time's measure. *Music Percept.* 4, 85–120. doi: 10.2307/40285353
- Colombetti, G., and Thompson, E. (2008). “The feeling body: toward an enactive approach to emotion,” in *Developmental Perspectives on Embodiment and Consciousness*, eds W. F. Overton, U. Muller, and J. L. Newman (New York, NY: Lawrence Erlbaum Ass.), 45–68.
- Coutinho, E., and Cangelosi, A. (2011). Musical emotions: predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. *Emotion* 11, 921–937. doi: 10.1037/a0024700
- Cross, I. (2009–2010). The evolutionary nature of musical meaning. *Musicae Sci.* 13, 179–200. doi: 10.1177/1029864909013002091
- Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Harper Collins.
- Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York, NY: Harcourt.
- Davidson, R., and Sutton, S. (1995). Affective neuroscience: the emergence of a discipline. *Curr. Opin. Neurobiol.* 5, 217–224. doi: 10.1016/0959-4388(95)80029-8
- de Sousa, R. (2008). “Epistemic feelings,” in *Epistemology and Emotions*, eds G. Brun, U. Doguoglu, and D. Kuenzle (Surrey: Ashgate), 185–204.
- Dick, F., Bates, E., Wulfek, B., Utman, J. A., Dronkers, N., and Gernsbacher, M. (2001). Language deficits, localization, and grammar: evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychol. Rev.* 108, 759–788. doi: 10.1037/0033-295X.108.4.759
- Dissanayake, E. (2008). If music is the food of love, what about survival and reproductive success? *Musicae Sci.* 2008, 169–195. doi: 10.1177/1029864908012001081
- Eder, A., Hommel, B., and De Houwer, J. (2007). How distinctive is affective processing? On the implications of using cognitive paradigms to study affect and emotion. *Cogn. Emot.* 21, 1137–1154.
- Eerola, T. (2011). Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres. *J. New Music Res.* 40, 349–366. doi: 10.1080/0929215.2011.602195
- Eerola, T. (2017). “Music and emotions,” in *Handbook of Systematic Musicology*, ed. S. Koelsch (Berlin: Springer).
- Eerola, T., Ferrer, R., and Alluri, V. (2012). Timbre and affect dimensions: evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Percept. Interdiscip. J.* 30, 49–70. doi: 10.1525/mp.2012.30.1.4
- Eerola, T., Friberg, A., and Bresin, R. (2013). Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Front. Psychol.* 4:487. doi: 10.3389/fpsyg.2013.00487
- Eerola, T., and Vuoskoski, J. (2013). A review of music and emotion studies: approaches, emotion models, and stimuli. *Music Percept.* 30, 307–340. doi: 10.1525/MP.2012.30.3.307
- Eerola, T., and Vuoskoski, K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychol. Music* 39, 18–49. doi: 10.1093/scan/nsv032
- Ellis, R., and Thayer, J. (2010). Music and autonomic nervous system (Dys)function. *Music Percept.* 27, 317–326.
- Falk, D. (2009). *Finding Our Tongues: Mothers, Infants and the Origins of Language*. New York, NY: Basic Books.
- Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press. doi: 10.1017/S0140525X0001921X
- Fodor, J. A. (1985). Précis of the modularity of mind. *Behav. Brain Sci.* 8, 1–42. doi: 10.1017/S0140525X0001921X
- Freyer, D. W., and Nicolay, C. (2000). “Fossil evidence for the origin of speech sounds,” in *The Origins of Music*, eds N. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press), 271–300.
- Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., et al. (2009). Universal recognition of three basic emotions in music. *Curr. Biol.* 19, 573–576. doi: 10.1016/j.cub.2009.02.058
- Gabrielsson, A. (2010). “Strong experiences with music,” in *Handbook of Music and Emotion. Theory, Research, Applications*, eds P. Juslin and J. Sloboda (Oxford: Oxford University Press), 547–574.
- Gabrielsson, A., and Lindström, S. (2003). Strong experiences related to music: a descriptive system. *Musicae Sci.* 7, 157–217. doi: 10.1177/102986490300700201
- Geissmann, T. (2000). “Gibbon songs and human music from an evolutionary perspective,” in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 103–123.
- Gorzałęńczyk, E. J., and Podlipniak, P. (2011). Human singing as a form of bio-communication. *Bio-Algorithms Med-systems* 7, 79–83. doi: 10.7208/chicago/9780226308760.001.0001
- Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2007). Listening to music as a re-creative process: physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Percept.* 24, 297–314. doi: 10.1525/mp.2007.24.3.297
- Griffiths, P. (1997). *What Emotions Really Are: The Problem of Psychological Categories*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226308760.001.0001
- Harrison, L., and Loui, P. (2014). Thrills, chills, frissons, and skin orgasms: toward an integrative model of transcendent psychophysiological experiences in music. *Front. Psychol.* 5:790. doi: 10.3389/fpsyg.2014.00790
- Hauser, M., and McDermott, J. (2003). The evolution of the music faculty: a comparative perspective. *Nat. Neurosci.* 6, 663–668. doi: 10.1038/nn1080
- Hauser, M. D. (1999). “The sound and the fury: primate vocalizations as reflections of emotion and thought,” in *The Origins of Music*, eds N. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press), 77–102.
- Hawk, S. T., van Kleef, G. A., Fischer, A. H., and van der Schalk, J. (2009). “Worth a thousand words”: absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion* 9, 293–305. doi: 10.1037/a0015178
- Honing, H., Cate, C., Peretz, I., and Trehub, S. E. (2015). Without it no music: cognition, biology and evolution of musicality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140088. doi: 10.1098/rstb.2014.0088
- Hurley, S. (1998). *Consciousness in Action*. London: Harvard University Press.

- Huron, D. (2003). "Is music an evolutionary adaptation?" in *The Cognitive Neuroscience of Music*, eds I. Peretz and R. Zatorre (Oxford: Oxford University Press), 57–75. doi: 10.1093/acprof:oso/9780198525202.003.0005
- Hutto, D., and Myin, E. (2013). *Radicalizing Enactivism. Basic Minds without Content*. Cambridge, MA: The MIT Press.
- Ishizu, T., and Zeki, S. (2014). A neurobiological enquiry into the origins of our experience of the sublime and the beautiful. *Front. Hum. Neurosci.* 8:891. doi: 10.3389/fnhum.2014.00891
- Johnstone, T., and Scherer, K. (2000). "Vocal communication of emotion," in *Handbook of Emotions*, eds M. Lewis and J. Haviland-Jones (New York, NY: The Guilford Press), 220–235.
- Jones, M. R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychol. Rev.* 83, 323–355. doi: 10.1037/0033-295X.83.5.323
- Jones, M. R., and Boltz, M. (1989). Dynamic attending and responding to time. *Psychol. Rev.* 96, 459–491. doi: 10.1037/0033-295X.96.3.459
- Juslin, P., and Laukka, P. (2004). Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. *J. New Music Res.* 33, 217–238. doi: 10.1080/0929821042000317813
- Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770
- Juslin, P. N., and Sloboda, J. A. (eds). (2010). *Music and Emotion. Theory, Research, Applications*. Oxford: Oxford University Press.
- Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music. *Behav. Brain Sci.* 5, 559–575.
- Justus, T., and Hutsler, J. (2005). Fundamental issues in the evolutionary psychology of music. Assessing innateness and domain specificity. *Music Percept.* 23, 1–27. doi: 10.4436/JASS.92008
- Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.
- Karpf, A. (2006). *The Human Voice*. New York, NY: Bloomsbury Publishing.
- Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–183. doi: 10.1038/nrn3666
- Laukka, P., Anger Elfenbein, H., Söder, N., Nordström, H., Althoff, J., Chui, W., et al. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Front. Psychol.* 4:353. doi: 10.3389/fpsyg.2013.00353
- Lavender, T., and Hommel, B. (2007). Affect and action: towards an event-coding account. *Cogn. Emot.* 21, 1270–1296. doi: 10.1080/02699930701438152
- LeDoux, J. E. (1989). Cognitive-emotional interactions in the brain. *Cogn. Emot.* 3, 267–289. doi: 10.1080/02699938908412709
- LeDoux, J. E. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York, NY: Simon and Schuster.
- Lehman, C., Welker, L., and Schiefenhövel, W. (2009–2010). Towards an ethology of song: a categorization of musical behaviour. *Musicae Sci.* 321–338.
- Levenson, R. (2003). "Autonomic specificity and emotion," in *Handbook of Affective Sciences*, eds R. Davidson, K. Scherer, and H. Goldsmith (New York, NY: Oxford University Press), 212–224.
- Leventhal, H., and Scherer, K. (1987). The relationship of emotion to cognition: a functional approach to a semantic controversy. *Cogn. Emot.* 1, 3–28. doi: 10.1080/02699938708408361
- Liljeström, S., Juslin, P. N., and Västfjäll, D. (2013). Experimental evidence of the roles of music choice, social context, and listener personality in emotional reactions to music. *Psychol. Music* 41, 579–599. doi: 10.1177/0305735612440615
- Lima, C. F., Castro, S. L., and Scott, S. K. (2013). When voices get emotional: a corpus of nonverbal vocalizations for research on emotion processing. *Behav. Res. Methods* 45, 1234–1245. doi: 10.3758/s13428-013-0324-3
- Livingstone, S., and Thompson, W. (2009–2010). The emergence of music from the Theory of Mind. *Musicae Sci.* 83–115.
- Lundqvist, L.-O., Carlsson, F., Hilmersson, P., and Juslin, P. N. (2009). Emotional responses to music: experience, expression, and physiology. *Psychol. Music* 37, 61–90. doi: 10.1177/0305735607086048
- Maiese, M. (2011). *Embodiment, Emotion, and Cognition*. New York, NY: Palgrave Macmillan. doi: 10.1057/9780230297715
- Malloch, S., and Trevarthen, C. (2009). "Musicality: communicating the vitality and interests of life," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 1–11.
- Matsumoto, D. L., and Ekman, P. (2009). "Basic emotions," in *Oxford Companion to Affective Sciences*, eds D. Sander and K. Scherer (Oxford: Oxford University Press), 69–72.
- McDermott, J., and Hauser, M. (2005). The origins of music; innateness, uniqueness, and evolution. *Music Percept.* 23, 29–59. doi: 10.1525/mp.2005.23.1.29
- Menary, R. (2010). Introduction to the special issue on 4E cognition. *Phenomenol. Cogn. Sci.* 9, 459–463. doi: 10.1007/s11097-010-9187-6
- Menon, V., and Levitin, D. (2005). The rewards of music listening: responses and physiological connectivity of the mesolimbic system. *Neuroimage* 28, 175–184. doi: 10.1016/j.neuroimage.2005.05.053
- Menon, V., Levitin, D., Smith, B., Lembke, A., Krasnow, B., Glaser, D., et al. (2002). Neural correlates of timbre change in harmonic sounds. *Neuroimage* 17, 1742–1754. doi: 10.1006/nimg.2002.1295
- Merchant, H., Grahm, J., Trainor, L., Rohrmeier, M., and Fitch, W. T. (2015). Finding the beat: a neural perspective across humans and non-human primates. *Philos. Trans. R. Soc. B Biol. Sci.* 370:20140093 doi: 10.1098/rstb.2014.0093
- Mithen, S. (2006). *The Singing Neanderthals. The Origin of Music, Language, Mind, and Body*. Cambridge, MA: Harvard University Press.
- Molino, J. (2000). "Toward an evolutionary theory of music and language," in *The Origins of Music*, eds N. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press), 165–176.
- Moors, A. (2007). Can cognitive methods be used to study the unique aspect of emotion: an appraisal theorist's answer. *Cogn. Emot.* 21, 1238–1269. doi: 10.1080/02699930701438061
- Nagel, F., Kopiez, R., Grewe, O., and Altenmüller, E. (2007). EmuJoy: software for continuous measurement of perceived emotions in music. *Behav. Res. Methods* 39, 283–290. doi: 10.3758/BF03193159
- Newman, J. D. (2007). Neural circuits underlying crying and cry responding in mammals. *Behav. Brain Res.* 182, 155–165.
- Niedenthal, P. (2007). Embodying emotion. *Science* 316, 1002–1005. doi: 10.1126/science.1136930
- Niedenthal, P., Barsalou, L., Winkielman, P., Krauth-Gruber, S., and Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personal. Soc. Psychol.* 9, 184–211. doi: 10.1207/s15327957pspr0903\_1
- Panksepp, J. (1995). The emotional sources of "Chills" induced by music. *Music Percept.* 13, 171–207. doi: 10.2307/40285693
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford University Press.
- Panksepp, J. (2005). Affective consciousness: core emotional feelings in animals and humans. *Conscious. Cogn.* 14, 30–80. doi: 10.1016/j.concog.2004.10.004
- Panksepp, J. (2009–2010). The emotional antecedents to the evolution of music and language. *Musicae Sci.* 13, 229–259. doi: 10.1177/1029864909013002111
- Panksepp, J., and Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behav. Process.* 60, 133–155. doi: 10.1016/S0376-6357(02)00080-3
- Pell, M. D., Paulmann, S., Dara, C., Alaseri, A., and Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: a comparison of four languages. *J. Phon.* 37, 417–435. doi: 10.1016/j.wocn.2009.07.005
- Peretz, I. (2001). "Listen to the brain: a biological perspective on musical emotions," in *Music and Emotion: Theory and Research*, eds P. N. Juslin and J. Sloboda (Oxford: Oxford University Press), 105–134.
- Peretz, I. (2006). The nature of music from a biological perspective. *Cognition* 100, 1–32. doi: 10.1016/j.cognition.2005.11.004
- Peretz, I., and Coltheart, M. (2003). Modularity of music processing. *Nat. Neurosci.* 6, 688–691. doi: 10.1038/nn1083
- Peretz, I., Vuvan, D., Lagrois, M.-É., and Armony J. L. (2015). Neural overlap in processing music and speech. *Philos. Trans. R. Soc. B Biol. Sci.* 370:20140090 doi: 10.1098/rstb.2014.0090
- Pinker, S. (1997). *How the Mind Works*. New York: W. W. Norton & Company.
- Rahman, R. A., and Sommer, W. (2008). Seeing what we know and understand: how knowledge shapes perception. *Psychon. Bull. Rev.* 15, 1055–1063. doi: 10.3758/PBR.15.6.1055

- Redondo, J., Fraga, I., Padrón, I., and Piñeiro, A. (2008). Affective range of sound stimuli. *Behav. Res. Methods* 40, 784–790. doi: 10.3758/BRM.40.3.784
- Reybrouck, M. (2001). Biological roots of musical epistemology: functional cycles, Umwelt, and enactive listening. *Semiotica* 134, 599–633. doi: 10.1515/semi.2001.045
- Reybrouck, M. (2005). A biosemiotic and ecological approach to music cognition: event perception between auditory listening and cognitive economy. *Axiomathes Int J. Ontol. Cogn. Sys.* 15, 229–266.
- Reybrouck, M. (2008). “The musical code between nature and nurture,” in *The Codes of Life: The Rules of Macroevolution*, ed. M. Barbieri (Springer: Dordrecht), 395–434.
- Reybrouck, M. (2013). “Musical universals and the axiom of psychobiological equivalence,” in *Topicality of Musical Universals / Actualité des Universaux Musicaux*, ed. J.-L. Leroy (Paris: Editions des Archives Contemporaines), 31–44.
- Reybrouck, M. (2014). Musical sense-making between experience and conceptualisation: the legacy of Peirce, Dewey and James. *Interdiscip. Stud. Musicol.* 14, 176–205.
- Reybrouck, M. (2015). Real-time listening and the act of mental pointing: deictic and indexical claims. *Mind Music Lang.* 2, 1–17.
- Reybrouck, M., and Brattico, E. (2015). Neuroplasticity beyond sounds: neural adaptations following long-term musical aesthetic experiences. *Brain Sci.* 5, 69–91. doi: 10.3390/brainsci5010069
- Robinson, J. (2009). “Aesthetic emotions (philosophical perspectives),” in *The Oxford Companion to Emotion and the Affective Sciences*, eds D. Sander and K. R. Scherer (New York, NY: Oxford University Press), 6–9.
- Russell, J. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev.* 110, 145–172. doi: 10.1037/0033-295X.110.1.145
- Russell, J. (2009). Emotion, core affect, and psychological construction. *Cogn. Emot.* 23, 1259–1283. doi: 10.1080/02699930902809375
- Russell, J., and Barrett, F. L. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *J. Pers. Soc. Psychol.* 76, 805–819. doi: 10.1037/0022-3514.76.5.805
- Sachs, M. E., Ellis, R. E., Schlaug, G., and Loui, P. (2016). Brain connectivity reflects human aesthetic responses to music. *Soc. Cogn. Affect. Neurosci.* 11, 884–891. doi: 10.1093/scan/nsw009
- Salimpoor, V. N., Zald, D. H., Zatorre, R., Dagher, A., and McIntosh, A. R. (2015). Predictions and the brain: how musical sounds become rewarding. *Trends Cogn. Sci.* 19, 1–6. doi: 10.1016/j.tics.2014.12.001
- Sander, D. (2013). “Models of emotion. The affective neuroscience approach,” in *The Cambridge Handbook of Human Affective Neuroscience*, eds J. Armony, and P. Vuilleumier (Cambridge: Cambridge University Press), 5–53. doi: 10.1017/CBO9780511843716.003
- Sauter, D. A., Eisner, F., Calder, A. J., and Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Q. J. Exp. Psychol.* 63, 2251–2272. doi: 10.1080/17470211003721642
- Scherer, K. (1993). Neuroscience projections to current debates in emotion psychology. *Cogn. Emot.* 7, 1–41. doi: 10.1080/02699939308409174
- Scherer, K. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *J. New Music Res.* 33, 239–251. doi: 10.1080/0929821042000317822
- Scherer, K., and Zentner, K. (2001). “Emotional effects of music: production rules,” in *Music and Emotion: Theory and Research*, eds P. Juslin and J. Sloboda (Oxford: Oxford University Press), 361–392.
- Scherer, K., and Zentner, M. (2008). Music evoked emotions are different – more often aesthetic than utilitarian. *Behav. Brain Sci.* 5, 595–596. doi: 10.1017/s0140525x08005505
- Scherer, K. R. (2003). Vocal communication of emotion: a review of research paradigms. *Speech Commun.* 40, 227–256. doi: 10.1016/S0167-6393(02)00084-5
- Schiavio, A., van der Schyff, D., Cespedes-Guevara, J., and Reybrouck, M. (2016). Enacting musical emotions. Musical meaning, dynamic systems, and the embodied mind. *Phenomenol. Cogn. Sci.* doi: 10.1007/s11097-016-9477-8
- Schneck, D., and Berger, D. (2010). *The Music Effect. Music Physiology and Clinical Applications*. London: Kingsley Publishers.
- Schubert, E. (2001). “Continuous measurement of self-report emotional response to music,” in *Music and Emotion: Theory and Research*, eds P. Juslin and J. Sloboda (New York, NY: Oxford University Press), 393–414.
- Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Percept. Interdiscip. J.* 21, 561–585. doi: 10.1525/mp.2004.21.4.561
- Schubert, E. (2013). Emotion felt by the listener and expressed by the music: literature review and theoretical perspectives. *Front. Psychol.* 4:837. doi: 10.3389/fpsyg.2013.00837
- Sperber, D. (1996). *Explaining Culture*. Oxford: Blackwell.
- Stern, D. N. (1985). *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*. New York, NY: Basic Books.
- Stern, D. N. (1999). “Vitality contours: the temporal contour of feeling as a basic unit for constructing the infant's social experience,” in *Early Social Cognition: Understanding Others in the First Month of the Life*, ed. Rochat, P. (Mahwah, NJ: Erlbaum), 67–90.
- Stewart, J., Gapenne, O., and Di Paolo, E. A. (eds). (2010). *Enaction: Toward a New Paradigm for Cognitive Science*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262014601.001.0001
- Storbeck, J., and Clore, G. (2007). On the interdependence of cognition and emotion. *Cogn. Emot.* 21, 1212–1237. doi: 10.1080/02699930701438020
- Striedter, G. F. (2005). *Principles of Brain Evolution*. Sunderland, MA: Sinauer Associates.
- Striedter, G. F. (2006). Précis of principles of brain evolution. *Behav. Brain Sci.* 29, 1–36. doi: 10.1017/S0140525X06009010
- 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
- Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenol. Cogn. Sci.* 4, 407–427. doi: 10.3389/fnint.2013.00015
- Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge: Harvard University Press.
- Thompson, W. F., and Balkwill L.-L. (2006). Decoding speech prosody in five languages. *Semiotica* 158, 407–424
- Tomkins, S. (1963). *Affect Imagery Consciousness: Vol. II, The Negative Affects*. New York: Springer.
- Tooby, J., and Cosmides, L. (1992). “The psychological foundations of culture,” in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, eds J. Barkow, L. Cosmides, and J. Tooby (Oxford: Oxford University Press), 19–136.
- Trainor, L. J., and Schmidt, L. A. (2003). “Processing emotions induced by music,” in *The Cognitive Neuroscience of Music*, eds I. Peretz and R. Zatorre (Oxford: Oxford University Press), 310–324.
- Trehub, S. (2003). The developmental origins of musicality. *Nat. Neurosci.* 6, 669–673. doi: 10.1038/nn1084
- Trehub, S. E., Becker, J., and Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philos. Trans. R. Soc. B Biol. Sci.* 370:20140096. doi: 10.1098/rstb.2014.0096
- Trost, W., Ethofer, T., Zentner, M., and Vuilleumier, P. (2012). Mapping aesthetic emotions in the brain. *Cereb. Cortex* 22, 2769–2783. doi: 10.1093/cercor/bhr353
- van der Zwaag, M., Westerink, J., and van den Broek, E. (2011). Emotional and psychophysiological responses to tempo, mode, and percussiveness. *Musicae Sci.* 15, 250–269. doi: 10.1177/1029864911403364
- Varela, F., Thompson, E., and Rosch, E. (1991). *The Embodied Mind*. Cambridge, MA: MIT Press.
- Vuoskoski, J., and Eerola, T. (2011). Measuring music-induced emotion A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Sci.* 15, 159–173. doi: 10.1177/1029864911403367
- Vuoskoski, J. K., and Eerola, T. (2012). Can sad music really make you sad? Indirect measures of affective states induced by music and autobiographical memories. *Psychol. Aesthet. Creat. Arts* 6, 204–213. doi: 10.1037/a0026937
- Vuoskoski, J. K., and Eerola, T. (2015). Extramusical information contributes to emotions induced by music. *Psychol. Music* 43, 262–274. doi: 10.1177/0305735613502373
- Wallin, N., Merker, B., and Brown, S. (eds). (2000). *The Origins of Music*. Cambridge, MA: The MIT Press.
- Witvliet, C., and Vrana, S. (1996). The emotional impact of instrumental music on affect ratings, facial EMG, autonomic response, and the startle reflex: effects of valence and arousal. *Psychophysiology* 33(Suppl. 1), 91.

Zentner, M., Grandjean, D., and Scherer, K. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* 8, 494–521. doi: 10.1037/1528-3542.8.4.494

**Conflict of Interest Statement:** 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.

*Copyright © 2017 Reybrouck and Eerola. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Global Sensory Qualities and Aesthetic Experience in Music

Pauli Brattico, Elvira Brattico\* and Peter Vuust

Center for Music in the Brain, Department of Clinical Medicine, Aarhus University and The Royal Academy of Music Aarhus/Aalborg, Aarhus, Denmark

## OPEN ACCESS

### Edited by:

Piotr Podlipniak,  
Adam Mickiewicz University in  
Poznań, Poland

### Reviewed by:

L. Robert Slevc,  
University of Maryland, College Park,  
USA

Dan Zhang,

Tsinghua University, China

Sasa Brankovic,

Clinical Center of Serbia, Serbia

### \*Correspondence:

Elvira Brattico  
elvira.brattico@clin.au.dk

### Specialty section:

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

**Received:** 30 November 2016

**Accepted:** 13 March 2017

**Published:** 05 April 2017

### Citation:

Brattico P, Brattico E and Vuust P  
(2017) Global Sensory Qualities and  
Aesthetic Experience in Music.  
*Front. Neurosci.* 11:159.  
doi: 10.3389/fnins.2017.00159

A well-known tradition in the study of visual aesthetics holds that the experience of visual beauty is grounded in global computational or statistical properties of the stimulus, for example, scale-invariant Fourier spectrum or self-similarity. Some approaches rely on neural mechanisms, such as efficient computation, processing fluency, or the responsiveness of the cells in the primary visual cortex. These proposals are united by the fact that the contributing factors are hypothesized to be global (i.e., they concern the percept as a whole), formal or non-conceptual (i.e., they concern form instead of content), computational and/or statistical, and based on relatively low-level sensory properties. Here we consider that the study of aesthetic responses to music could benefit from the same approach. Thus, along with local features such as pitch, tuning, consonance/dissonance, harmony, timbre, or beat, also global sonic properties could be viewed as contributing toward creating an aesthetic musical experience. Several such properties are discussed and their neural implementation is reviewed in the light of recent advances in neuroaesthetics.

**Keywords:** music aesthetics, neuroaesthetics, musical features, naturalistic paradigm, visual aesthetics

## INTRODUCTION

When the legendary music producer Phil Spector created the trademark “Wall of Sound” aesthetics during the 1960s, the point was not about music theory or song writing, or even about instrumentation, but something abstract yet firmly anchored in the world of sense: he wanted to create a saturated, dense sound that would be aesthetically appealing even when played out from the monoaural AM radio and jukebox devices of the time. Similar conclusions can be made on the basis of observations of audio and sound engineers who likewise work with abstract sonic notions that, somewhat paradoxically, refer to concrete sensory experiences. A guitar sound, for example, can be “thin” or “full”; a drum must be “singing out,” “wide-open,” “cool,” “not muffling,” “pretty tight,” to have “a little more of a smack” (Porcello, 2004, pp. 741–744).

Provided that such qualities are aesthetically important, and well-known and much used by musicians, what are they? To first coin a heuristic term, we propose to call them *global sensory qualities*. What we mean by saying that they are global is that they concern the “whole sound” distinct from any of its individual parts, instruments, harmony structure, intervals, melody, or tuning. Moreover, many or at least most of these musical qualities seem to refer to sensory qualities. For example, when a snare drum is characterized as “pretty tight,” the notion does not seem to single out a particular affective or cognitive property, let alone a property grounded in (Western or non-Western) music theory. From the context, it is clear that what is at stake is a snare drum sound not spread too wide in terms of its sensory-related acoustic dimensions (space and reverb, frequency, timbre, sustain) in order to “sit well” in the whole mix and thus to emerge distinctive

enough amongst the background of other materials. In short, the global sensory properties are both global properties, in that they concern the whole percept, but also sensory-based, since they seem to describe sensory qualities.

The premise of the present article is that global sensory qualities constitute an important yet neglected factor in a musical aesthetic experience, and could provide a fruitful avenue for research into the psychology and neurobiology of aesthetic perception. For instance, we propose that these global features are statistically extracted from the stimuli by the auditory system—or, perhaps more likely, by some subsystems (McDermott and Simoncelli, 2011; McDermott et al., 2013)—and then passed on to high-level processing, ultimately leading to the main outcomes of musical experience, namely aesthetic judgment, emotion and conscious liking, or preference (Cela-Conde et al., 2011; Brattico et al., 2013).

The idea itself is not new, especially what comes to visual aesthetics, but rarely applied to music. The notion that there are global visual sensory qualities triggering an aesthetic response has a long history, as argued for example by Bell (1914) in his theory that successful (visual) art involves a “significant form” leading to universal aesthetic experience and emotion. For Bell, the significant form, whose ultimate nature he left mysterious, consisted of “combinations and arrangements” of various visual elements such as lines, form and shapes. He wrote that “forms arranged and combined according to certain unknown and mysterious laws do move us in a particular way, and that it is the business of an artist so to combine and arrange them that they shall move us” (loc. 184).

Vision scientists have not shied away from searching for Bell’s significant formula for aesthetic experience, and, recently, a number of them have tried to locate the form in global sensory properties. Jacobs et al. (2016), for example, examined aesthetic judgments of various visual textures and argued that they correlate with global computational properties, such as the presence of lower spatial frequencies, oblique orientations, higher intensity variation, higher saturation, and overall redness. By examining industrial design and visual aesthetics, Hekkert (2006) proposed four sensory qualities that can increase the aesthetic appeal of an object: (i) maximum effect for minimum means (“economic computations are favored over more complex ones”); (ii) unity in variety (“ability to see regularities and patterns in complex observations”); (iii) most advanced, yet acceptable (“the correct balance between novelty and repetition”); (iv) and optimal match (“information from different sensory modalities should converge with each other”). Renoult et al. (2016) found out that the (algorithmically modeled) sparseness of the activity of simple cells in the primary visual cortex (V1) correlates with female face attractiveness when assessed by male participants, suggesting that there might be general, non-face recognition specific neuronal properties that factor into facial aesthetic evaluation. Spehar et al. (2015) reached similar conclusions by correlating visual sensitivity with the aesthetic properties of visual random patterns. Other candidates for global sensory properties that have been studied recently include processing fluency (Reber et al., 2004; Babel and McGuire, 2015; Forster et al., 2015), distribution of spectral frequency power (Menzel et al., 2015),

self-similarity and fractal properties (Taylor et al., 1999, 2011; Spehar et al., 2003; Hagerhall et al., 2004; Mureika et al., 2004; Graham and Field, 2007; Redies, 2007, 2015; Forsythe et al., 2011; Mallon et al., 2014).

Could similar properties play a role in determining aesthetic responses to music, and could this hypothetical causal relation be pinpointed accurately? In the following sections, we argue that this is likely the case and propose hypotheses to be tested in future research, complementing the current focus on more local factors derived from music theory. Indeed, global features constitute but one subset of auditory features relevant to music, along with others (e.g., pitch, timbre, intervals, harmony, melody, music syntax, and individual instruments), much studied both in connection with auditory processing in general (see e.g., Koelsch, 2011), but also in connection with aesthetic perception (for reviews, see Nieminen et al., 2011; Brattico and Pearce, 2013; Brattico et al., 2013; Hodges, 2016). Perhaps global sensory properties play even a special role in musical pieces of pop/rock/metal genres, in which harmony and voice leading rules are often violated but music producers follow specific professional principles toward reaching a defined aesthetic goal (Račić, 1981; Baugh, 1993; von Appen, 2007). Today almost all music is produced, recorded, reproduced and consumed electro-acoustically, and has become a ubiquitous experience in our everyday lives. Musical pieces that resemble classical music styles, such as film soundtracks (Huckvale, 1990) or computer game music (Bridgett, 2013), are today composed and produced with computers. While historically musical aesthetics has concentrated on the classical music genre, more recently also pop/rock and jazz music has received attention by aesthetic (von Appen, 2007; Juslin et al., 2016) and neuroaesthetic scholars (Limb and Braun, 2008; Janata, 2009; Berns et al., 2010; Brattico et al., 2011, 2015; Johnson et al., 2011; Montag et al., 2011; Pereira et al., 2011; Salimpoor et al., 2011, 2013; Zuckerman et al., 2012; Istok et al., 2013; Bogert et al., 2016). Indeed, even though “rock musicians never ask if a composition is aesthetically valuable,” they are still keen in evaluation “if it *sounds good*,” as observed by Račić (1981, p. 200, emphasis from the original). The study of aesthetics would be too narrowly construed if questions of what “sounds good” were ignored.

The same point can be made in the case of visual aesthetics. As pointed out by Redies (2015), the creation of visual beauty is not limited to any particular style, method, genre, or form, such as color, shape, luminance, texture, edges, or depth cues. A wide variety of materials can be used to create visually appealing objects. This suggests that the neural processes associated with aesthetic experience are not restricted to any particular feature (or corresponding neuronal circuits) or to a particular genre or style. We propose that the same might be true of music.

## GLOBAL AESTHETIC SENSORY QUALITIES

We argue that global computational properties play a role in music aesthetics, and provide an overview of what we consider some of the most relevant global sensory properties to be.

We also discuss previous research in the aesthetic of music that highlights the importance of such features. This review will be limited to global sensory properties, thus for the sake of clarity we ignore properties relating to culture, history or listeners' cognitive biases that are also supposed to play a role in a musical aesthetic experience (Chapman and Williams, 1976; McPherson and Schubert, 2004; Brattico, 2009–2010). The next section is dedicated to the discussion of the possible role of global properties in brain processing. As a provisional entry to this topic, note again that it is well-known that both musicians and non-musicians do in fact use global and “holistic” notions, such as “beautiful,” “melodious,” “rhythmic,” “touching,” “harmonic,” “peaceful,” “atmospheric,” “calming,” or “versatile” when describing the personal aesthetic value of music (Jacobsen, 2004; Istok et al., 2009). Most if not all of these concepts describe abstract impressionistic and holistic properties characterizing the piece as a whole, and are not strictly dependent on (although they might interact with) music-theory based local notions, such as intervals or chords. The same point can be further appreciated by noting that aesthetic perception is in no way tied to the Western music genres, but applies equally well to non-Western music. Indeed, when we look at art and aesthetics as a whole, it is true that “some kind of aesthetic activity is apparently a feature of all the 3,000 or so distinguishable cultures that are to be found on the earth's surface,” as observed by Berlyne (1971 p. 27). Hence, we believe that aesthetics or aesthetic theories should not be tied with any particular style, genre, or music-theoretical notion.

The key distinction between *global* and *local* features is best elucidated by first looking how they are used in the study of visual aesthetics, and then by extending the notion to the domain of music and auditory aesthetics. In the study of vision and visual beauty, local properties of an image constitute the individual parts of the image, such as local color patches, lines, shapes, contrast, textures, surfaces, or other visual elements. Such local elements can be either *formal*, consisting of various non-conceptual or non-representational forms, or *content-based*, consisting of elements that represent something else. Examples of the former elements are color patches, lines, and textures, of the latter faces and objects. Early processing of visual information is predominantly local, as each local point in an image is projected tonotopically to a point in a visual representation (Wurtz and Kandel, 2000). As the information processing continues, however, the local features are integrated into a whole percept, or Gestalt, that “puts each pictorial element in perceptual relation to the other elements in the artwork” (Redies, 2015, p. 6) and thus integrates the various local elements together. It is that whole Gestalt that, according to many vision researchers, is relevant to the appreciation of beauty (Ramachandran and Hirstein, 1999; Zeki, 1999). Thus, the “Global structure refers to statistical regularities in large parts of the image or in the entire image, for example the spatial frequency content of the image, the kurtosis of its luminance values, overall complexity of self-similarity” (Redies, 2015, p. 4). Hence, it is not generally possible to take a piece of art, break it into pieces and then reassemble it back in random order while automatically preserving its artistic qualities. Formulated in this way, the distinction between global and local properties becomes relative. A painting on a wall constitutes a

local feature of an even more global space, the whole wall. A modern artwork may consist of a red spot on a white background, making what in some other context would constitute a local feature a global one. These problems are kept under control by minimizing the impact of the context, for example, by framing and isolating the artwork in various ways from its natural surroundings and other objects of interest.

The global-local distinction elucidated above applies to music. In music, the local features can be best illustrated by the musical score, by separate tracks in a digital audio workstation (DAW), or by separating the performance of each band member from the rest, where each note/tone or interval appears in isolation and is mapped to the production of physical sound with certain timbre- and rhythmic characteristics during performance and/or recording. A note carries local information concerning timbre (instrument), dynamics (loudness), pitch, pitch changes (vibrato), duration and internal change (staccato, marcato, legato). The notes are further integrated into melodies and harmonic structures and relations that can still constitute local features. In a typical multi-instrument composition, several melodic themes are weaved together to create a sense of harmonic and melodic development. A local feature can be detached from the whole musical piece simply by muting it, or by muting a whole track in a sequencer. For instance, a melody can be changed, even dramatically, by changing the pitch or duration of just one note, and this produces fast reactions in the brain (such as the mismatch negativity, MMN, and the P3a responses) reflecting both an automatic processing of the change and the reorienting of involuntary attention toward the unexpected event. In turn, we propose that global musical features involve the composition as a whole, being synthesized from individual local features as they get summed into an integrated Gestalt. One can refer to the totality of all local features as the overall “musical texture.” Although it is possible to attend to each local part selectively, this is arguably not the norm and restricted to certain artificial contexts. The idea of removing, let alone freely reassembling, some parts from a composition is quite alien to the normal production and consumption of music. Thus, as in the case of visual art, we believe that it is the totality of all such elements that determine their artistic and aesthetic value. For instance, in the production of commercial-grade music global auditory features are manipulated during the final mastering process by using limiters, compressors, equalizers, and other dynamic and spectral processors. In the same vein, listening to any of the tracks or sounds in a musical piece in isolation will typically not lead to a positive, impressive aesthetic experience; it is their combined sum that will do that. Below we provide examples of aesthetically relevant music-specific global features.

## Distribution of Spectral Energy

An important aesthetic quality of music concerns the distribution and dynamics of its spectral energy. An aesthetically appealing sonic object is typically created by controlling the balance of its spectrum energy along several important dimensions such as (i) frequency, (ii) space, and (iii) time, as discussed below. “The goal” in sound engineering and mixing is “to get every aspect of the track to balance: every pitch and every noise; every

transient and every sustain; every moment in time and every region of the frequency spectrum” (Senior, 2011, loc. 4904). Orchestral and other groups of instrumentalists adhere to the same principle, explicitly, or implicitly. It is crucial that, even in a loud performance typical of rock music, for example, the instruments are balanced.

In the *frequency domain*, we propose that the crucial balance is achieved by ensuring that the musical information is distributed throughout the whole audible frequency spectrum, and that the signal-to-noise ratio for each meaningful package of musical information (i.e., instrument, singer, instrument group or, more generally, a perceived sound source) is good enough so that no lower or higher level auditory masking intervenes. Indeed, the idea that efficient coding plays a role in human perception is supported by empirical evidence. Listeners must be able to hear all instruments in a distinctive way (not as a fuzzy auditory mess) even if they focus attention only on one of them, and thus these instruments have to live inside their own “safe space” in the spatiotemporal spectrum to avoid frequency masking, even when the music is composed out of digital samples of instrument sounds. They must furthermore appear controlled and consistent. Unaesthetic dynamical changes, conflicts and overlaps are routinely cleaned up by using filters, equalizations, compressors, and other techniques. In addition, often pop/rock and jazz music thrives to fill in the whole frequency spectrum by having “bottom end” (bass, kick drum), “high end” (hihats, cymbals, high pitch sounds), and “middle range” (singers, guitars, snare drums) instruments playing simultaneously (Corozine, 2002). Systematic empirical evidence is scarce, but composers are aware that the complete lack of any of the here described components will lead into a distinctive impairment in the aesthetic quality of the overall sound.

The unaesthetic masking phenomenon referred to above might result from the biological architecture of the human auditory system. The auditory system works by decomposing the signal by several narrow cochlear filters, or critical bands, each spanning a relatively small frequency range. The number and constitution of these bands is derived from psychoacoustic masking experiments, so that they capture the upper bound on the human frequency discrimination ability (Zwicker, 1961; Moore, 2012: Ch. 3). For the most part the frequency range increases logarithmically as a function of the central frequency, and the amplitudes of the resulting filters undergo nonlinear basilar membrane compression such that they are less sensitive to higher amplitudes. Furthermore, the human ear is most sensitive to the middle frequencies around 1,500 Hz, while the sensitivity decreases for sounds with both lower and higher frequencies. The temporal resolution of the auditory system, however, surpasses that of the other senses. Indeed, temporal resolution is required in the processing of fast transients and other sound changes that occur in, e.g., natural speech (Plomp, 1964; Zatorre et al., 2002). Further processing takes place once the signal travels to the auditory cortex via several subcortical regions (Barbour and Wang, 2003). The implication is that there are limitations on how much frequency/temporal space each musical signal can occupy to be perceived distinctly and clearly by the human brain in relation to other, surrounding musical information. This is

especially relevant in the context of complex auditory signals, such as speech or music. Professional audio engineers’ music producers’ and composers’ aim for distinctiveness in the sound can be interpreted as suggesting that avoidance of low- and high-level auditory masking contributes to sonic aesthetic experience. The notion is global, however, because it concerns the musical piece as a whole: how distinct various instruments and musical signals are perceived in relation to each other.

In the *space domain*, several techniques such as panning, reverbs, filtering, delays, filtering, and pre-delays are manipulated to position the musical information distinctively within the spatial field. This positioning is achieved by modeling the way the human brain encodes spatial information from the acoustic signal (Zahorik, 2002). For example, when a musical instrument is embedded within a space by using an artificial or natural reverberation, a few milliseconds of pre-delay in the reverberation can change the perceived distance of the source: a reverb with no pre-delay will position the source to the back wall of the virtual space, while 20–30 ms pre-delay will bring it closer to the listener. This models the time the reflected (reverberated) sounds will normally lag behind the direct sound. Similar manipulations are used in experiments testing the neural abilities for discriminating sound sources. Notably, these abilities, relying on the fast elaboration of differences in the incoming signal as compared with the environment at the level of the auditory cortex are very sensitive to even small variations of spatial location (Colin et al., 2002; Roeber et al., 2003; Altmann et al., 2014). But the spatial interpretation of music is global in the sense that it concerns the relative position of the listener to that of the sound source and the environment, whether these are real or virtual. The spatial dimension is also used when positioning sound sources to different locations within a virtual space in order to keep the said sources sufficiently distinct from each other.

In the *temporal domain*, the dynamical qualities of individual instruments (e.g., transients) and the whole song structure are controlled to create a sense of music development and to adjust for the inevitable sensory habituation. “In a lot of cases in commercial music,” Senior (2011) observed, “you want to have enough repetition in the arrangement that the music is easily comprehensible to the general public. But you also want to continually demand renewed attention by varying the arrangement slightly in each section” (loc. 2523). For example, to maintain listeners’ attention one is advised to “provide some new musical or arrangement diversion every 3–5 s to keep listener riveted to the radio” (loc. 2592). Thus, the balance between repetition/regularity and novelty, much discussed in the study of aesthetics and supposedly following an inverted U-shape function (Berlyne, 1971), does not concern only rhythm (Vuust and Witek, 2014; Witek et al., 2014) or melody (Green et al., 2012), but is related to a change of any kind, including changes in the global musical texture.

## Musical Texture

The term “musical texture” refers to the way that local musical features, such as rhythm, melody, and harmony are integrated in a whole composition and, ultimately, into a whole Gestalt

percept in the listeners' brain (e.g., Meyer, 1956 p. 185–196). Texture is an elementary consideration in both arrangement and orchestration, processes that aim for crafting an aesthetic output from several local themes such as rhythm, melodies, counter-melodies, and harmony. The same four-way voicing, such as an arrangement for four saxophones, may have quite different textures if it is arranged in parallel compared to when the individual voices are allowed to cross one another. Music that strongly relies on music theory properties can benefit immensely from properties of the texture, as in the case of popular or film music, suggesting that texture alone can be a crucial component in determining an aesthetic response to music.

While the study of texture perception is a lively topic in the domain of vision, with by now a long tradition (e.g., Julesz, 1962), very little comparable research exists in the case of auditory modality. In one study, McDermott and Simoncelli (2011) constructed a physiologically realistic model of the auditory system, which they provided with samples of various repeating sound textures, such as rainstorms, insect swarms, river, and wind, and then used the model to extract biologically plausible time-averaged statistical properties from the textures. These statistical measures represent high-level descriptions of the sound source. They were used to synthesize the same texture sounds from white noise, and the results were compared against the natural sounds in an experiment by using human participants. Sound synthesis was either biologically realistic or unrealistic. The logic of the experiment was to use human performance as a way to benchmark the biological plausibility of the model. For example, when the synthetic sounds were indistinguishable from the natural samples by the human participants, it could be assumed that the generative model closely matched that of the human auditory system. When the participants noted marked differences with the original texture and the synthetic one, we can reason that the model did not mimic the human auditory system. A clear contrast emerged between realistic and unrealistic assumptions, suggesting that the human auditory system might indeed extract statistical properties of the sounds to encode and represent its global textural properties. For further experimental evidence that human auditory system utilizes time-averaged statistical processing to represent textures and other global features of sounds, see McDermott et al. (2013). In the latter study, the authors proposed a functional explanation for their findings, suggesting that statistical averaging is used by the auditory system to overcome memory limitations. The evidence that the auditory system uses statistical time-averages is encouraging for our hypothesis that part of the music aesthetic experience relies on global sensory properties, because it provides empirical justification for the claim that such global features could play a direct role also in auditory perception. These studies go further by proposing that there are neuronal populations within the auditory pathway that are specifically dedicated and tuned to detect global statistical properties of the auditory signal. This raises the possibility that the immediate aesthetic value in certain global sensory properties would be directly assessed by low-level modules in the brain, rather than being assembled only later when the isolated local features are merged into a whole

percept. Whatever the case, we encourage studies for testing the hypothesis that, as in the case of visual textures, musical texture would play a comparable role in the aesthetic perception of music.

## Expressivity

Another relevant global quality that affects the aesthetic appeal of a sonic object is its music-emotional impact or expressivity (Robinson, 1994; Gabrielsson and Juslin, 1996). While playing synthesized chord sequences or sinewave tones in isolation and in temporally exacting sequences can indeed evoke emotions and aesthetic judgments due to their ability to represent elementary harmony relations, there is a difference between fully mechanized, synthetic version and humanly played orchestral version of the same piece such that the latter will be perceived as more aesthetic than the former (Seashore, 1929). The “humanness” in the performance of a real human being is especially relevant to the perceived emotional character of the performance. This indicates that there are global sensory features that exhibit a direct causal relationship with human emotions and the “emotional centers” of the brain (Koelsch, 2014). What these features are remains elusive, but the study of visceral affective reactions to music, such as chills, has revealed that there indeed exist prototypical sonic qualities that tend to evoke strong emotional responses in listeners. Laeng et al. (2016), for example, mentions properties such as the beginning of a piece, an entry of an instrument or human voice, melodic appoggiaturas (“extra notes or ornaments”), dynamic changes in loudness, surprising harmonic changes, and sustained high-pitch tones of instruments or voice, among other techniques (see Sloboda, 1991; Panksepp, 1995; Gabrielsson and Juslin, 1996; Rickard, 2004; Grewe et al., 2007; Gabrielsson, 2011; Branković, 2013). If we compare raw mechanical and synthesized instrumentation to that of a real human performance, a complex of dynamic and timbral differences emerge such that latter contains a continuous stream of changes in dynamics (attack, sustain, release), pitch (vibrato, true legato), timbre and spectrum, pauses (breathing, bowing), and many others.

## Tempo and Mode

Researchers have shown that global properties such as tempo and mode (minor or major) influence preference and liking, possibly due to their association with basic emotions such as sadness and happiness (Hevner, 1935; Dalla Bella et al., 2001; Pallesen et al., 2003; Khalfa et al., 2005; Hunter et al., 2008; Schellenberg et al., 2008; Nieminen et al., 2012). Slow tempo and/or minor mode are associated with sadness, while fast tempo and/or major mode with happiness, the latter receiving more positive liking ratings (Husain et al., 2002). Tempo, meter and mode are global properties in the sense that they describe, not individual instruments or parts, but large segments of the compositions, or indeed the composition as a whole. Mode, for example, characterizes the underlying key (minor vs. major) upon which the composition, or a segment of the composition, is based on. It also describes the tonal center of the piece that the listener will expect the musical development to return periodically through tension and relaxation. When the mode is in major, the music

sounds happier overall than when it is in the minor key. Similarly, the meter of a song, be it, e.g., a waltz (3/4) or a march (2/4) fundamentally influences the mood of the song.

## Other Properties and Experimental Expectations

In addition to the examples above, there are other global properties that are known to affect the rewarding responses to music, such as exposure or familiarity (Heingartner and Hall, 1974; Bornstein, 1989; Peretz et al., 1998; Pereira et al., 2011) and groove (Janata et al., 2012; Sioros et al., 2014; Vuust and Witek, 2014; Vuust et al., 2014; Kilchenmann and Senn, 2015; Fitch, 2016). Exposure and familiarity, in particular, affect liking in an inverted U-shape function, so that repetition will first lead to increased preference but the effect disappears if too much repetition is administered (Green et al., 2012).

In sum, alongside the more local and analytical musical features, there are several types of global sensory properties that seem to play a role in the creation of an aesthetic experience of music. One group of properties involves the distribution and dynamics of spectral energy. In aesthetically appealing music, each instrument or meaningful musical signal should occupy its own spectral space in terms of its frequency-based, spatial and dynamical dimensions in order to control auditory masking. A closely related aspect of musical aesthetics is constituted by musical texture, which refers to the overall sound that results from the combination of its local parts. Arrangement and orchestration are two ways musical texture is created, with much of the consideration having to do with distinctiveness and hence ultimately spectral dynamics. We also discussed expressivity, mode, tempo, familiarity and groove, all linked to emotions, as other possible examples of aesthetically relevant global properties.

All the global sensory properties discussed above apply equally well to Western and non-Western music and musical styles. Thus, a spectrally and spatially rich musical texture can be generated by manipulating digital instruments in a modern studio, for instance, in Western style for a pop music project, as well as by producing Balinese gamelan music in its natural surroundings. This is reasonable, since aesthetic responses are not a privilege of Western music and thus should not be explained as outcomes of only one musical genre or style.

The hypothesis linking global properties to aesthetic responses renders itself naturally to empirical experimentation. For example, the balance in spectral energy distribution can be rigorously manipulated at the stimulus level. This can be achieved by removing and/or adding sonic components at specific locations within the spectrum, irrespective of their representational or other content. If our hypothesis is correct, then such manipulation should lead into prominent changes in, e.g., aesthetic pleasure of such objects irrespective of their higher-level content (i.e., comparison between Western and non-Western music). Another relevant consideration comes from the recent naturalistic paradigm, discussed in detail in the next section, that is suited for addressing global properties

particularly well. We return to the experimental issues in the section Sensory Aesthetics as Immersion and Arousal, where we discuss the present hypothesis from a neuroaesthetics viewpoint.

## THE NATURALISTIC PARADIGM FOR STUDYING GLOBAL SENSORY PROPERTIES

Some recent work toward analyzing musical stimuli in terms of their global sensory properties have been done thanks to the introduction of the naturalistic paradigm in music research. In this paradigm, the participants are required to listen attentively to a whole piece of music while their brain signal is measured. Afterwards, their brain signal is analyzed as a time-series in combination with qualities obtained by exploiting knowledge from music information retrieval (MIR), namely acoustic parameters that are relevant for identifying musical genres and extracting timbral, tonal, and rhythmic information from musical pieces. Specifically, the brain signal measured with functional magnetic resonance imaging (fMRI; Alluri et al., 2012, 2013; Burunat et al., 2016) and with electroencephalography (EEG; Poikonen et al., 2016a) has been analyzed by extracting acoustic variables from the music by using the MIR Toolbox (Lartillot and Toivainen, 2007). This approach is based on the assumption that global computational sensory properties in naturalistic musical stimuli provide a useful window not only to technological applications but also into our appreciation of music and its neural implementation. Most of the relevant properties in these studies are spectral in nature and concern the way in which auditory energy is distributed both in frequency- and time-domains (see **Table 1**). This approach itself is a derivative of a larger research agenda of MIR that is aimed at extracting musically relevant information from whole musical pieces by using computational and statistical techniques (see Peeters, 2004; Moffat et al., 2015). MIR algorithms extract global features from the audio signal that are furthermore distantly related to the global features we claim could be relevant to aesthetics.

Particularly, in Alluri et al. (2012) the authors asked the participants to consciously listen to a musical piece (*Adios Nonino* by Astor Piazzolla) while their brain activation was simultaneously observed by fMRI scanning. The brain scans were correlated with statistical properties extracted from the song, such as overall loudness, spectral centroid, high-energy-low energy ratio, spectral entropy, spectral flux, and tonal clarity (see **Table 1**). Once these features were extracted from the whole song, the original 25 features were reduced into 9 by performing a principal component analysis (PCA) on the resulting song-wide feature vector. The remaining cluster features were global features such as Fullness, Brightness, Timbral complexity, Rhythmic complexity, Key clarity, Pulse clarity, Event synchronicity, Activity, and Dissonance, of which two (Rhythmic complexity and Event Synchronicity) were removed as they did not correlate with participants' subjective assessment in a separate behavioral experiment. Of the remaining

**TABLE 1 | Acoustic features used in Alluri et al. (2012).**

Feature	Description
Loudness	Root mean square energy: the square root of sum of the squares of the amplitude.
Zero crossing rate	Number of time-domain zero crossings of the signal per time unit.
Spectral centroid	Geometric center on the frequency scale of the amplitude spectrum.
High-energy—low-energy ratio	Ratio of energy content below and above 1,500 Hz.
Spectral entropy	The relative Shannon entropy, which measures peaks in the auditory spectrum.
Spectral roll-off	Frequency below which contains 85% of the total energy.
Spectral flux	Measure of the temporal changes in the spectrum.
Spectral spread	Standard deviation of the spectrum.
Spectral flatness	Wiener entropy of the spectrum, which measures as the ratio of its geometrical mean to its arithmetical mean.
Sub-Band flux	Measures the fluctuation of frequency content in 10 octave-scaled sub-bands.
Roughness	Estimates the sensory dissonance.
Mode	Strength of major or minor mode.
Key clarity	Measures tonal clarity.
Fluctuation centroid	Estimates the average frequency of rhythmic periodicities.
Fluctuation entropy	Measures the rhythmic complexity.
Pulse clarity	An estimate of the clarity of the pulse.

*Of these, six clusters (Fullness, Brightness, Timbral complexity, Key clarity, Pulse clarity, Activity, and Dissonance) were created for the study by using principal component analysis (PCA). Detailed description of the features can be found from the original source and from the MIR Toolbox manual.*

six global sensory properties, the authors showed that their presence and absence in the musical stimuli indeed did correlate with brain activity. For example, the timbral features (Fullness, Brightness, Timbral complexity, and Activity) were associated positively with activity in the superior temporal gyrus (BA 22) bilaterally and the cerebellum, and negatively with several regions, such as the postcentral gyrus (BA 2, 3), the left precuneus (BA 7), and the inferior parietal gyrus (BA 40). The study shows that such global statistical features do play a role in the musical experience and are indeed meaningful from the point of view of processing of music in our brains.

It remains to be seen, however, whether this approach can be applied to the study of aesthetics. Although, the statistical properties used in our previous studies (Alluri et al., 2012, 2013; Burunat et al., 2016) may be too coarse to be directly relevant for aesthetics, in particular when it comes to the masking problems, the approach is consistent with the hypothesis advanced here. Moreover, the hypothesis that any of such properties were relevant to aesthetics can be tested empirically by correlating the presence of such properties to that of listeners' subjective liking. Promising initial attempts toward that direction, namely combining the fMRI timeseries with continuous or discrete ratings have been made by Trost et al. (2015) and Alluri et al. (2015).

Several challenges must also be met when applying this naturalistic paradigm. Although it allows researchers to use realistic music stimuli, the listening conditions are less than optimal, especially in a fMRI setting, in which noise saturation, low temporal resolution and the risk of false positives in the results (Eklund et al., 2016; Liu et al., 2017) pose considerable methodological challenges to our approach. Even if replicability of brain responses to musical features using the naturalistic paradigm has been shown (mainly for timbral features; Burunat et al., 2016), the concerns for applying the current approach to fMRI data might present bottlenecks that are hard to circumvent. A promising direction would be to utilize silent neurophysiological methodologies with millisecond temporal resolution, such as magnetoencephalography (MEG) and/or electroencephalography (EEG). Two papers have obtained neural correlates of MIR features using EEG signals (Poikonen et al., 2016a,b) and we are studying the application of this approach to MEG data, which allows also a spatial resolution that is almost comparable to that of fMRI. Moreover, we do not wish to imply that this methodology be restricted to brain-imaging settings. It may be applied to behavioral experiments, and indeed many studies done in the naturalistic paradigm do involve behavioral components. In such experiments, participants are asked to evaluate naturalistic stimuli continuously, for example, by providing on-line rating or feedback of the music they are listening (Coutinho and Dibben, 2012). Also, global sensory qualities of a naturalistic stimuli can be independently manipulated in behavioral experiments in order to examine the aesthetic effects of such variables. In our view, it is possible that the optimal results are obtained by utilizing a combination of behavioral and brain-imaging methods. In such hybrid paradigms, many methodological restrictions of purely brain-imaging paradigms can be circumvented by applying behavioral methods, while the brain-imaging studies can provide detailed anatomical, physiological and time-sensitive data unavailable by using behavioral methods alone.

## SENSORY AESTHETICS AS IMMERSION AND AROUSAL

In this section, we consider several possible neural explanations for the link between global sensory qualities and aesthetics. This approach is motivated by the fact that, if global sensory properties indeed are pertinent for the creation of an aesthetic experience, then there must be something in our brains, "some fundamental characteristics of the human nervous system" (Berlyne, 1971 p. 29), that explains that fact. What these fundamental characteristics are indeed constitutes a perennial problem of neuroaesthetics. The notion of global sensory properties could provide a contribution to this debate.

Historically, the search for a common aesthetic quality goes back at least to Bell's work on the aesthetics of art (Bell, 1914). Bell proposes that all art, and especially visual art, shares a universal time- and culture-independent "significant form" that is associated with aesthetic emotions. Bell thought that the form

arises from aesthetic laws pertaining to the configuration of visual features such as lines, shapes, and colors. For Bell, a crucial test for separating aesthetic art from other stimuli was its universality and time-independence: a genuine aesthetic art should be independent of time, culture, and era.

Zeki (2013) provides a modern interpretation of Bell's theory. He begins from the well-known organizational properties of our visual system, according to which the neuronal processing of visual stimuli is distributed over several quasi-independent modules in the brain, each processing its own specialized domain (movement, colors, lines, faces, direction, and such), and then proposes that each of these modules "have a certain, primitive, biologically derived combination [...] of elements for the attribute that it is specialized in processing, and that the aesthetic perception [...] is aroused when, in a composite picture, each of the specialized areas is activated preferentially" (p. 10). Aesthetic perception, according to this hypothesis, has its origin in a "preferential" activation pattern of the early sensory areas specialized in visual perception that will then lead into the activation of interest- and motivation-related brain areas and hence also to an experience of emotion, beauty, and preference (Sachs et al., 2016). This provides a neurobiological interpretation of Bell's original idea. The "significant form" would refer to the fact that some type of preferred activity occurs in various visual regions of the brain, as if each such module would have its own aesthetic principles. Artists are professionals who "create forms that activate the relevant visual areas either optimally or specifically [and] in a way that is different from that obtained by stimuli that lack the significant configuration" (Sachs et al., 2016 p. 9).

We see certain similarities to the case of music. Playing back a stereo track in mono takes away some aspect of its appeal, much in the same way as removing all reverb from a recording makes it dull and lifeless. We hypothesize that this may be because the neural systems wired to detect direction and distance of sound sources are not activated in a natural way, or they are not activated at all. If the spectral energy is further reduced by, say, removing musical material from frequencies below some threshold 500 Hz, the music becomes thinner and, again, loses part of its appeal. The neuronal systems registering lower frequencies receive no input, and therefore contribute nothing to the overall percept. The idea of avoiding too much repetition by introducing constant change derives from the same source: a dull, repeating music ceases to command our attention. In addition, if a musical piece performed by real human beings is replaced by machines mechanically playing sinewave instruments, then the performance loses some of its emotional connotations and, again, some neuronal processes linking auditory signals with emotions that would otherwise be engaged are not involved. Thus, as observed by Baugh (1993), rock music "aims at arousing and expressing feeling" (p. 23) in the listener, which we believe holds the key to sensory aesthetic experience. Music, like vision, is a composite of several qualities (direction, distance, depth, emotion) processed by semi-independent modules in our nervous system, while each such module responds to its own signature properties in the stimuli. It might be that, as in the case of vision, aesthetic appeal originates in a concerted and balanced

activation of all these modules. The global aesthetic properties in music, specifically, are aimed at optimizing the presence and balance of these qualities to keep different neural structures in the brain in a "preferential" activation and connectivity state. This brain state would, according to our hypothesis, lead to "immersion" or arousal in the listener, resulting in a rich, holistic experience (Brattico et al., 2013).

One way to refine this idea is to build on Berlyne's (1971) seminal work on aesthetics and arousal. The notion that aesthetic experience can be traced back to immersion, and especially arousal, was the cornerstone of Berlyne's work on the psychology and biology of aesthetics (e.g., Berlyne, 1971), who in turn followed much of the spirit of Fechner's (1876) pioneering work. Berlyne's main proposition was that the aesthetic experience, and aesthetic pleasure, derives from a change in organism's arousal level. The change could involve decrease (relaxing, tension reduction) or increase (excitement, expectation) of arousal level, and both could be triggered by several properties, among them novelty, surprise, complexity ambiguity for heightened arousal, and repetition, familiarity for reduced arousal levels. The global sensory qualities point toward the same direction. Thus, a sonic object evoking spatial and affective cognition, commanding the whole energy spectrum, and holding listeners' attention will lead to an immersive experience and continuous arousal: by introducing small changes and crafting a careful "building up" the artist creates a musical piece that avoids sensory habituation that would otherwise reduce its impact.

The Fechner-Berlyne approach has been subject to criticism. Their work belonged to the behaviorist-reductionist framework that sought to explain behavior in terms of bare stimulus-response principles. From such reductionist perspective, internal motivation, pleasure, or curiosity present themselves as near-paradoxical problems. An aesthetic object, in particular, is one that the organism is actively seeking to experience, and thus it presents a particularly difficult problem to explain. Berlyne's theory was an attempt to answer this problem. Many of his most strong critics, however, came from a different, humanistic-philosophical tradition involved with history, art criticism and philosophy, in which behaviorist problems played no meaningful role, and from which the whole enterprise appears unnaturally narrow (see Margolis, 1980, for an example). Today such criticism plays a much less significant role (Zeki, 2014; Bundgaard, 2015). The question of what motivates people, and makes objects desirable for them apart from their possible ecological functionality, is as relevant today as it was then. Further, the idea of explaining human behavior in terms of stimuli, brain physiology, and motoric responses cannot be substituted wholly by speculative philosophy, cultural relativism, or art history; modern neuroscience has a role in explaining human behavior. Indeed, there exists a small but active research program inside the neurosciences that can be characterized as "neuroscience of aesthetics" (for recent reviews see Jacobsen, 2006; Chatterjee, 2011; Brattico and Pearce, 2013; Orgs et al., 2013; Chatterjee and Vartanian, 2014, 2016; Pearce et al., 2016). Berlyne was in fact well-aware of the neuroscientific advances of his day, and documents such matters extensively in his work. At the same token, it is also clear that no neuroscientific or

naturalistic exploration can answer questions such as what, ultimately, is art, and what makes a piece of material constellation a genuine work of art instead of a, say, tool or random junk. This is because art is constituted by several non-appearance properties such as its history, intention, sincerity, and normativity, and not everything beautiful or appealing can be said to be art (Bundgaard, 2015). A naturalistic approach to aesthetics (Brown et al., 2011) will, therefore, pay a price in necessarily ignoring many aspects of art that we would regard as important in other contexts.

Berlyne (1971) proposed that the main categories of stimuli that can modulate arousal, and hence in his theory also involve aesthetic appreciation, fall into three distinction categories: psychophysical, ecological, and what he called “collative” (also “structural”). Psychophysical qualities refer to low-level sensory features and changes in such qualities. He mentions in this connection the fact that more intense stimuli are normally interpreted as more arousing. Ecological variables refer to stimuli that are directly associated, either innately or by means of learned association, with survival, pain, and pleasure. Finally, by the term collative or structural properties he means second-order properties that are arrived at by “summing up characteristics of several elements” (p. 69) that may be present simultaneously or could also be temporally distinct. Properties such as novelty, complexity ambiguity and surprisingness belong to this category. The global sensory properties discussed in the present work would, under this scheme, consist of a mixture of structural and psychophysiological properties: they are structural and global, in that they result from the summation of few or many individual qualities, but also sensory, in that they depend on the sensorium and are not constitutively affective or cognitive.

The immersion hypothesis, according to which aesthetic experience results from an activation of all or many brain regions specialized in the processing of the stimuli, leads to testable hypotheses, and empirical predictions. Vartanian and Goel (2004), for example, report that increased preference in the perception of visual art correlates with increased activation in the visual areas of the brain. If sensory immersion and arousal play a role in aesthetic perception, then the expected outcome is precisely that we should attest a positive correlation between aesthetic preference and the activation of the various brain areas involved in the processing of the stimuli. This prediction was also confirmed in a meta-analysis, likewise reporting an association between visual aesthetic experience and a wide-spread rather than localized brain activation (Boccia et al., 2016). In the case of music, the prediction is that the removal of relevant features, whether spatial, emotional, or spectral, should lead into a marked decrease both in the brain activation and in the aesthetic judgment. Crucially, our hypothesis predicts that this effect should not depend on local features, and should be observed entirely irrespective of musical genre, style, or (representational) content. If, in other words, the aesthetic balance in sensory qualities is achieved by means of immersion, itself based on the concurrent activation of the relevant brain regions, then what matters is the activation itself and not the particular local features present in the activating stimulus. This hypothesis could thus further be tested by invoking experimental top-down effects that

suffice to satisfy the activation condition without the presence of concrete stimulus.

However, this hypothesis predicts, if interpreted in a too simple way, that increasing the amplitude of any or all such features should always lead to increased liking. Oversaturated objects, such as overly loud music or pictures with bright colors, are not perceived as beautiful; instead, they can be perceived even as painful. Too much reverb, stereo widening or emotional expressivity makes the music incomprehensible and “wishy-washy.” This question has always puzzled those trying to understand aesthetic perception. Berlyne’s solution was to assume that stimulus levels beyond a certain moderate cutoff point would begin to activate “aversion systems” that are associated with a negative outcome (danger, unpleasantness). Zeki (2013) discusses this problem and points out that the determining factor cannot be the strength of the activation as such but, rather, there must be some quality in the original signal that prompts the positive response. He provides another interpretation of these results, according to which “it is not the strongest or maximal activity that correlates with preference but rather a specific activity that becomes optimal when stimuli of the right [aesthetic properties] are viewed” (p. 10). Hence, we are back at Bell’s mystery: there is an unknown quality in the stimulus that is preferred by the various regions in the brain specialized in processing that type of stimuli.

The case of music provides another possible interpretation. It is an established fact that the masking effect on one sound over another is amplified by the amplitude of the former. Thus, as the sound is increased in amplitude, the range of frequencies it will mask will also increase. Moreover, if we are presented with a piece of music in which one instrument is associated with overwhelming volume, our brains will attempt to adapt to the situation by attenuating the overall level. This will further reduce the perceived relative amplitudes of the rest of the musical information. Finally, the problem might not be as severe if the amplitude of all sound sources is increased in tandem, which corresponds to an increase in overall volume. Thus, we might be dealing, not with overall amplitude, but with relative amplitudes. It is possible that the reason why balanced performances instead of overly saturated ones are crucial for auditory aesthetics is because the former specifically avoids unaesthetic masking and thus keeps the musical sources distinct. This hypothesis could be tested experimentally. If the problem with amplitude concerns relative amplitudes and/or masking, then the same negative effect on aesthetic experience could be achieved by using other types of masking (noise masking) and/or also by decreasing an amplitude of a sound source relative to other sound sources.

There is another intriguing possibility. The causal relation between global sensory properties and aesthetics could be further captured in terms of the processing fluency hypothesis, as proposed in the domain of visual aesthetics (Reber et al., 2004; Babel and McGuire, 2015; Forster et al., 2015). For instance, if the crucial feature concerns the distinctiveness of each musical source in the absence of feature masking, then it is possible that the phenomenon reduces further to the notion of processing fluency (Reber et al., 2004), namely the relation between a positive aesthetic response and the ease of processing

in encoding and representing e.g., distinct sound sources. This hypothesis predicts that aiding the encoding of music and its sound sources by visual means, for example, by exhibiting the performance itself, should increase the aesthetic appeal of the piece irrespectively of whether the sound sources are overlapping or not. On the other hand, a dull, unsaturated but fluently processed sonic object might be less appealing than a complex one that incorporates the whole frequency and spatial spectrum, an obvious problem for the fluency hypothesis. This problem could be solved by combining the immersion hypothesis with the processing fluency hypothesis. Accordingly, perhaps an aesthetic appreciation requires a concerted and concurrent activation of all the relevant modules that participate in the processing of the stimulus, as assumed by the immersion hypothesis, but with the additional requirement that each module has to be able to process its input in a fluent and efficient manner, as assumed by the processing fluency theory. Under this hypothesis, the masking phenomenon linked with musical aesthetics would be interpreted as a distracting event that hinders fluent processing in any of the relevant submodules.

If we, instead, assume Zeki's hypothesis that there are "significant forms" that, by causing the various sensory submodules to enter their "preferred states," lead into aesthetic appreciation, then a rigorous definition of "significant form" is required. Zeki (2013) discusses the example of human faces in this connection. Humans have an inborn preference for perceiving, representing, and interpreting human faces, and there are specific neuronal resources dedicated to this task. These visual systems respond selectively to the properties of human faces and, moreover, some such features are perceived as more attractive than others. There are biological and evolutionary reasons why such preferences would inhabit our visual system, and the same phenomenon of "mate selection" is observed throughout the animal kingdom. The same argument can be found from several views concerning visual aesthetic, cited earlier in this paper. The idea is that the global sensory properties are shared with the biologically preferred visual images, such as natural landscapes or potential mates, which would then explain artistic preferences as a halo effect of the originally more mundane mechanism. While we do not wish to propose that all aesthetic perception derives from preferred tuning of the various sensory systems for mate selection, landscape detection, or healthy nutrition detection, this view provides a plausible argument for the *existence* of such mechanisms. Preference of certain types of mates, environments, foods, and tastes, for example, is something that our brains must be hardwired to do, although also learning and cultural exposure have an effect, while it is possible that such preferences spill over non-functionally to the perception of many types of objects, and even to abstract objects such as music. This hypothesis could be labeled as the *ecological hypothesis*. It has

been pursued in the domain of vision by examining whether global statistical sensory properties of ecological stimuli, such as natural landscapes or faces, lead into aesthetic experience when they are embedded in the context of abstract art objects or other visual stimuli. These experiments could be replicated in the case of music by extracting global statistical sensory properties from ecological sounds (wind, rain, human voice, crying, laughing) and replicating then synthetically in music or in music-type stimuli to determine if their aesthetic value can be modulated.

## CONCLUSIONS

We put forward a research agenda for studying holistic qualities of musical objects that likely play an important role in creating an aesthetic response in the listener. We propose that these global features are statistically extracted from the stimuli by our auditory system—or, rather, by some subsystems (McDermott and Simoncelli, 2011; McDermott et al., 2013)—and then passed on to high-level processing, ultimately leading to the main outcomes of a musical experience, namely aesthetic judgment, emotion and conscious liking, or preference (Cela-Conde et al., 2011; Brattico et al., 2013). A shift of paradigm from conventional studies using artificial stimulation, block design, and subtraction analysis methods toward novel naturalistic paradigms with non-conventional analysis methods based on MIR combined with brain time series is called upon to accurately measure and determine the effects of global properties on brain functioning and behavior. We also discussed several possible neuronal implementations of this general hypothesis: the immersion hypothesis, processing fluency hypothesis, and the ecological hypothesis. The immersion hypothesis claims that aesthetic experience results in a concerted activation of many or all critical brain regions involved in the processing of the stimuli, irrespectively of other stimulus content; the processing fluency requires that the stimuli can be processed effortlessly by the brain; and the ecological hypothesis contents that the modules have to enter into a "preferred" neural state that is further determined by ecological conditions. Another possibility is that they all play a role.

## AUTHOR CONTRIBUTIONS

PB and EB conceived the hypotheses of this paper. PB wrote most of the manuscript whereas EB wrote some parts of it. PV edited the manuscript and contributed to financing the work.

## ACKNOWLEDGMENTS

This work has been funded by the Danish National Research Foundation (project number DNR117).

## REFERENCES

- Alluri, V., Toiviainen, P., Burunat, I., Bogert, B., Numminen, J., and Brattico, E. (2015). Musical expertise modulates functional connectivity of limbic regions during continuous music listening. *Psychomusicology* 25, 443–454. doi: 10.1037/pmu0000124
- Alluri, V., Toiviainen, P., Jaaskelainen, 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
- Alluri, V., Toiviainen, P., Lund, T. E., Wallentin, M., Vuust, P., Nandi, A. K., et al. (2013). From Vivaldi to Beatles and back: predicting lateralized brain

- responses to music. *Neuroimage* 83, 627–636. doi: 10.1016/j.neuroimage.2013.06.064
- Altmann, C. F., Terada, S., Kashino, M., Goto, K., Mima, T., Fukuyama, H., et al. (2014). Independent or integrated processing of interaural time and level differences in human auditory cortex? *Hear. Res.* 312, 121–127. doi: 10.1016/j.heares.2014.03.009
- Babel, M., and McGuire, G. (2015). Perceptual fluency and judgments of vocal aesthetics and stereotypicality. *Cogn. Sci.* 39, 766–787. doi: 10.1111/cogs.12179
- Barbour, D., and Wang, X. (2003). Contrast tuning auditory cortex. *Science* 299, 1073–1075. doi: 10.1126/science.1080425
- Baugh, B. (1993). Prolegomena to any aesthetics of rock music. *J. Aesthet. Art Crit.* 51, 23–29. doi: 10.2307/431967
- Bell, C. (1914). *Art*. London: Chatto and Windus. All citations are from the Kindle Edition.
- Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. New York, NY: Appleton-Century-Crofts.
- Berns, G. S., Capra, C. M., Moore, S., and Noussair, C. (2010). Neural mechanisms of the influence of popularity on adolescent ratings of music. *Neuroimage* 49, 2687–2696. doi: 10.1016/j.neuroimage.2009.10.070
- Boccia, M., Sarbetti, S., Piccardi, L., Guariglia, C., Ferlazzo, F., Giannini, A. M., et al. (2016). Where does brain neural activation in aesthetic responses to visual art occur? Meta-analytic evidence from neuroimaging studies. *Neurosci. Biobehav. Rev.* 60, 65–71. doi: 10.1016/j.neubiorev.2015.09.009
- Bogert, B., Numminen-Kontti, T., Gold, B., Sams, M., Numminen, J., Burunat, I., et al. (2016). Hidden sources of joy, fear, and sadness: explicit versus implicit neural processing of musical emotions. *Neuropsychologia* 89, 393–402. doi: 10.1016/j.neuropsychologia.2016.07.005
- Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968–1987. *Psychol. Bull.* 106, 265–289. doi: 10.1037/0033-2909.106.2.265
- Branković, S. (2013). Neuroaesthetics and growing interest in “positive affect” in psychiatry: new evidence and prospects for the theory of informational needs. *Psychiatr. Danub.* 25, 97–107.
- Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., et al. (2011). A functional MRI study of happy and sad emotions in music with and without lyrics. *Front. Psychol.* 2:308. doi: 10.3389/fpsyg.2011.00308
- Brattico, E., Bogert, B., Alluri, V., Tervaniemi, M., Eerola, T., and Jacobsen, T. (2015). 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
- Brattico, E., Bogert, B., and Jacobsen, T. (2013). Toward a neural chronometry for the aesthetic experience of music. *Front. Psychol.* 4:206. doi: 10.3389/fpsyg.2013.00206
- Brattico, E., Brattico, P., and Jacobsen, T. (2009–2010). The origins of the aesthetic enjoyment of music - A review of the literature. *Music. Sci. Special Issue* 2009–2010, 15–31. doi: 10.1177/1029864909013002031
- Brattico, E., and Pearce, M. T. (2013). The neuroaesthetics of music. *Psychol. Aesthet. Creat. Arts* 7, 48–61. doi: 10.1037/a0031624
- Bridgett, R. (2013). “Contextualizing game audio aesthetics,” in *The Oxford Handbook of New Audiovisual Aesthetics*, eds J. Richardson, C. Gorbman, and C. Vernallis (Oxford: Oxford University Press), 563–571.
- Brown, S., Gao, X., Tisdelle, L., Eickhoff, S. B., and Liotti, M. (2011). Naturalizing aesthetics: brain areas for aesthetic appraisal across sensory modalities. *Neuroimage* 58, 250–258. doi: 10.1016/j.neuroimage.2011.06.012
- Bundgaard, H. (2015). Feeling, meaning, and intentionality—a critique of the neuroaesthetics of beauty. *Phenom. Cogn. Sci.* 14, 781–801. doi: 10.1007/s11097-014-9351-5
- Burunat, I., Toivainen, P., Alluri, V., Bogert, B., Ristaniemi, T., Sams, M., et al. (2016). The reliability of continuous brain responses during naturalistic listening to music. *Neuroimage* 124(Pt A), 224–231. doi: 10.1016/j.neuroimage.2015.09.005
- Cela-Conde, C. J., Agnati, L., Huston, J. P., Mora, F., and Nadal, M. (2011). The neural foundations of aesthetic appreciation. *Prog. Neurobiol.* 94, 39–48. doi: 10.1016/j.pneurobio.2011.03.003
- Chatterjee, A. (2011). Neuroaesthetics: a coming of age story. *J. Cogn. Neurosci.* 23, 53–62. doi: 10.1162/jocn.2010.21457
- Chatterjee, A., and Vartanian, O. (2014). Neuroaesthetics. *Trends Cogn. Sci.* 18, 370–375. doi: 10.1016/j.tics.2014.03.003
- Chatterjee, A., and Vartanian, O. (2016). Neuroscience of aesthetics. *Ann. N.Y. Acad. Sci.* 1369, 172–194. doi: 10.1111/nyas.13035
- Chapman, A. J., and Williams, A. R. (1976). Prestige effects and aesthetic experiences: adolescents' reactions to music. *Br. J. Soc. Clin. Psychol.* 15, 61–72. doi: 10.1111/j.2044-8260.1976.tb00007
- Colin, C., Radeau, M., Soquet, A., Dachy, B., and Deltenre, P. (2002). Electrophysiology of spatial scene analysis: the mismatch negativity (MMN) is sensitive to the ventriloquism illusion. *Clin. Neurophysiol.* 113, 507–518. doi: 10.1016/S1388-2457(02)00028-7
- Corozine, V. (2002). *Arranging Music for the Real World: Classical and Commercial Aspects*. Pacific, MO: Mel Bay.
- Coutinho, E., and Dikken, N. (2012). Psychoacoustic cues to emotion in speech prosody and music. *Cogn. Emot.* 27, 1–27. doi: 10.1080/02699931.2012.732559
- Dalla Bella, S., Peretz, I., Rousseau, L., and Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition* 80, B1–B10. doi: 10.1016/S0010-0277(00)00136-0
- Eklund, A., Nichols, T. E., and Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7900–7905. doi: 10.1073/pnas.1602413113
- Fechner, G. T. (1876). *Vorschule der Aesthetik [Experimental Aesthetics]*. Leipzig: Breitkopf & Härtel.
- Fitch, W. T. (2016). Dance, music, meter and groove: a forgotten partnership. *Front. Hum. Neurosci.* 10:64. doi: 10.3389/fnhum.2016.00064
- Forster, M., Gerger, G., and Leder, H. (2015). Everything's relative? Relative differences in processing fluency and the effects on liking. *PLoS ONE* 10:e0135944. doi: 10.1371/journal.pone.0135944
- Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C. J., and Sawey, M. (2011). Predicting beauty: fractal dimension and visual complexity in art. *Br. J. Psychol.* 102, 49–70. doi: 10.1348/000712610X498958
- Gabrielsson, A. (2011). “Emotions in strong experiences with music,” in *Handbook of Music and Emotion: Theory, Research, Applications*, eds P. N. Juslin and J. A. Sloboda (Oxford: Oxford University Press), 547–574.
- Gabrielsson, A., and Juslin, P. N. (1996). Emotional expression in music performance: between the performer's intention and the listener's experience. *Psychol. Music* 24, 68–91. doi: 10.1177/0305735696241007
- Graham, D. J., and Field, D. J. (2007). Statistical regularities of art images and natural scenes: spectra, sparseness and nonlinearities. *Spat. Vis.* 21, 149–164. doi: 10.1163/156856807782753877
- Green, A. C., Baerentsen, K. B., Stodkilde-Jørgensen, H., Roepstorff, A., and Vuust, P. (2012). Listen, learn, like! Dorsolateral prefrontal cortex involved in the mere exposure effect in music. *Neurol. Res. Int.* 2012:846270. doi: 10.1155/2012/846270
- Grewe, O., Nagel, F., Kopiez, K., and Altenmüller, E. (2007). Listening to music as a re-creative process: physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Percept.* 297, 297–314. doi: 10.1525/mp.2007.24.3.297
- Hagerhall, C. M., Purcell, T., and Taylor, R. (2004). Fractal dimension of landscape silhouette – outlines as a predictor of landscape preference. *J. Environ. Psychol.* 24, 247–255. doi: 10.1016/j.jenvp.2003.12.004
- Heingartner, A., and Hall, J. V. (1974). Affective consequences in adults and children of repeated exposure to auditory stimuli. *J. Pers. Soc. Psychol.* 29, 719–723. doi: 10.1037/h0036121
- Hekkert, P. (2006). Design aesthetics: principles of pleasure in design. *Psychol. Sci.* 48, 157–172.
- Hevner, K. (1935). The affective character of major and minor modes in music. *Am. J. Psychol.* 47, 103–118. doi: 10.2307/1416710
- Hodges, D. A. (2016). “The neuroaesthetics of music,” in *The Oxford Handbook of Music Psychology*, eds S. Hallam, I. Cross, and M. Thaut (Oxford: Oxford University Press), 247–262.
- Huckvale, D. (1990). Twins of Evil: an investigation into the aesthetics of film music. *Pop. Music* 9, 1–36. doi: 10.1017/S0261143000003718
- Hunter, P. G., Schellenberg, E. G., and Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cogn. Emot.* 22, 327–352. doi: 10.1080/02699930701438145
- Husain, G., Thompson, W. F., and Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Percept.* 20, 151–171. doi: 10.1525/mp.2002.20.2.151

- Istok, E., Brattico, E., Jacobsen, T., Krohn, K., Mueller, M., and Tervaniemi, M. (2009). Aesthetic responses to music: a questionnaire study. *Music. Sci.* 13, 183–206. doi: 10.1177/102986490901300201
- Istok, E., Brattico, E., Jacobsen, T., Ritter, A., and Tervaniemi, M. (2013). 'I love Rock 'n' Roll'—music genre preference modulates brain responses to music. *Biol. Psychol.* 92, 142–151. doi: 10.1016/j.biopsycho.2012.11.005
- Jacobs, R. H., Haak, K. V., Thumfart, S., Renken, R., Henson, B., and Cornelissen, F. W. (2016). Aesthetics by numbers: links between perceived texture qualities and computed visual texture properties. *Front. Hum. Neurosci.* 10:343. doi: 10.3389/fnhum.2016.00343
- Jacobsen, T. (2004). The primacy of beauty in judging the aesthetics of objects. *Psychol. Rep.* 94, 1253–1260. doi: 10.2466/pr0.94.3.1253-1260
- Jacobsen, T. (2006). Bridging the arts and sciences: a framework for the psychology of aesthetics. *Leonardo* 39, 155–162. doi: 10.1162/leon.2006.39.2.155
- Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cereb. Cortex* 19, 2579–2594. doi: 10.1093/cercor/bhp008
- Janata, P., Tomic, S. T., and Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *J. Exp. Psychol.* 141, 54–75. doi: 10.1037/a0024208
- Johnson, J. K., Chang, C. C., Brambati, S. M., Migliaccio, R., Gorno-Tempini, M. L., Miller, B. L., et al. (2011). Music recognition in frontotemporal lobar degeneration and Alzheimer disease. *Cogn. Behav. Neurol.* 24, 74–84. doi: 10.1097/WNN.0b013e31821de326
- Julesz, B. (1962). Visual pattern discrimination. *IRE Trans. Inform. Theory* 8, 84–92. doi: 10.1109/TIT.1962.1057698
- Juslin, P. N., Sakka, L., Barradas, G., and Liljeström, S. (2016). No accounting for taste? Idiographic models of aesthetic judgment in music. *Psychol. Aesthet. Creat. Arts* 10, 157–170. doi: 10.1037/aca0000034
- Khalfa, S., Schon, D., Anton, J. L., and Liegeois-Chauvel, C. (2005). Brain regions involved in the recognition of happiness and sadness in music. *Neuroreport* 16, 1981–1984. doi: 10.1097/00001756-200512190-00002
- Kilchenmann, L., and Senn, O. (2015). Microtiming in Swing and Funk affects the body movement behavior of music expert listeners. *Front. Psychol.* 6:1232. doi: 10.3389/fpsyg.2015.01232
- Koelsch, S. (2011). Toward a neural basis of music perception - a review and updated model. *Front. Psychol.* 2:110. doi: 10.3389/fpsyg.2011.00110
- Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666
- Laeng, B., Eidet, L. M., Sulutvedt, U., and Panksepp, J. (2016). Music chills: the eye pupil as a mirror to music's soul. *Conscious. Cogn.* 44, 161–178. doi: 10.1016/j.concog.2016.07.009
- Lartillot, O., and Toivainen, P. (2007). "A Matlab toolbox for musical feature extraction from audio," in *Paper Presented at the 10th International Conference on Digital Audio Effects (DAFx-07)* (Bordeaux).
- 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, C., Abu-Jamous, B., Brattico, E., and Nandi, A. K. (2017). Towards tunable consensus clustering for studying functional brain connectivity during affective processing. *Int. J. Neural Syst.* 27:1650042. doi: 10.1142/S0129065716500428
- Mallon, B., Redies, C., and Hayn-Leichsenring, G. U. (2014). Beauty in abstract paintings: perceptual contrast and statistical properties. *Front. Hum. Neurosci.* 8:161. doi: 10.3389/fnhum.2014.00161
- Margolis, J. (1980). *Art and Philosophy*. Atlantic Highlands, NJ: Humanities Press.
- McDermott, J. H., and Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron* 71, 926–940. doi: 10.1016/j.neuron.2011.06.032
- McDermott, J. H., Schemitsch, M., and Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nat. Neurosci.* 16, 493–498. doi: 10.1038/nn.3347
- McPherson, G. E., and Schubert, E. (2004). "Measuring performance enhancement in music," in *Musical Excellence: Strategies and Techniques to Enhance Performance*, ed A. Williamon (Oxford: Oxford University Press), 61–82.
- Menzel, C., Hayn-Leichsenring, G. U., Langner, O., Wiese, H., and Redies, C. (2015). Fourier power spectrum characteristics of face photographs: attractiveness perception depends on low-level image properties. *PLoS ONE* 10:e0122801. doi: 10.1371/journal.pone.0122801
- Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago, IL; London: University of Chicago Press.
- Moffat, D., Roman, D., and Reiss, J. D. (2015). "An evaluation of audio feature extraction toolboxes," in *Proceedings of the 18th International Conference on Digital Audio Effects (DAFx-15)* (Trondheim).
- Montag, C., Reuter, M., and Axmacher, N. (2011). How one's favorite song activates the reward circuitry of the brain: personality matters! *Behav. Brain Res.* 225, 511–514. doi: 10.1016/j.bbr.2011.08.012
- Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing*, 6th Edn. Bingley: Emerald.
- Mureika, J. R., Cupchik, G. C., and Dyer, C. C. (2004). Multifractal fingerprints in the visual arts. *Leonardo* 37, 53–56. doi: 10.1162/00240940472828139
- Nieminen, S., Istók, E., Brattico, E., and Tervaniemi, M. (2012). The development of the aesthetic experience of music: preference, emotions, and beauty. *Music. Sci.* 16, 372–391. doi: 10.1177/1029864912450454
- Nieminen, S., Istok, E., Brattico, E., Tervaniemi, M., and Huottilainen, M. (2011). The development of aesthetic responses to music and their underlying neural and psychological mechanisms. *Cortex* 47, 1138–1146. doi: 10.1016/j.cortex.2011.05.008
- Orgs, G., Hagura, N., and Haggard, P. (2013). Learning to like it: aesthetic perception of bodies, movements and choreographic structure. *Conscious. Cogn.* 22, 603–612. doi: 10.1016/j.concog.2013.03.010
- Pallesen, J. K., Brattico, E., and Carlsson, S. (2003). Emotional connotations of major and minor musical chords in musically untrained listeners. *Brain Cogn.* 51, 188–190. doi: 10.1016/S0278-2626(02)00541-9
- Panksepp, J. (1995). The emotional sources of "chills" induced by music. *Music Percept.* 13, 171–207. doi: 10.2307/40285693
- Pearce, M. T., Zaidel, D. W., Vartanian, O., Skov, M., Leder, H., Chatterjee, A., et al. (2016). Neuroaesthetics: the cognitive neuroscience of aesthetic experience. *Perspect. Psychol. Sci.* 11, 265–279. doi: 10.1177/1745691615621274
- Peeters, G. (2004). *A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project*. Technical Report, IRCAM.
- Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., and Brattico, E. (2011). Music and emotions in the brain: familiarity matters. *PLoS ONE* 6:e27241. doi: 10.1371/journal.pone.0027241
- Peretz, I., Gaudreau, D., and Bonnel, A.-M. (1998). Exposure effects on music preference and recognition. *Mem. Cogn.* 26, 884–902. doi: 10.3758/BF03201171
- Plomp, R. (1964). The ear as a frequency analyzer. *J. Acoust. Soc. Am.* 36, 1628–1636. doi: 10.1121/1.1919256
- Poikonen, H., Alluri, V., Brattico, E., Lartillot, O., Tervaniemi, M., and Huottilainen, M. (2016a). Event-related brain responses while listening to entire pieces of music. *Neuroscience* 312, 58–73. doi: 10.1016/j.neuroscience.2015.10.061
- Poikonen, H., Toivainen, P., and Tervaniemi, M. (2016b). Early auditory processing in musicians and dancers during a contemporary dance piece. *Sci. Rep.* 6:33056. doi: 10.1038/srep33056
- Porcello, T. (2004). Speaking of sound: language and the professionalization of sound-recording engineers. *Soc. Stud. Sci.* 34/5, 733–758. doi: 10.1177/0306312704047328
- Račić, L. (1981). On the aesthetics of rock music. *Int. Rev. Aesthet. Sociol.* 12, 199–202. doi: 10.2307/836562
- Ramachandran, V. S., and Hirstein, W. (1999). The science of art – A neurological theory of aesthetic experience. *J. Conscious. Stud.* 6, 15–51.
- 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
- Redies, C. (2007). A universal model of esthetic perception based on the sensory coding of natural stimuli. *Spat. Vis.* 21, 97–117. doi: 10.1163/156856807782753886
- Redies, C. (2015). Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Front. Hum. Neurosci.* 9:218. doi: 10.3389/fnhum.2015.00218
- Renoult, J. P., Bovet, J., and Raymond, M. (2016). Beauty is in the efficient coding of the beholder. *R. Soc. Open Sci.* 3:160027. doi: 10.1098/rsos.160027
- Rickard, N. S. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychol. Music* 32, 371–388. doi: 10.1177/0305735604046096

- Robinson, J. (1994). The expression and arousal of emotion in music. *J. Aesthet. Art Crit.* 52, 13–22. doi: 10.2307/431581
- Roeber, U., Widmann, A., and Schroger, E. (2003). Auditory distraction by duration and location deviants: a behavioral and event-related potential study. *Brain Res. Cogn. Brain Res.* 17, 347–357. doi: 10.1016/S0926-6410(03)00136-8
- Sachs, M. E., Ellis, R. J., Schlaug, G., and Loui, P. (2016). Brain connectivity reflects human aesthetic responses to music. *Soc. Cogn. Affect. Neurosci.* 11, 884–891. doi: 10.1093/scan/nsw009
- 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., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., and Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science* 340, 216–219. doi: 10.1126/science.1231059
- Schellenberg, E. G., Peretz, I., and Viellard, S. (2008). Liking for happy- and sad-sounding music: effects of exposure. *Cogn. Emot.* 22, 218–237. doi: 10.1080/02699930701350753
- Seashore, C. E. (1929). A base for the approach to quantitative studies in the aesthetics of music. *Am. J. Psychol.* 39, 141–144. doi: 10.2307/1415406
- Senior, M. (2011). *Mixing Secrets for the Small Studio*. Amsterdam: Elsevier. Kindle Edition.
- Sioros, G., Miron, M., Davies, M., Gouyon, F., and Madison, G. (2014). Syncopation creates the sensation of groove in synthesized music examples. *Front. Psychol.* 5:1036. doi: 10.3389/fpsyg.2014.01036
- Sloboda, J. A. (1991). Music structure and emotional response: some empirical findings. *Psychol. Music* 19, 110–120. doi: 10.1177/0305735691192002
- Spehar, B., Clifford, C., Newell, B., and Taylor, R. P. (2003). Universal aesthetic of fractals. *Comput. Graph.* 37, 813–820. doi: 10.1016/S0097-8493(03)00154-7
- Spehar, B., Wong, S., van de Klundert, S., Lui, J., Clifford, C. W., and Taylor, R. P. (2015). Beauty and the beholder: the role of visual sensitivity in visual preference. *Front. Hum. Neurosci.* 9:514. doi: 10.3389/fnhum.2015.00514
- Taylor, R. P., Micolich, A. P., and Jonas, D. (1999). Fractal analysis of Pollock's drip paintings. *Nature* 399, 422–422. doi: 10.1038/20833
- Taylor, R. P., Spehar, B., Van Donkelaar, P., and Hagerhall, C. M. (2011). Perceptual and physiological responses to jackson pollock's fractals. *Front. Hum. Neurosci.* 5:60. doi: 10.3389/fnhum.2011.00060
- Trost, W., Fruhholz, S., Cochrane, T., Cojan, Y., and Vuilleumier, P. (2015). Temporal dynamics of musical emotions examined through intersubject synchrony of brain activity. *Soc. Cogn. Affect. Neurosci.* 10, 1705–1721. doi: 10.1093/scan/nsv060
- Vartanian, O., and Goel, V. (2004). Neuroanatomical correlates of aesthetic preference for paintings. *Neuroreport* 15, 893–897. doi: 10.1097/01.wnr.0000118723.38067.d6
- von Appen, R. (2007). On the aesthetics of popular music. *Music Therapy Today* 8, 5–25.
- 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
- Vuust, P., Gebauer, L. K., and Witek, M. A. (2014). “Neural underpinnings of music: the polyrhythmic brain,” in *Neurobiology of Interval Timing, Advances in Experimental Medicine and Biology* 829, eds H. Merchant and V. de Lafuente (New York, NY: Springer), 339–356.
- Witek, M. A., Clarke, E. F., Wallentin, M., Kringelbach, M. L., and Vuust, P. (2014). Syncopation, body-movement and pleasure in groove music. *PLoS ONE* 9:e94446. doi: 10.1371/journal.pone.0094446
- Wurtz, R. H., and Kandel, E. R. (2000). “Central visual pathways,” in *Principles of Neural Science, 4th Edn.* eds E. R. Kandel, J. H. Schwartz, and T. M. Jessell (New York, NY: McGraw-Hill), 523–547.
- Zahorik, P. (2002). Assessing auditory distance perception using virtual acoustics. *J. Acoust. Soc. Am.* 111, 1832–1846. doi: 10.1121/1.1458027
- Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46. doi: 10.1016/S1364-6613(00)01816-7
- Zeki, S. (1999). *Inner Vision - An Exploration of Art and the Brain*. Oxford: Oxford University Press.
- Zeki, S. (2013). Clive Bell's “Significant Form” and the neurobiology of aesthetics. *Front. Hum. Neurosci.* 7:730. doi: 10.3389/fnhum.2013.00730
- Zeki, S. (2014). Neurobiology and the humanities. *Neuron* 84, 12–14. doi: 10.1016/j.neuron.2014.09.016
- Zuckerman, M., Levy, D. A., Tibon, R., Reggev, N., and Maril, A. (2012). Does this ring a bell? Music-cued retrieval of semantic knowledge and metamemory judgments. *J. Cogn. Neurosci.* 24, 2155–2170. doi: 10.1162/jocn\_a\_00271
- Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands. *J. Acoust. Soc. Am.* 33, 248. doi: 10.1121/1.1908630

**Conflict of Interest Statement:** 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.

Copyright © 2017 Brattico, Brattico and Vuust. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Constituents of Music and Visual-Art Related Pleasure – A Critical Integrative Literature Review

Marianne Tiihonen<sup>1,2\*</sup>, Elvira Brattico<sup>2\*</sup>, Johanna Maksimainen<sup>3</sup>, Jan Wikgren<sup>4</sup> and Suvi Saarikallio<sup>1</sup>

<sup>1</sup> Finnish Centre for Interdisciplinary Music Research, Department of Music, Art and Culture Studies, University of Jyväskylä, Jyväskylä, Finland, <sup>2</sup> Center for Music in the Brain, Department of Clinical Medicine, Aarhus University and The Royal Academy of Music, Aarhus/Aalborg, Aarhus, Denmark, <sup>3</sup> Max Planck Institute for Empirical Aesthetics, Department of Music, Frankfurt, Germany, <sup>4</sup> Centre for Interdisciplinary Brain Research, Department of Psychology, University of Jyväskylä, Jyväskylä, Finland

## OPEN ACCESS

### Edited by:

Piotr Podlipniak,  
Adam Mickiewicz University  
in Poznań, Poland

### Reviewed by:

Dan Lloyd,  
Trinity College, United States  
Sofia Dahl,  
Aalborg University, Denmark

### \*Correspondence:

Marianne Tiihonen  
marianne.t.tiihonen@student.jyu.fi  
Elvira Brattico  
elvira.brattico@clin.au.dk

### Specialty section:

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

**Received:** 16 November 2016

**Accepted:** 03 July 2017

**Published:** 20 July 2017

### Citation:

Tiihonen M, Brattico E,  
Maksimainen J, Wikgren J and  
Saarikallio S (2017) Constituents  
of Music and Visual-Art Related  
Pleasure – A Critical Integrative  
Literature Review.  
Front. Psychol. 8:1218.  
doi: 10.3389/fpsyg.2017.01218

The present literature review investigated how pleasure induced by music and visual-art has been conceptually understood in empirical research over the past 20 years. After an initial selection of abstracts from seven databases (keywords: pleasure, reward, enjoyment, and hedonic), twenty music and eleven visual-art papers were systematically compared. The following questions were addressed: (1) What is the role of the keyword in the research question? (2) Is pleasure considered a result of variation in the perceiver's internal or external attributes? (3) What are the most commonly employed methods and main variables in empirical settings? Based on these questions, our critical integrative analysis aimed to identify which themes and processes emerged as key features for conceptualizing art-induced pleasure. The results demonstrated great variance in how pleasure has been approached: In the music studies pleasure was often a clear object of investigation, whereas in the visual-art studies the term was often embedded into the context of an aesthetic experience, or used otherwise in a descriptive, indirect sense. Music studies often targeted different emotions, their intensity or anhedonia. Biographical and background variables and personality traits of the perceiver were often measured. Next to behavioral methods, a common method was brain imaging which often targeted the reward circuitry of the brain in response to music. Visual-art pleasure was also frequently addressed using brain imaging methods, but the research focused on sensory cortices rather than the reward circuit alone. Compared with music research, visual-art research investigated more frequently pleasure in relation to conscious, cognitive processing, where the variations of stimulus features and the changing of viewing modes were regarded as explanatory factors of the derived experience. Despite valence being frequently applied in both domains, we conclude, that in empirical music research pleasure seems to be part of core affect and hedonic tone modulated by stable personality variables, whereas in visual-art research pleasure is a result of the so called conceptual act depending on a chosen strategy to approach art. We encourage an integration of music and visual-art into to a multi-modal framework to promote a more versatile understanding of pleasure in response to aesthetic artifacts.

**Keywords:** music, visual-art, pleasure, reward, enjoyment, aesthetic experience

## INTRODUCTION

When considering human behavior in general, striving for pleasure and reward seems to be an integral part of human behavior and a driving force in humans and in animals (Kringelbach and Berridge, 2010a,b). Indeed, pleasure, including positive and negative affect, is related to processes crucial for survival and adaptive functions; it is involved in the regulation of procreation, food intake and motivation, also it is considered a core affect in some of the main emotion models (Russel, 1980; Ledoux, 2000; Barrett, 2006; Nesse, 2012). Thus, it seems that we continuously evaluate the sensory input from our environment according to our internal states of needs and desires (Cabanac, 1971). Regarding types of pleasure, Berridge and Kringelbach (2008) separated basic pleasures (sensory and social) from those of higher-order (monetary, artistic, altruistic, musical, and transcendent), considering arts in general as higher-order pleasures. Yet, it has also been suggested that music and visual-art are not restricted to that of the higher-order pleasure. Brattico and Pearce (2013) advocated for a distinction between *immediate sensory pleasure* and *reflective process of enjoyment* in regard to music. It also seems to be common to many models of visual-art to integrate low-level feature analysis relying on the visual sensory system (bottom-up) and higher-order ways to give meanings to artworks (top-down) when aiming to explain experiences derived from visual-art (Pelowski et al., 2016). Indeed, already in Fechner's (1876) "Vorschule der Ästhetik" the research of visual and auditory elements was under the same label of "aesthetics from below," where the bottom-up mechanisms of music and visual-art were considered eventually to explain the top-down mechanisms of art enjoyment in general. Since then the scientific take on the influence of music and visual-art has not only become broader, causing the fields to split into several smaller sub-disciplines, but the empirical research on music and on visual-art has grown apart, and the term aesthetic experience seems to be more characteristic to the research on objects and artifacts perceived visually (Hargreaves and North, 2010; Brattico et al., 2013; Hodges, 2016). Despite overlapping research questions, and the assumption that the same components (perception, production, response and interaction) govern both, pleasure derived from music and visual-art, and art appreciation in general (Chatterjee, 2011; Bullo and Reber, 2013), empirical research of visual-art and music has had relatively little dialog with each other in the recent years.

Considering the omnipresent audio-visual culture, we live in, we acknowledge that most likely also aesthetic objects, such as music and visual-art, are likely to be integrated into our lives in an interactive loop involving the environment and the derived pleasure and emotions. As already mentioned above, positive and negative affect are known to have adaptive functions, and positive affect in particular has consequences in daily life for planning and constructing cognitive and emotional resources (Lyubomirsky et al., 2005; Fredrickson et al., 2008). The purpose of this review is not only to contribute to the unification of the two fields, but also to provide a better starting point for the growing research investigating how music and visual-art in general impact the everyday life, such as

enhancement of living environments and well-being. In addition, research on pleasure and reward is a valuable contribution in affective neuroscience when doing research on affect-based psychopathologies such as eating disorders, obsession, depression and drug addiction (Berridge, 2003). We believe that research on art-induced pleasure has a position in the endeavor of elucidating the psychological constituents behind the human behavior underlying pleasure. Yet, we do not want to advocate solely for a naturalistic approach, according to which the appraisal of art objects does not need to be separated from that of any other object (Brown et al., 2011). A recent paper discussing the past and the future of neuroaesthetics, recognized three different emphases in the cognitive science of aesthetic experiences: cognitive neuroscience of aesthetics, cognitive neuroscience of art and cognitive neuroscience of beauty (Pearce et al., 2016). Following this categorisation, we focus on the neuroscience of art in general, and suggest an approach to sensory multimodality through the concept of pleasure which we consider suitable for two reasons. First, it is expected that when focusing on the term pleasure, studies dealing with the cognitive and the emotional aspects of music and visual-art engagement will be reached. Recent literature on reward demonstrate that pleasure is a much more complex phenomenon than mere hedonic response, both on the conceptual and on the functional levels (Kringelbach and Berridge, 2009; Leknes and Tracey, 2010; Smith et al., 2010). Indeed, reward seems to be constructed of different psychological components which have been characterized as affect, motivation and learning, which can further be delineated into comprising elements of affective and cognitive processes, such as wanting based on cognitive incentives and incentive salience, learning based on cognitive and associative learning, and affect consisting of explicit feelings and implicit affective reactions (Berridge, 2003). Second, it is expected that studies focusing on affective experiences, other than those of intense aesthetic experience or peak emotions, will also be captured. A qualitative thematic analysis was chosen to approach the research question in order to recognize patterns, similarities and differences in the chosen aspects of the data. The goal of this review is to understand how pleasure has been conceptualized, either directly or indirectly, in recent empirical research on music and visual-art, to eventually enable the emergence of cognitive neuroscience of art.

Here we applied the keywords of "pleasure," "reward," "enjoyment" and "hedonic" to evaluate how empirical music and visual-art research have approached pleasure empirically. The focus was set on the selected methods and variables, yet due to the large variability of the roles of the keywords in each paper, their positioning in the context of the research questions were investigated in further detail. Regarding the focus – pleasure – of this review, we are aware of the terminological importance of preference, expertise, beauty, liking and valence in the fields of visual-art research and music psychology (Silvia, 2008; Rentfrow and McDonald, 2010; Ishizu and Zeki, 2011). Yet, those terms were not included as keywords because they were considered either as too specific, or too controversial to be paralleled with pleasure (see, e.g., Bundgaard, 2015; Pearce et al., 2016). Indeed, we chose the keywords to reflect the universality of pleasure, without being too much rooted into either of the

disciplines, such as beauty is rooted in neuroaesthetics, where it is used to describe the feelings an aesthetic experience can evoke, and also the perceptual features of an aesthetic object (Bundgaard, 2015). Also, liking was seen here more or less as a synonym for preference, which is often in music studies related to genre specific studies dealing with background variables, such as self-esteem, age, sex and socio-economic variables, not necessarily related to the experiential features of enjoyment (North, 2010; Corrigan and Schellenberg, 2015). Also, valence is an extremely frequently used standardized measure applied in many psychological studies. Had valence been included, it is to be expected that the focus of the review would shifted away from the experiential pleasure resulting in a very large amount of papers, exceeding the scope of the review. Also, for the sake of clarity, we aimed to define this review terminologically by focusing on music and visual-art as objects of empirical research, instead of aesthetics, or aesthetic experiences in general. Indeed, the research on visual-art is closely related to aesthetics, yet aesthetics as such comprised of a multi-disciplinary field of research and is, as a concept, not well defined and thus remains outside the scope of this review (Carroll, 2000). Because the history of empirical studies on music and visual-art is long and characterized by different research trends and emphasis (Hargreaves and North, 2010), we decided to limit the scope of the review to the recent 20 years.

## MATERIALS AND METHODS

### Literature Search and Selection

The following databases were searched for literature: APA, Jstor, PubMed, Science Direct, Scopus, Web of Science, and Nelli. We followed a procedure illustrated and described below (**Figure 1**). For a more detailed walk-through, please see the Appendix I. The literature search consisted of several steps of inclusion and exclusion, and it consisted of systematically developing different types of filters while searching the literature. Also, searches were conducted by using an asterisk (e.g., *pleasur\**, instead of *pleasure*) to not to oversee papers with language-based variability in the use of the key-words. The purpose of this strategy was firstly, to have an overview of the literature of both fields of interest, and secondly, to avoid losing relevant literature or overlooking crucial terminology. The first applied filter we call the normative filter, indicating that all papers which fulfilled the criteria of the wide set of keywords were searched. Thus, the data sampling strategy was comprehensive and included all the fields provided by each database search engine.

Conclusively, reports on empirical studies that focused on pleasurable, hedonic, enjoyable or rewarding experience of music or visual-art were included. Studies were also included if any of these pleasure-synonymous concepts were embedded in the context of an aesthetic experience. This resulted in 59 theoretical and empirical papers. Of these 59 references, only papers reporting on empirical studies were included, resulting in 20 music and 11 visual-art papers. For the sake of readability, hereafter the term “pleasure” is used to refer to the other keywords of enjoyment, hedonic, and reward as well.

### Data Extraction

The same core information from each paper was extracted and tabulated into a spreadsheet consisting of general publication background data (author names, journal name, year of publication, sample characteristics) and specific data extracted to answer the research questions. As far as possible, the data were copied directly as they were stated in the corresponding article, and the tabulated data were then used as a source for drawing further conclusions and categorizations for the subsequent synthesis and analysis.

### Data Synthesis

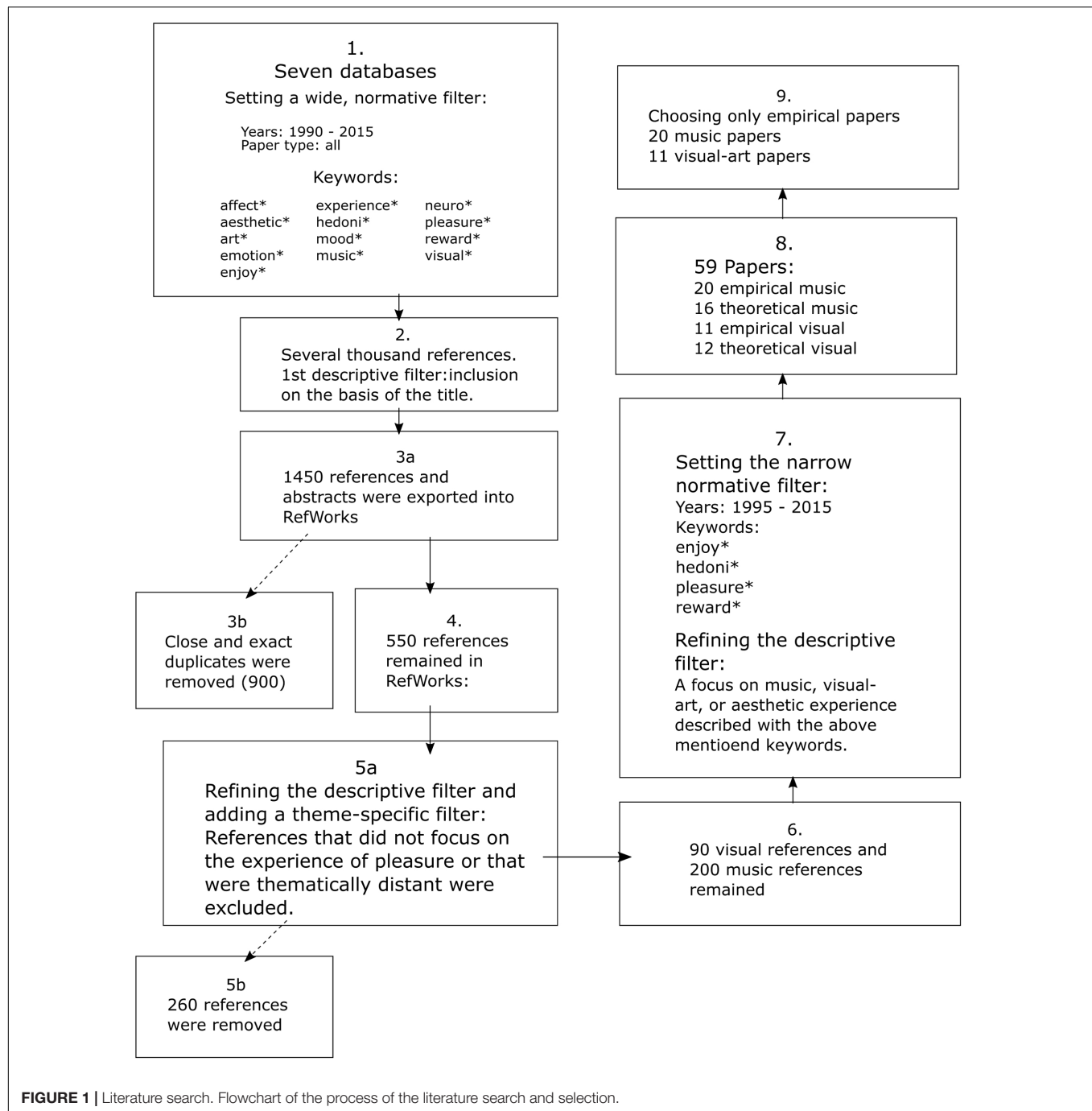
Here, the synthesis was conducted mostly in a narrative form to identify patterns in the data, and to strive for a more holistic understanding of the conceptualization of pleasure (Rumrill and Fitzgerald, 2001), yet in order to support the findings the tabulated aspects were also quantified. Since this review aims to understand how pleasure has been approach in empirical research, it was decided to focus on inspecting the taken methodologies and variables. Despite the systematic appliance of the filters, while searching the literature, a great variance among the papers regarding the keywords could still be detected. It is due to this reason that the role of the keyword and the type of the research question were further categorized. Thus, the decision to focus the synthesis on the two other aspects – role of the keyword and type of the research question – emerged from the included papers, that is, they were not predetermined.

#### The Role of the Keyword in Relation to the Research Question

The papers were first categorized according to two different positions of the keyword, either as direct or indirect. If the role of the keyword was considered direct, the keyword was clearly the object of inquiry. Whereas, if the role of the keyword was not the main target of the inquiry but, rather, an attribute of the main object of research, it was considered to be indirect. Here it should be noted that because the term aesthetic experience was included, the role of the keyword was considered indirect if it was used in that context (e.g., Belke et al., 2010 where the main term is aesthetic or art appreciation, yet it is constantly described with terms of hedonic or pleasure).

#### Type of Research Question

The types of research questions were divided into three categories: External factor-driven, internal factor-driven, and impact-driven. The studies in the first category posed questions in which external factors were considered to influence the internal state of the perceiver (e.g., how musical expressivity influences the derived pleasure). In the second category, the experience was investigated from the perspective of the subject (e.g., the experience depended on the perceiver's personality). Finally, some studies used the experience of pleasure to investigate other phenomena, and these questions were labeled as impact-driven questions (e.g., the influence of music-induced pleasure on learning outcomes). The questions were categorized on the basis of how the research question was postulated in the corresponding paper without consideration of single variables



of the experimental setting. Because many papers had several questions, the question could be categorized under two types, both internal- and external-driven questions. Therefore, more than one type of research question was tabulated for each paper.

### Methods and Main Variables

The methods applied in each paper were tabulated according to the following criteria: “Neuroimaging” refers to methods of brain imaging and brain neurophysiology. “Behavioral” refers to tasks

given to the participants, usually consisting of music listening or picture viewing, and the subsequent rating of the stimuli. “Questionnaire, Interview” refers to studies using online or pen and paper questionnaires or interviews. “Physiological measures” refers to objective, psychophysiological measurements such as heart rate. Only a maximum of two methods were tabulated for each paper.

Additionally, the main variables of each study were tabulated to obtain more detailed information on the variables measured. Because most studies used a large variety of different variables,

only the most frequently used ones were categorized and discussed in relation to the enlisted methods (see Appendix II).

## Analysis

In the analysis, we aim to identify aspects of music and visual-art-induced pleasure that are missing, incomplete, or poorly represented in the literature (Torraco, 2005). The tabulated results are inspected as an entirety on the experiential level in reference to stimulus features, perceiver attributes, cognitive-perceptual, and emotional attributes. Finally, the results are discussed in the light of pleasure conceptualisation in the interdisciplinary literature of philosophy and affective neuroscience as introduced in the beginning of the review. Further it is also discussed, whether pleasure is learned or instinctual, biological or cultural, universal or individual, and whether pleasure is a result of action or whether it facilitates the pursuit of actions (Sizer, 2013; Matthen, 2017).

## RESULTS

Altogether 20 papers were found in the music domain, and 11 papers in the visual-art domain. In both fields, the majority of the papers were published after the year 2008. The extracted information is tabulated below. In the **Table 1** the role of the keyword (direct or indirect) is assigned to the corresponding field of either music or visual-art. In the **Table 2** the applied methods (brain physiology, questionnaire and interview, behavioral and psychophysiology) are cross-tabulated with the questions types (external, internal, impact or external and internal) for each domain.

### Role of the Keyword

As **Table 1** shows, the majority of the music papers had pleasure as a clear object of investigation. Examples of pleasure clearly being the object of investigation were, e.g., musical reward responses, music reward experiences, and reward circuitry of the brain (Montag et al., 2011; Mas-Herrero et al., 2013, 2014). Among the music papers, only in three studies the role of the keyword was said to be indirect. The rewarding aspects of music-evoked sadness, emotional rewards of music, and reward-related activation are examples of the indirect use of keywords (Zentner et al., 2008; Chapin et al., 2010; Taruffi and Koelsch, 2014).

Because the term aesthetic experience was used frequently in the visual-art papers, the keyword was often embedded in the aesthetic context. The keyword had a direct role in a minority of the papers. The indirect keywords were used to describe concepts such as aesthetic pleasure, aesthetic experience, beauty, pictorial perception and aesthetic appreciation. In addition, the terms *aesthetic experience* and *pleasure* or *aesthetic pleasure* were

occasionally used interchangeably. In one of the two articles in which the keywords could be said to also be the objects, the focus was on intrinsic reward manifested in neural correlates (Lacey et al., 2011). The second article focused on the so-called hedonic principle, which was considered to be the underlying mechanism of motivation to spend a certain amount of time viewing pictures (Kron et al., 2014).

A clear difference between music and visual-art papers was the use of the actual keywords. Sixteen of the 20 music papers included reward- and/or pleasure-related terminology, whereas hedonic and enjoyment-related terms were a clear minority, used in only four of 20 articles. In regard to the keywords of the visual-art papers, the terms pleasure and hedonic were the most frequently used terms, whereas the term reward played a central role in only one of the studies, in which it also was the object of the research (Lacey et al., 2011).

### Question Type

**Table 2** illustrates the findings related to the type of question asked in the examined literature. Most external factor-driven papers (five music papers and one visual-art paper) investigated neural correlates or neural mechanisms underlying pleasure. For example, the aim was to test whether limbic and paralimbic brain areas were activated during passive music listening when participants were not given an explicit instruction to focus on emotions (Brown et al., 2004); or to map out neural mechanisms underlying mildly and intensely pleasurable music (Blood and Zatorre, 2001). The visual-art study sought to determine whether the activation of the reward circuitry took place solely from the process of recognizing that an image is artistic rather than non-artistic in nature (Lacey et al., 2011). The remainder of the external factor-driven questions aimed to recognize the quality and frequency of the reported emotions, and how these emotions could be categorized (Zentner et al., 2008; Taruffi and Koelsch, 2014) or whether liking depended on the order in which the stimuli were heard (Parker et al., 2008).

In the internal factor-driven music papers, the variables that depended on the perceiver's attributes were arousal, familiarity, anticipation, musical knowledge and, most of all, personality traits. For example, research investigated individual variation in the experience of reward caused by money or music (Mas-Herrero et al., 2013, 2014); or whether familiarity and arousal correlated with pleasure (van den Bosch et al., 2013). Among the visual-art papers, one of the studies implementing an internal factor-driven approach tested whether the process of perception (ambiguous vs. non-ambiguous portraits) itself depended on the aesthetic experience of the viewer (Boccia et al., 2015). The second paper investigated whether emotions influenced aesthetic experience (Marković, 2010).

Visual-art papers typically included both question types. The experiments were designed to test several different variables according to the stimulus features, the perceiver, and their correlation. The relationship between the internal and external factors was thematized in several research questions. For example, a study conducted by Cupchik et al. (2009) aimed to investigate, on one hand, how different modes of viewing (aesthetic vs. pragmatic viewing mode) paintings influenced the experience

**TABLE 1** | Tabulation of the results based on the role of the key word.

Role of the keyword	Music	Visual-art
Direct	17	2
Indirect	3	9

**TABLE 2** | Cross-tabulation of the results based on the role of the key word, question type and applied research methods.

Question type	Total		Brain physiology		Questionnaire and Interview		Behavioral		Psychophysiology	
External	M5	V1	M3	V1	M1	V0	M1	V0	M3	V1
Internal	M8	V4	M2	V2	M3	V1	M4	V0	M5	V3
Impact	M3	V0	M0	V0	M3	V0	M0	V0	M2	V0
Extr. and Intr.	M4	V6	M1	V3	M3	V2	M0	V1	M1	V5

The letter M preceding the numbers refers to music, and the letter V to visual-art.

and, on the other hand, how the experience depended on the structural (soft edges vs. hard edges) content of the paintings. The impact-driven questions of the music papers addressed learning, stress, attitude and music information seeking and how these factors were related to pleasure (e.g., Gold et al., 2013; Perlovsky et al., 2013).

## Methods and Main Variables

Both fields used functional magnetic resonance imaging (fMRI) most frequently (e.g., Menon and Levitin, 2005; Montag et al., 2011; Jacobs et al., 2012; Boccia et al., 2015). Also, it is notable that in music studies it was common to apply questionnaires and interviews, and physiological measures, whereas these were a clear methodological minority in the visual-art papers. However, as visible from the cross-tabulation of **Table 2**, most studies applied more than one method, which is why comparing the different methods is hard and the subsequent discussion is more interesting when considering the taken variables as well (see Appendix II for more details). In the following, we aim to provide a characterisation of the common combinations of variables and methods typical in both fields of interest.

The main variables of the imaging methods common to the music papers were the neural correlates of reward and intense pleasure or liking (e.g., Blood and Zatorre, 2001; Menon and Levitin, 2005; Montag et al., 2011; Salimpoor et al., 2011), whereas the visual-art papers addressed the difference between basic visual processing and aesthetic emotional processing, hence imagining the brain more broadly focusing on brain areas involved in pictorial processing (e.g., Jacobs et al., 2012; Kreplin and Fairclough, 2013). One of the visual-art studies addressed a question similar to those addressed in the music studies: whether the artistic status of a picture alone can activate the reward center in the brain (Lacey et al., 2011). The variables of the studies combining imaging and viewing and rating-based behavioral tasks varied including naturalness, beauty and roughness; valence and complexity; liking; classification between artistic and non-artistic statuses; aesthetic preference; reaction time or familiarity, demonstrating that in addition to perception modes, the influence of stimulus features was measured.

The most common variable in the music studies was valence, including its different variations from liking to disliking or from pleasing to not pleasing (a total of 12 studies: e.g., Parker et al., 2008; Salimpoor et al., 2011). Arousal was also frequently measured (in a total of eight studies) (e.g., Salimpoor et al., 2011; Mas-Herrero et al., 2014). In visual-art studies, valence or a similar dimension was measured in five studies (e.g., Vessel et al., 2012; Kron et al., 2014). In addition to mere liking or enjoyment,

visual-art studies implemented more complex measures such as beauty, endorsement, aesthetic preference, and emotional movement (e.g., Lacey et al., 2011; Hager et al., 2012; Jacobs et al., 2012; Vessel et al., 2012). Arousal was measured in only one study (Kron et al., 2014). The music studies used character inventories, such as Behavioral Inhibition/Behavioral Approach System (BIS/BAS) or Temperament and Character Inventory (TCI), to mention a few (Montag et al., 2011; Mas-Herrero et al., 2013), and questionnaires to address the listening background or music preference (Garrido and Schubert, 2011; Gold et al., 2013). In visual-art papers, frequently addressed modes or judgmental aspects were artistic vs. non-artistic, pragmatic vs. aesthetic, emotional introspection vs. external object identification and evaluative vs. emotional components.

In sum, a common method used in both fields was brain imaging. Furthermore, when the object of research was reward or pleasure, the object was mainly thought to consist of self-reports based on valence and on psychophysiological measurements (in music studies) or different modes of judgment or perception (in visual-art studies), which had neural correlates as their reference. In the visual-art field, subjective perception was highlighted without additional objective measures. This approach was used to investigate the degree to which pleasure or the aesthetic experience depended on varying modes of perception. Thus, the subjective preparedness and focus of attention were considered the starting points for the whole experience. Music studies used both subjective and objective measures: The conscious, subjective valence and the objectively quantified parameters – such as activation of the reward circuitry or psychophysiological parameters – were required for an experience to be considered pleasurable or rewarding. Few studies aimed to test whether the stimuli used could activate reward-related brain circuitry without conscious listening or viewing.

Valence and related measures were variables that were commonly examined in both fields. Visual-art studies additionally used complex experiential and stimulus-derived descriptors, whereas the music studies collected person-derived data on background, personality and music consumption. In music studies, the more frequent use of psychophysiological measures indicates that arousal was addressed more often. In the visual-art studies, pictures of paintings, drawings or photos were used as stimuli. In all studies, the stimuli were selected by the experimenter, and many studies mixed abstract and representational stimuli. Also, one production task was given where the participants were instructed to depict affectively expressive content (Takahashi, 1995). In contrast, in the music studies, the frequent use of different questionnaires revealed

the lack of real-time music stimuli, since these studies relied on retrospective memory retrieval and on participants' conception of their own identity as music consumers: typically, these studies aimed at developing an instrument or at identifying induced emotions. One questionnaire study implemented music listening as part of the data collection (Vuoskoski and Eerola, 2011). With two exceptions (Blood and Zatorre, 2001; Montag et al., 2011), all music stimuli were pre-selected, either by a separate group of participants or by the experimenters.

## DISCUSSION

Overall, the reviewed papers demonstrate a great variety in the ways in which music and visual-art papers address pleasure. The **Figure 2** below was constructed to illustrate and structure the results in regard to stimulus properties (A), perceiver attributes (B), cognitive perceptual attributes (C), and emotional attributes (D). The **Figure 2** was constructed around the above-mentioned features to open the results of the review in the experiential context. Thus, rather than further discussing the experimental settings such as variables and methods, with the **Figure 2**, we hope to synthesize the most prominent features characterizing the experience of listening to music or viewing art, prevalent in both domains. This way we wish to lead the discussion to the more in-depth analysis of the results. Each of the above-mentioned aspects of the examined literature is discussed below. Please, see the Appendix II for the detailed tabulation of the data.

### (A) Stimulus Properties

Stimulus properties refer to different audible or visible qualities of music and visual-art. Here the comparison showed that in visual-art research, the role of the stimulus was emphasized in a very versatile manner. By contrast, music research emphasized the perceiver's personal background and biographical factors, which are visible when inspecting the perceiver attributes (B), and cognitive perceptual attributes (C).

### (B) Perceiver attributes

Perceiver attributes refer to the individual and biographical qualities of the perceiver. Only the music research addressed listener attributes using various types of character inventories and collected data on biographical information.

### (C) Cognitive perceptual attributes

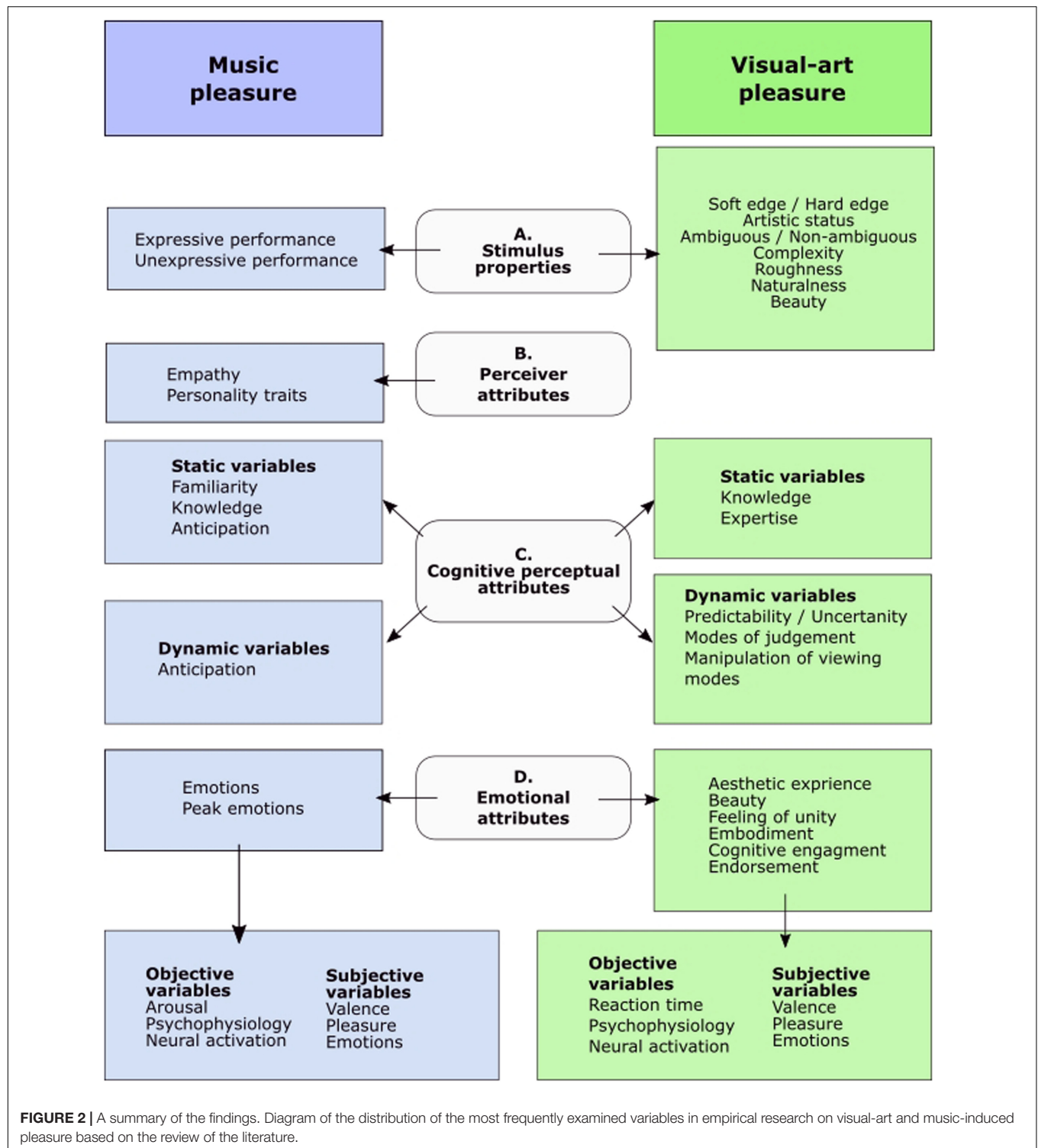
This level refers to the cognitive process of perceiving the stimulus. Here, instead of comparing the two fields in regard to the methods and their variables, we aimed to summarize the results by categorizing the variables in regard to the very fundamental differences among music and visual-art. Namely, music evolving in time and visual-art being static, and spatially distributed. Static variables refer to variables that accumulate over time (e.g., as a result of learning), are more biographical, and are relatively unchangeable features of the perceiver. Dynamic variables refer to attributes that can be consciously manipulated (as

in visual-art research, e.g., viewing mode) or that strongly depend on the corresponding stimulus (e.g., anticipation based on the temporal evolvement of a certain musical piece). Indeed, in the field of visual-art, the range of dynamic variables is much larger, giving the perceiver an active role as an interpreting subject. Thus, it seems that whereas music evolves in time, the applied measures are static, and visual-art which is spatially distributed and temporally static, is investigated more by using variables prone to change and conscious manipulation. This approach, in which the person categorizes and actively interprets information, has also been recognized in emotion research, for example, by Barrett (2006). She called this the "conceptual act" (as opposed to emotions as "natural kind entities"). Specifically, she stated that emotions emerge as a result of people applying their previously acquired knowledge to process and categorize sensory information. Conclusively, many experimental set-ups relied on the perceiver's ability to vary the mode of viewing art and recorded whether this changed the resulting experience. Instead of highlighting personality traits or general background, such research considered the viewer as an active participant in the experience through his or her perceptual and interpretational input during the actual viewing situation. By implementing these various modes of perception, and by changing the stimulus features, scholars often attempted to capture the degree to which the derived experience depended on the judgmental or experiential/emotional mode.

### (D) Emotional attributes

Here, it becomes evident that both fields addressed emotional dimensions of an experience by applying subjective and objective measures. Research conducted in the field of music focused generally on emotions – including also negative emotions – whereas visual-art research often approached pleasurable experience by using rather complex, abstract, and evaluative terminology such as endorsement and being moved. To approach the different types of emotions and experiences, both fields measured the degree of experienced valence. Valence and arousal are dimensions that are commonly applied in emotion psychology to characterize different emotional qualities. For example, Feldman Barrett and Russell (1999) postulated that valence and arousal are independent of each other, and that both have independent polarities. Indeed, it in the visual-art field, arousal was not commonly used as a dimension of a pleasurable experience. This was also evident in the lack of physiological measurements, which are typically applied to measure arousal. In the music studies the applied arousal measures were usually objective psychophysiological measurements, even though arousal can also be applied, e.g., in the form of questionnaires as a subjective self-report.

In music papers, a typical underlying conceptualization was intrinsic reward, which is discussed as a dimension in appraisal theories. Intrinsic pleasantness represents a rather early



reaction in the unfolding chain of events of appraisal, and it is considered to determine the fundamental reaction to an already detected stimulus encouraging avoidance or approach (Ellsworth and Scherer, 2003). First, many papers aimed to demonstrate that music is indeed intrinsically rewarding. Second, the interaction between cortical and subcortical brain regions was

investigated to elaborate how one derives pleasure from abstract sounds.

In summary both fields do represent in the philosophical literature of sensory affect prevalent anti-representational view, in that they separate the experience from the objective features of the stimulus, such that the locus of affect is indeed the experience

of the individual, and that the phenomenology of the sensation is not explained by the stimulus features (Aydede and Fulkerson, *forthcoming*). In the visual-art field the conceptualisation of sensory affect can be inspected in the light of attitudinal or externalist theories. In accordance to these theories, the pleasant sensation to the sensory features of the stimulus, together with a mental attitude – such as desiring, wanting, preferring and liking – construct the composite state of a pleasurable sensation. Crucial here is the idea that sensory pleasure is strongly connected to mental states without having an intrinsic qualia, and thus it is causally connected to the current state of the person (Aydede and Fulkerson, *forthcoming*). In visual-art the explanatory power to the differences in the experience is given to knowledge, intentionality, history and time. According to an imperative view in the philosophy of sensory affect, sensory information presents in itself command-like information to the organism, which informs the organism to action or to retain from an action. Thus, sensory information are considered as motivational states (Aydede and Fulkerson, *forthcoming*). This kind understanding of pleasure seems prevalent in papers which address the stimulation of the reward center of the brain. Yet an approach more refined and closer to the understanding of affective neuroscience seems to be the psychofunctionalist view, according to which incoming sensory information is valued in causal and functional roles such that the information still holds motivational components, yet it is integrated to the mental economy of the perceiver.

## CONCLUSION

This literature review aimed to understand how pleasure derived from music and visual-art had been understood conceptually, either directly or indirectly, in empirical research during the past 20 years. The papers were analyzed in qualitative terms, instead of a quantitative meta-analysis, due to the small amount of papers and due to the large variability in the operationalisation of the key words. The distinction between direct and indirect keyword use is a good example of qualitative comparison, where the papers being reviewed guide the question formulation, which might mean that the formulation of the research question can change during the review process. It turned out that in particular in the visual-art papers pleasure was a very vaguely used term that is, many times it was not a clear object of investigation but rather, it was a characterisation of the researched phenomenon. In our view, an informative quantitative meta-analysis would have required more common nominators and less divergence among the papers. The first findings emerged already during the literature search that, after having applied descriptive, theme-specific and normative filters, started from approximately 200 papers in music, and 90 papers in visual-art and, after refining the keywords and filters, ended up with 20 in music and 11 visual-art studies. The clearly smaller amount of visual-art papers in comparison to the music papers, is a clear demonstration that the phenomenon of interest – pleasure – had a different position in visual-art research. This is also highlighted by the fact that the keywords in the visual-art papers were frequently embedded into

the context of an aesthetic experience. Yet, as demonstrated in the literature search flowchart, the ratio between the fields was more balanced when the theoretical papers were also included. This is an indication that pleasure has a more concise role in theories and models of visual-art than in the equivalent empirical research.

Next to the literature search, the actual synthesis confirmed the above-discussed findings. Music and visual-art studies showed an emphasis on different keywords (reward and pleasure in music research, hedonic and pleasure – embedded in aesthetic experience – in visual-art) and appointed different roles for the keywords (more direct in music, indirect in visual-art), thus demonstrating that pleasure is not a scientifically unanimously defined, nor a conceptually clear object of investigation. Indeed, the process of choosing the correct keywords was a result of several discussions, thus also highlighting the definitional issues related, on one hand to the phenomenon of interest, and on the other hand, on the differences between the two fields. The focus of this review was not aesthetic experiences as such, yet had we included beauty as a keyword, and had papers solely focusing on aesthetic experience, without a clear connection to pleasure, also been included, then the balance between the papers would have been different. Whereas the term aesthetic experience is prevalent in the field of visual-art, a similarly important term in the field of music is the term “peak emotion” or “strong emotion” which often investigate the psychophysiological chills, also known as goose pimples (See, e.g., Gabrielsson, 2010; Grewe et al., 2011). Nevertheless, chills, nor the specific terminology related to the peak emotions were included as keywords because they, too, would have been too specific compared with the more general terms related to pleasure. Also, characteristic to chills is that they may occur in response to unpleasant events, which would have stretched the scope of the review. We assume that the reason why the concept of pleasure seems to play a larger and a more direct role in the empirical music research than in the empirical visual-art research lies in the different backgrounds of the disciplines. The prevalence of the term “reward” in the music studies can probably be traced down to the field of affective neuroscience, where it typically refers to the activation of the reward circuitry of the brain and is concerned with mapping the neural basis of mood and emotional processing of the brain (Dalglish, 2004). The history of empirical research on music and visual-art is long, yet the scope of the review was short, comprising the past 20 years of research to only include relatively recent literature. During this time the term neuroaesthetics was coined (Ishizu and Zeki, 2011) (see also Zeki, 1999), which is a sub-discipline of cognitive neuroscience, focusing on understanding how the brain processes pictorial information and beauty, and which biological functions underlie these processes; the degree to which a good pictorial organization underlies aesthetic experiencing; and how an aesthetic experience becomes a conscious one (Di Dio and Vittorio, 2009; Chatterjee, 2011). Indeed, rather than searching for the correlates in the reward center of the brain, neuroaesthetics has been more concerned with finding common nominators among the stimuli which are artistically appreciated and liked (Ramachandran and Hirstein, 1999), thus possibly explaining the difference in the use of the keywords. In contrast, the background of the music papers lies in emotion psychology,

which most likely explains why pleasure was often discussed and investigated in emotion related terms. The fact the music studies did not address the variation of the stimulus features in similar scale as the visual-art is surprising, considering the fact the question about the link between musical features and the corresponding emotions has been a traditional topic in music psychology. Yet one of the fundamental differences between the art forms is the fact that they employ different sensory systems and also, they are culturally integrated in our daily lives in a different manner. This difference might lie in the cultural significance of our visual perception as our dominant sense and that we are most accustomed to extracting semantic meaning from and ascribing it to visual representations.

We can conclude that music research conceptualized pleasure by using elements of core affect or hedonic tone (valence and arousal) (Feldman Barrett and Russell, 1999; Russell and Barrett, 1999) and intrinsic reward. In particular, the idea of music being able to activate the reward center and the use of psychophysiological measures refer to the idea of music-induced pleasure being biological, rather than culture and context specific in nature. It seems, as if musical pleasure was more involved in the homeostasis of the organism, having an access to the parts of the nervous system which are not subjected to volitional control of the person such as autonomic nervous system and limbic structures of the brain. This aspect is also highlighted when inspecting musical pleasure in terms of the survival circuits and functions related to that, such as motivation, emotions, reinforcement and arousal (LeDoux, 2012). Although an element of core affect – valence – was also common in visual-art research, the derived pleasure was considered to emerge as a result of the conceptual act (Barrett, 2006). That is, the experience is dependent on the perceiver's active interpretation and attribution of meaning, referring to a more culture and context specific understanding of pleasure (see, e.g., Bulot and Reber, 2013). It seems that visual-art pleasure was conceptualized more as an act of information processing consisting of the duality of feature processing and representation (Marr, 1982). Inspecting the results on the dichotomy of learning and instinct, it seems that in both domains it was rather learning-, than instinct-based factors that were dominant. With some variance, both discussed expertise, familiarity and anticipation, which can be seen as examples of accumulative learning (Silvia, 2008; Huron, 2010). Also, both domains highlighted the importance of individuality over universality in response to the stimuli, yet different aspects were highlighted. Music research focused on subject-driven parameters such as familiarity, biographical background and personality, which seem to be rather stable features and inaccessible to voluntary modulation of the perceiver. Whereas in the field of visual-art, the experience was particularly conceived a result of a conscious, and

an active process of interpretation, depending on dynamic variables subjected to the level of expertise and personal control.

As demonstrated in the beginning of this review pleasure and the human desire for pleasure facilitates mental processes and behavior. In literature on pleasure, it has been discussed whether pleasure facilitates the pursuit of an activity, or whether it is the result of an activity (Sizer, 2013; Matthen, 2014). Mainly due to the fact that pleasure had such a variant role in the papers reviewed here, no conclusion about such a relationship could be made. Yet, exactly the questions how art-induced pleasure and reward mediate human behavior and mental processes, or how different pleasure systems (Berridge and Kringelbach, 2015) underlie pleasurable experiences are particularly intriguing ones, and indeed, have been highlighted in the recent literature (Chatterjee and Vartanian, 2016; Pearce et al., 2016). Ultimately, with this review we wish to encourage future empirical research to approach pleasure and its mediating role for cognition and affect from the multimodal perspective of music and visual-art. Yet, as long as music and visual-art research are not integrated and they lack a shared framework, the research on sensory multimodality will remain difficult and restricted (Marin, 2015; Hodges, 2016). Also, we hope that future comparative research would reveal certain modality-specific characteristics in emotional responses to music and visual-art, leading to a more realistic and versatile understanding of enjoyment, not only on the conceptual, but also on the sensory level.

## AUTHOR CONTRIBUTIONS

SS, JW, JM, and MT defined the scope of the review (keywords, inclusion- and exclusion criteria), the goal and the purpose of the paper. Additionally, JW, SS, and JM commented on the text of the paper. Also, SS is the thesis supervisor of the first author and she provided methodological support as well. EB commented on the paper, provided discussion input and was involved in the process of writing as well.

## FUNDING

We would like to thank Kone Foundation for funding this review (grant number: 32881-9).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01218/full#supplementary-material>

## REFERENCES

- Aydede, M., and Fulkerson, M. (forthcoming). "Reasons and theories of sensory affect," in *The Nature of Pain*, eds D. Bain, M. Brady, and J. Corns.
- Barrett, L. F. (2006). Solving the emotion paradox: categorization and the experience of emotion. *Pers. Soc. Psychol. Rev.* 10, 20–46. doi: 10.1207/s15327957pspr1001\_2
- Belke, B., Leder, H., Strobach, T., and Carbon, C. C. (2010). Cognitive fluency: high-level processing dynamics in art appreciation. *Psychol. Aesthet. Creat. Arts* 4, 214–222. doi: 10.1037/a0019648
- Berridge, K. C. (2003). Pleasures of the brain. *Brain Cogn.* 52, 106–128. doi: 10.1016/S0278-2626(03)00014-9
- Berridge, K. C., and Kringelbach, M. L. (2008). Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology* 199, 457–480. doi: 10.1007/s00213-008-1099-6
- Berridge, K. C., and Kringelbach, M. L. (2015). Pleasure systems in the brain. *Neuron* 86, 646–664. doi: 10.1016/j.neuron.2015.02.018
- 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
- Boccia, M., Nemmi, F., Tizzani, E., Guariglia, C., Ferlazzo, F., Galati, G., et al. (2015). Do you like Arcimboldo's? Esthetic appreciation modulates brain activity in solving perceptual ambiguity. *Behav. Brain. Res.* 278, 147–154. doi: 10.1016/j.bbr.2014.09.041
- Brattico, E., Bogert, B., and Jacobsen, T. (2013). Toward a neural chronometry for the aesthetic experience of music. *Front. Psychol.* 4:206. doi: 10.3389/fpsyg.2013.00206
- Brattico, E., and Pearce, M. (2013). The neuroaesthetics of music. *Psychol. Aesthet. Creat. Arts* 7, 48–61. doi: 10.1037/a0031624
- Brown, S., Gao, X., Tisdelle, L., Eickhoff, S. B., and Liotti, M. (2011). Naturalizing aesthetics: brain areas for aesthetic appraisal across sensory modalities. *Neuroimage* 58, 250–258. doi: 10.1016/j.neuroimage.2011.06.012
- Brown, S., Martinez, M. J., and Parsons, L. M. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. *Neuroreport* 15, 2033–2037. doi: 10.1097/00001756-200409150-00008
- Bullot, N. J., and Reber, R. (2013). The artful mind meets art history: toward a psycho-historical framework for the science of art appreciation. *Behav. Brain Sci.* 36, 123–137. doi: 10.1017/S0140525X12000489
- Bundgaard, P. F. (2015). Feeling, meaning, and intentionality: a critique of the neuroaesthetics of beauty. *Phenomenol. Cogn. Sci.* 14, 781–801. doi: 10.1007/s11097-014-9351-5
- Cabanac, M. (1971). Physiological role of pleasure. *Science* 173, 1103–1107. doi: 10.1126/science.173.4002.1103
- Carroll, N. (2000). Art and the domain of the aesthetic. *Br. J. Aesthet.* 40, 191–208. doi: 10.1093/bjaesthetics/40.2.191
- Chapin, H., Jantzen, K., Kelso, J. A. S., Steinberg, F., and Large, E. (2010). Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLoS ONE* 5:e13812. doi: 10.1371/journal.pone.0013812
- Chatterjee, A. (2011). Neuroaesthetics: a coming of age story. *J. Cogn. Neurosci.* 23, 53–62. doi: 10.1162/jocn.2010.21457
- Chatterjee, A., and Vartanian, O. (2016). Neuroscience of aesthetics. *Ann. N. Y. Acad. Sci.* 1369, 172–194. doi: 10.1111/nyas.13035
- Corrigall, K. A., and Schellenberg, E. G. (2015). "Liking music: genres, contextual factors, and individual differences," in *Art, Aesthetics, and the Brain*, eds P. H. Joseph, M. Nadal, F. Mora, L. F. Agnati, and C. J. Cela-Conde (Oxford: Oxford University Press), 263–284. doi: 10.1093/acprof:oso/9780199670000.003.0013
- Cupchik, G. C., Vartanian, O., Crawley, A., and Mikulis, D. J. (2009). Viewing artworks: contributions of cognitive control and perceptual facilitation to aesthetic experience. *Brain Cogn.* 70, 84–91. doi: 10.1016/j.bandc.2009.01.003
- Dalgleish, T. (2004). The emotional brain. *Nat. Rev. Neurosci.* 5, 583–589. doi: 10.1038/nrn1432
- Di Dio, C., and Vittorio, G. (2009). Neuroaesthetics: a review. *Curr. Opin. Neurobiol.* 19, 682–687. doi: 10.1016/j.conb.2009.09.001
- Ellsworth, P. C., and Scherer, K. R. (2003). "Appraisal processes in emotion," in *Handbook of Affective Sciences*, eds R. Davidson, K. R. Scherer, and H. H. Goldsmith (New York, NY: Oxford University Press), 572–595.
- Fechner, G. T. (1876). *Vorschule der Ästhetik*. Leipzig: Breitkopf und Hartel.
- Feldman Barrett, L., and Russell, J. A. (1999). The structure of current affect: controversies and emerging consensus. *Curr. Dir. Psychol. Sci.* 8, 10–14. doi: 10.1111/1467-8721.00003
- Fredrickson, B. L., Cohn, M. A., Coffey, K. A., Pek, J., and Finkel, S. M. (2008). Open hearts build lives: positive emotions, induced through loving-kindness meditation, build consequential personal resources. *J. Pers. Soc. Psychol.* 95, 1045–1062. doi: 10.1037/a0013262
- Gabrielsson, A. (2010). "Strong experiences with music," in *Handbook of Music and Emotion: Theory, Research and Applications*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 547–574.
- Garrido, S., and Schubert, E. (2011). Individual differences in the enjoyment of negative emotion in music: a literature review and experiment. *Music Percept.* 28, 279–296. doi: 10.1525/mp.2011.28.3.279
- Gold, B. P., Frank, M. J., Bogert, B., and Brattico, E. (2013). Pleasurable music affects reinforcement learning according to the listener. *Front. Psychol.* 4:541. doi: 10.3389/fpsyg.2013.00541
- Grewe, O., Katzur, B., Kopiez, R., and Altenmüller, E. (2011). Chills in different sensory domains: frisson elicited by acoustical, visual, tactile and gustatory stimuli. *Psychol. Music* 39, 220–239. doi: 10.1177/0305735610362950
- Hager, M., Hagemann, D., Danner, D., and Schankin, A. (2012). Assessing aesthetic appreciation of visual artworks—the construction of the art reception survey (ARS). *Psychol. Aesthet. Creat. Arts* 6, 320–333. doi: 10.1037/a0028776
- Hargreaves, D. J., and North, A. C. (2010). "Experimental aesthetics and liking for music," in *Handbook of Music and Emotion: Theory, Research and Applications*, eds P. N. Juslin and J. A. Sloboda (Oxford: Oxford University Press), 515–546.
- Hodges, D. A. (2016). *The Neuroaesthetics of Music*, eds S. Hallam, I. Cross, and M. Thaut. Oxford: Oxford University Press, doi: 10.1093/oxfordhb/9780198722946.013.20
- Huron, D. (2010). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge: The MIT Press.
- Ishizu, T., and Zeki, S. (2011). Toward a brain-based theory of beauty. *PLoS ONE* 6:e21852. doi: 10.1371/journal.pone.0021852
- Jacobs, R. H. A. H., Renken, R., and Cornelissen, F. W. (2012). Neural correlates of visual aesthetics - beauty as the coalescence of stimulus and internal state. *PLoS ONE* 7:e31248. doi: 10.1371/journal.pone.0031248
- Kreplin, U., and Fairclough, S. H. (2013). Activation of the rostromedial prefrontal cortex during the experience of positive emotion in the context of esthetic experience. An fNIRS study. *Front. Hum. Neurosci.* 7:879. doi: 10.3389/fnhum.2013.00879
- Kringelbach, M. L., and Berridge, K. C. (2009). Towards a functional neuroanatomy of pleasure and happiness. *Trends Cogn. Sci.* 13, 479–487. doi: 10.1016/j.tics.2009.08.006
- Kringelbach, M. L., and Berridge, K. C. (eds). (2010a). *Pleasures of the Brain*. Oxford: Oxford University Press.
- Kringelbach, M. L., and Berridge, K. C. (2010b). The neuroscience of happiness and pleasure. *Soc. Res.* 77, 659–678. doi: 10.1016/j.biotechadv.2011.08.021
- Secreted Kron, A., Pilkiw, M., Goldstein, A., Lee, D. H., Gardhouse, K., and Anderson, A. K. (2014). Spending one's time: the hedonic principle in ad libitum viewing of pictures. *Emotion* 14, 1087–1101. doi: 10.1037/a0037696
- Lacey, S., Hagtvædt, H., Patrick, V. M., Anderson, A., Stilla, R., Deshpande, G., et al. (2011). Art for reward's sake: visual art recruits the ventral striatum. *Neuroimage* 55, 420–433. doi: 10.1016/j.neuroimage.2010.11.027
- LeDoux, J. (2012). Rethinking the emotional brain. *Neuron* 73, 653–676. doi: 10.1016/j.neuron.2012.02.004
- Ledoux, J. E. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23, 155–184. doi: 10.1146/annurev.neuro.23.1.155
- Leknes, S., and Tracey, I. (2010). "Pain and pleasure: masters of mankind," in *Pleasures of the Brain*, eds M. L. Kringelbach and K. C. Berridge (Oxford: Oxford University Press), 320–335.
- Lyubomirsky, S., King, L., and Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychol. Bull.* 131, 803–855. doi: 10.1037/0033-2909.131.6.803
- Marin, M. M. (2015). Crossing boundaries: toward a general model of neuroaesthetics. *Front. Hum. Neurosci.* 9:443. doi: 10.3389/fnhum.2015.00443
- Marković, S. (2010). Aesthetic experience and the emotional content of paintings. *Psihologija* 43, 47–64. doi: 10.2298/PSI1001047M

- Marr, D. (1982). *A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262514620.001.0001
- Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., and Rodriguez-Fornells, A. (2013). Individual differences in music reward experiences. *Music Percept.* 31, 118–138. doi: 10.1525/mp.2013.31.2.118
- Mas-Herrero, E., Zatorre, R. J., Rodriguez-Fornells, A., and Marco-Pallarés, J. (2014). Dissociation between musical and monetary reward responses in specific musical anhedonia. *Curr. Biol.* 24, 699–704. doi: 10.1016/j.cub.2014.01.068
- Matthen, M. (2014). How to explain pleasure. *Br. J. Aesthet.* 54, 477–481. doi: 10.1093/aesthj/ayu033
- Matthen, M. (2017). The pleasure of art. *Australas. Philos. Rev.* 1, 6–28. doi: 10.1080/24740500.2017.1287034
- Menon, V., and Levitin, D. J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *Neuroimage* 28, 175–184. doi: 10.1016/j.neuroimage.2005.05.053
- Montag, C., Reuter, M., and Axmacher, N. (2011). How one's favorite song activates the reward circuitry of the brain: personality matters! *Behav. Brain Res.* 225, 511–514. doi: 10.1016/j.bbr.2011.08.012
- Nesse, R. M. (2012). Natural selection and the elusiveness of happiness. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 1333–1347. doi: 10.1093/acprof:oso/9780198567523.003.0001
- North, A. C. (2010). Individual differences in musical taste. *Am. J. Psychol.* 123, 199–208. doi: 10.5406/amerjpsyc.123.2.0199
- Parker, S., Bascom, J., Rabinovitz, B., and Zellner, D. (2008). Positive and negative hedonic contrast with musical stimuli. *Psychol. Aesthet. Creat. Arts* 2, 171–174. doi: 10.1037/1931-3896.2.3.171
- Pearce, M. T., Zaidel, D. W., Vartanian, O., Skov, M., Leder, H., Chatterjee, A., et al. (2016). Neuroaesthetics: the cognitive neuroscience of aesthetic experience. *Perspect. Psychol. Sci.* 11, 265–279. doi: 10.1177/1745691615621274
- Pelowski, M., Markey, P. S., Luring, J. O., and Leder, H. (2016). Visualizing the impact of art: an update and comparison of current psychological models of art experience. *Front. Hum. Neurosci.* 10:160. doi: 10.3389/fnhum.2016.00160
- Perlovsky, L., Cabanac, A., Bonniot-Cabanac, M. C., and Cabanac, M. (2013). Mozart effect, cognitive dissonance, and the pleasure of music. *Behav. Brain Res.* 244, 9–14. doi: 10.1016/j.bbr.2013.01.036
- Ramachandran, V. S., and Hirstein, W. (1999). The science of art A neurological theory of aesthetic experience. *J. Conscious. Stud.* 6, 15–51. doi: 10.1179/174327908X392906
- Rentfrow, P. J., and McDonald, J. A. (2010). "Preference, personality and emotion," in *Handbook of Music and Emotion: Theory, Research and Applications*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 669–696.
- Rumrill, P. D., and Fitzgerald, S. M. (2001). Using narrative literature reviews to build a scientific knowledge base. *Work* 16, 165–170.
- Russel, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714
- Russell, J. A., and Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *J. Pers. Soc. Psychol.* 76, 805–819. doi: 10.1037/0022-3514.76.5.805
- 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
- Silvia, P. J. (2008). Knowledge-based assessment of expertise in the arts: exploring aesthetic fluency. *Psychol. Aesthet. Creat. Arts* 1, 247–249. doi: 10.1037/1931-3896.2.1.33
- Sizer, L. (2013). The two facets of pleasure. *Philos. Top.* 41, 215–236. doi: 10.5840/philtopics201341110
- Smith, K. S., Mahler, S. V., Pecina, S., and Berridge, K. C. (2010). "Hedonic hotspots: generating sensory pleasure in the brain," in *Pleasures of the Brain*, eds M. L. Kringelbach and K. C. Berridge (Oxford: Oxford University Press), 27–49.
- Takahashi, S. (1995). Aesthetic properties of pictorial perception. *Psychol. Rev.* 102, 671–683. doi: 10.1037/0033-295X.102.4.671
- 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
- Torraco, R. J. (2005). Writing integrative literature reviews: guidelines and examples. *Hum. Resour. Dev. Rev.* 4, 356–367. doi: 10.1177/1534484305278283
- van den Bosch, I., Salimpoor, V. N., and Zatorre, R. J. (2013). Familiarity mediates the relationship between emotional arousal and pleasure during music listening. *Front. Hum. Neurosci.* 7:534. doi: 10.3389/fnhum.2013.00534
- Vessel, E. A., Gabrielle Starr, G., and Rubin, N. (2012). The brain on art: intense aesthetic experience activates the default mode network. *Front. Hum. Neurosci.* 6:66. doi: 10.3389/fnhum.2012.00066
- Vuoskoski, J. K., and Eerola, T. (2011). Measuring music-induced emotion: a comparison of emotion models, personality biases, and intensity of experiences. *Music. Sci.* 15, 159–173. doi: 10.1177/102986491101500203
- Zeki, S. (1999). *Inner Vision: An Exploration of Art and the Brain*. New York, NY: Oxford University Press.
- Zentner, M., Grandjean, D., and Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* 8, 494–521. doi: 10.1037/1528-3542.8.4.494

**Conflict of Interest Statement:** 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.

Copyright © 2017 Tiihonen, Brattico, Maksimainen, Wikgren and Saarikallio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Arousal Rules: An Empirical Investigation into the Aesthetic Experience of Cross-Modal Perception with Emotional Visual Music

Irene Eunyoung Lee<sup>1,2\*</sup>, Charles-Francois V. Latchoumane<sup>3</sup> and Jaeseung Jeong<sup>1,4\*</sup>

<sup>1</sup> Communicative Interaction Lab, Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology, Daejeon, South Korea, <sup>2</sup> Beat Connectome Lab, Sonic Arts & Culture, Yongin, South Korea, <sup>3</sup> Center for Cognition and Sociality, Institute for Basic Science, Daejeon, South Korea, <sup>4</sup> Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea

## OPEN ACCESS

### Edited by:

Mark Reybrouck,  
KU Leuven, Belgium

### Reviewed by:

Lutz Jäncke,  
University of Zurich, Switzerland  
Luc Nijls,  
Ghent University, Belgium

### \*Correspondence:

Irene Eunyoung Lee  
irenelee@sonicart.co  
Jaeseung Jeong  
jsjeong@kaist.ac.kr

### Specialty section:

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

**Received:** 14 October 2016

**Accepted:** 09 March 2017

**Published:** 04 April 2017

### Citation:

Lee IE, Latchoumane C-FV and Jeong J (2017) Arousal Rules: An Empirical Investigation into the Aesthetic Experience of Cross-Modal Perception with Emotional Visual Music. *Front. Psychol.* 8:440. doi: 10.3389/fpsyg.2017.00440

Emotional visual music is a promising tool for the study of aesthetic perception in human psychology; however, the production of such stimuli and the mechanisms of auditory-visual emotion perception remain poorly understood. In Experiment 1, we suggested a literature-based, directive approach to emotional visual music design, and inspected the emotional meanings thereof using the self-rated psychometric and electroencephalographic (EEG) responses of the viewers. A two-dimensional (2D) approach to the assessment of emotion (the valence-arousal plane) with frontal alpha power asymmetry EEG (as a proposed index of valence) validated our visual music as an emotional stimulus. In Experiment 2, we used our synthetic stimuli to investigate possible underlying mechanisms of affective evaluation mechanisms in relation to audio and visual integration conditions between modalities (namely congruent, complementation, or incongruent combinations). In this experiment, we found that, when arousal information between auditory and visual modalities was contradictory [for example, active (+) on the audio channel but passive (−) on the video channel], the perceived emotion of cross-modal perception (visual music) followed the channel conveying the stronger arousal. Moreover, we found that an *enhancement effect* (heightened and compacted in subjects' emotional responses) in the aesthetic perception of visual music might occur when the two channels contained contradictory arousal information and positive congruency in valence and texture/control. To the best of our knowledge, this work is the first to propose a literature-based directive production of emotional visual music prototypes and the validations thereof for the study of cross-modally evoked aesthetic experiences in human subjects.

**Keywords:** visual music and emotion, aesthetic experience, auditory-visual perception, art and emotion, aesthetic perception, cross-modal integration, auditory-visual integration, music and emotion

## INTRODUCTION

Previous psychological studies have revealed that movies are effective emotion inducers (Gross and Levenson, 1995; Rottenberg et al., 2007), and several studies have used films as emotional stimuli to investigate the biological substrates of affective styles (Wheeler et al., 1993; Krause et al., 2000). In addition, modern neuroimaging and neurophysiological techniques commonly use representational pictures (such as attractive foods, smiling faces, landscapes, and so on) from the International Affective Picture System (IAPS) as visual sources, and excerpts of classical music as auditory sources (Baumgartner et al., 2006a,b, 2007). Several studies have examined the combined influence of pictures and music but, to our knowledge, researchers have not used specifically composed, emotion-targeting cross-media content to elicit human emotion. It has been proposed that composing syncretistic artwork involves more complex types of practice by exploiting the *added values*, specific audio-visual responses that arise due to the synthesis of sounds and images, that occur as the result of naturally explicit A/V cues (Chion, 1994; Grierson, 2005). Hence, the practice of visual music, which composes abstract animations that simulate the *aesthetic purity of music* (Brougher et al., 2005), is naturally intermedia and has a synergy with modern digital auditory-visual (A/V) media. Thus, the affective value of visual music can embody important interactions between perceptual and cognitive aspects of the dual channels. Therefore, using abstract visual music to study emotion remains a promising but uninvestigated tool in human psychology.

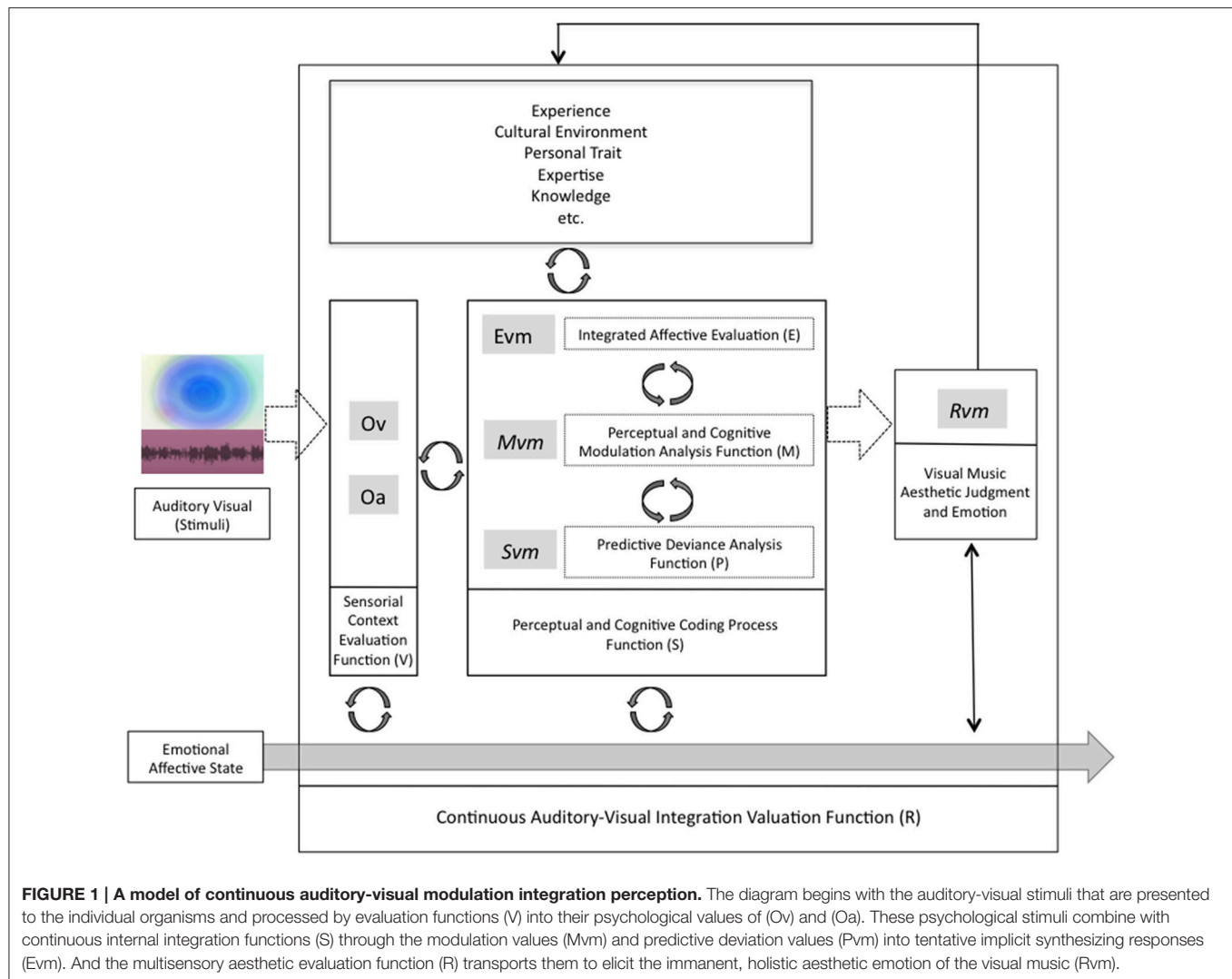
To understand the underlying mechanisms of the audio-visual aesthetic experience, we focused on the assessment of intrinsic structural and contextual aspects of stimuli, and the perceived aesthetic evaluations to formalize the process whereby visual music can suggest target emotions. Although we consider it premature to discuss the mechanism of the entire aesthetic experience within our study, we proposed a model of Continuous Auditory-Visual Modulation Integration Perception (**Figure 1**) to aid in the explanation of our research question. Our model is derived from the *Information Integration Theory* (Anderson, 1981), the *Functional Measurement Diagram* (Somsaman, 2004), and the *Information-Processing Model of Aesthetic Experience* (Leder et al., 2004), and describes how A/V integration might affect the overall aesthetic experience of visual music (**Figure 1**). Briefly explaining the three theories, the essential hypothesis of information integration theory requires three functions when a task involves certain responses from multiple stimuli: valuation (represented psychological values of multiple stimuli), integration (being combined into single psychological values), and response (being converted into an observable response) functions. The functional measurement diagram describes the perception of emotions in multimedia in the context of audio and visual information in relation to the information integration theory. And, the information-processing model of aesthetic experience describes how aesthetic experiences are involved with cognitive and affective processing and the formation of aesthetic judgments in a number of stages.

To investigate the perception of emotion via visual music stimuli, we started by examining the association of information-integration conditions (conformance, complementation, and contest) between two modalities (audio and video) and aesthetic perception (for details about integration conditions, see Cook, 1998, p. 98–106; Somsaman, 2004, p. 42–43). It is known that music/sound alters the meaning of particular aspects of a movie/visual stimulus (Marshall and Cohen, 1988; Repp and Penel, 2002), and that the ability of music to focus attention on a visual object stems from structural and semantic congruencies (Boltz et al., 1991; Bolivar et al., 1994). As researchers have recently emphasized that the semantic congruency between pairs of auditory and visual stimuli to enhance behavioral performance (Laurienti et al., 2004; Taylor et al., 2006), we hypothesized that the enhancement effect of visual music, by way of added value, might rely on the *congruency* of emotional information between unimodal channels (such as comparable valence and arousal information between auditory and visual channels). Therefore, our null hypothesis anticipates to see an enhancement effect of behavioral responses on the perception of emotion in the conformance combination condition of visual music due to the added value from the congruent emotional information provided by cross-modality. Accordingly, such a hypothesis might provide an indication of possible cases in which added values result in an enhancement effect via a cross-modal modulation evaluation process of the dual unimodal channel interplays as a *functional-information-integration* apparatus.

Our aim in this study, as a principled research paradigm, is a careful investigation of the aesthetic experience of auditory-visual stimuli in relation to the dual-channel combination conditions and the enhancement effect approaches from auditory-visual integrations by using emotional visual music prototypes. To examine our hypothesis, we conducted two different experiments, namely:

- (1) The creation of emotional visual music stimuli and the assessment of perceived emotional meanings of the stimuli according to modalities, and
- (2) The investigation of our hypothesis regarding the enhancement effect in relation to the integration conditions.

For Experiment 1, we first designed three positive emotional visual music stimuli by producing audio and video content in accordance with literature-based formal property directions to evoke target-emotions. We then presented our compositions as unimodal (audio only or visual only) or cross-modal (visual music) stimuli to subject groups and surveyed psychometric affective responses (self-rating of emotion), from which three indices were derived (evaluation, activity, and potency, which are equivalent to valence, arousal, and texture/control, respectively). In this experiment, we focused on a two-dimensional (2D) representation (valence and arousal) to validate the affective meaning of our visual music in accordance with the *circumplex model of affect* (Posner et al., 2005). Finally, we examined electroencephalography (EEG) responses as physiological internal representations of valence (in other words, as represented by frontal alpha asymmetry) to the representation of visual music as a partial emotional validation



of visual music. In our main experiment (Experiment 2), we investigated the auditory-visual aesthetic experience, with a particular focus on added-value effects that result in affective enhancement as a functional-information-integration apparatus. To investigate this, we included two additional visual music compositions (negative in valence) created by a solo media artist to our visual music stimuli. We separated the unimodal stimuli from the five original visual music stimuli into independent audio-only and visual-only channel information (named A1–A5 and V1–V5, respectively), and assessed the affective emotion thereof via our subjects' self-ratings (as in Experiment 1). Finally, we cross-matched the unimodal information of altered visual music, forming three combination conditions (conformance, complementation, and contest) of multimedia combinations (Figure S1), and compared the aesthetic affective responses of viewers (self-rated) to investigate our enhancement effect hypothesis using the nine visual music stimuli (five original visual music stimuli and four altered visual music stimuli).

## EXPERIMENT 1: EMOTIONAL VISUAL MUSIC DESIGN AND VALIDATION

In this experiment, we explained the scheme for the construction of each modality of the three visual music stimuli, and we assessed the subjects' perceptions of our target emotion via a 2D representation (valence vs. arousal using indices constructed from the self-ratings). We also assessed the electrophysiological response of the participants relative to valence perception (positive vs. negative) and EEG frontal alpha asymmetry to validate perceived and potential emotion elicitation. The aim of this first experiment was to validate the construction paradigm and the emotion-targeting properties of our abstract visual music movies.

### Methods Participants

For the preliminary experiment, we recruited 16 people from the Graduate School of Culture Technology (GSCT) in the Korean

Advanced Institute of Science and Technology (KAIST) and from Chungnam University, Daejeon, South Korea. Advertisements to recruit aesthetic questionnaire participants for two different groups were posted on the bulletin boards to collect native Korean-language speakers with no special training or expertise in the visual arts or in music (people who could not play any musical instruments or who only played instruments as a hobby and had < 3 years' of lessons). The two groups were the unimodal behavioral survey group and the cross-modal survey with EEG recording group. The unimodal survey group subjects [mean age = 26.12, sex (m:f) = 4:4, minimum age = 24, maximum age = 29] attended individual survey sessions at an allocated time. Each subject attended the experimental session for about 20 min, and had the option to enter a draw to win an on-line shopping voucher worth \$20 in return for their participation. All participants completed a questionnaire detailing their college major, age, gender, and musical or visual background. The participants had various kinds of degrees, such as broadcasting, computational engineering, bio-neuro science, and business management, while no subject was a musician or an art major. The cross-modal survey group subjects [mean age = 22.5, sex (m:f) = 4:4, minimum age = 20, maximum age = 26] attended the experimental session in the morning and spent about 60 min including preparation, recording the EEG and the behavioral survey session, and after experiment debriefing. The subjects were given \$30 per hour as compensation. All participants were certified as not suffering from any mental disorder, not having a history of drug abuse of any sort, and provided a signed consent form after receiving an explanation about the purpose of and procedure for the experiment. In our cross-modal group, no subject was a music or an art major. The study was approved by the Institutional Review Board of KAIST.

### Stimuli Composition

The notion of the existence of some universal criteria for beauty (Etcoff, 2011), together with the consideration of *formalist* (aesthetic experience relies on the intrinsic sensual/perceptual beauty of art) and *contextual* (aesthetic experience depends on the intention/concept of the artist and the circumstances of display) theories (for reviews, see Shimamura, 2012; Redies, 2015) makes it possible to create the intrinsic positive valence of auditory-visual stimuli based on an automatic evaluation of esthetic qualities. Hence, we constructed abstract animations synchronized with music that could convey jointly equivalent positive emotional connotations to assess positive emotion-inducing visual music. In comparison to existing visual music, the new stimuli with a directive design could be advantageous for conducting emotion-inducing empirical research because they allow for the inspection of structural and contextual components, and provide information about production to a certain extent. They can also balance production quality differences among stimuli more easily than when using existing artworks. Furthermore, they remove the familiarity effect that seems to be a critical factor in the listener's emotional engagement with music (Pereira et al., 2011).

To conceptualize the target emotion and to discriminate among the perceived affective responses to the stimuli, we took a dimensional approach, which identifies emotions based on the placement on a small number of dimensions and allows a dimensional structure to be derived from the response data (for a review of the prominent approaches to conceptualizing emotion, see Sloboda and Juslin, 2001, p. 76–81). We chose and characterized three “positive target-emotions” (happy, relaxed, and vigorous) that can be distinctly differentiated from each other when placed on a 2D plane similar to the circumplex model (Posner et al., 2005). Please see **Figure S2**.

- i) Happy: high-positive valence and neutral-active arousal.
- ii) Relaxed: mid-positive valence and mid-passive arousal.
- iii) Vigorous: neutral-positive valence and high-active arousal.

We then focused on previously published empirical studies that examined the iconic relationships between the separate structural properties of visual stimuli and emotion (Takahashi, 1995; Somsaman, 2004) and auditory stimuli and emotion (Gabrielsson and Lindstrom, 2001; Sloboda and Juslin, 2001) to customize our directive production guidelines for our team of creative artists. By considering constructivist theory (Mandler, 1975, 1984), we assembled important structural components that matched the targeted emotional expressions via a broad but non-exhaustive review of previous research on the correlation of emotional elicitation in auditory (**Table 1**) and visual (**Table 2**) cues. Based on the idea that not only the formal properties of stimuli but also the intention of the artist and the circumstances can affect aesthetic experience to a large extent (see Redies, 2015), we further briefed our creative artists about the use of stimuli in the experiment. The creative artists cooperated fully to create visual music content conveying targeted emotions to the viewers by complying with the directive guidelines to output them as positive affection-inducing stimuli.

### Audio-only stimuli production








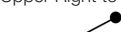
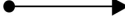
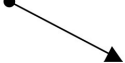

Three 60-s, non-lyrical (containing no semantic words), and consonant music pieces were created by a music producer based on the directive design guidelines (**Table 1**). Basic literature-based information for musical structure properties, such as harmony, modality, melody, metrical articulation, intensity, rhythm, instrument, and tempo, were suggested (for reviews, see Bunt and Pavlicevic, 2001; Gabrielsson and Lindstrom, 2001). However, the artist appointed to create emotion-inducing music noted particular details regarding the decisions. For example, the research team suggested modality directions based on reviewed studies (such as “Major” for “happy”), and the artist chose the tonic center of the key (F#), which can be a basic sub-element of the modality. All the music was created in a digital audio workstation that was equipped with hardware sound modules, such as a Kurzweil K2000, Roland JV-2080, and Korg TR-Rack, and music recording/sequencing programs, such as Digidesign's Pro Tools, Motu's Digital Performer, and Propellerhead's Reason, as well as other audio digital signal processing (DSP) plug-ins, such as Waves Gold Bundle. The final outputs were

TABLE 1 | Summary of musical design structural guidelines selected from reviewed studies and artists' decisions.

Structural component	Happy	Relaxed	Vigorous	Studies linked to the correlation of structural components and emotion
Harmony	Consonant	Consonant	Consonant (with harmonic tensions such as b9, 9, 11, and 13, for example)*	For comprehensive reviews for studies of correlations between musical expressions and emotions, see (Burt and Pavlicevic, 2001; Gabrielsson and Lindstrom, 2001).
Tonal Modality	Major (in F#)*	minor (in C)* moving to Major (Eb)* Alternation	minor (in D)*	
Melodic intervals	Wide Range, High-Pitched	Narrow Range, Low-Pitched	Complex (narrow to wide) Range, Large Intervals	
Pitch level	High	Mid	High+Low	
Loudness and loudness variation	Small	Soft	Loud	
Articulation	few/no change	few/no change	rapid changes	
Tempo and pulse	Staccato	Legato	Complex (staccato + legato)	
	Firm pulse, fast and lively (112 bpm, Dotted Tempo, 4/4 meter)*	Relaxed pulse, gentle and slow pulse (around 60 bpm, 4/4 meter)*	Firm pulse, variedly quick and rapid (from 174 to 87 bpm, 4/4 meter)*	
Adjectives for timbre and texture (depicted by main instruments and musical style)	Bright and playful (female vocal, latin percussive)*	Calm, serene, with slow breathing (korean daegum flute, ethnic)*	Tension and alleviation (electronic synthesizer lead, drum and bass)*	
Pitch contour and melodic direction	Slightly Ascending at the Ending	Descending toward Ending	Building Up & Down	
Motive phrase melodic motion	Intervallic leaps, descending pitch contour	Stepwise + Leaps, ascending pitch contour	From stepwise to intervallic leaps, high complexity	

Prototypical directive guidelines provided information about basic settings in important musical structural expressions to generate the associated target emotional meaning in the music. \*Sub-element of the component chosen by the artist considering the target emotion of the music production.

**TABLE 2 | Summary of visual design structural guidelines selected from reviewed studies and artists' decisions.**

Structural component	Happy	Relaxed	Vigorous	Studies of related variables of visual components**
<b>Thematic milieus (adjective)</b>	Warm, delight, cheerful, joyous	Calm, soothing, serene, tranquil	Exploding, lively, sprightly, exciting	Lundholm, 1921; Hevner, 1935
<b>Main motive object shape</b>	Water droplet: circular form  (Spherical)	Water waves: S-curve flows  Then (Kaleidoscopic, Oval, Rhombus, Hexagon)*	Water Vortex:  up heaving (Symmetrical Pattern, Petaloid)*	Hevner, 1935; Lyman, 1979; Takahashi, 1995
<b>Main object movement (secondary movement)*</b>	Repetition of circles	Descending S-curves then focusing in*	Expanding out then repetition of rising bursts*	Poffenberger and Barrows, 1924; Takahashi, 1995
<b>Main color pallets (secondary colors)*</b>	Red, orange, yellow  (Green, Blue)*	Blue, green, purple, white, brown 	Blue, green, yellow, white  (Pink)*	Hevner, 1935; Wexner, 1954; Lyman, 1979
<b>Object kinetics and moving vectors</b>	Uphill Slant (Lower Left to Upper Right),  Upper Right to Lower Left 	Horizontal (Left To Right),  Downhill Slant (Upper Left to Lower Right) 	Vertical Upward,  Expansion, Fast Cuts, Contrasts	Block, 2013; Zettl, 2013
<b>Editorial scene rhythm</b>	Medium	Slow	Quick	Block, 2013; Zettl, 2013
<b>Sequence transition</b>	Strong and fast changes, jump cuts, long exposure of objects	Slow and weak changes, smooth and dissolve, small number of scenes	Contrasting between different images, jump cuts, many scene changes	Block, 2013; Zettl, 2013

Prototypical directive guidelines provided information about basic settings in important visual and animation structural expressions to generate the associated target emotional meaning in the abstract animation. \*Sub-element of the component chosen by the artist considering the target emotion of the music production. \*\*Selective studies vary in methodological approaches to investigating the influence of the related visual component and its expressions regarding emotion.

exported as digitized sound files (44.1 k sampling rate, 16-bit stereo).

### Video-only stimuli production

Three colored, full-motion, and abstract animations that included forms (shapes), movement directions, rhythm, colors, thematic milieus, scene-changing velocities, and animation characteristic directions were created based on the guidelines (Table 2). The collaborative team consisted of two visual artists, one of whom designed the image layout and illustration with shapes and colors, while the other created the animated motion graphics for it. As with the audio stimuli, we suggested directive guidelines for the overall important structural factors to the artists, and they noted detailed sub-elements. To create the animations, the artists used a digital workstation equipped with Adobe Photoshop, Premier, Max/MSP with Jitter, After Effects,

and a Sony DCR-HC1000. The final visual stimuli consisted of QuickTime movie (.mov) files that were encoded using Sorenson Video 3 (400 × 300, NTSC, 29.97 non-drop fps).

### Visual music integration

For each visual music integration, the motion graphic artist arranged the synchronization of visual animations to its comparable music (for example, happy visual animation synchronized to happy music) while taking the directive motion guidelines (Table 2) into account. In other words, the movements of the visual animation components, sequence changes, kinetics, and scene-change speed evolved over time in accordance with the directive guidelines while incorporating a good accompaniment to the compatible formal and contextual changes in the music (such as accents in rhythms, climax/high points, the start of a new section, and so on). An illustrative

overview of the three visual music stimuli is provided in **Figure 2**, and the content is available online at <https://vimeo.com/user32830202>.

## Procedure

Prior to each experiment, regardless of the modalities, all participants were fully informed that they were participating in a survey for a study investigating aesthetic perception for a scientific research project in both written and verbal forms. We distributed a survey kit, which included information about the purpose of the survey, questions to be answered by the participants regarding demographic information, major subjects, and music/art educational backgrounds (if they played any instruments, how long they had experienced musical/visual art training), and an affective questionnaire form for each presentation. Our affective questionnaire consisted of 13 pairs of scales that were generated by referencing a previous study (Brauchli et al., 1995) that used nine-point scale ratings for bipolar sensorial adjectives to categorize the emotional meanings of the perceived temporal affective stimuli (**Figure S3**). We also explained verbally to our participants how to respond to the emotion-rating tasks on the bipolar scales. The consent of all subjects was obtained before taking part in the study, and the data were analyzed anonymously. All our survey studies, irrespective of modal conditions, were exempt from the provisions of the Common Rule because any disclosure of identifiable information outside of the research setting would not place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

## Unimodal presentation

For the unimodal group, we conducted 20-min experimental sessions with each individual participant in the morning (10–11 a.m.), afternoon (4–5 p.m.), or evening (10–11 p.m.) time slots, depending on the subject's availability. A session consisted of a rest period (5 s) followed by the stimuli presentation (60 s), repeated in sequence for all stimuli. Three audio-only and three visual-only stimuli were shown to subjects via a 19-inch LCD monitor (Hewlett Packard LE1911) with a set of headphones (Sony MDR-7506). The stimuli presentation was in pseudo-random order by altering the play order of modality blocks (audio only or video only) and by varying individual stimuli sequences within each block to obtain the emotional responses of subjects for all unimodal stimuli while avoiding an unnecessary modality difference effect; for example, V1-V3-V2-A1-A2-A3 for subject 1, A2-A1-A3-V2-V1-V3 for subject 2, V3-V2-V1-A3-A2-A1 for subject 3, and so on. The audio-only stimuli accompanied simple black screens, and the video-only stimuli had no sound on the audio channels. Each subject completed the questionnaire while watching or listening to the stimulus (see **Figure S4a**). Subjects were not given a break until the completion of the final questionnaire in an effort to preserve the mood generated and to avoid unexpected mood changes caused by the external environment.

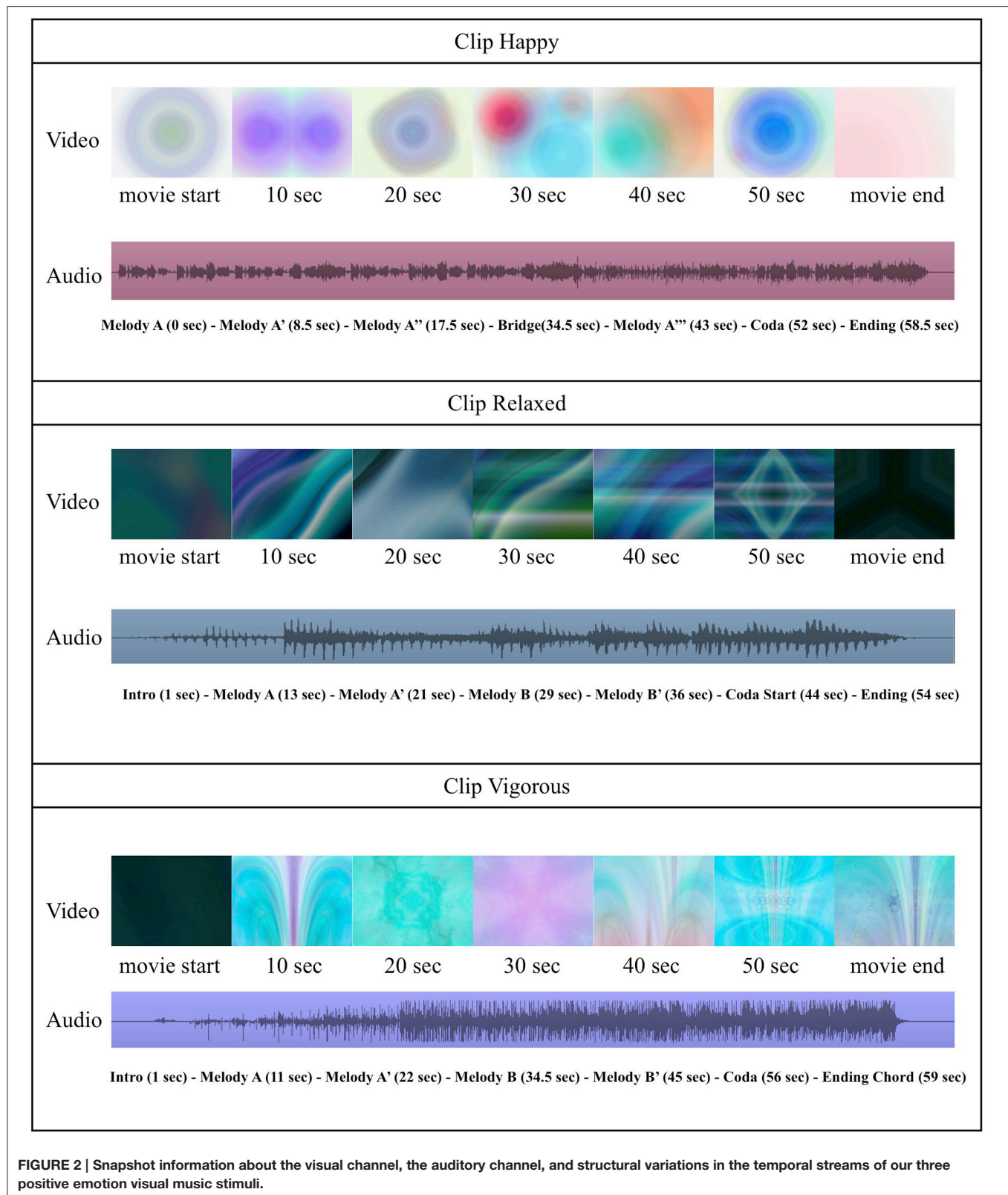
## Cross-modal presentation and EEG recordings

The cross-modal tests were performed in the morning (9–12 a.m.) and, excluding the preparation and debriefing, the EEG recording and behavioral response survey procedures took ~15–20 min per subject. During each session, participants ( $n = 8$ ) were seated in a comfortable chair with their heads positioned in custom-made head holders placed on a table in front of the chair (Hard PVC support with a soft cushion to reduce head movement and to maintain a normal head position facing the screen). Each participant was presented with the visual music, had their heads facing the computer screen (the head to screen distance was 60 cm from a 19-inch LCD monitor; Hewlett Packard LE1911), and was given a set of headphones (Sony MDR-7506). Three visual music stimuli were shown in pseudo-random order to avoid sequencing effects. Resting EEGs were recorded for 20 s as a baseline before each watching session to allow for a short break before the subjects moved on to the next stimulus, thus avoiding excessive stress and providing a baseline time calculation for the resting alpha frequency peak (see Section Data Analysis, EEG analysis). The subjects answered the survey relating to emotion after completing the EEG recording of watching the visual music stimulus (**Figure S4b**) to allow for complete focus on the presentation of the stimuli and to avoid unnecessary noise artifacts in the EEG recordings. EEG recordings were digitalized at a frequency of 1,000 Hz over 17 electrodes (Ag/AgCl, Neuroscan Compumedics) that were attached according to the 10–20 international system with reference and ground at left and right earlobes (impedance  $< 10\text{ k ohms}$ ), respectively. Artifacts resulting from eye movements and blinking were eliminated based on the recording of eye movements and eye blinking (HEO, VEO), using the Independent Component Analysis (ICA) in the EEGLAB Toolbox<sup>®</sup> (Delorme and Makeig, 2004). Independent components with high amplitude, high kurtosis and spatial distribution in the frontal region (obtained through the weight topography of ICA components) were visually inspected and removed when identified as eye movement/blinking contaminations. Other muscle artifacts related to head movements were identified via temporal and posterior distribution of ICA weights, as well as via a high-frequency range (70–500 Hz). The EEG recordings were filtered using a zero-phase IIR Butterworth bandpass filter (1–35 Hz). All of the computer analyses and EEG processing were performed using MATLAB<sup>®</sup> (Mathworks, Inc.).

## Data Analysis

### Emotion index and validity

Although the 13 pairs of bipolar ratings from the survey could provide useful information about the stimuli independently of each other, there was a need to divide them into smaller dimensions to identify emotions based on their position in a small number of dimensions. While the circumplex model is known to capture fundamental aspects of emotional responses, in order to avoid losing important aspects of the emotional process as a result of dividing them into too many dimensions, we identified three indices of evaluation, activity, and potency. We referred to the *Semantic Differential Technique* (Osgood et al., 1957), and extracted the three indices by calculating mean values



of the ratings of four pairs per index from the original 13 pairs used in our surveys. Specifically, the evaluation index assimilates “valence,” and its value was obtained from the mean ratings

of the happy-sad, peaceful-irritated, comfortable-uncomfortable, and interested-bored scales. The activity factor represents “arousal,” and is the average of the tired-lively, relaxed-tense,

dull-energetic, and exhausted-fresh scales. The potency factor reflects “control-related,” and was derived from the unsafe-safe, unbalanced-balanced, not confident-confident, and light-heavy scales. We did not include the calm-restless scale in the activity index because we found (after completion of the surveys) that there was a discrepancy in the Korean translation, which showed “unmatched” for the bipolarity pairing. The final indices (evaluation, activity, and potency) were rescaled from the nine-point scales to a range of  $[-1, 1]$ , as shown in **Figure S5**.

### EEG analysis

For the EEG, we adopted a narrow-band approach based on the Individual Alpha Frequency (IAF; Klimesch et al., 1998; Sammler et al., 2007) rather than a fixed-band approach (for example, a fixed alpha band of 8–13 Hz). This approach is known to reduce inter-subject variability by correcting the true range of a narrow-band based on an individual resting alpha frequency peak. In other words, prior to watching each clip (subjects resting for 20 s with their eyes focused on a cross [+]), the baseline IAF peak of each subject was calculated (clip 1:  $10.7 \pm 1.5$  Hz; clip 2:  $10.1 \pm 0.8$  Hz; clip 3:  $10.1 \pm 1.3$  Hz). The spectral power of the EEG was calculated using a fast-Fourier transform (FFT) method for consecutive and non-overlapping epochs of 10 s (for each clip, independent baseline and clip presentation). In order to reduce inter-individual differences, the narrow band ranges were corrected using the IAF that was estimated prior to each clip according to the following formulas: Theta ( $[0.4-0.6] \times \text{IAF}$  Hz), lowAlpha1 ( $[0.6-0.8] \times \text{IAF}$  Hz), lowerAlpha2 ( $[0.8-1.0] \times \text{IAF}$  Hz), upperAlpha band ( $[1.0-1.2] \times \text{IAF}$  Hz) and Beta ( $[1.2-2] \times \text{IAF}$  Hz). From the power spectral density, the total power in each band was calculated and then log transformed ( $10\log_{10}$ , results in dB) in order to reduce skewness (Schmidt and Trainor, 2001). The frontal alpha power asymmetry was used as a valence indicator (Coan and Allen, 2004), and was calculated for the first 10 seconds of recording ( $10\log_{10}$ ) to monitor emotional responses during the early perception of the visual music (Bachorik et al., 2009) and the delayed activation of the EEGs (also associated with a delayed autonomic response; Sammler et al., 2007). For the overall topographic maps over time (considering all 17 channels; EEGLAB topoplot), the average power in each band divided by the average baseline power in the same band was plotted for each 10 s epoch from the baseline to the end of the presentation (subject-wise average of  $10\log_{10}(P/P_{\text{base}})$ , where  $P$  is the power and  $P_{\text{base}}$  is the baseline power in the band; see **Figure 4** and **Figure S6**). All statistical analyses were performed on the log-transformed average power estimates.

### Statistical analysis

To inspect the perceived emotional meanings of each unimodal presentation, we compared the means of the evaluation, activity, and potency indices (three factors of “affection” as dependent variables) across the audio only and across the video only. We performed a multivariate analysis of variance (two-way repeated measure MANOVA) using two categories as between subject factors (independent variables), “modality” (two levels: audio only and visual only) and “target-emotion” (three levels: happy,

relaxed, and vigorous). This analysis examines three different null hypotheses:

- (1) “Modality” will have no significant effect on “emotional assessment” (evaluation, activity, and potency),
- (2) “Target-emotion” will have no significant effect on “emotional assessment,” and
- (3) The interaction of “modality” and “emotion” will have no significant effect on “emotional assessment.”

We then performed follow-up one-way repeated measure analysis of variance (one-way repeated measure ANOVA) tests for the three indices combined with the same within-subject factor “target-emotion” for each modality. We checked typical assumptions of a MANOVA, such as normality, equality of variance, multivariate outliers, linearity, multicollinearity, and equality of covariance matrices and, unless stated otherwise, the test results met the required assumptions. ANOVAs that did not comply with the sphericity test (Mauchly’s test of sphericity) were reported using the Hyunh-Feldt correction. *Post-hoc* multiple comparisons for the repeated measure factor “clip” were performed using a paired *t*-test with a Bonferroni correction.

To inspect the perceived emotional meanings of the visual music stimuli, we conducted a one-way ANOVA to examine the means of evaluation, activity, and potency indices (three factors of “affection” as dependent variables) for each “target-emotion” clip. Considering the mixed subject groups (identical subjects in audio only and video only, but different subjects in the visual music group), we then conducted non-parametric Kruskal-Wallis *H*-tests (K-W test) to see if “modality” would have an effect on “emotional assessment” (evaluation, activity, and potency) in the same “target-emotion” group (“happy” target-emotion stimuli: A1, V1, and A1V1), and to examine whether the interaction of “modality” and “emotion” would have an effect on “emotional assessment.”

For the frontal asymmetry study of the EEGs, we used a symmetrical electrode pair, F3 and F4, for upperAlpha band powers (see EEG Analysis Section for a definition of upperAlpha band from IAF estimation) to calculate lateral-asymmetry indices by power subtraction (power of F4 upperAlpha—power of F3 upperAlpha). We performed a Pearson’s correlation between the valence index (evaluation) and the frontal upperAlpha asymmetry in order to validate our index representation of valence through electrophysiology. For each clip, we performed a paired *t*-test between F4 and F3 upperAlpha values to quantify the frontal alpha asymmetry.

All statistical tests were performed using the Statistical Package for Social Science (SPSS) version 23.0.

## Results

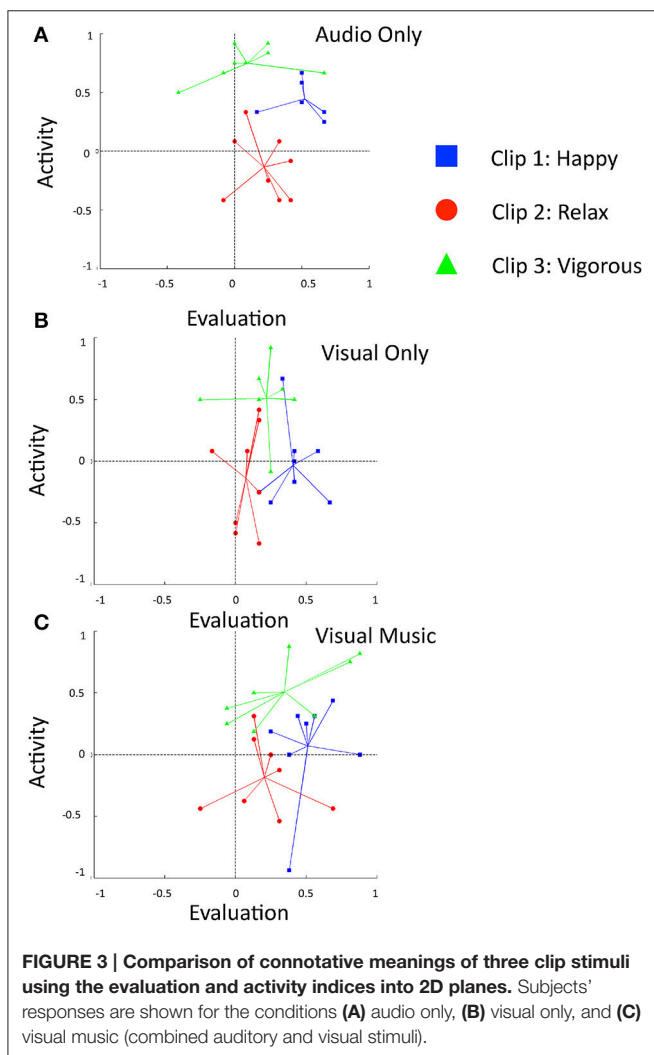
### Unimodal Stimuli

#### Audio only

We obtained the affective characteristic ratings of each auditory-only stimulus, as shown in **Table 3** and **Figure 3**. A one-way ANOVA was conducted to compare the effect of the presentations on the three indices. The result showed that the effect of the clips on the levels of all three indices was significant;

**TABLE 3 | Mean and standard deviation comparison for evaluation, activity, and potency indices of all three stimuli in three different modalities.**

Stimuli	N	Evaluation		Activity		Potency	
		Mean	SD	Mean	SD	Mean	SD
Audio Only Clip 1 (A1)	8	0.60	0.17	0.54	0.16	0.27	0.15
Visual Only Clip 1 (V1)	8	0.47	0.16	0.06	0.33	0.08	0.36
A1V1 (A1V1)	8	0.51	0.20	0.07	0.43	0.34	0.17
Audio Only Clip 2 (A2)	8	0.32	0.18	−0.06	0.28	0.32	0.29
Visual Only Clip 2 (V2)	8	0.15	0.11	−0.02	0.42	0.08	0.36
A2V2 (A2V2)	8	0.13	0.27	−0.27	0.24	−0.11	0.24
Audio Only Clip 3 (A3)	8	0.18	0.30	0.81	0.12	−0.28	0.27
Visual Only Clip 3 (V3)	8	0.28	0.25	0.59	0.27	0.26	0.22
A3V3 (A3V3)	8	0.34	0.37	0.51	0.27	0.13	0.35



evaluation [ $F_{(2, 21)} = 7.360, p = 0.004, \eta^2 = 0.412$ ], activity [ $F_{(2, 21)} = 39.170, p = 0.000, \eta^2 = 0.789$ ], and potency [ $F_{(2, 21)} = 14.948, p = 0.000, \eta^2 = 0.571$ ]. We found that clip 1 received a significantly higher rating in the evaluation index than did clip

3 ( $p = 0.003$ ). The clip 2 received a significantly lower rating in the activity index than did clip 1 ( $p = 0.000$ ), and clip 3 ( $p = 0.000$ ). Clip 3 received a significantly higher rating in the activity index ( $p = 0.045$ ) and a lower rating in the potency index ( $p = 0.001$ ) than did clip 1, and a significantly higher rating in the activity index ( $p = 0.000$ ) and lower indexing in the potency index ( $p = 0.000$ ) than did clip 2. The results indicate that the valence level (evaluation value) of all three auditory stimuli was perceived as positive, with the clip 1 “happy” showing the highest positive level of valence. The variations in the activity index showed “relaxed” as the lowest (mean =  $-0.06$ , neutral-passive) and “vigorous” as the highest (mean =  $0.81$ , high-active) arousal levels, respectively.

### Video only

For the visual-only content, we obtained the mean values as shown in Table 3 and Figure 3. A One-way ANOVA test showed that the effect of the clips on the levels of the three indices was significant in the evaluation [ $F_{(2, 21)} = 6.262, p = 0.007, \eta^2 = 0.373$ ] and activity indices [ $F_{(2, 21)} = 7.488, p = 0.004, \eta^2 = 0.416$ ]. The evaluation index of clip 1 obtained a significantly higher rating than did clip 2 ( $p = 0.006$ ). The activity index showed clip 3 received a significantly highest rating than did clip 1 ( $p = 0.017$ ) and clip 2 ( $p = 0.005$ ). No significantly different effect of the clips was reported for the potency index. The results indicate that all visual-only animations were perceived positively at valence level; however, there were no clear distinctions of the perceived emotional information (evaluation, activity, and potency) among the three animations (V1~V3), as was partially shown in the audio clips (A1~A3). The animation “vigorous” showed the highest activity (mean =  $0.59$ , high-active) rating, while both the animations “happy” and “relaxed” showed activity ratings that indicated low-level arousal experienced by the subjects for these two animations (mean =  $0.06$ , neutral-passive and mean =  $-0.02$ , neutral-passive, respectively).

We conducted two-way repeated measure MANOVA tests with the three indices and two between-subject factors (“modality” and “target-emotion”), and Mauchly’s test indicated that the assumption of sphericity had been violated,  $\chi^2_{(2)} = 14.244, p = 0.001$ ; therefore, degrees of freedom were corrected

**TABLE 4 | Results of Two-Way Repeated Measure MANOVA using the Huynh-Feldt Correction following a violation in the assumption of sphericity.**

Effect	Type III sum of squares	df	Mean square	F	Sig.	Partial eta squared
AestheticEmotion	1.527	1.785	0.855	11.018	0.000	0.208
AestheticEmotion * Modality	0.324	1.785	0.182	2.340	0.109	0.053
AestheticEmotion * Emotion	4.933	3.570	1.382	17.796	0.000	0.459
AestheticEmotion * Modality * Emotion	0.984	3.570	0.257	3.548	0.013	0.145
Error(Affection)	5.821	74.978	0.078			

The analysis checks for effects of variables in modality and target emotion categories on affection information (evaluation, activity, and potency).

using Huynh-Feldt estimates of sphericity ( $\varepsilon = 0.77$ ), and the results are shown in **Table 4**. The results of the  $2 \times 3$  MANOVA with modality (audio only, visual only) and target-emotion (happy, relaxed, vigorous) as between-subject factors show a main effect of emotion, [ $F_{(3.57, 74.98)} = 17.80, p = 0.000, \eta_p^2 = 0.459$ ], and an interaction between modality and emotion [ $F_{(3.57, 74.98)} = 3.548, p = 0.013, \eta_p^2 = 0.145$ ]. The results indicate that “target-emotion” and the interaction of “modality and target-emotion” have significantly different effects on evaluation, activity, and potency levels.

## Cross-Modal Stimuli

### Behavioral response

The ratings for the cross-modal subjects were obtained, and were distributed as shown in **Table 5** and **Figure 3**. The simple placement of evaluation and activity values on the 2D plane illustrates the characteristics of our stimuli at a glance, and we can see that they indicate positive in emotional valence (mean evaluation index  $\geq 0$ ), irrespective of the modality of the presentation (**Figure 3**). Clip 1, “happy,” targeted a high-positive valence and neutral-active arousal emotion, and indicated the comparable perception of the subjects (evaluation:  $0.51 \pm 0.20$ ; activity:  $0.07 \pm 0.43$ ); clip 2 “relaxed,” a mid-positive valence and mid-passive arousal, displayed a tendency similar to the viewers’ responses (evaluation:  $0.13 \pm 0.32$ ; activity:  $-0.27 \pm 0.24$ ), while clip 3, “vigorous,” showed a neutral-positive valence and high-active arousal (evaluation:  $0.34 \pm 0.37$ ; activity:  $0.51 \pm 0.27$ ).

We performed a multivariate analysis of variance using a category as the independent variable “target-emotion” (three levels: happy, relaxed, and vigorous), with evaluation, activity, and potency indices as dependent variable. The test showed a statistically significant main effect of the clip (target-emotion) on the activity index [ $F_{(2, 23)} = 11.285, p = 0.000, \eta_p^2 = 0.518$ ] and on the potency index [ $F_{(2, 23)} = 6.499, p = 0.006, \eta_p^2 = 0.382$ ]. A *post-hoc* test indicated that clip 3, “vigorous,” showed significantly higher activity rating values than did clip 1 “happy” ( $p = 0.042$ ) and clip 2 “relaxed” ( $p = 0.000$ ). Clip 1, “happy,” showed significantly higher potency values than did clip 2, “relaxed” ( $p = 0.005$ ). The means of the evaluation index shows that the clip “happy” had the highest evaluation value ( $M = 0.51$ , high-positive level), and the clip “relaxed” had the lowest ( $M = 0.12$ , mid-positive). The activity index in the visual music showed that the clip “vigorous” had the highest activity value ( $M = 0.51$ , high-active in arousal).

Further inspection of the perceived emotional information (evaluation, activity, and potency) on the interactions of “target emotion” and “modality” was conducted via a K–W test. The test results showed that there was no statistically significant difference among the three different modalities on the evaluation index. However, there were statistically significant differences on activity scores for the “happy” stimuli among the different modalities,  $\chi^2_{(2)} = 11.409, p = 0.003$ , with a mean rank index score of 19.250 for A1, 10.130 for A1V1, and 8.130 for V1, and for the “vigorous” stimuli,  $\chi^2_{(2)} = 6.143, p = 0.046$ , with a mean rank index score of 17.440 for A3, 9.190 for A3V3, and 10.880 for V3. Three statistically significant differences in potency scores were reported. The “happy” stimuli in different modalities showed  $\chi^2_{(2)} = 6.244, p = 0.044$ , with a mean rank index score of 14.060 for A1, 15.880 for A1V1, and 7.560 for V1; the “relaxed” stimuli,  $\chi^2_{(2)} = 6.687, p = 0.035$ , had a mean rank index score of 17.310 for A2, 8.250 for A2V2 and 11.940 for V2, while the “vigorous” stimuli showed  $\chi^2_{(2)} = 8.551, p = 0.014$ , with a mean rank index of 6.560 for A3, 15.130 for A2V2, and 15.810 for V2 (**Table 5**).

### EEG response

The temporal dynamic of visual music creates complex responses in EEGs, with spatial, spectral, and temporal dependency for each clip (**Figure S6**). In order to remain within the scope of this study, we focused on a well-known EEG-based index of valence. It has been proposed that the frontal alpha asymmetry (EEG) could show close association with perceived valence in the early stage of the presentation of stimuli (Kline et al., 2000; Schmidt and Trainor, 2001; van Honk and Schutter, 2006; Winkler et al., 2010). We used the known neural correlates of emotion (frontal asymmetry in alpha power) to assess the internal responses of our subjects and to provide a partial physiological validation of our targeted-emotion elicitation from watching visual music. We found statistically significant correlations between evaluation and the frontal upperAlpha difference (evaluation-upperAlpha Subtraction:  $r_{(24)} = 0.505, p = 0.012$ ; **Figure 4**). We also found a significant difference between F4 and F3 upperAlpha power [ $t_{(7)} = 5.2805; p = 0.001$ ; paired *t*-test] for clip 1, the visual music stimulus showing the most positive response in valence. Clips 2 and 3 did not show significant differences in the frontal upperAlpha band power (**Figure 5A**).

In addition to the frontal asymmetry, we noted a qualitative average increase in frontal theta power and a decrease in power for lowAlpha2, compared to the baseline in the early presentation

**TABLE 5 | Results of the Kruskal-Wallis Test comparing the effect of our designed emotional stimuli on the three perceived emotion assessment indices (evaluation, activity, and potency).**

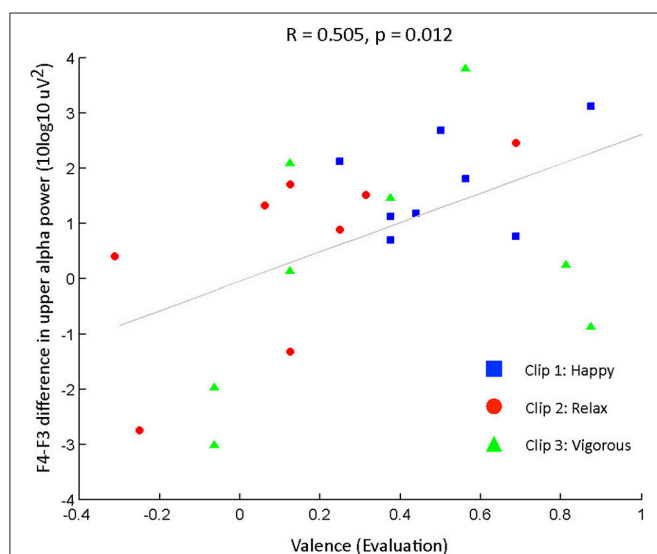
Emotion assessment indices		N	Mean	Std. Deviation	Min	Max	Percentiles		
							25th	50th (Median)	75th
DESCRIPTIVE STATISTICS (NPar Test)									
Clip 1 Target-Emotion 'Happy'	Evaluation	24	0.526	0.178	0.250	0.880	0.375	0.531	0.688
	Activity	24	0.224	0.387	−0.940	0.750	0.000	0.281	0.438
	Potency	24	0.237	0.204	−0.060	0.630	0.063	0.250	0.359
	Total	24	24.330	26.811	1.000	61.000	1.000	11.000	61.000
Clip 2 Target-Emotion 'Relaxed'	Evaluation	24	0.198	0.227	−0.310	0.690	0.063	0.219	0.359
	Activity	24	−0.117	0.328	−0.560	0.560	−0.422	−0.188	0.125
	Potency	24	0.086	0.344	−0.500	0.690	−0.172	−0.031	0.359
	Total	24	28.670	25.481	2.000	62.000	2.000	22.000	62.000
Clip 3 Target-Emotion 'Vigorous'	Evaluation	16	0.268	0.306	−0.310	0.880	0.063	0.281	0.469
	Activity	16	0.635	0.256	0.000	0.940	0.500	0.719	0.813
	Potency	16	0.000	0.339	−0.750	0.690	−0.250	−0.031	0.250
	Total	16	33.000	25.022	3.000	63.000	3.000	33.000	63.000
Emotion assessment indices		Clip ID	N	Mean Rank	Chi-Square	df	Asymp. Sig.	Effect Size (η²)	
RANKS (KRUSKAL-WALLIS TEST)									
Clip 1 Target-Emotion 'Happy'	Evaluation	A1	8	16.060	3.227	2	0.199	0.140	
		A1V1	8	11.380					
		V1	8	10.060					
		Total	24						
	Activity*	A1	8	19.250	11.409	2	0.003	0.496	
		A1V1	8	10.130					
		V1	8	8.130					
		Total	24						
	Potency*	A1	8	14.060	6.244	2	0.044*	0.271	
		A1V1	8	15.880					
		V1	8	7.560					
		Total	24						
Clip 2 Target-Emotion 'Relaxed'	Evaluation	A2	8	16.380	3.674	2	0.159	0.160	
		A2V2	8	10.560					
		V2	8	10.560					
		Total	24						
	Activity	A2	8	14.250	2.379	2	0.304	0.103	
		A2V2	8	9.380					
		V2	8	13.880					
		Total	24						
	Potency*	A2	8	17.310	6.687	2	0.035	0.291	
		A2V2	8	8.250					
		V2	8	11.940					
		Total	24						

(Continued)

TABLE 5 | Continued

	Emotion assessment indices	Clip ID	N	Mean Rank	Chi-Square	df	Asymp. Sig.	Effect Size ( $\eta^2$ )
Clip 3 Target-Emotion 'Vigorous'	Evaluation	A3	8	10.060	1.441	2	0.487	0.063
		A3V3	8	13.630				
		V3	8	<b>13.810</b>				
		Total	24					
	Activity*	A3	8	<b>17.440</b>	6.143	2	0.046	0.267
		A3V3	8	9.190				
		V3	8	10.880				
		Total	24					
	Potency*	A3	8	6.560	8.551	2	0.014	0.372
		A3V3	8	15.130				
		V3	8	<b>15.810</b>				
		Total	24					

The rank-based non-parametric analysis was used to check the rank-order of the three modalities (audio only, video only, visual music) and to determine if there were statistically significant differences between two (or more) modalities. For further investigation, a higher value in the mean rank indicates the upper rank; for example, in the "happy" clip group, A1 (Mean rank = 16.06) is higher in rank than the other two modalities, V1 (Mean rank = 10.06) and A1V1 (Mean rank = 11.38) on the evaluation index. \* Indicates a significant difference  $p < 0.05$ , while bold type indicates the highest rank per category.

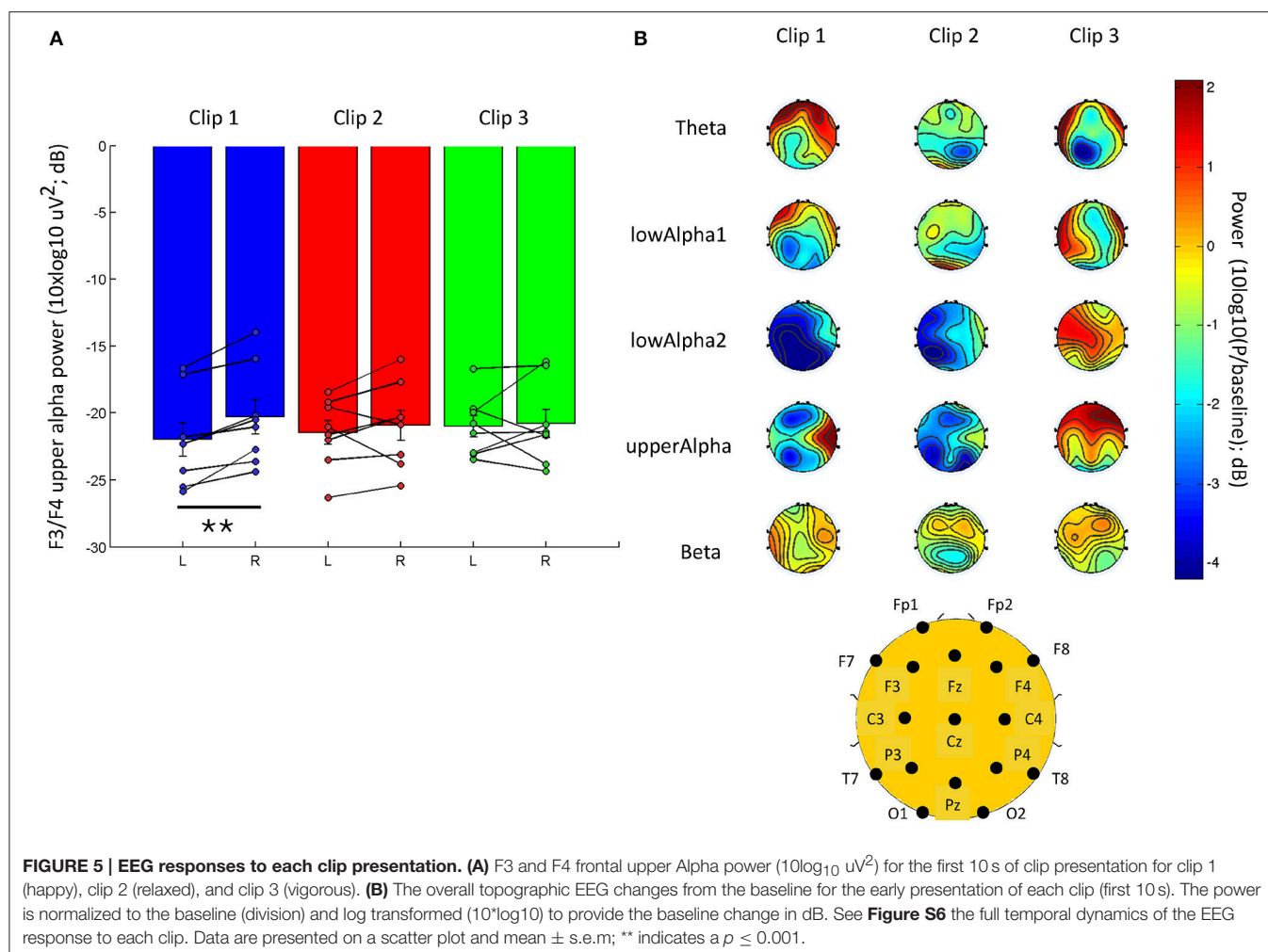


**FIGURE 4 | Scatter plot and correlation between evaluation rating (valence) and upperAlpha frontal asymmetry.** Upper Alpha power was estimated for the early presentation of clips (10 s), and was log transformed ( $10\log_{10} \text{uV}^2$ ). Frontal asymmetry was estimated as the difference of F3-F4 of log transform of upperAlpha band power [ $10\log_{10}(P)$ ]. Pearson's correlation was used ( $N = 8$  subjects).

phase of clip 1 (first 10 s; **Figure 5B**). Subjects' responses during the early presentation of clip 2 (relaxed) exhibited a more symmetrical activity in the lower frequency bands (theta and lowAlpha1), as well as in beta power. Clip 3 (vigorous) elicited a stronger average increase in power. Notably, clip 3 showed more remarkable increase in frontal upperAlpha and lowAlpha2 power, compared to baseline recording.

## Discussion

In this experiment, we plotted the valence and arousal information of our designed clips on a well-known 2D plane with the means of the extracted evaluation and activity indices from psychometric self-rated subject surveys. The statistical analysis result indicates that our directive design of visual music delivered positive emotional meanings to the viewers with somewhat different valence, arousal, and texture information when compared to each other. The 2D plane illustration indicated that our constructed emotional visual music videos delivered positive emotional meanings to viewers that were similar to the intended target emotions. Comparisons of same target-emotional stimuli between modalities indicated that the likelihood of assessing valence (evaluation) information was similar for all three modality groups, while assessing the arousal (activity) and texture/control (potency) information was not similar. In other words, the target-emotion designs of audio, video, and visual music were likely to have similar information on valence levels with differences on arousal and texture/control levels. In addition, the frontal EEG asymmetry response during the early phase (first 10 s) of the visual music presentation correlated with the perceived valence index, which supports a possible relationship between valence ratings and the physiological response to positive emotion elicitation in our subjects. In our view, this preliminary experiment might provide a basis for the composition of synthetically integrated abstract visual music as unitary structural emotional stimuli that can elicit autonomic sensorial-perception brain processing; hence, we propose visual music as prototypic emotional stimuli for a cross-modal aesthetic perception study. The result of this experiment suggests the possibility of authenticating the continued investigation of aesthetic experiences using affective visual music stimuli as functions of information integration. However, the addition of negative emotional elements in



experiment 2 seemed necessary to investigate how the auditory-visual integration condition might affect the overall aesthetic experience of visual music and the affective information interplay among modalities.

## EXPERIMENT 2: AUDIO-VISUAL INTEGRATION EXPERIMENT

To conduct the experiment, we first included two additional visual music stimuli created by a solo multimedia artist (who did not participate in the design of Experiment 1) within the negative spectrum of valence. We then confirmed the perceived emotional responses of participants watching our emotional visual music, and divided them into three indices, namely evaluation (valence), activity (arousal), and potency (texture/control). Lastly, we studied the audio-visual integration effect by cross-linking unimodal information from the original visual music to retain stimuli for all three integration conditions: conformance, complementation and contest (Section Stimuli Construction), and to investigate the enhancement effect of perceived emotion by synthetic cross-modal stimuli.

## Methods

To examine the association of different information-integration conditions between two modalities (audio and video) on aesthetic perception and the enhancement effect (to test our hypothesis regarding the integration enhancement effect), we conducted surveys involving audio-only, video-only, original visual music, and altered visual music stimuli. By using an analogous method in Experiment 1, we conducted psychometrical surveys of participants and assessed the evaluation, activity and potency factors across unimodal and cross-modal stimuli. We then compared the emotional information in unimodal channels and visual music, and checked for the presence of added values resulting from the A/V integration.

## Participants

We recruited participants from the Dongah Institute of Media and Arts (DIMA) in Anseong, Korea, and Konyang University in Nonsan, South Korea for the four subject groups in our audio-visual integration experiment. Students from the DIMA enrolled in Home Recording Production or Video Production classes served as subjects in exchange for extra credit. Students from Konyang University enrolled in general education classes

in Thought and Expression or Meditation Tourism served as voluntary subjects. Voluntary participants from Konyang University had the option of entering a draw to win a coffee voucher worth \$10 in return for their participation. The groups were composed as follows:

- Audio-only group of 33 video production majoring students [mean age = 22.33, sex (m:f) = 10:23, minimum age = 21, maximum age = 25];
- Visual-only subjects of 27 students majoring in music [mean age = 25.15, sex (m:f) = 14:13, minimum = 21, maximum age = 28];
- The original visual music group was composed of 42 students majoring in cultural events and tourism [mean age = 22.50, sex (m:f) = 22:20, minimum age = 21, maximum age = 29], and
- The altered visual music group was composed of 53 students who were studying majors that included fashion, optometry, tourism, business, computer engineering, economics, psychology, and the like [mean age = 22.32, sex (m:f) = 24:29, minimum age = 20, maximum age = 27].

Considering that our directive design stimuli were based on the concept of universal beauty in art, but that the survey subjects did not include participants from different cultures, the difference in the number of subjects in the groups was not critical ( $n \geq 27$  in all four groups). The survey kits and questionnaires were identical to the ones used in Experiment 1, except for the addition of a pleasant-unpleasant pair as a simple positive-negative affect response (see **Figure S2**). As in Experiment 1, subjects were fully informed, in both written and verbal forms, that they were participating in a survey for a study investigating aesthetic perception for scientific purposes. The consent of all subjects was obtained orally before taking part in the study, and the data were analyzed anonymously. Most of the visual music subjects (including both original and altered groups) were not taking majors related to music or visual art, with the exception of nine students (one digital content, one fashion design, two visual design, one cultural video production, and four interior design).

### Stimuli Construction

As briefly explained previously, we extracted the unimodal stimuli from five original visual music videos to divide them into independent audio-only and visual-only channel information (named A1–A5 and V1–V5, respectively). Based on the evaluated emotion from the unimodal information, we then cross-linked the clips to create altered visual music. This method was inspired by the three *basic models of multimedia* which is postulated by Cook (1998) and *combinations of visual and audio information based on the two-dimensional (2D) model* by Somsaman (2004). The result was nine cross-modal stimuli (five original and four altered clips) in three combinations of conditions: *conformance* (agreement between valence and arousal information from the video and audio channels), *complementation* (partial agreement between valence and arousal information from the video and audio channels) and *contest* (conflict between valence and arousal information from the video and audio channels) to use in our experiment (**Figure S1**).

**TABLE 6 | Visual music stimuli combination conditions information in the current study.**

	Audio Only (Valence, Arousal)				
	P = Positive, N = Negative, H = High, L = Low				
	A1 (P, H)	A2 (P, H)	A3 (P, H)	A4 (N, H)	A5 (N, H)
Visual Only (Valence, Arousal) P = Positive, N = Negative, H = High, L = Low	V1 (P, L)	$\alpha$ 2(C, I)			$\beta$ 3(I, I)
	V2 (P, L)	$\alpha$ 2(C, I)		$\beta$ 3(I, I)	
	V3 (P, H)		$\alpha$ 1(C, C)*		
	V4 (N, H)	$\beta$ 2(I, C)		$\alpha$ 1(C, C)	
	V5 (N, H)	$\beta$ 2(I, C)			$\alpha$ 1(C, C)

Clip name and the emotional contexts (valence and arousal information) of each unimodal clip are indicated in bold types. Cross-binding codes indicate auditory-visual (A–V) combination conditions for all of the visual music stimuli in the current study. For example,  $\beta$  2(I, C) where A2 combines with V5 indicates that it is an altered combination in contest condition with incongruent relationship in valence and congruent relationship in arousal between auditory and visual channels' emotional information. The emotional meanings were validated using similar evaluation and activity indices from the unimodal stimuli survey results in Experiment 1 and Experiment 2.  $\alpha$ , original combination;  $\beta$ , altered combination; 1, Conformance; 2, Complementation; 3, Contest; \*, control.

### Procedure

#### Unimodal

Surveys for the audio-only and visual-only groups were conducted on the same day at 11 a.m. and 1 p.m., and all the subjects in each group sat together in a large, rectangular audio studio classroom ( $\sim 6 \times 6.5$  m) that was equipped with a wall-mounted screen ( $122 \times 92$  cm) A/V system and a pair of Mackie HRmk2 active monitor speakers. The audio-only group heard sound clips (60 s each) individually with a 20-s rest between presentations, ordered A2–A3–A4–A1–A5, and the visual-only group watched video clips (60 s each) that were ordered V2–V4–V3–V1–V5, with a 20-s rest between each presentation. Each student evaluated the emotional qualities of each stimulus by answering the questionnaires while listening to or watching the stimuli, and each session lasted  $\sim 10$  min.

#### Cross-modal

We performed the psychometric rating experiments in the afternoon for each group at 1 and 3 p.m., and five visual music clips were played in each survey. All the subjects sat together in a general classroom (approximate room size  $7 \times 12$  m) that was equipped with a wall-mounted, large-screen ( $305 \times 229$  cm) projection A/V system and a pair of Mackie SRM450 speakers. For the original visual music group, we played the stimuli in the order A4V4–A2V2–A5V5–A1V1–A3V3. For the altered visual music group, we played the stimuli in the order A5V1–A4V2–A1V4–A2V5–A3V3 (the control clip was the same as A3V3). Taking into account the temporal contextual changes that occur during the presentation of clips, the students responded the self-assessment questionnaires after watching each clip.

## Data Analysis

### Reliability and competency tests of the indices

Before investigating the added value (enhancement effect) of cross-modal perception via visual music, we checked the reliability of our three emotional aspect factors (evaluation, activity, and potency) by using the data from all conditions in the experiment. The result showed a fairly high reliability value for the three indices (Cronbach's  $\alpha = 0.728$ ,  $N = 775$ ), and the test indicated the omission of the activity index could lead to a higher Cronbach's alpha value ( $\alpha = 0.915$ ). The inter-relation correlation matrix values among the indices were evaluation-activity (0.295114), evaluation-potency (0.843460), and activity-potency (0.223417). A Pearson's correlation test indicated that three emotional indices were strongly associated: [evaluation-activity:  $r_{(775)} = 0.295$ ,  $p < 0.000$ , evaluation-potency:  $r_{(775)} = 0.843$ ,  $p < 0.000$ , and activity-potency:  $r_{(775)} = 0.223$ ,  $p < 0.000$ ]. An independent-sample  $t$ -test of three indices was conducted to compare the unimodal group and the cross-modal group in the evaluation, activity, and potency level conditions. The result indicated no significant differences between unimodal ( $n = 300$ ) and cross-modal groups ( $n = 475$ ), except in the activity values of unimodal ( $0.216 \pm 0.313$ ) and cross-modal ( $0.139 \pm 0.279$ ) conditions;  $t_{(773)} = 3.587$ ,  $p = 0.000$ ,  $d = 0.260$ . The results indicated higher mean activity values for audio only ( $0.342 \pm 0.219$ ) than for visual only ( $0.062 \pm 0.341$ );  $t_{(219, 71)} = 8.267$ ,  $p = 0.000$ ,  $d = 0.975$ . In addition, we found higher mean activity values for the original group ( $0.189 \pm 0.230$ ) than for the altered group [ $0.099 \pm 0.306$ ];  $t_{(473)} = 3.528$ ,  $p = 0.000$ ,  $d = 0.333$ ]. To further check the competency of the activity indices, we checked the responses to our control clip (A3V3) of the original ( $n = 42$ ) and altered cross-modal ( $n = 53$ ) survey groups. As Levene's test for equality of variances also revealed no significant differences in the two groups' assessments of the control clip, this provides some evidence that the equal variance assumption is satisfied on the univariate level. We found no statistically significant differences among groups of the three indices, as determined by a one-way ANOVA: evaluation [ $F_{(1, 93)} = 0.051$ ,  $p = 0.822$ ,  $d = 0.047$ ], activity [ $F_{(1, 93)} = 0.061$ ,  $p = 0.806$ ,  $d = 0.045$ ], and

potency [ $F_{(1, 93)} = 0.942$ ,  $p = 0.336$ ,  $d = 0.199$ ]. The descriptive statistical results are shown in **Table 7** and **Table S1**.

### Statistical analysis

To investigate the perception of emotion in the information-integration conditions (conformance, complementation, and contest) between two modalities (audio and video), we performed a separate one-way ANOVA analysis for each survey group (five audio only, five video only, five original video music, and five altered visual music). *Post-hoc* multiple comparisons of significant ANOVA results were then performed using the Bonferroni correction. Levene's test results for equality of variances were recorded when violated. All statistical tests were performed using the Statistical Package for Social Science (SPSS) version 23.0. The silhouette-clustering index was used as a measure of clustering. A score close to 1 indicates a compact and well-separated cluster, while a score close to  $-1$  indicates a cluster with a large spread and/or poor separability. The silhouette analysis was performed using MATLAB<sup>®</sup> (Mathworks, Inc.).

## Results

### Emotional Meanings of Clips

Tests of normality, sphericity (Mauchly's test of sphericity), equality of covariance matrices (Box's M), and multicollinearity (Pearson's correlation) for the three indices indicated that our data have violations in assumptions check for skewness, kurtosis, sphericity, equality of covariance, and correlations to validate the use of parametric ANOVA or MANOVA tests across three modalities (audio only, video only, and audio-visual). Hence, for comparisons across modalities, we opted for the non-parametric alternative, K-W test. In order to focus on the investigation of "information-integration conditions," we report only critical results related to cross-modal stimuli in the manuscript, and other detailed investigations of multiple K-W tests and one-way MANOVA tests using three indices (evaluation, activity, potency) as dependent variables. "Modality" (audio only, video only, and cross modal), "clip" (A1, A2, V1, V2, A3V3, A5V1, and so on), "synchronization" (original vs. altered), or any interactions (of

**TABLE 7 | Mean and Standard Deviation Comparison between Original and Altered Groups for Evaluation, Activity, and Potency Indices of the Control Clip (A3V3).**

		<i>N</i>	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Evaluation	Original group	42	0.226	0.306	0.047	0.131	0.322	−0.438	0.938
	Altered group	53	0.212	0.292	0.040	0.132	0.293	−0.500	0.875
	Total	95	0.218	0.297	0.030	0.158	0.279	−0.500	0.938
Activity	Original group	42	0.238	0.216	0.033	0.171	0.305	−0.313	0.750
	Altered group	53	0.251	0.284	0.039	0.173	0.330	−0.625	1.000
	Total	95	0.245	0.255	0.026	0.193	0.297	−0.625	1.000
Potency	Original group	42	0.109	0.373	0.058	−0.008	0.225	−0.938	0.938
	Altered group	53	0.179	0.335	0.046	0.087	0.272	−0.625	0.813
	Total	95	0.148	0.352	0.036	0.076	0.220	−0.938	0.938

modality, clip, and synchronization) as independent variables will be provided separately (Table S2).

Table 8 shows the indices' values (evaluation, activity, and potency) for each mode of presentation, namely unimodal stimuli and the auditory-visual integrated stimuli. The valence values for each clip were found to have similar values to those of auditory stimuli in the original groups (A1V1, A2V2, and A3V3 had positive valence; in other words, the evaluation index  $> 0$ ; A4V4 and A5V5 had negative valence, as the evaluation index  $< 0$  with small difference in variance; A1V1 ( $0.66 \pm 0.20$ ), A2V2 ( $0.43 \pm 0.27$ ), A3V3 ( $0.23 \pm 0.31$ ), A4V4 ( $-0.42 \pm 0.28$ ), and A5V5 ( $-0.20 \pm 0.33$ ).

To inspect perceived emotional information (evaluation, activity, and potency) in interactions of "synchronization" (original vs. altered), "modality" (audio only, visual only, and cross modal), and "clip" (A1 ~ A5V1), we conducted several K-W tests. The test results showed that there were statistically significant differences among the conditions caused by different synchronization, modalities, and clips on evaluation, activity, and potency indices (all  $p \leq 0.001$ , except  $p < 0.05$  for clip 3 (for potency), clip 4 (for evaluation and potency), and clip 5 for evaluation and activity), except clip 3 for evaluation ( $p = 0.389$ ) and clip 5 for potency ( $p = 0.069$ ), as shown in Table 9.

To further investigate the significant differences shown in Table 9, we conducted several other K-W tests among different modalities in the same clip groups (for example, A1 vs. V1, A1 vs. A1V1, A1 vs. A1V4, A1V1 vs. A1V4, and so on). However, in order to focus on investigating the added-value effects that result in affective enhancement, we only report the results of the comparison between a high-responsive unimodal (audio only) vs. original visual music (Table 10), and unimodal (audio only) vs. altered visual music (Table 11). The comparison between auditory-channel only and the original synchronization cross-modal stimuli results indicated a statistically significant difference in potency scores between A2 and A2V2,  $\chi^2_{(1)} = 6.173$ ,

$p = 0.013$ , with a mean rank index score of 30.970 for A1 and 43.520 for A1V1. The A3 and A3V3 comparison showed a significant difference in the activity index,  $\chi^2_{(1)} = 23.984$ ,  $p = 0.000$ , with a mean rank index score of 51.850 for A3 and 27.120 for A3V3. The A4 and A4V4 comparison showed a significant difference in the activity index,  $\chi^2_{(1)} = 33.707$ ,  $p = 0.000$ , with a mean rank index score of 54.440 for A4 and 25.080 for A4V4, while the A5 vs. A5V5 comparison indicated a significant difference in the evaluation index,  $\chi^2_{(1)} = 5.791$ ,  $p = 0.016$ , with a mean rank index score of 30.630 for A5 and 42.740 for A5V5. No statistically different effect of clips (A1 or A1V1) on any of three indices (evaluation, activity, and potency) was shown. In four out of five comparison cases, the cross-modal stimuli showed a higher mean rank over audio-only in the same group in the evaluation index; A1V1 (mean rank = 40.420) vs. A1 (mean rank = 34.920), A2V2 (mean rank = 41.850) vs. A2 (mean rank = 33.110), A3V3 (mean rank = 41.200) vs. A3 (mean rank = 33.920), and A5V5 (mean rank = 42.740) vs. A5 (mean rank = 30.630).

The results of the comparison between auditory-channel only and altered (cross-binding) synchronization cross-modal stimuli revealed eight (out of 12) cases of statistically significant differences on the evaluation, activity, and potency indices (Table 11). In all clip groups, the cross-modal stimuli showed a lower mean rank than audio-only in the same group on the evaluation index. A more detailed investigation of the effects of the individual clips on three indices was conducted via a one-way ANOVA, and the result thereof will be provided separately (Table S3).

Similarly to Figure 3, the scatter plot representation of the original five visual music clips in 2D quadrants (valence and arousal dimensions) per modality (visual only, audio only, and visual music) is shown in Figure 6 as a simple representation of the overall emotion information derived from the clips. Emotional meaning aspects of each stimuli using the

**TABLE 8 | Means and STDs of three indices describing the emotional meanings of each clip stimulus in statistics: the highest absolute value in valence (evaluation) is indicated in bold type.**

Modality	Indices	Clip 1	Clip 2	Clip 3	Clip 4	Clip 5
Audio only	Evaluation	0.59 $\pm$ 0.25 (A1)	0.33 $\pm$ 0.21 (A2)	0.16 $\pm$ 0.23 (A3)	-0.30 $\pm$ 0.24 (A4)	-0.35 $\pm$ 0.26 (A5)
	Activity	0.29 $\pm$ 0.15 (A1)	0.18 $\pm$ 0.22 (A2)	0.50 $\pm$ 0.20 (A3)	0.45 $\pm$ 0.22 (A4)	0.29 $\pm$ 0.15 (A5)
	Potency	0.52 $\pm$ 0.24 (A1)	0.22 $\pm$ 0.27 (A2)	0.04 $\pm$ 0.29 (A3)	-0.49 $\pm$ 0.21 (A4)	-0.23 $\pm$ 0.28 (A5)
Visual only	Evaluation	0.40 $\pm$ 0.35 (V1)	0.10 $\pm$ 0.37 (V2)	<b>0.26 <math>\pm</math> 0.31 (V3)</b>	-0.21 $\pm$ 0.29 (V4)	-0.20 $\pm$ 0.40 (V5)
	Activity	-0.05 $\pm$ 0.30 (V1)	-0.17 $\pm$ 0.35 (V2)	0.21 $\pm$ 0.21 (V3)	0.24 $\pm$ 0.30 (V4)	0.09 $\pm$ 0.35 (V5)
	Potency	0.35 $\pm$ 0.34 (V1)	0.00 $\pm$ 0.33 (V2)	0.31 $\pm$ 0.37 (V3)	-0.28 $\pm$ 0.27 (V4)	-0.39 $\pm$ 0.37 (V5)
Original visual music	Evaluation	<b>0.66 <math>\pm</math> 0.20 (A1V1)</b>	<b>0.43 <math>\pm</math> 0.27 (A2V2)</b>	0.23 $\pm$ 0.31 (A3V3)	<b>-0.42 <math>\pm</math> 0.28 (A4V4)</b>	-0.20 $\pm$ 0.33 (A5V5)
	Activity	0.26 $\pm$ 0.15 (A1V1)	0.17 $\pm$ 0.25 (A2V2)	0.24 $\pm$ 0.22 (A3V3)	0.03 $\pm$ 0.27 (A4V4)	0.25 $\pm$ 0.16 (A5V5)
	Potency	0.51 $\pm$ 0.25 (A1V1)	0.37 $\pm$ 0.25 (A2V2)	0.11 $\pm$ 0.37 (A3V3)	-0.48 $\pm$ 0.31 (A4V4)	-0.27 $\pm$ 0.34 (A5V5)
Altered visual music	Evaluation	0.41 $\pm$ 0.37 (A1V4)	0.06 $\pm$ 0.32 (A2V5)	0.21 $\pm$ 0.29 (A3V3)	-0.36 $\pm$ 0.32 (A4V2)	<b>-0.41 <math>\pm</math> 0.27 (A5V1)</b>
	Activity	0.25 $\pm$ 0.20 (A1V4)	-0.01 $\pm$ 0.28 (A2V5)	0.25 $\pm$ 0.28 (A3V3)	-0.11 $\pm$ 0.23 (A4V2)	0.11 $\pm$ 0.35 (A5V1)
	Potency	0.28 $\pm$ 0.38 (A1V4)	0.07 $\pm$ 0.28 (A2V5)	0.18 $\pm$ 0.34 (A3V3)	-0.37 $\pm$ 0.25 (A4V2)	-0.37 $\pm$ 0.25 (A5V1)

Data are described as mean  $\pm$  SD.

**TABLE 9 | Results of the Kruskal-Wallis Test comparing the effect of clips with similar emotional meanings across modalities and synchronizations of the three perceived emotion assessment indices (evaluation, activity, and potency).**

Emotion Assessment Indices		N	Mean	Std. Deviation	Min	Max	Percentiles		
							25th	50th (Median)	75th
DESCRIPTIVE STATISTICS (NPar TEST)									
Clip 1 Group	Evaluation	155	0.515	0.321	−0.560	1.000	0.375	0.563	0.750
	Activity	155	0.209	0.233	−0.560	0.630	0.125	0.250	0.375
	Potency	155	0.404	0.328	−0.880	0.940	0.250	0.438	0.625
	Total	155	18.610	20.115	1.000	61.000	11.000	14.000	14.000
Clip 2 Group	Evaluation	155	0.224	0.335	−0.560	1.000	0.000	0.250	0.438
	Activity	155	0.050	0.303	−0.750	0.750	−0.125	0.063	0.250
	Potency	155	0.172	0.312	−0.630	0.940	−0.063	0.188	0.438
	Total	155	25.740	18.849	2.000	62.000	22.000	25.000	25.000
Clip 3 Group	Evaluation	155	0.214	0.286	−0.500	1.000	0.000	0.250	0.438
	Activity	155	0.293	0.259	−0.630	1.000	0.125	0.250	0.500
	Potency	155	0.153	0.351	−0.940	0.940	−0.063	0.188	0.375
	Total	155	32.180	18.717	3.000	63.000	33.000	34.000	34.000
Clip 4 Group	Evaluation	155	−0.340	0.294	−1.000	0.630	−0.563	−0.375	−0.125
	Activity	155	0.108	0.329	−0.750	0.880	−0.125	0.125	0.375
	Potency	154	−0.416	0.295	−0.940	0.810	−0.625	−0.438	−0.250
	Total	155	38.280	19.531	4.000	64.000	42.000	42.000	44.000
Clip 5 Group	Evaluation	154	−0.305	0.322	−1.000	0.560	−0.516	−0.313	−0.063
	Activity	155	0.185	0.282	−0.690	1.000	0.000	0.188	0.375
	Potency	154	−0.316	0.307	−1.000	0.560	−0.500	−0.344	−0.125
	Total	155	44.730	21.274	5.000	65.000	51.000	51.000	55.000
Emotion Assessment Indices		Clip ID	N	Mean Rank	Chi-Square	df	Asymp. Sig.	Effect Size (η²)	
RANKS (KRUSKAL-WALLIS TEST)									
Clip 1 Group	Evaluation**	A1	33	87.110	18.206	3	0.000	0.118	
		A1V1	42	98.200					
		A1V4	53	64.390					
		V1	27	62.170					
		Total	155						
	Activity**	A1	33	90.180	27.684	3	0.000	0.180	
		A1V1	42	85.060					
		A1V4	53	85.640					
		V1	27	37.130					
		Total	155						
	Potency**	A1	33	92.890	16.127	3	0.001	0.105	
		A1V1	42	92.520					
		A1V4	53	61.710					
		V1	27	69.190					
		Total	155						
Clip 2 Group	Evaluation**	A2	33	92.300	36.510	3	0.000	0.237	
		A2V2	42	105.670					
		A2V5	53	55.020					

(Continued)

TABLE 9 | Continued

Emotion Assessment Indices		Clip ID	N	Mean Rank	Chi-Square	df	Asymp. Sig.	Effect Size ( $\eta^2$ )
	Activity**	V2	27	62.590	26.277	3	0.000	0.171
		Total	155					
		A2	33	97.790				
		A2V2	42	93.810				
		A2V5	53	68.490				
		V2	27	47.890				
	Potency**	Total	155		31.111	3	0.000	0.202
		A2	33	83.800				
		A2V2	42	107.130				
		A2V5	53	62.240				
		V2	27	56.540				
Clip 3 Group	Evaluation	Total	155		3.019	3	0.389	0.020
		A3	33	66.970				
		A3V3	42	80.830				
		A3V3c	53	78.600				
		V3	27	<b>85.890</b>				
	Activity**	Total	155		31.326	3	0.000	0.203
		A3	33	116.110				
		A3V3	42	67.930				
		A3V3c	53	70.920				
		V3	27	60.980				
	Potency*	Total	155		10.298	3	0.016	0.067
		A3	33	61.650				
		A3V3	42	73.940				
		A3V3c	53	81.310				
		V3	27	97.800				
Clip 4 Group	Evaluation*	Total	155		10.020	3	0.018	0.065
		A4	33	85.830				
		A4V2	53	72.420				
		A4V4	42	66.180				
		V4	27	<b>97.760</b>				
	Activity**	Total	155		66.513	3	0.000	0.432
		A4	33	124.860				
		A4V2	53	48.070				
		A4V4	42	67.400				
		V4	27	95.960				
	Potency**	Total	155		12.611	3	0.006	0.082
		A4	32	68.090				
		A4V2	53	82.140				
		A4V4	42	64.350				
		V4	27	100.000				
Clip 5 Group	Evaluation*	Total	154		13.786	3	0.003	0.090
		A5	32	71.410				

(Continued)

TABLE 9 | Continued

Emotion Assessment Indices	Clip ID	N	Mean Rank	Chi-Square	df	Asymp. Sig.	Effect Size ( $\eta^2$ )
Activity*	A5V1	53	62.760	14.845	3	0.002	0.096
	A5V5	42	<b>94.370</b>				
	V5	27	87.410				
	Total	154					
	A5	33	97.380				
	A5V1	53	67.800				
	A5V5	42	86.980				
	V5	27	60.370				
	Total	155					
	A5	33	90.940				
Potency	A5V1	53	70.310	7.101	3	0.069	0.046
	A5V5	42	83.520				
	V5	26	65.370				
	Total	154					

The rank-based non-parametric analysis was used to examine the rank-order of four clips in a group (audio only, video only, original visual music vs. altered visual music), and to determine whether there were statistically significant differences between two (or more) modalities. A higher value in mean rank indicates that the upper rank (for example, for the clip group 1, A1V1 (Mean rank = 98.200) is higher in rank than are the other three clips A1 (Mean rank = 87.110), V1 (Mean rank = 10.06) or A1V4 (Mean rank = 64.390) on the evaluation index. \* Indicates a significant difference  $p < 0.05$ , \*\* indicates  $p < 0.001$ , and bold type indicates the highest rank in valence (evaluation).

evaluation, activity and potency indices, and following the three media combination conditions (congruence, complementation, contest), are illustrated as bar graphs in **Figure 7**.

### Compactness Check

We used the silhouette index to quantify the quality of clustering for each mode of presentation within the 2D plane (**Figure 6**). For each clip, the silhouette value was obtained from each subject's response by using the evaluation and activity indices (**Table 9**), and returned a value representing the compactness (cohesion in emotional rating) and separation (differentiation from other clip responses) of the subjects' responses for different modality or different spatio-temporal combinations (such as separating original auditory-visual synchronizations into altered combinations) for each clip per modality.

The clustering analysis shown in **Table 12** indicates that the audio only category ( $\text{sil} = 0.085$ ) and original visual music category ( $\text{sil} = 0.045$ ) were better clustered overall compared to the visual only category ( $\text{sil} = -0.076$ ) and the altered visual music category ( $\text{sil} = -0.066$ ). In particular, clip 1 showed the best clustering score of all the clips for the audio-only and visual music categories, indicating that its emotional connotation was best recognized by the subjects.

We then checked the compactness of each clip based on the average distances between the mean rating of each clip (centroid) and the subject rating using Euclidean distance and the indices evaluation, activity, and potency as dimensions; this estimation was performed for each clip and for each modality. The A1V1 and A2V2 visual music integrations were revealed to be the most compact scores (clustering and average distribution) in all other modal categories.

## Discussion

### Enhancement Effect Hypothesis Verification

We used three major indices (evaluation, activity, and potency) to thoroughly quantify the aesthetic/emotional meanings of abstract synthetic visual music clips in this study. To identify any proof of the enhancement effect, we looked for heightened mean or median scores for auditory-visual compared to other comparable unimodal scores for subjects' responses on the evaluation index because it is valence factor, which indicates the "likeness" characteristic of emotion. We also considered compactness via a clustering analysis and an average distance assessment to inspect the enhancement effect in relation to congruency in valence and arousal information by inspecting media combination conditions (congruence, complementation, and contest).

In our study, three clips were categorized in the congruence condition (A3V3, A4V4, and A5V5), four clips in complementation, and two clips in the contest condition (see **Table 6**, **Figure 7**, and **Figure S1**). These three clips exhibited no indication of higher emotional perceptive responses in both mean and median comparisons in the valence level (**Tables 8, 10**). However, with regard to the median rank, A3V3 and A5V5 showed higher mean ranks than did the other modalities (**Table 10**, bold type). With regard to clustering, only A4V4 had a higher score compared to other modalities (**Table 12**), and none of them showed good compactness (**Table 13**). None of these three clips ranked higher than did the other modalities for more than two of the four different assumptions (mean comparison, median rank comparison, clustering score, and compactness score) to check an enhancement effect. Therefore, our result seems to contradict our original hypothesis that the enhancement effect would be reliant on the congruent media combination

**TABLE 10 | Results of a Kruskal-Wallis Test comparing the effect of the emotional meaning of clips between the auditory modality and the cross-modality (original synchronization) on the three perceived emotion assessment indices (evaluation, activity, and potency).**

Emotion Assessment Indices		N	Mean	Std. Deviation	Min	Max	Percentiles		
							25th	50th (Median)	75th
DESCRIPTIVE STATISTICS (NPar TEST)									
Clip 1 Group	Evaluation	75	0.632	0.226	−0.060	1.000	0.500	0.688	0.813
	Activity	75	0.275	0.148	−0.060	0.630	0.188	0.250	0.375
	Potency	75	0.513	0.242	−0.250	0.880	0.375	0.500	0.688
	Total	75	6.600	4.997	1.000	11.000	1.000	11.000	11.000
Clip 2 Group	Evaluation	75	0.388	0.250	−0.130	1.000	0.188	0.375	0.563
	Activity	75	0.173	0.235	−0.500	0.750	0.063	0.188	0.313
	Potency	75	0.307	0.268	−0.250	0.940	0.125	0.313	0.500
	Total	75	13.200	9.995	2.000	22.000	2.000	22.000	22.000
Clip 3 Group	Evaluation	75	0.197	0.275	−0.440	0.940	0.000	0.188	0.375
	Activity	75	0.352	0.245	−0.310	0.810	0.188	0.375	0.500
	Potency	75	0.079	0.340	−0.940	0.940	−0.125	0.063	0.313
	Total	75	19.800	14.992	3.000	33.000	3.000	33.000	33.000
Clip 4 Group	Evaluation	75	−0.369	0.268	−1.000	0.190	−0.500	−0.438	−0.188
	Activity	75	0.212	0.324	−0.630	0.810	0.000	0.250	0.500
	Potency	74	−0.487	0.268	−0.940	0.440	−0.688	−0.500	−0.375
	Total	75	26.400	19.989	4.000	44.000	4.000	44.000	44.000
Clip 5 Group	Evaluation	74	−0.267	0.308	−1.000	0.500	−0.500	−0.219	−0.063
	Activity	75	0.271	0.156	−0.130	0.560	0.188	0.250	0.375
	Potency	75	−0.254	0.313	−1.000	0.500	−0.500	−0.250	−0.063
	Total	75	33.000	24.986	5.000	55.000	5.000	55.000	55.000
Emotion Assessment Indices		Clip ID	N	Mean Rank	Chi-Square	df	Asymp. Sig.	Effect Size (η²)	
RANKS (KRUSKAL-WALLIS TEST)									
Clip 1 Group	Evaluation	A1	33	34.920	1.183	1	0.277	0.016	
		A1V1	42	40.420					
		Total	75						
	Activity	A1	33	39.650	0.344	1	0.558	0.005	
		A1V1	42	36.700					
		Total	75						
	Potency	A1	33	38.060	0.000	1	0.983	0.000	
		A1V1	42	37.950					
		Total	75						
Clip 2 Group	Evaluation	A2	33	33.110	2.994	1	0.084	0.040	
		A2V2	42	41.850					
		Total	75						
	Activity	A2	33	39.200	0.180	1	0.672	0.002	
		A2V2	42	37.060					
		Total	75						
	Potency*	A2	33	30.970	6.173	1	0.013	0.083	
		A2V2	42	43.520					
		Total	75						
Clip 3 Group	Evaluation	A3	33	33.920	2.074	1	0.150	0.028	
		A3V3	42	41.200					

(Continued)

TABLE 10 | Continued

	Emotion Assessment Indices	Clip ID	N	Mean Rank	Chi-Square	df	Asymp. Sig.	Effect Size ( $\eta^2$ )
Clip 4 Group	Activity**	Total	75					
		A3	33	51.850	23.984	1	0.000	0.324
		A3V3	42	27.120				
	Potency	Total	75					
		A3	33	34.800	1.274	1	0.259	0.017
		A3V3	42	40.510				
	Evaluation*	Total	75					
		A4	33	<b>43.670</b>	4.027	1	0.045	0.054
		A4V4	42	33.550				
	Activity**	Total	75					
		A4	33	54.440	33.707	1	0.000	0.456
		A4V4	42	25.080				
Clip 5 Group	Potency	Total	75					
		A4	32	38.660	0.164	1	0.685	0.002
		A4V4	42	36.620				
	Evaluation*	Total	74					
		A5	32	30.630	5.791	1	0.016	0.079
		A5V5	42	<b>42.740</b>				
	Activity	Total	74					
		A5	33	42.110	2.129	1	0.145	0.029
		A5V5	42	34.770				
	Potency	Total	75					
		A5	33	39.500	0.281	1	0.596	0.004
		A5V5	42	36.820				
		Total	75					

The rank-based non-parametric analysis was used to examine the rank-order of the three modalities (audio only, video only, vs. visual music), and to determine whether there were statistically significant differences between two (or more) modalities. A higher value in mean rank indicates the upper rank (for example, in the clip group 1, A1V1 (Mean rank = 40.420) is higher in rank than is A1 (Mean rank = 34.920) on the evaluation index. \* Indicates significant difference  $p < 0.05$ , \*\* indicates  $p < 0.001$ , and bold type indicates the highest rank in valence (evaluation).

condition in valence and arousal throughout our experiment results.

However, in this study, we found a great likelihood of valence and texture/control (evaluation and potency) information assessments of the visual music linked to the polarity of the auditory channel information in the contest conditions (Figure 7). Although the means and median ranks of evaluation, activity, and potency of visual music fluctuated between the corresponding indices' values of the two unimodal channels in most cases (Tables 8, 9, and Figure 7), we observed a few possible heightened emotional perception cases (for example, positive-augmentation or negative-diminishment on emotional aspect factors). "A1V1" showed a positively augmented mean value in the evaluation (valence), the highest median rank in the evaluation (valence), and the strongest compactness (both clustering and average distance) compared to its comparable

unimodal clips (Tables 8, 9, 12, 13); "A2V2" showed a positively augmented mean value for evaluation, the highest media rank in the evaluation, and the smallest average distance (highest compactness) compared to unimodal stimuli. "A1V1" and "A2V2" both belong to the complementation combination condition with congruency in valence (positive) and incongruency in arousal [active (+) in audio and passive (−) in visual]. "A4V4," in the conformance condition, showed more negative values in evaluation, and the highest clustering density (Tables 8, 9). "A5V1," in a contest condition, showed more increased negative values in the evaluation and potency indices, but its clustering density and general mean-distance results indicated that perceived emotions were widespread compared to corresponding dual unimodal sources. For all of the other integrations, the visual music emotional perception evaluation rate showed various degrees in levels

**TABLE 11 | Results of the Kruskal-Wallis Test comparing the effect of the emotional meaning of clips between auditory modality and cross-modality (altered synchronization) on the three perceived emotion assessment indices (evaluation, activity, and potency).**

Emotion Assessment Indices		<i>N</i>	Mean	Std. Deviation	Min	Max	Percentiles		
							25th	50th (Median)	75th
<b>DESCRIPTIVE STATISTICS (NPar TEST)</b>									
Clip 1 Group	Evaluation	86	0.480	0.336	−0.560	1.000	0.375	0.563	0.688
	Activity	86	0.264	0.182	−0.500	0.630	0.188	0.281	0.375
	Potency	86	0.371	0.352	−0.880	0.940	0.188	0.438	0.625
	Total	86	9.010	6.359	1.000	14.000	1.000	14.000	14.000
Clip 2 Group	Evaluation	86	0.160	0.310	−0.560	0.880	−0.016	0.188	0.375
	Activity	86	0.064	0.275	−0.750	0.500	−0.125	0.125	0.313
	Potency	86	0.127	0.286	−0.560	0.940	−0.063	0.125	0.313
	Total	86	16.170	11.250	2.000	25.000	2.000	25.000	25.000
Clip 4 Group	Evaluation	86	−0.339	0.289	−0.940	0.630	−0.563	−0.375	−0.125
	Activity	86	0.108	0.352	−0.750	0.810	−0.188	0.031	0.438
	Potency	85	−0.423	0.288	−0.940	0.810	−0.563	−0.438	−0.313
	Total	86	27.420	18.587	4.000	42.000	4.000	42.000	42.000
Clip 5 Group	Evaluation	85	−0.389	0.262	−0.880	0.250	−0.594	−0.438	−0.188
	Activity	86	0.182	0.300	−0.690	0.810	−0.063	0.250	0.438
	Potency	86	−0.315	0.266	−1.000	0.440	−0.500	−0.313	−0.188
	Total	86	33.350	22.501	5.000	51.000	5.000	51.000	51.000

Emotion Assessment Indices		Clip ID	<i>N</i>	Mean Rank	Chi-Square	<i>df</i>	Asymp. Sig.	Effect Size ( $\eta^2$ )
<b>RANKS (KRUSKAL-WALLIS TEST)</b>								
Clip 1 Group	Evaluation*	A1	33	<b>51.300</b>	5.267	1	0.022	0.062
		A1V4	53	38.640				
		Total	86					
	Activity	A1	33	44.620	0.110	1	0.740	0.001
		A1V4	53	42.800				
		Total	86					
	Potency**	A1	33	54.210	9.897	1	0.002	0.116
		A1V4	53	36.830				
		Total	86					
Clip 2 Group	Evaluation**	A2	33	<b>57.450</b>	16.803	1	0.000	0.198
		A2V5	53	34.810				
		Total	86					
	Activity**	A2	33	54.240	9.982	1	0.002	0.117
		A2V5	53	36.810				
		Total	86					
	Potency*	A2	33	51.470	5.502	1	0.019	0.065
		A2V5	53	38.540				
		Total	86					
Clip 4 Group	Evaluation	A4	33	<b>48.200</b>	1.905	1	0.168	0.022
		A4V2	53	40.580				
		Total	86					
	Activity**	A4	33	67.650	50.346	1	0.000	0.592
		A4V2	53	28.460				
		Total	86					

(Continued)

TABLE 11 | Continued

	Emotion Assessment Indices	Clip ID	N	Mean Rank	Chi-Square	df	Asymp. Sig.	Effect Size ( $\eta^2$ )
Clip 5 Group	Potency	Total	86					
		A4	32	38.330	1.858	1	0.173	0.022
		A4V2	53	45.820				
	Evaluation	Total	85					
		A5	32	<b>46.410</b>	0.984	1	0.321	0.012
		A5V1	53	40.940				
	Activity*	Total	85					
		A5	33	52.480	6.994	1	0.008	0.082
		A5V1	53	37.910				
	Potency*	Total	86					
		A5	33	51.360	5.357	1	0.021	0.063
		A5V1	53	38.600				
		Total	86					

The rank-based non-parametric analysis was used to examine the rank-order of the three modalities (audio only, video only, vs. visual music), and to determine whether there were statistically significant differences between two (or more) modalities. A higher value in mean rank indicates the upper rank; (for example, in the clip group 1, A1 (Mean rank = 51.300) is higher in rank than is A1V4 (Mean rank = 38.640) on the evaluation index. \* Indicates significant difference  $p < 0.05$ , \*\* indicates  $p < 0.001$ , and bold type indicates the highest rank per category.

of integration formations (Figure 7). Overall, these results indicate that synchronizing audio and video information in the complementation combination condition could show instances of heightened perceptive emotion (enhancement effect) of multimodal perception in our study.

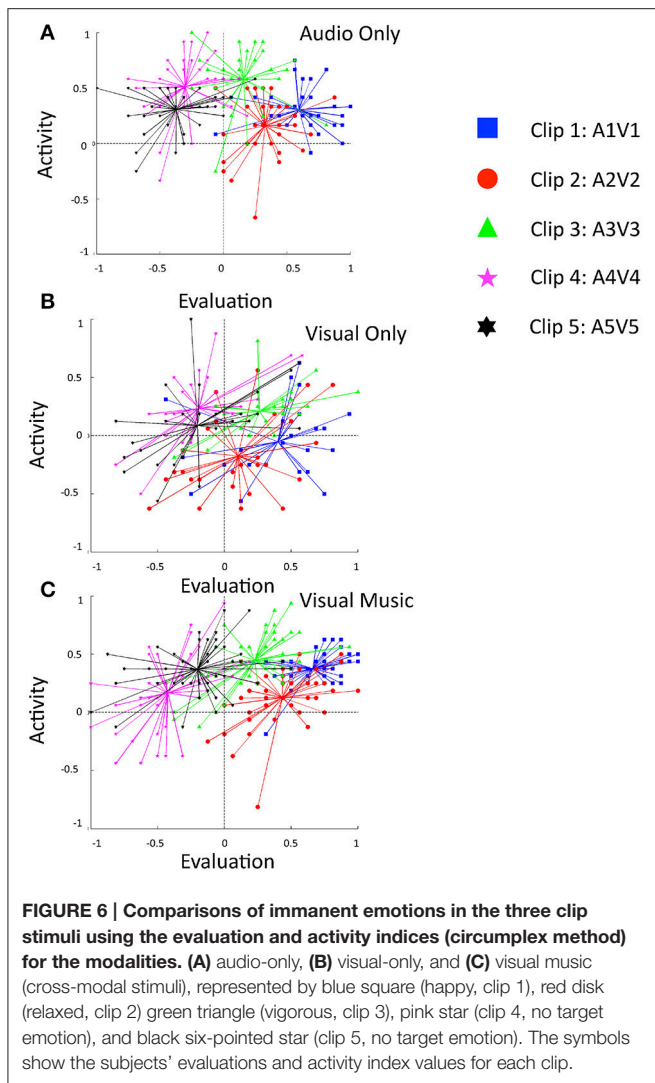
### Cross-Modal Interplays

In our investigations, we observed a few notable results: First, we observed an *auditory dominant polarization* tendency, consistent with a few previous studies demonstrating auditory dominance over abstract vision in the temporal and perceptual processing of multimodalities (Marshall and Cohen, 1988; Repp and Penel, 2002; Somsaman, 2004). Since visual music is an abstract animation representing the purity of music by nature, it may be imperative for visual music to have audio channel information conveying stronger affective meanings (evaluation, activity, and potency values) via visual channel information. Nonetheless, the differences in emotional information on arousal (activity index) between the auditory and visual modalities left the auditory channel emotion information with a higher arousal value, hence dominating the visual channel and transferring affective meaning toward the overall emotional perception of cross-modal perception (the absolute mean value of activity for the auditory channel always showed a greater level than did the visual channel in *all* nine visual music presentations, and the mean rank of audio only was always higher than was visual only, as seen in Tables 8, 9).

Second, from the compactness response and silhouette analysis, we found that the overall perceptual grouping of “gestalt” (whole beauty) or “schema” (conceptual frameworks for expectations) in auditory-visual domains showed disperse

responses as a result of the altered, cross-matched stimuli compared to the original integrations. This indicates the interplay of semantic and structural congruency (sharing temporal accent patterns) between auditory and visual information in forming the focus of attention in cross-modal perception as cognitive psychologists implied good gestalt principles (see Cohen, 2001, p. 260). In our finding, the spatiotemporal information in the arbitrary cross-matching could not assemble into good synchronous groupings in structural and semantic features with temporal cues and movements (e.g., tempo, directions, dynamic); hence, it impeded creating better interplay focus of attention compared to the original stimuli. In particular, the arbitrary integration by cross-matching audio and video channel information from different sources created semantic and structural asynchronous distraction in multisensory perceptual grouping thus resulting in “worse” aesthetic emotion (e.g., A5V1).

Finally, we found two cases of the positive enhancement effect in aesthetic perception resulting from the functions of information integration (auditory and visual) in this experiment, namely A1V1 and A2V2. They had contrary polarity for activity values [audio only activity (+) vs. visual only activity (−)], but evaluation (and potency) exhibited congruency in their polarity [all positive (+)]. This suggests that positive congruency in valence (and texture/control) information with uncompetitive discrepancy in arousal levels between the visual and audio channels might trigger the enhancement effect in aesthetic emotion appraisals (see Figure 8). This finding possibly relates to the art rewarding hypothesis which a state of uncertainty recovers into predictable patterns resulting to rewarding effect of increased expectedness (Van de Cruys and Wagemans, 2011).



The congruency in valence and texture/control aspects in our A1V1 and A2V2 might, for example, implicitly stabilize the conjoint gestalt or schema that people use to form expectations or predictions whereas the substantial differences in the arousal aspect between the two modalities might implicitly emphasize elements that alter the continuous coding of predictive errors and recovering to predictable patterns. In other words, we assume that violations of prediction (predictive error) resulted from differences in any part where the two channels information return to a state of rewarding due to the formation of stable gestalt/schema. Other three original visual music stimuli did not indicate a strong enhancement effect in mean valence levels or median ranks although some showed improved compactness in silhouette index compared to its comparable unimodal stimuli (e.g., A4V4).

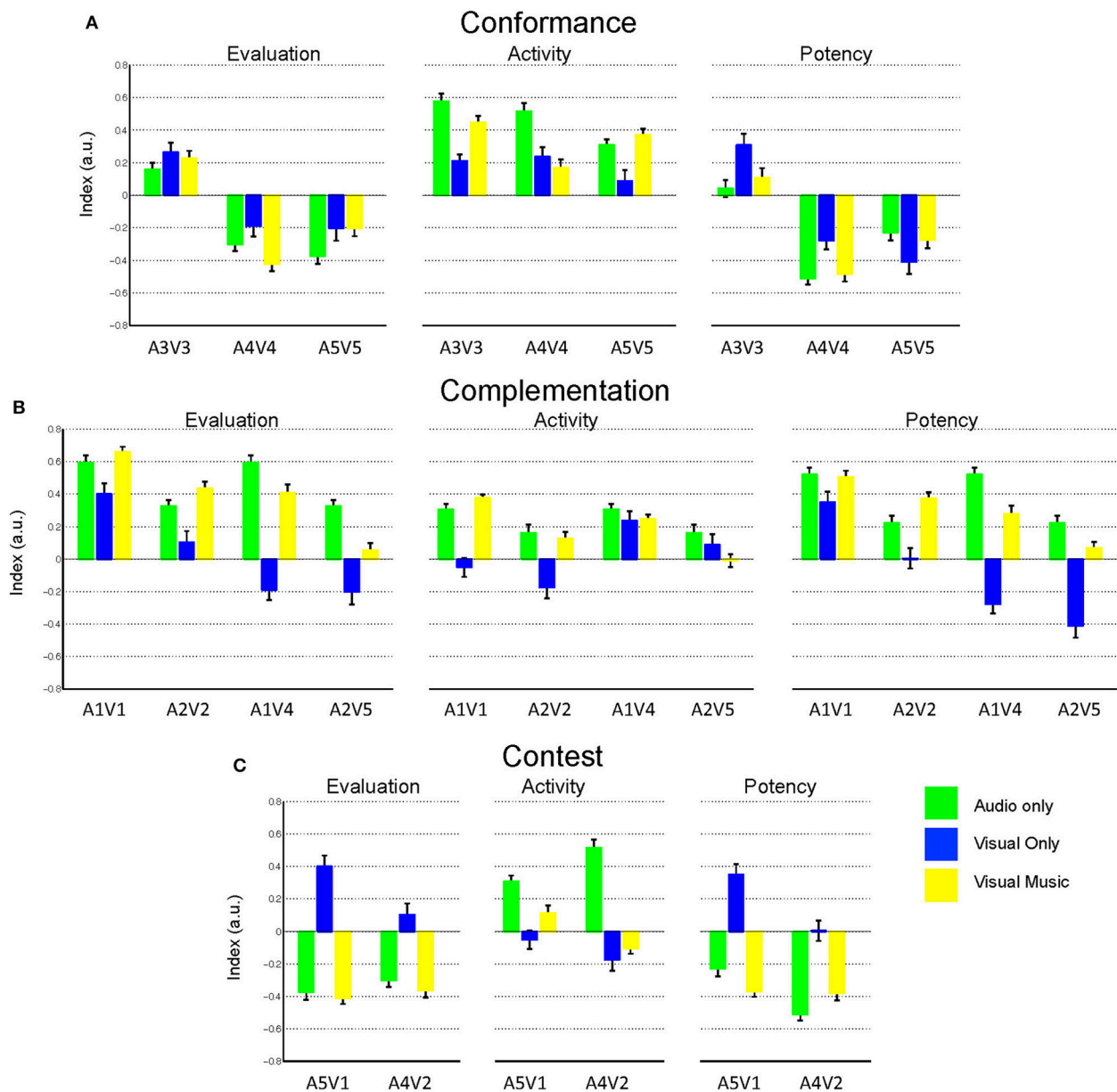
## GENERAL DISCUSSION

Our study inspected the two unimodal channel interplays as a functional-information-integration apparatus by examining

an enhancement effect in cross-modal evaluation processes of emotional visual music experience. During the two experiments, we could see that visual music can embody various structural variables that may cause important interactions between perceptual and cognitive aspects of the dual channels. The directive design guidelines we used to create target emotional stimuli in Experiment 1 (Tables 1, 2) indicate numerous parameters that can be used when creating visual music. The use of three extracted indices (evaluation, activity, and potency) to appraise ambiguous emotional meanings was effective to assess both the artwork and the subjects' responses in the study. It was also encouraging that we found a positive correlation between the evaluation index and the lateral frontal upperAlpha power differences in this study. Despite the small number of subjects in the study, the promising results from the frontal alpha asymmetry and its correlation with valence might encourage the inclusion of a wider range of physiological measurements to study the complex interactions between the external sensorial/perceptual context and the internal cognitive modulation/coding mechanism of aesthetic experience (see Figure 9 for an illustration of the partial potential interactions of sensorial context and cognitive coding, and Figure S5 for the temporal dynamics of spectral and spatial activation in EEG).

Our findings demonstrated the role of and interplays between the valence and arousal information in emotional evaluation of the auditory-visual aesthetic experience. The common effects of congruent positive valence between auditory and visual domains refer to good quality in gestalt/schema formation. Stronger arousal level of auditory channel information not only outweighed the visual channel information in making affective (valence and texture) judgments of visual music, but also ensued uncompetitive focus/attention to result increased states of positively predictable patterns. Hence, arousal levels and conditions hold a key role in modulating the excitement of affective emotion perception in visual music experience. Consequently, taking both dimensions of emotion (valence and arousal) into account is necessary to determine whether abstract auditory-visual stimuli carry strong, distinctive emotional meanings in particular.

As suggested by scholars in the field of aesthetic judgment and perception, emotion studies using works of art require particular insightful appraisal tools that should differ from depictive or prepositional expressions, such as pictures or languages (Takahashi, 1995). Hence, to infer emotional meanings from and to inspect the aesthetic perceptions of visual music as affective visual music, a careful choice of assessment factors for aesthetic perceptions of abstract visual music stimuli is crucial. More efforts to design refined and uncomplicated assessment apparatus for visual music perception would be challenging, yet worthy. The limitation of our mono-cultural background (college students in South Korea) makes it unable to generalize our findings as a truly representative universal aesthetic perception rule. Therefore, considering the findings in conjunction with all possible media integration conditions (Figure S1) will help to identify universal principles in visual music aesthetic



**FIGURE 7 | Bar graph summary of the immanent emotion for the nine visual music stimuli using evaluation, activity, and potency for the combination conditions of (A) conformance, (B) complementation and (C) contest.** Data are shown as mean  $\pm$  s.e.m. The mean increase of evaluation index was observed for clips A1V1 and A2V2 compared to their respective unimodal responses.

**TABLE 12 | Silhouette clustering index values using 2D factors (evaluation and activity) rating indices for the data presented in Figure 6.**

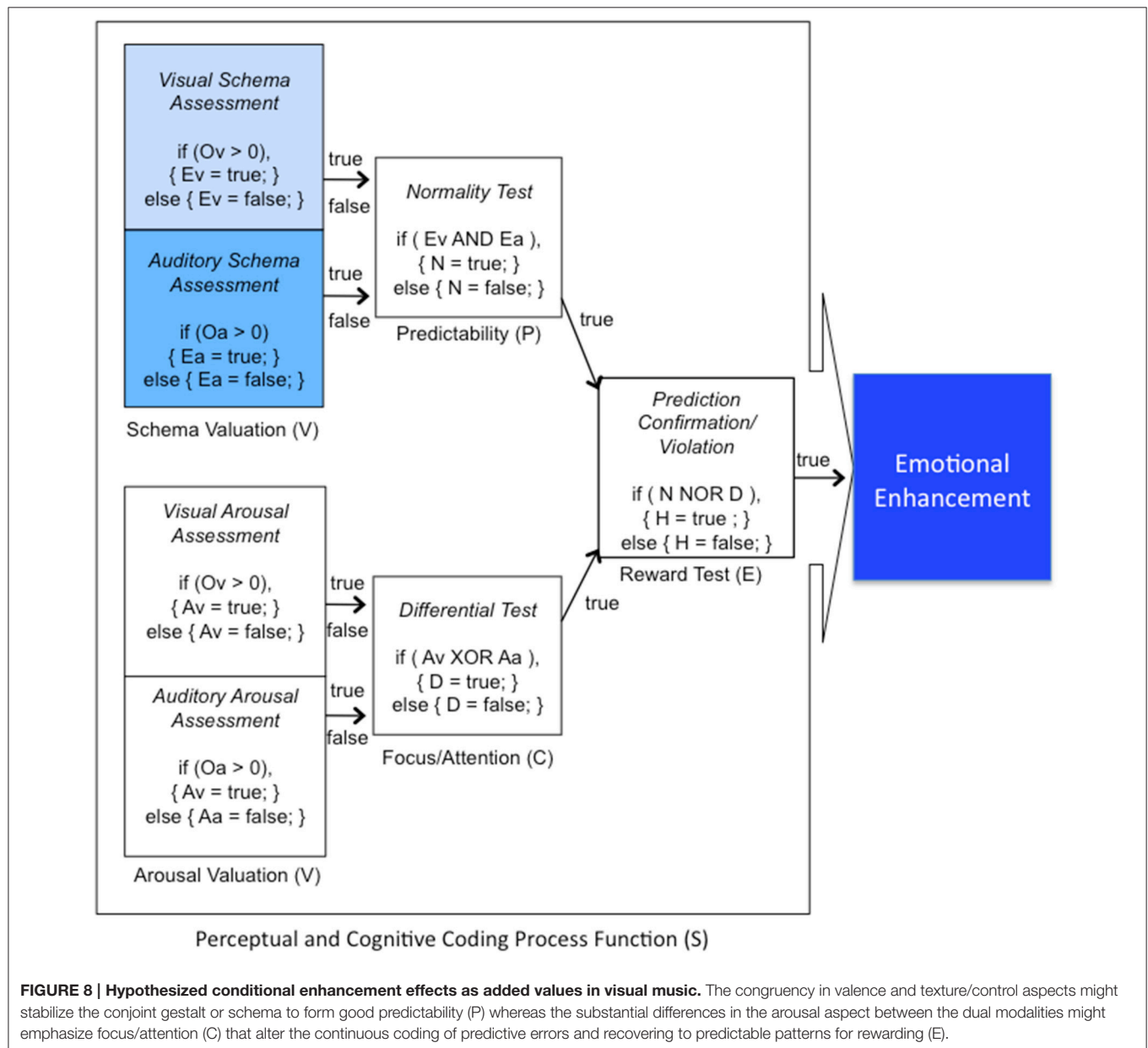
	Clip 1	Clip 2	Clip 3	Clip 4	Clip 5	Mean
Audio Only	0.274048	<b>-0.044834</b>	0.177347	-0.101890	<b>0.123215</b>	<b>0.085577</b>
Visual Only	-0.103724	-0.172770	<b>0.209722</b>	-0.013282	-0.301105	-0.076232
Original Visual Music	<b>0.538540</b> (A1V1)	-0.134250 (A2V2)	-0.181865 (A3V3)	<b>0.053864</b> (A4V4)	-0.047418 (A5V5)	0.045774
Altered Visual Music	(A1V4) 0.047593	(A2V5) -0.192984	(A3V3)* -0.121569	(A4V2) 0.048313	(A5V1) -0.111249	-0.065979

Silhouette values are given for individual clips and overall clustering quality. Silhouette values range within  $[-1, 1]$ , with 1 indicating the best clustering and  $-1$  the worst clustering. Bold values indicate the best clustering values relative to each clip (row-wise comparison). \*, control clip.

**TABLE 13 | Compactness of response.**

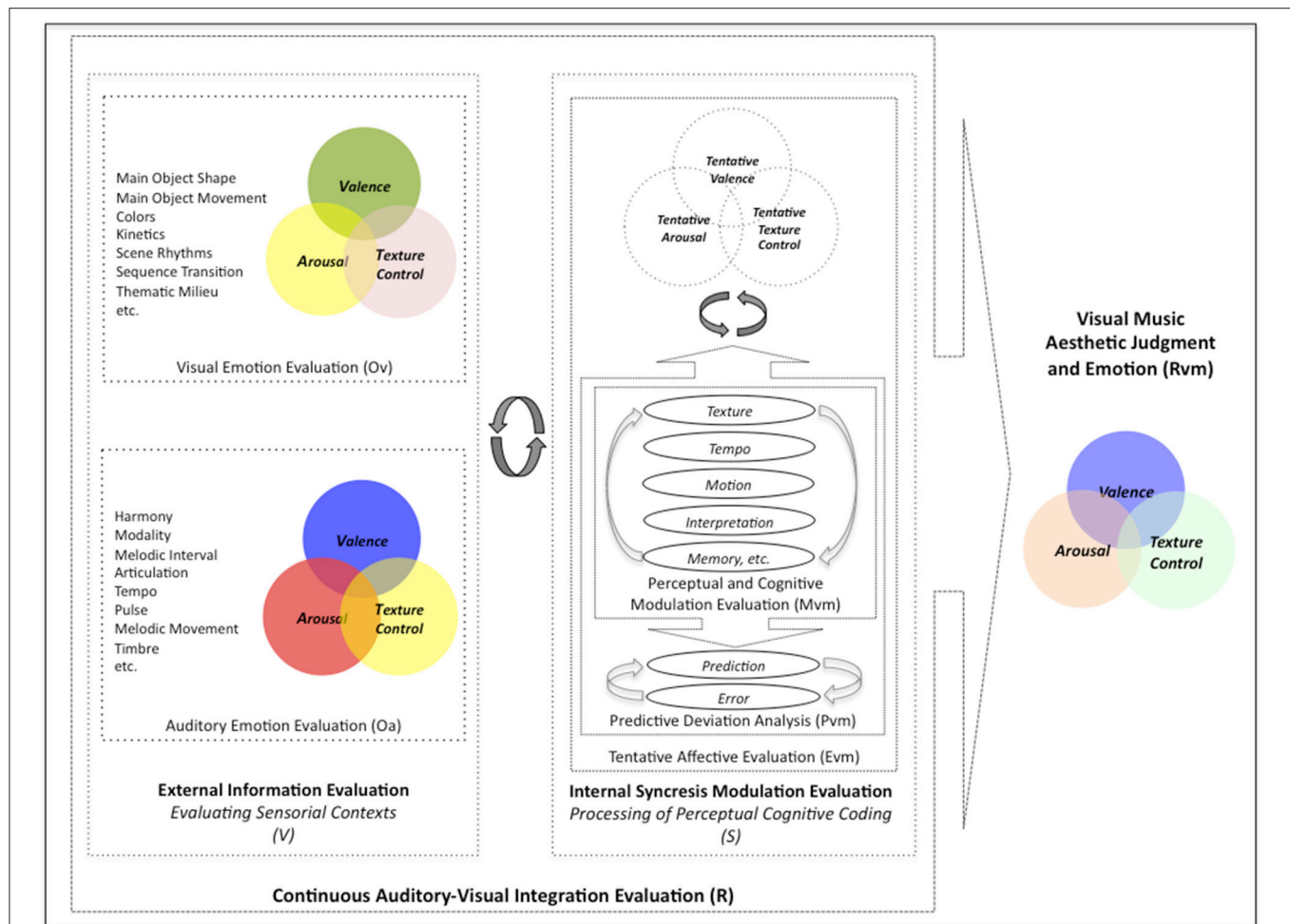
	Clip 1	Clip 2	Clip 3	Clip 4	Clip 5	All Clips
Audio only	0.34628204	0.394233616	<b>0.40508022</b>	<b>0.391176521</b>	<b>0.390453802</b>	<b>0.38544524</b>
Visual only	0.499628207	0.555422101	0.437046321	0.464087266	0.576930643	0.506622908
Original visual music	<b>0.295774755</b> (A1V1)	<b>0.393163965</b> (A2V2)	0.468533303 (A3V3)*	0.479346853 (A4V4)	0.466745095 (A5V5)	0.420712794
Altered visual music	0.454134911 A1V4	0.44759437 (A2V5)	0.460930243 (A3V3)*	0.419427517 (A4V2)	0.448666851 (A5V1)	0.44759437

Average distances to centroid using the evaluation, activity, and potency indices. The centroid for each clip was estimated first, and the average distance of each subject's response for the same clip was then estimated. Closer to zero indicates more compactness. Bold values indicate the closest compactness to relative centroid of each clip. \*, control clip.



perception. In addition, further investigations of how the interaction of three aspects of emotional meaning (valence, arousal, and texture/control) affect aesthetic emotion whilst

considering temporal factors of visual music may also expand the understanding of the perceptual process in aesthetic experiences.



**FIGURE 9 | A hypothesized functional-information-integration process within our model of continuous auditory-visual modulation integration perception.** Visual Music aesthetic experience requires multi-sensory perceptual, cognitive judgment, and processing that involve a number of continuous processing stages such as affective analysis (in valence, activity, and potency), perceptual and cognitive modulation analysis, and predictive deviation analysis.

## CONCLUSION

There has been a significant development in theories and experiments that explain the process of aesthetic perception and experience during the past decade (for a review, see Leder and Nadal, 2014; nonetheless, research studies on emotion elicitation have long relied on inflexible, static, or non-intentionally designed uni/cross-modal stimuli. Due to the lack of sufficient research evidence, aesthetic researchers have been calling for more sophisticated investigations of the interplay of perceptual and cognitive challenges with using novel, complex, and multidimensional stimuli (Leder et al., 2004; Redies, 2015). The need for new empirical approaches in aesthetic science requires an extensive amount of principled research effort to study the numerous components of emotion and competencies via several measurements, as Scherer and Zentner (2001) explained. The investigations of the process whereby art evokes emotion using a novel attempt in cross-modal aesthetic studies hence necessitate extensive research efforts with certain measurement competencies as important aspects.

Our empirical study of audio-visual aesthetic perception has a cross-disciplinary approach including music, visual aspects, aesthetics, neuroscience, and psychology and takes more of a holistic than an elementary approach; this is unconventional in several ways when compared to classical, disciplined paradigms. Initially, strong demands from commercial industries for practical psychotherapeutic contents cued our research team to bring artistic issues into the science laboratory. It inspired us to create artwork with verified, literature-based correlations with positive emotions, and to find ways to validate the ambiguous nature of visual music via the observable assessment of suitable measurements to translate it into psychological and cognitive science investigations. The aesthetic experience is known to have three components (artist, artwork, and beholder), and our study involved all three aspects; however, psychological aesthetic studies have historically been related to how art evokes an emotional response from viewers instead of exploring the factors that motivate individuals to produce art (see Shimamura, 2012, p. 24). When we demonstrated a basis for the directive production of emotional visual music to our artists, they

understandably complied with the intention of the emotional stimuli production (creating target-emotion eliciting visual music), and took the directive settings of the structural/formal components of visual stimuli and music (see **Tables 2, 3**) into account in their artistic activity (which involves well-developed and highly complex cognitive processes). Hence, we can claim that our emotional visual music stimuli constitute at least three predominant approaches in experimental aesthetic theories—expressionist, contextual, and formalist. Several theories have suggested models that explain human aesthetic perception and judgment processes (Leder et al., 2004; Chatterjee and Vartanian, 2014; Leder and Nadal, 2014; Redies, 2015), calling for more diverse empirical investigations that adopt various kinds of approaches. However, using real artwork in empirical research has generated disappointing results, although it is an interesting topic for artists and psychologists, and there is a need to extend previous approaches in emotional aesthetics to understand hedonic properties, cognitive operations, and greater compositional potential (for a review, see Leder et al., 2004). However, through our study, we believe that we have determined that the composition of abstract visual clips with directive design could cover a range of emotions, which can be assessed by evaluation, activity, and potency indices, and has the potential to be used as stimuli for more complex continuous response measures. Hence, we posit that properly controlled, well-designed visual music stimuli may be useful for future psychological and cognitive research studying the continuous reciprocal links between affective experience and cognitive processing, and specifically to understand how collective abstract expressions stimulate a holistic experience for audiences. In particular, because visual music has temporal narratives, it could be useful for future research to inspect the temporal dynamics of brain activity, skin conductance responses, changes in respiration or skin temperature as objective (autonomic) measures of emotional experiences in holistic information processing of the subject's state in relation to auditory and visual perception property controls. If possible, constructing a database of visual music with emotional meanings that provides a standardized set of abstractive auditory visual stimuli with accessible controls of various contextual parameters might be beneficial for future aesthetic emotion and aesthetic appreciation studies. The use of validated holistic stimuli and structural property controls may allow for investigations of integration synthesizing functions with semantic and syntax processing in auditory-visual aesthetic evaluation mechanisms.

To the best of our knowledge, our study is the first to propose a paradigm for the composition of abstract visual music with emotional validation at the unimodal, cross-modal, psychological and neurophysiological levels. Based on the findings of this study, we suggest that controlled, affective visual music can be a useful tool for investigating cognitive processing in affective aesthetic appraisals.

## ETHICS STATEMENT

The study was approved by the Institutional Review Board of Korea Advanced Institute of Science and Technology. All our subjects were fully informed that they were participating

in a survey for a study investigating aesthetic perception for a scientific research in both written and verbal forms. All participants signed a written informed consent form prior to engaging in the experiment.

## AUTHOR CONTRIBUTIONS

The conception or design of the work: IL, CL, and JJ. The acquisition, analysis, or interpretation of data for the work: IL and CL. Drafting the work: IL. Revising it critically for important intellectual content: IL and CL. Final approval of the version to be published: IL, CL, and JJ. Supervising the work overall: JJ. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: IL, CL, and JJ.

## FUNDING

The research and creation of the abstract visual music (positive) contents used in this study were financed by the company Amore Pacific Corporation, 181 Hanggangro-2-ga, Yongsan-gu, Seoul, South Korea. The financial support included the research staff's remuneration, artists' remuneration, other technical facilities for the creation of multimedia contents and electrophysiological recordings subjects remuneration.

## ACKNOWLEDGMENTS

The authors would like to thank Jangsub Lee and Hyosub Lee (visual animation artists for "V1," "V2," and "V3"), aRing (audio artist for "A1," "A2," and "A3"), Jinwon Lee (a.k.a. Gajaeal, visual music artist for "A4V4" and "A5V5"), and Jiyun Shin, Soyun Song, and Seongmin for assisting with the researching the production guidelines for visual music content. We would also like to thank Jaewon Lee, Dongil Chung, and Mookyung Han for assisting with the EEG data recording and analysis. The abstract visual music contents (A1V1, A2V2, and A3V3) were used for a marketing campaign comprising a series of on-line (Internet) advertisements for a cosmetic skin product from Amore Pacific Corporation; however, our research on the newly created abstract visual music was not on the orders of Amore Pacific. The campaign using our three positive visual music productions was an outstanding marketing success in South Korea, with over 250,000 people viewing the advertisements and over 20,000 people downloading the content (Sekyung Choi and Yoon, 2008).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.00440/full#supplementary-material>

**Figure S1 | Possible situations in three (conformance, complementation, and contest) conditions of cross-modal combination.**

**Figure S2 | Target emotion characteristics on the 2D (valence and arousal) plane illustration.**

**Figure S3 | Survey questionnaire with 13 pairs of bipolar sensorial adjectives.** The questionnaire was initially designed in US English and was presented to the participants in the Korean language. The unpleasant-pleasant pair was only additional in experiment 2 surveys.

**Figure S4 | Experiment design for stimuli clip presentation and rating.**

**Figure S5 | Composition of the three emotional aspect indices (evaluation, activity, and potency) and the method of rating conversion.** A total of 12 pairs of bipolar ratings were used to extract the evaluation, activity, and potency indices. The indices were rescaled from the nine-point scales to a range of  $[-1, 1]$  as shown.

**Figure S6 | Temporal and topographic responses to visual music presentations.** Average normalized power for 10-s non-overlapping epochs during the presentation of clip 1(a), clip 2(b), and clip 3(c)'s visual music. EEG

power is estimated using a FFT method for each 10-s non-overlapping periods from baseline to end of clip presentation. The two first epochs (20 s) are averaged to provide a baseline power for each frequency range (theta, lowAlpha1, lowAlpha2, UpperAlpha, and Beta); all power values are then normalized according to baseline power and 10log10 transform (dB). The topographic positioning of EEG leads is shown in the inset (bottom right corner). Time (bottom label) indicates the center of the 10-s epoch. The dashed line represents the start of the clip presentation after baseline resting.

**Table S1 | Independent T-test result of two groups' responses to the control clip.**

**Table S2 | Unreported results of K-W test across modalities results.**

**Table S3 | One-way ANOVA analysis results for comparing clips within the same modality.**

## REFERENCES

- Anderson, N. H. (1981). *Foundations of Information Integration Theory*. San Diego, CA: Academic Press.
- Bachorik, J. P., Bangert, M., Loui, P., Larke, K., Berger, J., Rowe, R., et al. (2009). Emotion in motion: investigating the time-course of emotional judgments of musical stimuli. *Music Percept. Interdisc. J.* 26, 355–364. doi: 10.1525/mp.2009.26.4.355
- Baumgartner, T., Esslen, M., and Jäncke, L. (2006a). From emotion perception to emotion experience: emotions evoked by pictures and classical music. *Int. J. Psychophysiol.* 60, 34–43. doi: 10.1016/j.ijpsycho.2005.04.007
- Baumgartner, T., Lutz, K., Schmidt, C. F., and Jäncke, L. (2006b). The emotional power of music: how music enhances the feeling of affective pictures. *Brain Res.* 1075, 151–164. doi: 10.1016/j.brainres.2005.12.065
- Baumgartner, T., Willi, M., and Jäncke, L. (2007). Modulation of corticospinal activity by strong emotions evoked by pictures and classical music: a transcranial magnetic stimulation study. *Neuroreport* 18:261. doi: 10.1097/WNR.0b013e328012272e
- Block, B. (2013). *The Visual Story: Creating the Visual Structure of Film, TV and Digital Media*. Burlington, MA: CRC Press.
- Bolivar, V. J., Cohen, A. J., and Fentress, J. C. (1994). Semantic and formal congruency in music and motion pictures: effects on the interpretation of visual action. *Psychomusicology* 13:28.
- Boltz, M., Schulkind, M., and Kantra, S. (1991). Effects of background music on the remembering of filmed events. *Mem. Cogn.* 19, 593–606. doi: 10.3758/BF03197154
- Brauchli, P., Michel, C. M., and Zeier, H. (1995). Electrocutaneous, autonomic, and subjective responses to rhythmic audio-visual stimulation. *Int. J. Psychophysiol.* 19, 53–66. doi: 10.1016/0167-8760(94)00074-O
- Brougher, K., Strick, J., Wiseman, A., Zilcher, J., and Mattis, O. (2005). *Visual Music: Synaesthesia in Art and Music Since 1900*. London: Thames & Hudson.
- Bunt, L., and Pavlicevic, M. (2001). *Music and Emotion: Perspectives from Music Therapy*. New York, NY: Oxford University Press.
- Chatterjee, A., and Vartanian, O. (2014). Neuroaesthetics. *Trends Cogn. Sci.* 18, 370–375. doi: 10.1016/j.tics.2014.03.003
- Chion, M. (1994). *Audio-Vision: Sound on Screen*. New York, NY: Columbia University Press.
- Coan, J. A., and Allen, J. J. (2004). Frontal EEG asymmetry as a moderator and mediator of emotion. *Biol. Psychol.* 67, 7–50. doi: 10.1016/j.biopsycho.2004.03.002
- Cohen, A. J. (2001). "Music as a source of emotion in film," in *Music and Emotion: Theory and Research*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 249–272.
- Cook, N. (1998). *Analysing Musical Multimedia*. New York, NY: Oxford University Press.
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Etcoff, N. (2011). *Survival of the Prettiest: The Science of Beauty*. New York, NY: Anchor Books.
- Gabrielsson, A., and Lindstrom, E. (2001). "The influence of musical structure on emotional expression," in *Music and Emotion: Theory and Research*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 223–248.
- Grierson, M. S. (2005). *Audiovisual Composition*. Doctoral Thesis, University of Kent at Canterbury.
- Gross, J. J., and Levenson, R. W. (1995). Emotion elicitation using films. *Cogn. Emot.* 9, 87–108. doi: 10.1080/02699939508408966
- Havner, K. (1935). Experimental studies of the affective value of colors and lines. *J. Appl. Psychol.* 19:385. doi: 10.1037/h0055538
- Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., and Schwaiger, J. (1998). Induced alpha band power changes in the human EEG and attention. *Neurosci. Lett.* 244, 73–76. doi: 10.1016/S0304-3940(98)00122-0
- Kline, J. P., Blackhart, G. C., Woodward, K. M., Williams, S. R., and Schwartz, G. E. (2000). Anterior electroencephalographic asymmetry changes in elderly women in response to a pleasant and an unpleasant odor. *Biol. Psychol.* 52, 241–250. doi: 10.1016/S0301-0511(99)00046-0
- Krause, C. M., Viemerö, V., Rosenqvist, A., Sillanmäki, L., and Aström, T. (2000). Relative electroencephalographic desynchronization and synchronization in humans to emotional film content: an analysis of the 4-6, 6-8, 8-10 and 10-12 Hz frequency bands. *Neurosci. Lett.* 286, 9–12. doi: 10.1016/S0304-3940(00)01092-2
- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., and Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Exp. Brain Res.* 158, 405–414. doi: 10.1007/s00221-004-1913-2
- Leder, H., and Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments: the aesthetic episode—developments and challenges in empirical aesthetics. *Brit. J. Psychol.* 105, 443–464. doi: 10.1111/bjop.12084
- Leder, H., Belke, B., Oeberst, A., and Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *Brit. J. Psychol.* 95, 489–508. doi: 10.1348/0007126042369811
- Lundholm, H. (1921). The affective tone of lines: experimental researches. *Psychol. Rev.* 28, 43. doi: 10.1037/h0072647
- Lyman, B. (1979). Representation of complex emotional and abstract meanings by simple forms. *Percept. Mot. Skills* 49, 839–842. doi: 10.2466/pms.1979.49.3.839
- Mandler, G. (1975). *Mind and Emotion*. Huntington, NY: Krieger Publishing Company.
- Mandler, G. (1984). *Mind and Body: Psychology of Emotion and Stress*. New York, NY: WW Norton.
- Marshall, S. K., and Cohen, A. J. (1988). Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Percept.* 6, 95–112. doi: 10.2307/40285417
- Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Urbana, IL: University of Illinois Press.
- Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., and Brattico, E. (2011). Music and emotions in the brain: familiarity matters. *PLoS ONE* 6:e27241. doi: 10.1371/journal.pone.0027241

- Poffenberger, A. T., and Barrows, B. (1924). The feeling value of lines. *J. Appl. Psychol.* 8:187. doi: 10.1037/h0073513
- Posner, J., Russell, J. A., and Peterson, B. S. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* 17, 715–734. doi: 10.1017/S0954579405050340
- Redies, C. (2015). Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Front. Hum. Neurosci.* 9:218. doi: 10.3389/fnhum.2015.00218
- Repp, B. H., and Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 1085–1099. doi: 10.1037/0096-1523.28.5.1085
- Rottenberg, J., Ray, R. R., and Gross, J. J. (2007). “Emotion elicitation using films,” in *The Handbook of Emotion Elicitation and Assessment*, eds J. A. Coan and J. J. B. Allen (New York, NY: Oxford University Press), 9–28.
- Sammler, D., Grigutsch, M., Fritz, T., and Koelsch, S. (2007). Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology* 44, 293–304. doi: 10.1111/j.1469-8986.2007.00497.x
- Scherer, K. R., and Zentner, M. R. (2001). “Emotional effects of music: production rules,” in *Music and Emotion: Theory and Research*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 361–392.
- Schmidt, L. A., and Trainor, L. J. (2001). Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cogn. Emot.* 15, 487–500. doi: 10.1080/02699930126048
- Sekyung Choi, S. L., and Yoon, E. (2008). “Research of analysis on future outlook and labor demands in new media advertising industry in Korea,” in *KBI Research*, ed J. Park (Seoul: Korea Creative Content Agency).
- Shimamura, A. P. (2012). “Toward a science of aesthetics: issues and ideas,” in *Aesthetic Science: Connecting Minds, Brains, and Experience*, eds A. P. Shimamura and S. E. Palmer (New York, NY: Oxford University Press), 3–28.
- Sloboda, J. A., and Juslin, P. N. (2001). “Psychological perspectives on music and emotion,” in *Music and Emotion: Theory and Research*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 71–104.
- Somsaman, K. (2004). *The Perception of Emotions in Multimedia: An Empirical Test of Three Models of Conformance and Contest*. Doctoral Thesis, Case Western Reserve University.
- Takahashi, S. (1995). Aesthetic properties of pictorial perception. *Psychol. Rev.* 102, 671–683. doi: 10.1037/0033-295X.102.4.671
- Taylor, K. I., Moss, H. E., Stamatakis, E. A., and Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8239–8244. doi: 10.1073/pnas.0509704103
- Van de Cruys, S., and Wagemans, J. (2011). Putting reward in art: a tentative prediction error account of visual art. *Iperception* 2, 1035–1062. doi: 10.1068/i0466aap
- van Honk, J., and Schutter, D. J. (2006). From affective valence to motivational direction. The frontal asymmetry of emotion revised. *Psychol. Sci.* 17, 963–965. doi: 10.1111/j.1467-9280.2006.01813.x
- Wexner, L. B. (1954). The degree to which colors (hues) are associated with mood-tones. *J. Appl. Psychol.* 38:432. doi: 10.1037/h0062181
- Wheeler, R. E., Davidson, R. J., and Tomarken, A. J. (1993). Frontal brain asymmetry and emotional reactivity: a biological substrate of affective style. *Psychophysiology* 30, 82–89. doi: 10.1111/j.1469-8986.1993.tb03207.x
- Winkler, I., Jäger, M., Mihajlovic, V., and Tsoneva, T. (2010). Frontal EEG asymmetry based classification of emotional valence using common spatial patterns. *World Acad. Sci. Eng. Technol.* Available online at: [https://www.researchgate.net/publication/265654104\\_Frontal\\_EEG\\_Asymmetry\\_Based\\_Classification\\_of\\_Emotional\\_Valence\\_using\\_Common\\_Spatial\\_Patterns](https://www.researchgate.net/publication/265654104_Frontal_EEG_Asymmetry_Based_Classification_of_Emotional_Valence_using_Common_Spatial_Patterns)
- Zettl, H. (2013). *Sight, Sound, Motion: Applied Media Aesthetics*. Boston, MA: Cengage Learning.

**Conflict of Interest Statement:** 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.

Copyright © 2017 Lee, Latchoumane and Jeong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Pitch Syntax Violations Are Linked to Greater Skin Conductance Changes, Relative to Timbral Violations – The Predictive Role of the Reward System in Perspective of Cortico–subcortical Loops

Edward J. Gorzelańczyk<sup>1,2,3,4</sup>, Piotr Podlipniak<sup>5\*</sup>, Piotr Walecki<sup>6</sup>, Maciej Karpiński<sup>7</sup> and Emilia Tarnowska<sup>8</sup>

<sup>1</sup> Department of Theoretical Basis of Bio-Medical Sciences and Medical Informatics, Nicolaus Copernicus University Collegium Medicum, Bydgoszcz, Poland, <sup>2</sup> Non-Public Health Care Center Sue Ryder Home, Bydgoszcz, Poland, <sup>3</sup> Medseven—Outpatient Addiction Treatment, Bydgoszcz, Poland, <sup>4</sup> Institute of Philosophy, Kazimierz Wielki University, Bydgoszcz, Poland, <sup>5</sup> Institute of Musicology, Adam Mickiewicz University in Poznań, Poznań, Poland, <sup>6</sup> Department of Bioinformatics and Telemedicine, Jagiellonian University Collegium Medicum, Krakow, Poland, <sup>7</sup> Institute of Linguistics, Adam Mickiewicz University in Poznań, Poznań, Poland, <sup>8</sup> Institute of Acoustics, Adam Mickiewicz University in Poznań, Poznań, Poland

## OPEN ACCESS

### Edited by:

Gavin M. Bidelman,  
University of Memphis, USA

### Reviewed by:

Stefanie Andrea Hutka,  
University of Toronto, USA  
Christopher J. Smalt,  
MIT Lincoln Laboratory, USA

### \*Correspondence:

Piotr Podlipniak  
podlip@poczta.onet.pl

### Specialty section:

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

**Received:** 30 November 2016

**Accepted:** 29 March 2017

**Published:** 18 April 2017

### Citation:

Gorzelańczyk EJ, Podlipniak P,  
Walecki P, Karpiński M and  
Tarnowska E (2017) Pitch Syntax  
Violations Are Linked to Greater Skin  
Conductance Changes, Relative  
to Timbral Violations – The Predictive  
Role of the Reward System  
in Perspective of Cortico–subcortical  
Loops. *Front. Psychol.* 8:586.  
doi: 10.3389/fpsyg.2017.00586

According to contemporary opinion emotional reactions to syntactic violations are due to surprise as a result of the general mechanism of prediction. The classic view is that, the processing of musical syntax can be explained by activity of the cerebral cortex. However, some recent studies have indicated that subcortical brain structures, including those related to the processing of emotions, are also important during the processing of syntax. In order to check whether emotional reactions play a role in the processing of pitch syntax or are only the result of the general mechanism of prediction, the comparison of skin conductance levels reacting to three types of melodies were recorded. In this study, 28 subjects listened to three types of short melodies prepared in Musical Instrument Digital Interface Standard files (MIDI) – tonally correct, tonally violated (with one out-of-key – i.e., of high information content), and tonally correct but with one note played in a different timbre. The BioSemi ActiveTwo with two passive Nihon Kohden electrodes was used. Skin conductance levels were positively correlated with the presented stimuli (timbral changes and tonal violations). Although changes in skin conductance levels were also observed in response to the change in timbre, the reactions to tonal violations were significantly stronger. Therefore, despite the fact that timbral change is at least as equally unexpected as an out-of-key note, the processing of pitch syntax mainly generates increased activation of the sympathetic part of the autonomic nervous system. These results suggest that the cortico–subcortical loops (especially the anterior cingulate – limbic loop) may play an important role in the processing of musical syntax.

**Keywords:** pitch syntax, prediction, cortico–subcortical loops, skin conductance, timbre

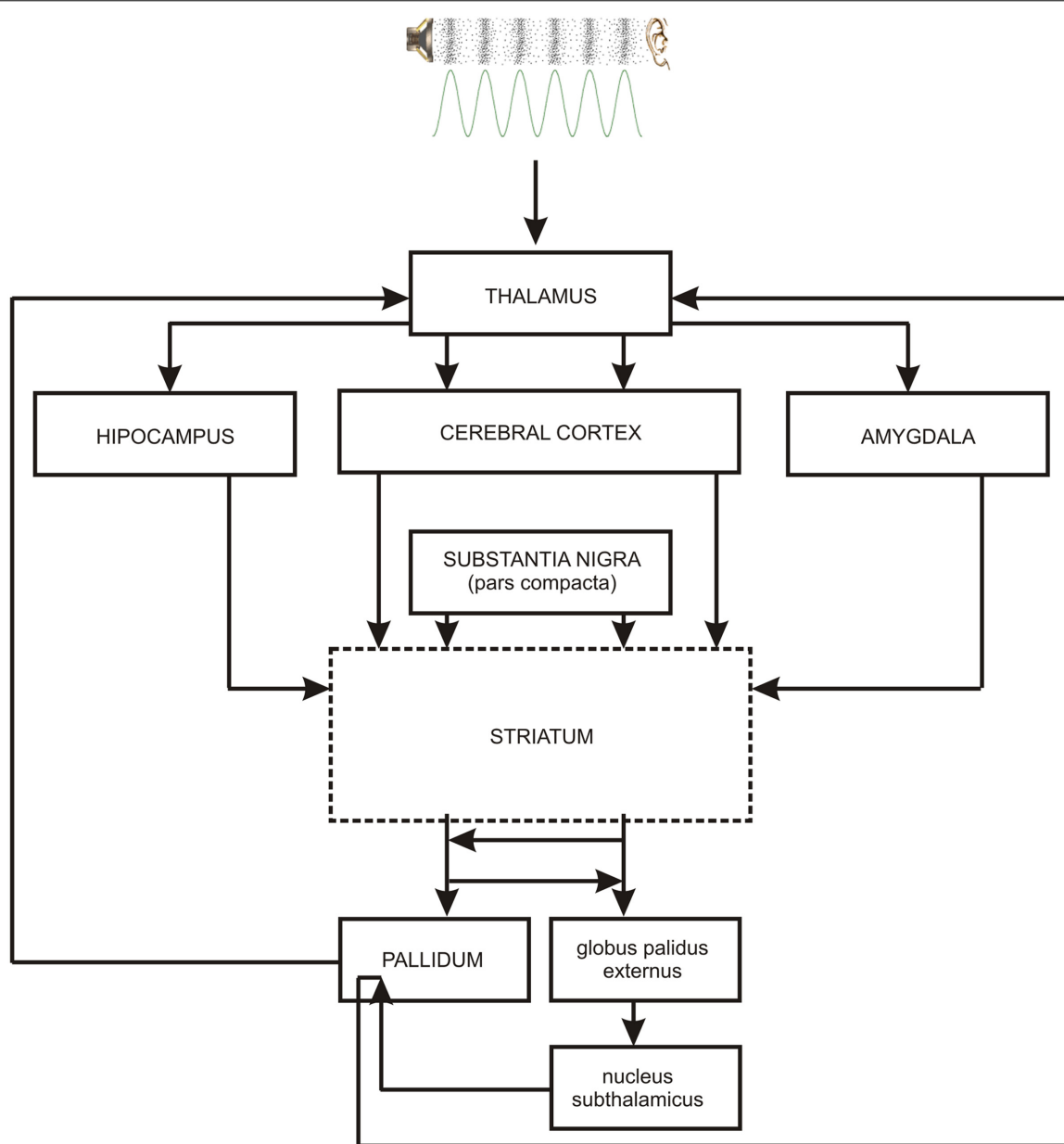
## INTRODUCTION

Tonal music is a natural, and complex syntactic system (Lerdahl and Jackendoff, 1983) based on implicitly learned norms (Tillmann et al., 2000; Tillmann, 2005). The perception of pitch structure as hierarchically organized discrete units (pitch classes) is an important part of syntactic processing in music (Krumhansl, 1990; Krumhansl and Cuddy, 2010; Lerdahl, 2013). During this process, the recognition of each pitch class in the context of other pitch classes is accompanied by subtle emotional sensations known as tension, uncertainty, stability, completion, and power, etc. which are often referred to as ‘tonal qualia’ (Huron, 2006; Margulis, 2012). According to Huron (Huron, 2006), positive emotions (e.g., the tonal qualia of resolution or completeness often described as pleasure, contentment etc.) are the result of limbic reward for accurate predictions, whereas negative emotions (e.g., the tonal qualia of tension or incompleteness often described as uncomfortable, jarring, anxious etc.) are elicited in case our predictions are inaccurate. This claim is in line with classic Darwinian rules (Darwin, 1859) as emotions actually enable the adaptation of behavior to particular circumstances (Darwin, 1872; Panksepp, 1998). Accordingly animals, including mammals, are able to continuously predict changes in their environment. More accurate prediction implies more probable survival. Therefore, various emotions are elicited depending on the different probabilities of occurring events. Since the probabilities of pitch class occurrences depend on the context of other pitch classes, then various levels of prediction accuracy result in slightly diverse emotional sensations. Because the reward system plays a crucial role in the prediction of perceived stimuli by the means of emotional control (Abler et al., 2006; Juckel et al., 2006), the concept of cortico-subcortical loops may help to provide a basic explanation of these phenomena (Gorzelańczyk, 2011). The main structure of the reward system is the ventral striatum, which is a crucial part of the limbic loop (Abler et al., 2006). The limbic system controls the functions of the autonomic nervous system (Mujica-Parodi et al., 2009; Tavares et al., 2009) and the endocrine system (Herman et al., 2005; Krüger et al., 2015; Tasker et al., 2015). The activity of the latter influences skin conductance (Dawson et al., 2007; Zhang et al., 2016) and other physiological parameters (Wood et al., 2014; Schröder et al., 2015).

As the occurrence of out-of-key notes in tonal melody is usually very improbable (Pearce and Wiggins, 2006; Pearce et al., 2010; Hansen and Pearce, 2014), the response of the reward system should be stronger than in the case of ‘in-key’ notes. In fact, a number of studies indicate that the perception of a violated tonal structure leads to measurable somatic reactions such as changes of skin conductance (Steinbeis et al., 2006; Koelsch et al., 2008b). There are also studies that show the modulation in the activity of the amygdala (Koelsch et al., 2008a; Mikutta et al., 2015), the orbitofrontal cortex (Tillmann et al., 2003; Koelsch et al., 2005; Mikutta et al., 2015), the inferior frontal gyrus, the orbital frontolateral cortex, the anterior insula, the ventrolateral premotor cortex, the anterior and posterior areas of the superior temporal gyrus, the superior temporal sulcus, the supramarginal gyrus (Koelsch et al., 2005), the inferior

frontal cortex (Tillmann et al., 2003, 2006), the parahippocampal gyrus and the cingulate cortex (Omigie et al., 2015), depending on the expectedness of musical pitch structure. Information processed by various cortico-subcortical loops partly overlap in the striatum where it is exchanged between particular circuits (**Figure 1**). The certain parts of the striatum (the caudatus, the putamen, the nucleus accumbens) are connected to specific areas of the cerebral cortex. Sensorimotor cortices are connected with the putamen, association cortices are connected to the caudatus, and limbic cortices and the amygdala are connected with the nucleus accumbens (**Figure 2**). The striatum, especially the ventral striatum (nucleus accumbens) (Kruse et al., 2016), as well as the amygdala, and the orbitofrontal cortex (Tabbert et al., 2006) are strictly physiologically connected to the autonomic nervous system which controls the vegetative functions of the body (the cardiovascular system, the endocrine system, and the digestive system) including blood circulation in the skin, the activity of sweat and sebaceous glands, and smooth muscle tension in the skin (**Figure 3**). These physiological responses can cause changes in the electrical conductance of the skin (Johnson and Corah, 1963; Grove, 1992; Alonso et al., 1996; Klucken et al., 2009; Benedek and Kaernbach, 2010; Boucsein, 2014). This can explain why the perception of violated tonal structure can lead to measurable somatic reactions such as changes in skin conductance (Steinbeis et al., 2006; Koelsch et al., 2008b). Therefore, it is not surprising that the modulation of the activity in the amygdala (closely linked to the nucleus accumbens and the limbic system) (Koelsch et al., 2008a; Mikutta et al., 2015) changes value of skin conductance. The latero-orbito-frontal and limbic loops are particularly important in the control of executive functions (Alexander et al., 1986; Royall et al., 2002; Haber, 2003), which may explain why the activation of the orbitofrontal cortex (Tillmann et al., 2003; Koelsch et al., 2005; Mikutta et al., 2015) can change skin conductance. The fact that the anterior cingulate loop is responsible for correcting behavior following a mistake (Peterson et al., 1999) and that the parahippocampal gyrus is strictly connected to the limbic system, are consistent with observations that the activation of the parahippocampal gyrus and the cingulate cortex (Omigie et al., 2015) are related to the changing expectedness of the components of musical pitch structure.

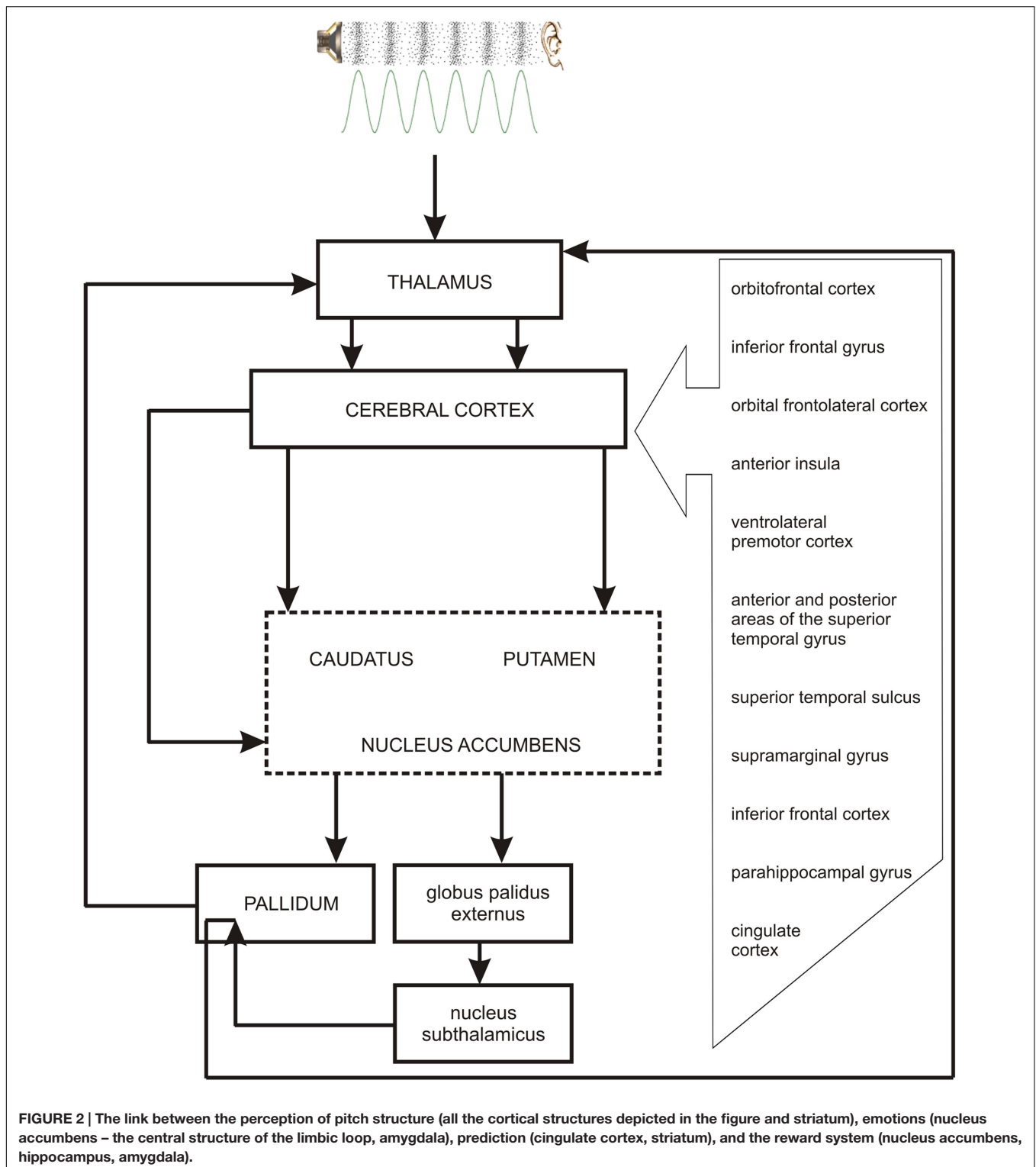
Even a very simple melody is a complex stimulus, composed of sounds characterized by not only the fundamental frequency ( $F_0$ ), which mainly influences the sensation of pitch (Stainsby and Cross, 2008), but also by other acoustic parameters. Our predictions also include some of these parameters. For example, spectral centroid, attack time, and spectral irregularity influence the sensation of timbre in music (McAdams and Giordano, 2008). Although it may seem that timbre in music is not perceptively organized in a hierarchical way, the specific characteristics of spectral (e.g., spectral centroid) and temporal (e.g., attack time) cues allow for the discrimination between different categories of timbres, e.g., between the timbre of the flute and the piano. These categories, similar to pitch, occur in music with different probabilities. For example, the probability that a melody played in a particular timbre will suddenly change (e.g., from piano to



**FIGURE 1 | The main brain structures of cortico-subcortical loops: striatum – cerebral cortex – pallidum.** PALLIDUM = globus pallidus internalis + substantia nigra reticulata + ventral part of the globus pallidus. Contiguous brain structures (hippocampus, amygdala) are also shown. The broken line indicates the main structure of every loop.

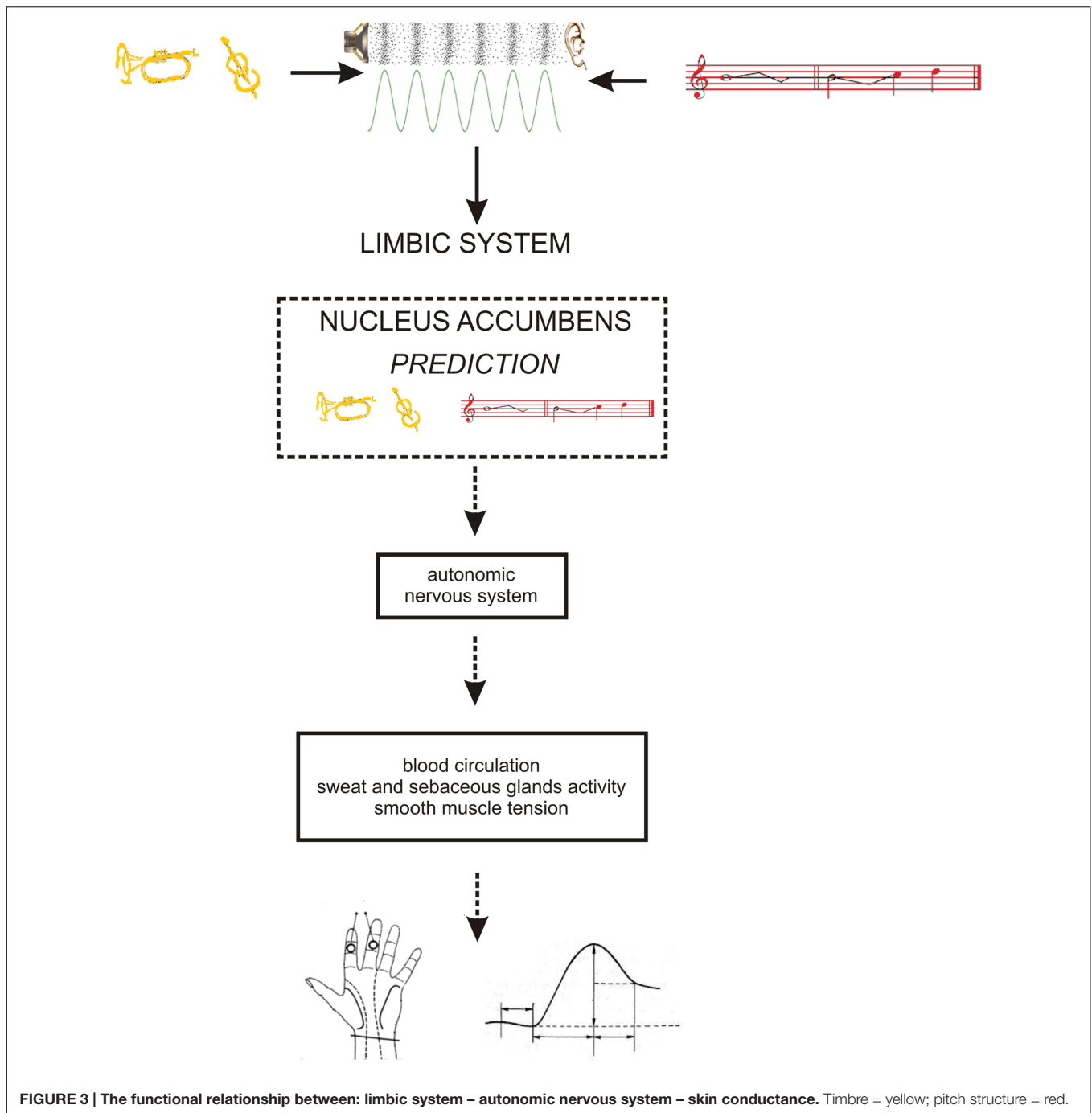
flute) seems to be even lower than the occurrence of an out-of-key note. After all, the vast majority of musical tunes which are experienced by listeners living in Western society are sung or played by the same singer or the same musical instrument. Of course, some spectral changes can occur even when a melody is sung or played by one musical source. However, from a cognitive point of view these spectral deviations do not break the perceptual congruity of a percept as belonging to one timbral category. If a human reaction to sound probabilities depends solely on the general mechanism of prediction, then timbral

changes should cause a reaction of the autonomic nervous system at least as strong as that to an out-of-key note. Apart from that, in contrast to a small change in pitch (out-of-key note), changes in timbre cause a violation in the auditory stream (Wessel, 1979; Gregory, 1994; Huron, 2016) which should also elicit a stronger reaction of the autonomic nervous system than in the case when the auditory stream is preserved when the whole melody is played in the same timbre. We hypothesize that pitch structure in music is processed by a domain-specific circuit which differs functionally from that responsible for the predictions



of timbre. Although both circuits implement the predictive mechanisms of cortico-subcortical loops (Gorzelańczyk, 2011), only the activity of the pitch class prediction circuit results in syntactical organization of perceived sounds in music which allows for the recognition of a recursive relationship (Woolhouse

et al., 2016). According to behavioral observations, emotional reaction related to prediction is a specific part of the recognition of pitch hierarchy (Huron, 2006). This specificity should be possible to recognize by measuring the somatic markers of autonomic nervous system activity. In other words, the reaction



of the autonomic nervous system to an out-of-key note should be different from the reaction to an in-key note and to a change of timbre.

## MATERIALS AND METHODS

### Stimuli

Six simple tonal melodies were prepared as MIDI files without any dynamics and tempo changes in order to avoid additional

expressive content. Each melody lasted for 6–12 s. The key signatures of all melodies were chosen randomly in order to avoid the exposure effect as a result of the latent memory of pitch. From the MIDI files of these melodies six additional MIDI files were generated so that one note in each basic melody was changed into an out-of-key note. In order to avoid possible interpretation of the results as just a reaction to interval change or scale degree, different intervals leading to out-of-key notes and different scale degrees of changed pitches were chosen. Because musical rhythm and meter is also processed in the brain by the means of predictive



FIGURE 4 | Tonally correct melodies.

coding (Vuust and Witek, 2014), the locations of out-of-key notes were placed randomly in the bars in order to exclude the possible influence of the same metrical stress (or lack of stress) on skin conductance reactions. Apart from this, six further MIDI files were prepared. This time the timbre of one note in each basic melody was changed instead of the fundamental frequency. The notes of the basic melodies in which the timbre was changed were exactly the same notes to those previously changed into out-of-key. Each change was made after 8 to 12 notes of each melody. As a result there were three versions of six melodies: tonally correct (Figure 4), tonally violated – with an out-of-key note (Figure 5), and tonally correct but with one note played in a different timbre (Figure 6).

## Method

Twenty eight musically untrained (people without any formal musical education and who do not play any instrument) medical students (18 women, 10 men; age: mean = 20.21;  $SD = 1.55$ ) were studied. The research was conducted on people who voluntarily participated in the study. Prior to testing the subjects were informed that some melodies had been modified by replacing one note with another note, an out-of-key note, and some others by changing the timbre of one note. The subjects were then asked to listen for and focus their attention on tonal violation in the stimuli in order to concentrate the attention of the subjects on the stimuli. Before each session, the subjects listened to one example of each type of melody so that they understood what the task would involve. Skin conductance was measured with the ActiveTwo (Biosemi®) biopotential measurement system and two passive Nihon Kohden electrodes were placed on the medial phalanges of the index and middle finger of the subject's non-dominant hand. The current used was 1  $\mu A$  at 16 Hz, synchronized with a sampling rate of 8.192 kHz. The resolution of measurements was 1 nanoSiemens [ $nS$ ]. The duration of each experimental session (for one person)

was ca. 30 min. Individual tests were performed in the same location for all participants. After attaching the electrodes, each participant remained in a quiet and dark place for 15 min. This was for the purpose of calming any emotional arousal and, simultaneously, as an adaptation period, allowing for the equilibration of hydration and sodium at the interface between the skin and the electrode gel. The experimental procedure was based on the presentation of one MIDI file composed of 18 randomly ordered melodies separated by 1 s pauses. Each melody belonged to one of the following groups: tonally correct, tonally violated, and tonally correct but with one note played in a different timbre. Electrodermal activity was continuously measured during the entire time of listening to each melody. For each melody, the difference between the initial value and the maximum value of conductivity was calculated. This study was carried out in accordance with the recommendations of the Bioethics Committee of the Nicolaus Copernicus University in Torun at Collegium Medicum in Bydgoszcz No. KB 416/2008 on 17.09.2008, with written informed consent from all subjects in accordance with the Declaration of Helsinki.

## RESULTS

We observed three types of skin-galvanic activity (Figure 7). First was the activity associated with the perception of in-key notes. This activity had a low maximum amplitude (mean = 247.26 [ $nS$ ];  $SD = 53.44$  [ $nS$ ]). The second reaction was associated with the notes played in a different timbre, the maximum amplitude (mean = 855.32 [ $nS$ ];  $SD = 144.79$  [ $nS$ ]). The third response, associated with the tonally violated notes, was characterized by a high maximum amplitude (mean = 1311.57 [ $nS$ ];  $SD = 231.12$  [ $nS$ ]). Importantly, the subjects differed in the frequency of responses to changed notes. In other words, they did not respond to all out-of-key notes and notes played in a different timbre equally often. Additionally, we observed that the reactions to



FIGURE 5 | Tonally violated melodies. Out-of-key notes are red.



FIGURE 6 | Tonally correct melodies with one note played in a different timbre (flute instead of piano timbre). The notes played in a different timbre are in yellow.

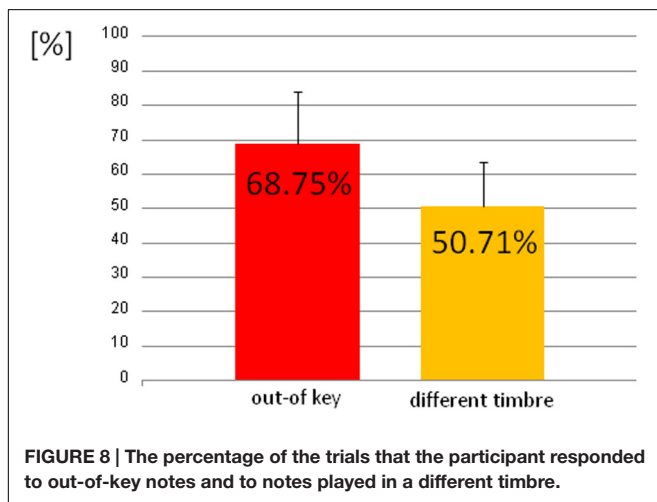
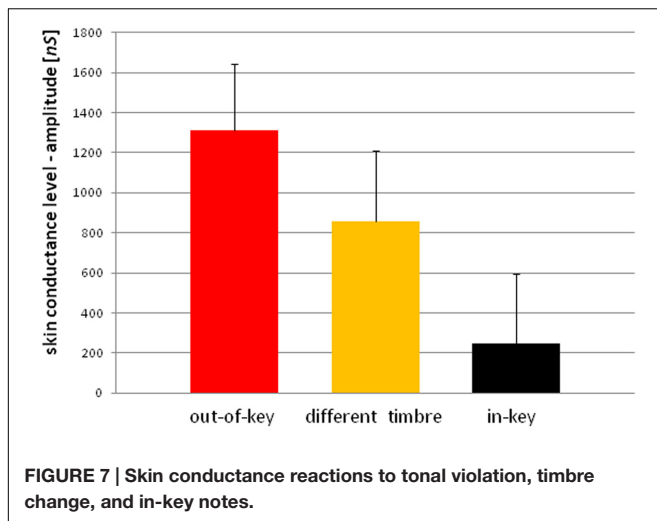
out-of-key notes were more frequent (68.75% of responses) than to notes played in a different timbre (50.71% of responses) (Figure 8).

Because the data did not have a normal distribution (Shapiro-Wilk  $W = 0.90410$ ,  $p = 0.01429$ ) we used the 'Wilcoxon matched-pairs signed-ranks test.' The 'Wilcoxon matched-pairs signed-ranks test' indicates a statistically significant ( $p < 0.05$ ) difference between the mean maximum amplitude of reactions after in-key notes and the mean maximum amplitude of reactions after out-of-key notes and the notes played in a different timbre. A statistically significant ( $t$ -test for dependent means,  $p < 0.05$ ) difference was also found between reactions after out-of-key notes and notes played in a different timbre as well as between the frequency of

reactions to out-of-key notes and to notes played in a different timbre.

## DISCUSSION

It has been observed that the mean skin conductance value in response to tonal violations is significantly higher in musical stimuli than to a change in timbre. The higher percentage of responses to tonal violations in comparison to responses to the changes in timbre has also been observed. Interestingly, the electro-dermal responses to tonal violations are more frequent and have a higher value in contrast to the electro-dermal response to timbral changes, although listeners reported after



the study that they had been more conscious of the timbral changes in comparison to tonal violations. This may suggest that the recognition strategy for tonal violations is different from the one employed for the timbral change recognition despite the fact that their attention was guided toward the tonal changes (instructions). The observed difference also suggests that the biological significance of tonal violations is higher in comparison to the recognition of the timbral change in music, which suggests a potential biological adaptive value of pitch structure (Podlipniak, 2016). The fact that timbre as a source cue is more evolutionarily salient than key or tempo of melody (Schellenberg and Habashi, 2015) indicates that its change should cause a greater physiological reaction than a change in pitch. Our results show, however, that the change of pitch class which is unexpected in terms of pitch structure leads to a greater physiological response. This result suggests that pitch structure may be evolutionarily important too. Whilst listening to a melody, information about the congruency of the listened tune with schematic expectations seems to become more important than the cues of the sound source.

Such an effect supports the claim that pitch structure is a part of the species-specific form of vocalization rather than just a culturally changeable tradition of composing sounds on the basis of pitch. In fact, the perception of species-specific forms of vocalization by mammals (Gouzoules and Gouzoules, 2000; Juslin and Scherer, 2008) and birds (Rothenberg et al., 2014) usually contains an emotional/motivational component. Therefore, the obtained results can support the view that pitch structure is an important part of the vocal communication of *Homo sapiens*. Note that the pitch structure here does not refer to the varying perceptive dimension of pitch which is homologous to speech intonation, but to the cognitive dimension of pitch composed of discrete units (pitch classes) which is analogous to digitized sounds in speech, e.g., phonemes (Jackendoff, 2008). An interesting way to compare the biological importance of pitch and timbre would be to measure reactions to timbral and pitch class changes in different environmental conditions. We suspect that different environments (e.g., listening to melodies in the relatively safe environment of a laboratory versus more dangerous circumstances such as in a forest at night) can influence the physiological reactions of the listener. In a less safe environment timbre as a cue of the sound source should be more biologically important than in a laboratory.

Taking into account that skin conductance changes are one of the measurable reactions of the autonomic nervous system tone (mostly the sympathetic part) controlled by the limbic system and the reward system, it can be assumed that the reward system is more susceptible to pitch class changes than to timbral changes. Although, in contrast to timbre, pitch structure is assumed to be a cognitive dimension of music (Barrett and Janata, 2016) it is also possible that the recognition of pitch structure is related to the activity of the subcortical structures which are rarely studied. Therefore, it seems that the most promising way to explore and understand the processing of pitch structure is a model which takes into account the role of cortico-subcortical loops. From physiological perspective, it is possible to infer from the obtained results that the change of timbre is less important for the fluency of tonal music than tonal violation.

Although not all participants responded as equally often to “out-of-key” notes and timbral changes, the skin conductance responses to “out-of-key” notes were significantly more frequent than to the changes of timbre. This may be a result of the instruction delivered before the study. However, if there was a lack of instruction, different subjects’ attentions, depending on possible different listening strategies chosen by the subjects, could influence their reactions. However, according to many observations, unconscious neural activity precedes and influences conscious decisions (Soon et al., 2013). Therefore, it is more probable that the changes of pitch structure were more conspicuous. The fact that people did not react to all instances of change (tonal violation and timbral change) can be explained by either the fact that they did not focus equally well to the presented stimuli or that not all tonal violations and timbral changes, respectively, were equally prominent for them.

The limited range of timbre changes introduced to our stimuli has resulted in some limitations to our conclusions and generalizations. This opens the debate about different possible

reactions to other timbre changes. Since different timbres can be interpreted by the nervous system as the labels of different sound sources which may be dangerous or attractive, i.e., emotionally significant, it is probable that changes of notes into such emotionally significant timbres will cause greater changes of skin conductance levels than emotionally neutral timbres. Thus, in further studies, responses to different timbre changes should be measured. Another possibility would be to modify the primary sounds or entire melodies, e.g., by using various playback rates or reverse the sounds in the time domain.

Another serious issue is the cultural influence on perception strategies used by humans in music listening. Because the role of pitch in speech depends on culture as in the case of tonal and non-tonal languages (Ge et al., 2015), it is possible that the importance of pitch syntax and timbre in music can also differ depending on culture. For example the phenomenon of 'kuchi shōga' – a Japanese sound symbolism used as an acoustic-icomic mnemonic system (Hughes, 2000) – necessitates elaborate sensitivity to timbral features which can influence the perception of musical timbre by people skilled in this method. In order to address this question adequately, intercultural studies are necessary. A similar question is related to the possible influence of musical training on the sensitivity to timbral features. Some researchers suggest that musical training can cause musicians to process timbre using their brain network which is specialized for music (Marozeau et al., 2013). If this is true then musicians, especially those who are more familiar with contemporary music in which timbral changes are more salient than pitch structure, should be more sensitive to timbral changes. Since the processing of spectral and temporal cues is important for the processing of the phonological aspects of speech (Shannon et al., 1995; Xu et al., 2005) it is possible that the reactions of individuals to the change in timbre may differ and the results would correlate with the acoustic properties of their mother tongue. What is more, non-musicians who use tonal language (e.g., Mandarin) can process acoustical stimuli like musicians (trained and exposed to western music) (Bidelman et al., 2013), so performing such a study on different groups of people (e.g., musicians) should be the next step of this research. In further research, we intend to employ paired musical and speech stimuli in order to directly compare responses to tonal and phonotactic violations.

An interesting question has been whether there is any visual analog for the observed results. For example, syntax in language can cross modalities, which is evident in the case of sign language. Although certain scholars claim that music can also be cross-modal, e.g., as an expression of body movements (Lewis, 2013),

in our opinion it is only possible in restricted elements of music. While rhythm can be expressed in dance by the means of movements there is nothing resembling the experience of pitch syntax in other domains of human perception. In other words, tonal relations seem to be unique and specific only to the auditory modality. However, it would be interesting to investigate human physiological reactions to syntactic violations in speech, sign language and music.

## AUTHOR CONTRIBUTIONS

EG: Substantial contributions to the conception and design of the study and interpretation of data for the work; writing the work and revising it critically for neurobiological content. PP: Substantial contributions to the conception and design of the study and interpretation of data for the work; writing the work and revising it critically for musicological and psychological content. PW: Substantial contributions to the acquisition, analysis, and interpretation of data for the work; writing the work and revising it critically for psychological content. MK: Substantial contributions to the acquisition, analysis, and interpretation of data for the work as well as to the conception and design of the study; writing the work and revising it critically for psycholinguistic and psychomusicological content. ET: Substantial contributions to the acquisition, analysis, and interpretation of data for the work; writing the work and revising it critically for acoustic content. All the authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All authors contributed to the final approval of the version to be submitted.

## FUNDING

Program of the Polish Ministry of Health financed by Gambling Problems Funds contracted by the National Bureau for Drug Prevention – under the project: Carrying out research to increase knowledge in the field of addiction grant number 68/H/E/2015 of 02.03.2015 and 6/H/E/K/2016 of 04.01.2016.

## ACKNOWLEDGMENTS

We would like to thank the reviewers for their critical and helpful suggestions and comments. We would also like to thank Peter Kośmider-Jones for proofreading and his language consultation.

## REFERENCES

- Abler, B., Walter, H., Erk, S., Kammerer, H., and Spitzer, M. (2006). Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *Neuroimage* 31, 790–795. doi: 10.1016/j.neuroimage.2006.01.001
- Alexander, G. E., DeLong, M. R., and Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.* 9, 357–381. doi: 10.1146/annurev.ne.09.030186.002041
- Alonso, A., Meirelles, N. C., Yushmanov, V. E., and Tabak, M. (1996). Water increases the fluidity of intercellular membranes of stratum corneum: correlation with water permeability, elastic, and electrical resistance properties. *J. Invest. Dermatol.* 106, 1058–1063. doi: 10.1111/1523-1747.ep12338682
- Barrett, F. S., and Janata, P. (2016). Neural responses to nostalgia-evoking music modeled by elements of dynamic musical structure and individual differences in affective traits. *Neuropsychologia* 91, 234–246. doi: 10.1016/j.neuropsychologia.2016.08.012

- Benedek, M., and Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* 190, 80–91. doi: 10.1016/j.jneumeth.2010.04.028
- Bidelman, G. M., Hutka, S., and Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PLoS ONE* 8:e60676. doi: 10.1371/journal.pone.0060676
- Boucsein, W. (2014). *Electrodermal Activity*. Igarss 2014. Boston, MA: Springer, doi: 10.1007/978-1-4614-1126-0
- Darwin, C. (1859). *On the Origin of the Species*, Vol. 5. London: John Murray.
- Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London: John Marry, 374. doi: 10.1037/h0076058
- Dawson, M. E., Schell, A. M., Filion, D. L., and Berntson, G. G. (2007). "The Electrodermal System," in *Handbook of Psychophysiology*, eds J. T. Cacioppo, L. G. Tassinary, and G. Berntson (Cambridge: Cambridge University Press), 157–181. doi: 10.1017/CBO9780511546396.007
- Ge, J., Peng, G., Lyu, B., Wang, Y., Zhuo, Y., Niu, Z., et al. (2015). Cross-language differences in the brain network subserving intelligible speech. *Proc. Natl. Acad. Sci. U.S.A.* 112, 2972–2977. doi: 10.1073/pnas.1416000112
- Gorzelańczyk, E. J. (2011). "Functional anatomy, physiology and clinical aspects of basal ganglia," in *Neuroimaging for Clinicians – Combining Research and Practice*, ed. J. F. P. Peres (Rijeka: InTech), 89–106.
- Gouzoules, H., and Gouzoules, S. (2000). Agonistic screams differ among four species of macaques: the significance of motivation-structural rules. *Anim. Behav.* 59, 501–512. doi: 10.1006/anbe.1999.1318
- Gregory, A. H. (1994). Timbre and auditory streaming. *Music Percept.* 12, 161–174. doi: 10.2307/40285649
- Grove, G. L. (1992). Skin surface hydration changes during a mini regression test as measured in vivo by electrical conductivity. *Curr. Ther. Res.* 52, 835–837. doi: 10.1016/S0011-393X(05)80462-X
- Haber, S. N. (2003). The primate basal ganglia: parallel and integrative networks. *J. Chem. Neuroanat.* 26, 317–330.
- Hansen, N. C., and Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Front. Psychol.* 5:1052. doi: 10.3389/fpsyg.2014.01052
- Herman, J. P., Ostrander, M. M., Mueller, N. K., and Figueiredo, H. (2005). Limbic system mechanisms of stress regulation: hypothalamo-pituitary-adrenocortical axis. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 29, 1201–1213. doi: 10.1016/j.pnpbp.2005.08.006
- Hughes, D. (2000). No nonsense: the logic and power of acoustic-iconic mnemonic systems. *Ethnomusicol. Forum* 9, 93–120. doi: 10.1080/09681220008567302
- Huron, D. B. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. London: The MIT Press.
- Huron, D. B. (2016). *Voice Leading: The Science behind a Musical Art*. London: MIT Press.
- Jackendoff, R. (2008). Parallels and nonparallels between language and music. *Music Percept.* 26, 195–204.
- Johnson, L. C., and Corah, N. L. (1963). Racial differences in skin resistance. *Science* 139, 766–767. doi: 10.1126/science.139.3556.766
- Juckel, G., Schlagenhauf, F., Koslowski, M., Filonov, D., Wüstenberg, T., Villringer, A., et al. (2006). Dysfunction of ventral striatal reward prediction in schizophrenic patients treated with typical, not atypical, neuroleptics. *Psychopharmacology* 187, 222–228. doi: 10.1007/s00213-006-0405-4
- Juslin, P. N., and Scherer, K. R. (2008). "Vocal expression of affect," in *The New Handbook of Methods in Nonverbal Behavior Research*, eds J. A. Harrigan, R. Rosenthal, and K. R. Scherer (Oxford: Oxford University Press), 65–135. doi: 10.1093/acprof:oso/9780198529620.003.0003
- Klucken, T., Kagerer, S., Schweckendiek, J., Tabbert, K., Vaitl, D., and Stark, R. (2009). Neural, electrodermal and behavioral response patterns in contingency aware and unaware subjects during a picture-picture conditioning paradigm. *Neuroscience* 158, 721–731. doi: 10.1016/j.neuroscience.2008.09.049
- Koelsch, S., Fritz, T., and Schlaug, G. (2008a). Amygdala activity can be modulated by unexpected chord functions during music listening. *Neuroreport* 19, 1815–1819. doi: 10.1097/WNR.0b013e32831a8722
- Koelsch, S., Fritz, T., Schulze, K., Alsop, D., and Schlaug, G. (2005). Adults and children processing music: an fMRI study. *Neuroimage* 25, 1068–1076. doi: 10.1016/j.neuroimage.2004.12.050
- Koelsch, S., Kilches, S., Steinbeis, N., and Schelinski, S. (2008b). Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *PLoS ONE* 3:e2631. doi: 10.1371/journal.pone.0002631
- Krüger, O., Shiozawa, T., Kreifelts, B., Scheffler, K., and Ethofer, T. (2015). Three distinct fiber pathways of the bed nucleus of the stria terminalis to the amygdala and prefrontal cortex. *Cortex* 66, 60–68. doi: 10.1016/j.cortex.2015.02.007
- Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195148367.001.0001
- Krumhansl, C. L., and Cuddy, L. L. (2010). "A theory of tonal hierarchies in music," in *Music Perception*, eds M. R. Jones, R. R. Fay, and A. N. Popper (New York, NY: Springer), 51–87. doi: 10.1007/978-1-4419-6114-3\_3
- Kruse, O., Stark, R., Klucken, T., Neuroscience, S., Kruse, O., and Neuroscience, S. (2016). Neural correlates of appetitive extinction in humans. *Soc. Cogn. Affect. Neurosci.* doi: 10.1093/scan/nsw157 [Epub ahead of print].
- Lerdahl, F. (2013). "Musical syntax and its relation to linguistic syntax," in *Language, Music, and the Brain*, ed. M. A. Arbib (London: The MIT Press), 257–272. doi: 10.7551/mitpress/9780262018104.003.0010
- Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. London: The MIT Press.
- Lewis, J. (2013). "A cross-cultural perspective on the significance of music and dance to culture and society. Insight from BaYaka Pygmies," in *Language, Music, and the Brain*, ed. M. A. Arbib (London: The MIT Press), 45–65. doi: 10.7551/mitpress/9780262018104.003.0002
- Margulis, E. H. (2012). *On Repeat: How Music Plays the Mind*. Oxford: Oxford University Press.
- Marozeau, J., Innes-Brown, H., and Blamey, P. J. (2013). The effect of timbre and loudness on melody segregation. *Music Percept.* 30, 259–274. doi: 10.1525/MP.2012.30.3.259
- McAdams, S., and Giordano, B. L. (2008). "The perception of musical timbre," in *The Oxford Handbook of Music Psychology*, Vol. 1, eds S. Hallam, I. Cross, and M. Thaut (Oxford: Oxford University Press), 72–80. doi: 10.1093/oxfordhb/9780199298457.013.0007
- Mikutta, C. A., Dürschmid, S., Bean, N., Lehne, M., Lubell, J., Altorfer, A., et al. (2015). Amygdala and orbitofrontal engagement in breach and resolution of expectancy: a case study. *Psychomusicology* 25, 357–365. doi: 10.1037/pmu0000121
- Mujica-Parodi, L. R., Korgaonkar, M., Ravindranath, B., Greenberg, T., Tomasi, D., Wagshul, M., et al. (2009). Limbic dysregulation is associated with lowered heart rate variability and increased trait anxiety in healthy adults. *Hum. Brain Mapp.* 30, 47–58. doi: 10.1002/hbm.20483
- Omigie, D., Pearce, M., and Samson, S. (2015). "Intracranial evidence of the modulation of the emotion network by musical structure," in *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music*, eds J. Ginsborg, A. Lamont, M. Phillips, and S. Bramley (Manchester: Royal Northern College of Music).
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*, Vol. 4. New York, NY: Oxford University Press.
- Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., and Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *Neuroimage* 50, 302–313. doi: 10.1016/j.neuroimage.2009.12.019
- Pearce, M. T., and Wiggins, G. A. (2006). Expectation in melody: the influence of context and learning. *Music Percept.* 23, 377–405.
- Peterson, B. S., Skudlarski, P., Gatenby, J. C., Zhang, H., Anderson, A. W., and Gore, J. C. (1999). An fMRI study of Stroop word-color interference: evidence for cingulate subregions subserving multiple distributed attentional systems. *Biol. Psychiatry* 45, 1237–1258.
- Podlipniak, P. (2016). The evolutionary origin of pitch centre recognition. *Psychol. Music* 44, 527–543. doi: 10.1177/0305735615577249
- Rothenberg, D., Roeske, T. C., Voss, H. U., Naguib, M., and Tchernichovski, O. (2014). Investigation of musicality in birdsong. *Hear. Res.* 308, 71–83. doi: 10.1016/j.heares.2013.08.016
- Royall, D. R., Lauterbach, E. C., Cummings, J. L., Reeve, A., Rummans, T. A., Kaufer, D. I., et al. (2002). Executive control function: a review of its promise and challenges for clinical research. A report from the Committee on Research of the American Neuropsychiatric Association. *J. Neuropsychiatry Clin. Neurosci.* 14, 377–405. doi: 10.1176/jnp.14.4.377

- 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
- Schröder, O., Schriewer, E., Golombek, K. S., Kürten, J., Lohmann, H., Schwindt, W., et al. (2015). Impaired autonomic responses to emotional stimuli in autoimmune limbic encephalitis. *Front. Neurol.* 6:250. doi: 10.3389/fneur.2015.00250
- Shannon, R. V., Zeng, F.-G. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303–304. doi: 10.1126/science.270.5234.303
- Soon, C. S., He, A. H., Bode, S., and Haynes, J.-D. (2013). Predicting free choices for abstract intentions. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6217–6222. doi: 10.1073/pnas.1212218110
- Stainsby, T., and Cross, I. (2008). “The perception of pitch,” in *The Oxford Handbook of Music Psychology*, Vol. 1, eds S. Hallam, I. Cross, and M. Thaut (Oxford: Oxford University Press), 47–58. doi: 10.1093/oxfordhob/9780199298457.013.0005
- Steinbeis, N., Koelsch, S., and Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: evidence from subjective, Physiological, and Neural Responses. *J. Cogn. Neurosci.* 18, 1380–1393. doi: 10.1162/jocn.2006.18.8.1380
- Tabbert, K., Stark, R., Kirsch, P., and Vaitl, D. (2006). Dissociation of neural responses and skin conductance reactions during fear conditioning with and without awareness of stimulus contingencies. *Neuroimage* 32, 761–770. doi: 10.1016/j.neuroimage.2006.03.038
- Tasker, J. G., Chen, C., Fisher, M. O., Fu, X., Rainville, J. R., and Weiss, G. L. (2015). Endocannabinoid regulation of neuroendocrine systems. *Int. Rev. Neurobiol.* 125, 163–201. doi: 10.1016/bs.irn.2015.09.003
- Tavares, R. F., Corrêa, F. M. A., and Resstel, L. B. M. (2009). Opposite role of infralimbic and prelimbic cortex in the tachycardiac response evoked by acute restraint stress in rats. *J. Neurosci. Res.* 87, 2601–2607. doi: 10.1002/jnr.22070
- Tillmann, B. (2005). Implicit investigations of tonal knowledge in nonmusician listeners. *Ann. N. Y. Acad. Sci.* 1060, 100–110. doi: 10.1196/annals.1360.007
- Tillmann, B., Bharucha, J. J., and Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychol. Rev.* 107, 885–913. doi: 10.1037/0033-295X.107.4.885
- Tillmann, B., Janata, P., and Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cogn. Brain Res.* 16, 145–161. doi: 10.1016/S0926-6410(02)00245-8
- Tillmann, B., Koelsch, S., Escoffier, N., Bigand, E., Lalitte, P., Friederici, A. D., et al. (2006). Cognitive priming in sung and instrumental music: activation of inferior frontal cortex. *Neuroimage* 31, 1771–1782. doi: 10.1016/j.neuroimage.2006.02.028
- Vuust, P., and Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Front. Psychol.* 5:1111. doi: 10.3389/fpsyg.2014.01111
- Wessel, D. (1979). Timbre space as a musical control structure. *Comput. Music J.* 3, 45–52. doi: 10.2307/3680283
- Wood, K. H., Ver Hoef, L. W., and Knight, D. C. (2014). The amygdala mediates the emotional modulation of threat-elicited skin conductance response. *Emotion* 14, 693–700. doi: 10.1037/a0036636
- Woolhouse, M., Cross, I., and Horton, T. (2016). Perception of nonadjacent tonic-key relationships. *Psychol. Music* 44, 802–815. doi: 10.1177/0305735615593409
- Xu, L., Thompson, C. S., and Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *J. Acoust. Soc. Am.* 117, 3255–3267. doi: 10.1121/1.1886405
- Zhang, S., Mano, H., Ganesh, G., Robbins, T., and Seymour, B. (2016). Dissociable learning processes underlie human pain conditioning. *Curr. Biol.* 26, 52–58. doi: 10.1016/j.cub.2015.10.066

**Conflict of Interest Statement:** 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.

Copyright © 2017 Gorzelańczyk, Podlipniak, Walecki, Karpiński and Tarnowska. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Pleasure Evoked by Sad Music Is Mediated by Feelings of Being Moved

Jonna K. Vuoskoski<sup>1,2\*</sup> and Tuomas Eerola<sup>2,3</sup>

<sup>1</sup> Faculty of Music, University of Oxford, Oxford, UK, <sup>2</sup> Department of Music, University of Jyväskylä, Jyväskylä, Finland,

<sup>3</sup> Department of Music, Durham University, Durham, UK

Why do we enjoy listening to music that makes us sad? This question has puzzled music psychologists for decades, but the paradox of “pleasurable sadness” remains to be solved. Recent findings from a study investigating the enjoyment of sad films suggest that the positive relationship between felt sadness and enjoyment might be explained by feelings of being moved (Hanich et al., 2014). The aim of the present study was to investigate whether feelings of being moved also mediated the enjoyment of sad music. In Experiment 1, 308 participants listened to five sad music excerpts and rated their liking and felt emotions. A multilevel mediation analysis revealed that the initial positive relationship between liking and felt sadness ( $r = 0.22$ ) was fully mediated by feelings of being moved. Experiment 2 explored the interconnections of perceived sadness, beauty, and movingness in 27 short music excerpts that represented independently varying levels of sadness and beauty. Two multilevel mediation analyses were carried out to test competing hypotheses: (A) that movingness mediates the effect of perceived sadness on liking, or (B) that perceived beauty mediates the effect of sadness on liking. Stronger support was obtained for Hypothesis A. Our findings suggest that – similarly to the enjoyment of sad films – the aesthetic appreciation of sad music is mediated by being moved. We argue that felt sadness may contribute to the enjoyment of sad music by intensifying feelings of being moved.

**Keywords:** sad music, being moved, music-induced emotion, empathy, liking, beauty

## OPEN ACCESS

### Edited by:

Sonja A. Kotz,  
Maastricht University, Netherlands

### Reviewed by:

Thomas Jacobsen,  
Helmut Schmidt University, Germany  
Sofia Dahl,  
Aalborg University, Denmark

### \*Correspondence:

Jonna K. Vuoskoski  
jonna.vuoskoski@music.ox.ac.uk

### Specialty section:

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

**Received:** 27 November 2016

**Accepted:** 09 March 2017

**Published:** 21 March 2017

### Citation:

Vuoskoski JK and Eerola T (2017)  
The Pleasure Evoked by Sad Music Is  
Mediated by Feelings of Being  
Moved. *Front. Psychol.* 8:439.  
doi: 10.3389/fpsyg.2017.00439

## INTRODUCTION

Why do people sometimes enjoy listening to music that makes them sad? The paradox of “pleasurable sadness” has attracted significant research interest among music psychology scholars in recent years (for a review, see Sachs et al., 2015), but the puzzle remains to be solved. A body of empirical work has shown that listening to nominally sad music induces a multifaceted emotional response that is not clearly negative or positive (e.g., Vuoskoski et al., 2012; Kawakami et al., 2013; Taruffi and Koelsch, 2014), and that there are certain personality variables that are consistently associated with the enjoyment of sadness-inducing music (e.g., Garrido and Schubert, 2011; Vuoskoski et al., 2012; Taruffi and Koelsch, 2014; Eerola et al., 2016). Furthermore, sad music has been shown to induce sadness-related biases in memory and judgment (Vuoskoski and Eerola, 2012, 2015), suggesting that listening to sad music is indeed able to induce ‘genuine’ sad affective states in listeners.

During the past two decades, multiple theoretical accounts for the ‘sadness paradox’ have been proposed by different scholars. Schubert (1996) postulated that, although positive and negative emotions are typically connected to pleasure and displeasure (respectively), the connection between negative emotion and displeasure gets inhibited in an aesthetic context such as music listening. However, this account does not explain why certain nominally negative music-induced emotions such as sadness are enjoyed, while others such as fear are not (see Vuoskoski et al., 2012). Huron (2011) has further extended the idea that music-induced sadness is disconnected from the negative real-world implications and displeasure that are typically associated with experiences of sadness, and proposed that the pleasure sometimes experienced while listening to sad music might be related to the adaptive, consoling physiological responses (such as the release of prolactin) triggered by a sad affective state. While this account fits together nicely with empirical findings linking greater intensity of felt sadness with greater enjoyment (e.g., Vuoskoski et al., 2012; Eerola et al., 2016), direct empirical evidence for the potential role of prolactin and other hormones is currently still lacking. Moreover, Juslin (2013) has criticized the fact that the proposed prolactin/consolation effect is actually an ‘after-effect’ rather than a pleasurable experience of listening to sadness-inducing music, and that in this account “there is no ‘pleasurable sadness,’ there is only pleasure following sadness” (Juslin, 2013; p. 258).

Juslin (2013), on the other hand, has proposed that the enjoyment of sadness-inducing music might have nothing to do with sadness itself, but that sad music is experienced as pleasurable simply because it is aesthetically pleasing or ‘beautiful’. Indeed, empirical studies have documented strong positive correlations between perceived sadness and perceived beauty (at least in Western music tradition; Eerola and Vuoskoski, 2011), and some of the most intense aesthetic listening experiences have been brought about by sad music (Gabrielsson and Lindström, 1993; Eerola and Peltola, 2016). But rather than solving the paradox, Juslin’s proposal brings us back to the original problem: What exactly makes sad music so profoundly ‘beautiful,’ and is ‘beautiful’ not just another way of describing pleasurable stimuli?

The concept of ‘beauty’ is central to the aesthetic appreciation of music (Istók et al., 2009; Brattico and Pearce, 2013), although aesthetic experiences comprise other components as well. In the broader context of music-related aesthetic experiences, Brattico et al. (2013) propose distinguishing between aesthetic emotions (e.g., feelings of awe, enjoyment, and interest), aesthetic judgments (the appraisal of beauty, proficiency, and other aesthetic dimensions), and conscious liking (involving decisional, evaluative processes). They argue that conscious liking succeeds aesthetic emotions and judgments in the temporal domain, and might even occur independently of them. Preliminary empirical work suggests that conscious liking, perceived beauty and pleasantness are highly inter-correlated constructs ( $r_s = 0.56\text{--}0.87$ ; Eerola and Vuoskoski, 2011), but indeed not identical. Although significant advances have been made in uncovering the neural underpinnings of music-induced pleasure (including aesthetic ‘chills’; Blood and Zatorre, 2001; Salimpoor et al., 2011) and in distinguishing between the brain

structures involved in liking and the perception of happiness and sadness (Brattico et al., 2016), the constituents of musical ‘beauty’ are not yet well understood (e.g., Brattico et al., 2013).

A new clue for the ‘sadness paradox’ may be offered by recent findings from the field of film studies. Hanich et al. (2014) investigated the enjoyment of sad films using 38 film clips as stimuli, and found that the initial positive relationship between felt sadness and enjoyment was entirely mediated by feelings of being moved. Experiences of ‘being moved’ are still not fully understood due to a lack of psychological research, but Menninghaus et al. (2015; see also Kuehnast et al., 2014) have recently offered a comprehensive account of the phenomenon: On the basis of multiple exploratory studies, they concluded that feelings of being moved are typically evoked by critical life and relationship events such as birth, death, and marriage, but also by exposure to artworks, nature, and music. Importantly, the two main emotional ingredients of being moved appear to be sadness and joy. The instances of joy and/or sadness that give rise to the special emergent feeling of being moved seem to have certain common characteristics: Events eliciting feelings of being moved are often characterized by high compatibility with prosocial norms and self-ideals, and the person experiencing the emotion is typically unable to affect the event or its outcome. Wassiliwizky et al. (2015) extended the findings of Hanich et al. (2014) by investigating emotional and aesthetic responses to sad and joyful film clips. In line with Hanich et al. (2014), they found that the positive relationship between felt sadness and enjoyment was entirely mediated by feelings of being moved, but no mediation effect was found for the relationship between felt joy and enjoyment. Crucially, they also found that participants often used empathy-related words such as “compassionate” and “sympathetic” to describe their emotional responses to the sad film clips, corroborating the link between sadly moving scenarios and prosocial norms and ideals (see Menninghaus et al., 2015). Indeed, Menninghaus et al. (2015) have proposed that feelings of being moved may serve a social bonding function by activating the value of social bonds and prosocial behavior, and thus those high in socially responsive traits such as empathy may be especially prone to experience feelings of being moved.

Interestingly, Wassiliwizky et al. (2015) also found that feelings of being moved were the best predictor of the likelihood of experiencing aesthetic chills; pleasurable bodily sensations commonly described as a ‘spreading gooseflesh’ (cf. Panksepp, 1995). Feelings of sadness were also positively associated with the likelihood of chills, but there was no statistically significant association between felt joy and chills. Similar findings have also been obtained in the context of music listening, where sad music has been found to be significantly more likely to evoke aesthetic chills than happy music (Panksepp, 1995). Although it has been more than a decade since Konecni (2005) outlined aesthetic awe, being moved, and chills as the three central aesthetic responses (awe being the rarest and most profound of the three, and chills being the most common), feelings of ‘being moved’ have rarely been explicitly studied in the context of music listening. A recent study by Eerola et al. (2016), however, explored the structure of emotions experienced in response to nominally sad, unfamiliar music, and found that feelings of being moved played

a central role in enjoyable feelings of music-induced sadness. More specifically, they found that feelings of sadness, being moved, and liking all loaded strongly on the same latent emotion factor that they subsequently labeled 'Moving sadness,' and that experiences of 'Moving sadness' were significantly predicted by trait empathy. However, it is not yet known whether feelings of being moved might actually mediate the positive relationship between felt sadness and enjoyment as in the case of sad films (Hanich et al., 2014; Wassiliwizky et al., 2015).

## The Present Study

The aim of the present study was to investigate the hypothesis that feelings of being moved would mediate the positive effect of felt sadness on enjoyment in the context of music listening. Enjoyment was operationalized as 'liking' for the music. Furthermore, since trait empathy has been repeatedly implicated in the enjoyment of sad music (e.g., Garrido and Schubert, 2011; Vuoskoski et al., 2012; Taruffi and Koelsch, 2014; Eerola et al., 2016) and in the psychological phenomenon of being moved (Menninghaus et al., 2015; Wassiliwizky et al., 2015), it was hypothesized that trait empathy would contribute to feelings of being moved evoked by sad music. Experiment 1 explored the potential mediating role of 'being moved' and the contribution of trait empathy in an online listening setting. Experiment 2 was carried out in a laboratory, and was designed to untangle the complex relationships between perceived sadness, movingness, beauty, and liking.

## EXPERIMENT 1

### Method

#### Ethics Statement

The experimental protocol was approved by the Ethics Committee of the University of Jyväskylä, Finland. All participants gave their written, informed consent, and the study was carried out in accordance with the approved guidelines.

#### Participants

Three hundred and thirty-eight participants from different countries took part in an online experiment. After deleting partial answers and outliers (i.e., those participants whose inter-group correlation was 2 SDs below the mean inter-group correlation), we were left with 308 participants (239 female) aged 18–75 ( $M = 31.7$ ,  $SD = 9.7$ ). Participants were recruited by distributing the experiment link on social media (Facebook, Twitter, and Reddit). Two hundred two participants (61.2%) were Finnish, 10.6% were American, 5.2% British, 4.5% German, and 18.5% were other nationalities.

#### Stimulus Material

The stimuli were selected by a panel of five expert judges (including the authors). With the aim to find a variety of music examples that would successfully convey sadness, each panel member selected three musical examples from different genres. The resulting 15 music examples were then rated by all panel members using the same set of rating scales as used

by the participants of the main experiment (see Procedure – section below). Out of these 15 music examples, four examples were chosen for the main experiment on the basis that they conveyed differing (yet sufficient) levels of sadness as well as varying degrees of movingness and other emotional qualia (e.g., peacefulness and anxiety). The four selected examples were *Oblivion* (composed by Astor Piazzolla, performed by Stjepan Hauser with the Zagreb Philharmonic Orchestra), *Darkness* (by Lacrimosa), *Something I Can Never Have* (by Nine Inch Nails), and *Together We Will Live Forever* (by Clint Mansell), representing different genres (classical, film music, and gothic and industrial rock). Two of the examples contained lyrics (*Darkness* and *Something I Can Never Have*). Furthermore, a fifth piece, *Discovery of the Camp* (composed by Michael Kamen), was included on the basis that it has successfully been used in previous studies to induce sadness and feelings of being moved (Vuoskoski and Eerola, 2012; Eerola et al., 2016). Two-minute excerpts of each of these five pieces were used as the stimuli in the experiment.

### Measures

Two subscales of The Interpersonal Reactivity Index (IRI; Davis, 1983), Fantasy and Empathic Concern, were used to measure participants' trait empathy. The IRI is a widely used, multifaceted measure of trait empathy, and the two aforementioned subscales have been repeatedly implicated in studies investigating individual differences in the enjoyment of sadness-inducing music (Garrido and Schubert, 2011; Vuoskoski et al., 2012; Eerola et al., 2016). We also included a measure of trait emotional contagion, The Emotional Contagion Scale (ECS; Doherty, 1997), since recent work has shown the ECS to be – in addition to Fantasy and Empathic Concern – one of the best predictors of the enjoyment of unfamiliar, sadness-inducing music (Eerola et al., 2016).

### Procedure

The experiment was carried out online using the *Qualtrics*-platform. In order to minimize self-selection bias related to music-induced emotions (and sadness in particular), participants were recruited by promising them individualized feedback on their personality traits (based on their trait empathy scores). Participants were told that they would hear some music in the experiment, but emotions were not mentioned in the study advertisement. The five musical stimuli were presented in a random order, and the participants were instructed to listen to the entire excerpt before giving their ratings. They were asked to wear headphones if possible to ensure optimal sound quality. The experiment was programmed in such a way that the participants could not move to the next excerpt before 1 min had passed. The participants were asked to rate how much they liked each excerpt, and to describe their emotional reaction ('How did you feel when you listened to the music?') using seven adjective scales (*sad*, *melancholic*, *moved*, *in awe*, *peaceful*, *anxious*, and *powerless*). The selection of rating scales was based on a previous study that explored the underlying factor structure of emotional responses to sad-sounding music (Vuoskoski and Eerola, submitted), as the objective was to provide the participants with a selection

of scales that would satisfactorily reflect the range and type of emotions typically experienced in response to sad music. The scale extremes were labeled “Does not describe my emotional reaction at all” and “Describes my emotional reaction very well.” The liking and emotion ratings were given using slider scales ranging from 0 to 100. The participants were also asked to rate the familiarity of the music excerpts (on a 4-point scale). After listening to all five music excerpts and reporting their felt emotions, liking, and familiarity, the participants filled in the trait empathy questionnaires.

## Results

### Descriptive Statistics

The mean familiarity ratings for the five music excerpts ranged from 1.13 to 1.60, (on a scale from 1 to 4) indicating that the excerpts were unfamiliar to the majority of participants. The mean ratings of liking, felt sadness and being moved given to the five music excerpts are displayed in **Figure 1**, demonstrating the variability of the felt emotions and liking responses evoked by the different stimuli. In order to explore the general pattern of associations among the ratings of felt emotion and liking, Pearson correlation coefficients were calculated for each participant using their raw ratings, and then averaged over participants (see **Table 1** for the correlation matrix). Because of the high number of correlations and the descriptive nature of the analysis, we have refrained from making inferences regarding the statistical significance of the correlations. The correlations revealed a strong positive association between liking and being moved ( $r = 0.69$ ), a moderate correlation between felt sadness and being moved ( $r = 0.29$ ), and a small correlation between liking and felt sadness ( $r = 0.22$ ), thus providing grounds for a mediation analysis.

### Mediation Analysis

The hypothesis that the enjoyment of sadness-inducing music (i.e., the positive association between felt sadness and liking ratings) is mediated by feelings of being moved was tested through a multilevel (1–1–1) mediation analysis, following the method documented by Bauer et al. (2006). Essentially, this approach – also used by Hanich et al. (2014) – provides all the necessary information for evaluating the hypothesized causal effects of the mediation model by combining the dependent variable (liking) and the mediator (being moved) into a single stacked response variable, and running a mixed model with selection variables for the DV and mediator to toggle between models. Multilevel mediation analysis was used because of the structure of the rating data, which represented a nested structure with the five music excerpts (at Level 1) nested within participants at Level 2; the model included random slopes and random intercepts for participants. Confidence interval for the mediation (indirect) effect was calculated using the method presented by Preacher and Selig (2010). The analyses were carried out in R using the lme4-package (Bates et al., 2015).

The paths, coefficients, and random-slope plots of the multilevel mediation analysis are visualized in **Figure 2**. The total effect of felt sadness on liking was significant (path c;  $\beta = 0.25$ ,

$t = 8.46$ ). The effect of felt sadness on being moved (path a;  $\beta = 0.43$ ,  $t = 13.82$ ), and the effect of being moved on liking (path b;  $\beta = 0.68$ ,  $t = 32.85$ ) were also significant. However, when feelings of being moved were controlled for, the direct effect of felt sadness on liking became non-significant (path c';  $\beta = -0.042$ ,  $t = -1.67$ ). The estimated indirect effect of felt sadness on liking (mediated by feelings of being moved) was 0.30 (95% CI [0.25; 0.34]), suggesting that the positive relationship between sadness and liking was entirely mediated by feelings of being moved.

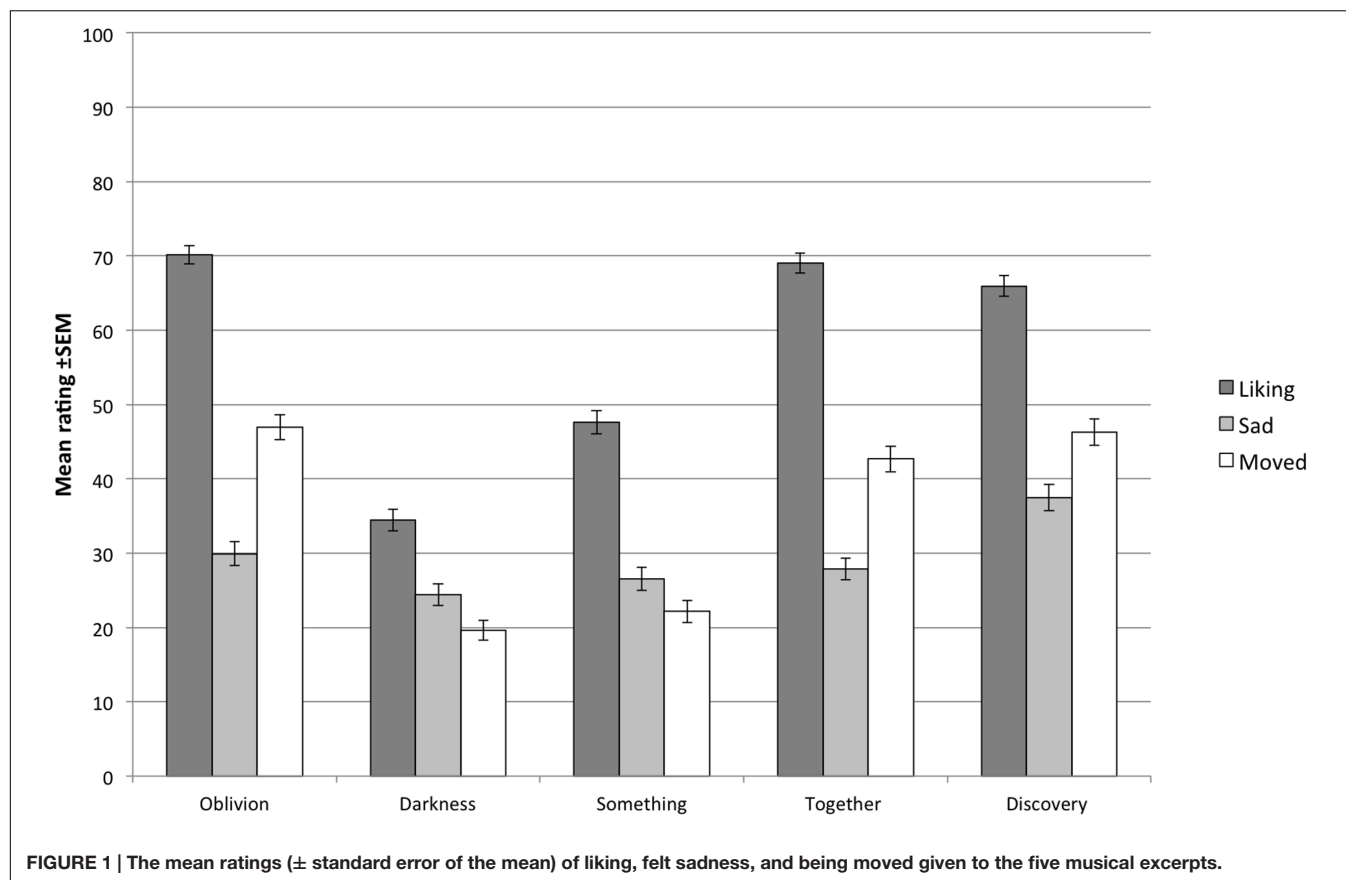
If we adopt a broader view of felt sadness and include melancholy into an aggregate measure of felt sadness (felt sadness + felt melancholy), the pattern of coefficients in the mediation analysis remains relatively unchanged: The total effect of felt sadness on liking becomes somewhat stronger (path c;  $\beta = 0.38$ ,  $t = 11.65$ ) as does the effect of felt sadness on being moved (path a;  $\beta = 0.56$ ,  $t = 18.41$ ). The effect of being moved on liking remains very similar (path b;  $\beta = 0.62$ ,  $t = 27.49$ ), and when feelings of being moved are controlled for, the direct effect of felt sadness on liking becomes non-significant once again (path c';  $\beta = 0.046$ ,  $t = 1.58$ ). The estimated indirect effect of felt sadness on liking was practically unchanged (0.33; 95% CI [0.29; 0.38]), further confirming that the positive effect of felt sadness on liking is mediated fully by feelings of being moved.

### Individual Differences

Finally, we explored the hypothesis that trait empathy would contribute to feelings of being moved, as well as the possibility that trait empathy might modulate the relationships between felt sadness, being moved, and liking. Fantasy, Empathic Concern, and Emotional Contagion were all significantly correlated with mean ratings of felt sadness ( $r_s = 0.18$ – $0.23$ ,  $p < 0.001$ – $0.01$ ) and being moved (all  $r_s = 0.25$ ,  $p < 0.001$ ), but only Empathic Concern was significantly correlated with mean liking ratings ( $r = 0.15$ ,  $p < 0.01$ ). However, none of the trait empathy variables were significantly correlated with the individual slope coefficients extracted from the first multilevel mediation analysis, suggesting that – although trait empathy may contribute to the overall intensity of felt sadness and feelings of being moved (and thus liking) – it does not modulate the relationships between the variables.

## Discussion

In line with previous findings obtained using film clips (Hanich et al., 2014; Wassiliwizky et al., 2015), we found that the initial positive relationship between felt sadness and liking was entirely mediated by feelings of being moved. The results were almost identical regardless of the type of felt sadness used as the independent variable; ratings of felt sadness, or an aggregate of felt sadness and felt melancholy. The close similarities in the patterns of mediation obtained in the present study and in those by Hanich et al. (2014) and Wassiliwizky et al. (2015) are especially remarkable when the differences in the operationalization of ‘enjoyment’ are taken into consideration. In the present study, we used conscious liking as the dependent measure, while Hanich et al. (2014) and Wassiliwizky et al. (2015)



used the degree of ‘wanting to see the entire film’ as a proxy for enjoyment (a decision for which they provide a well-argued explanation). Nevertheless, the closely replicated path coefficients in the mediation models suggest that both operationalizations seem to tap into the same broader construct of ‘enjoyment.’

As hypothesized, measures of trait empathy (Fantasy, Empathic Concern, and Emotional Contagion) were positively correlated with the mean ratings of being moved. Trait empathy was also associated with ratings of felt sadness, although only Empathic Concern was significantly correlated with mean liking ratings. These results corroborate the findings of previous studies (e.g., Garrido and Schubert, 2011; Vuoskoski et al., 2012; Taruffi and Koelsch, 2014; Kawakami and Katahira, 2015),

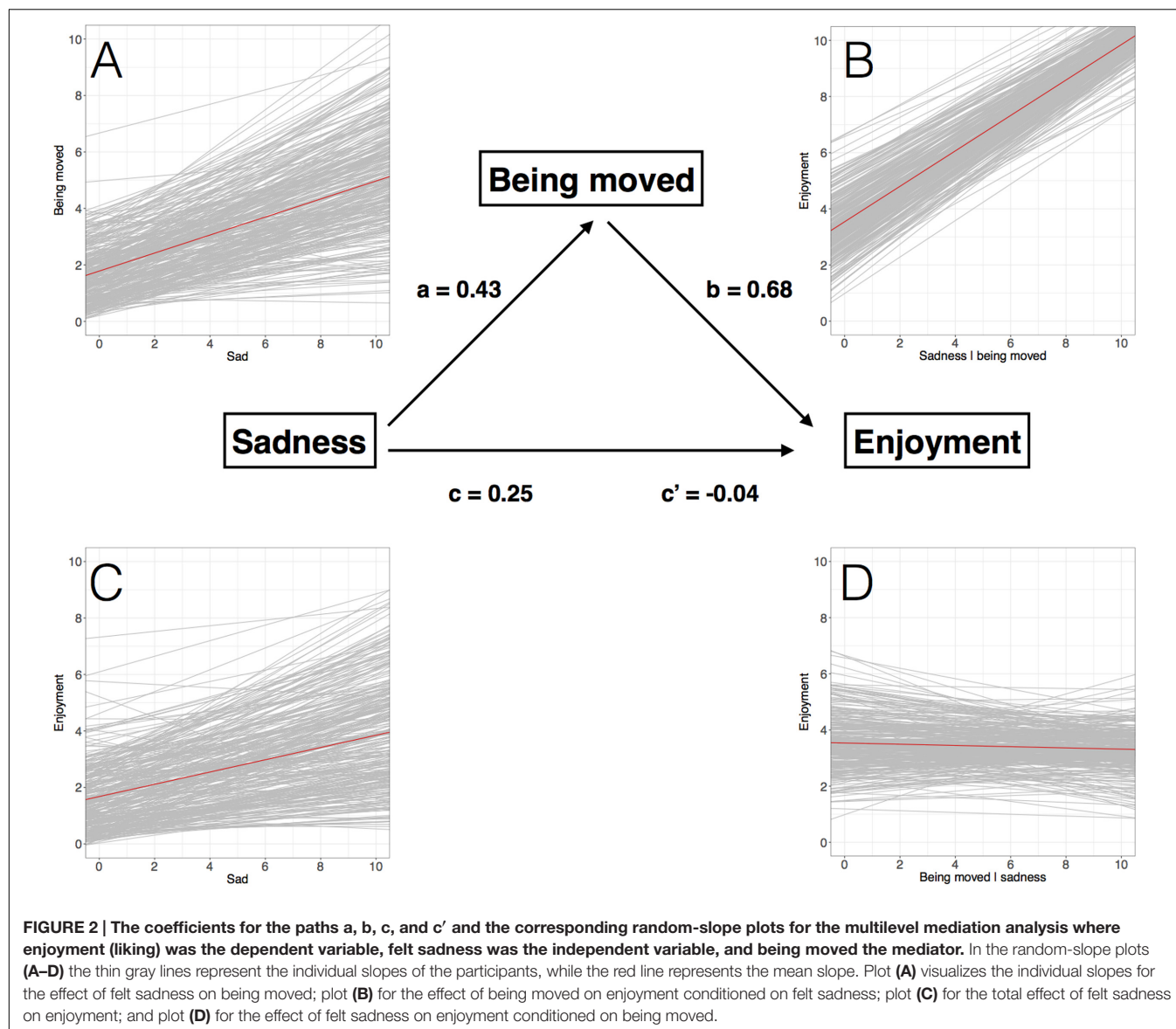
most notably those of Eerola et al. (2016), who found that Fantasy, Empathic Concern, and Emotional Contagion were the best predictors of experiences of ‘Moving sadness.’ However, we did not find any association between trait empathy and the individual slope coefficients extracted from the multilevel mediation model, suggesting that trait empathy did not modulate the relationships between felt sadness, being moved, and liking.

The findings and conclusions of Experiment 1 are subject to certain limitations and considerations. First, as the experiment was carried out on an online platform, we did not have any control over the listening situation or the amount of attention that participants paid to the listening and rating tasks. Second,

**TABLE 1 |** Pearson correlations coefficients between ratings of felt emotion and liking.

	Liking	Sad	Moved	Melanc.	Anxious	Powerless	Peaceful
Sad	0.22						
Moved	0.69	0.29					
Melancholic	0.36	0.52	0.39				
Anxious	−0.29	0.24	−0.20	−0.03			
Powerless	−0.05	0.25	−0.01	0.19	0.24		
Peaceful	0.56	0.09	0.48	0.26	−0.36	−0.09	
In awe	0.52	0.11	0.55	0.21	−0.17	−0.04	0.36

The correlations were first calculated within each participant, and then averaged across participants.



in order to prevent experiment fatigue and drop-outs, we only used a small number (5) of stimuli. This prevented any systematic variation of stimulus features, although the selected stimuli were deliberately intended to evoke varying degrees of liking and being moved. Thus, Experiment 2 was designed to address these limitations, and to further explore the hypothesized mediating role of being moved in the enjoyment of sad music.

## EXPERIMENT 2

The aim of Experiment 2 was to try and elucidate the interconnections of perceived sadness, movingness, beauty, and liking in a laboratory setting. Previous work has shown perceived sadness and beauty to be highly correlated ( $r = 0.59$ ; Eerola and Vuoskoski, 2011), but it is not known whether this covariance is inherent to the two phenomena, or whether

they just happen to be correlated in the Western music corpus. Furthermore, it is not fully understood what qualities or features contribute to perceived beauty in the context of music listening. We set out to select musical material where levels of sadness and beauty would vary as independently as possible, as this would enable us to investigate Juslin's (2013, p. 258) claim that "It is not that the sadness *per se* is a source of pleasure, it only happens to occur together with a percept of beauty." Specifically, we wanted to test two competing hypotheses: (A) that movingness mediates the effect of perceived sadness on liking, or (B) that perceived beauty mediates the effect of sadness on liking. We also wanted to investigate whether movingness might mediate the positive relationship between perceived sadness and beauty. The decision was made to focus on perceived rather than felt emotion, as emotion perception can be reliably measured using relatively short stimuli (see e.g., Eerola and Vuoskoski, 2011, 2012). The shorter

duration allowed us to use a larger number (27) of stimuli, and thus systematically vary levels of perceived sadness and beauty.

## Method

### Ethics Statement

The experimental protocol was approved by the Ethics Committee of the University of Jyväskylä, Finland. All participants gave their written, informed consent, and the study was carried out in accordance with the approved guidelines.

### Participants

The participants of Experiment 2 were 19 music students from the University of Jyväskylä (studying musicology or music education) aged 20–45 ( $M = 24.74$ ,  $SD = 5.50$ , 15 female).

### Stimulus Material

The stimuli were 27 short film music excerpts (duration: 13–26 s;  $M = 17.56$ ,  $SD = 3.27$ ) that were selected from a pool of 403 excerpts with pre-existing ratings of perceived emotion (360 excerpts from Eerola and Vuoskoski, 2011, and 43 excerpts from an unpublished dataset;  $n = 9$ ). Pre-existing ratings of perceived beauty were present for a subset of 110 excerpts from Eerola and Vuoskoski (2011), but not for the remaining excerpts. For these 293 excerpts, the beauty ratings were estimated using a regression model based on ratings of perceived emotion (built using the dataset of 110 excerpts from Eerola and Vuoskoski, 2011). Based on the actual and estimated mean ratings of perceived beauty and sadness, we selected 27 examples where levels of beauty and sadness would vary as independently as possible; low, medium, and high levels of both in a  $3 \times 3$  factorial design; three excerpts per combination. In the selected set of stimuli, 24 excerpts were from the set of Eerola and Vuoskoski (2011), while three excerpts were from the unpublished set (see Supplementary Material for the list of excerpts).

### Procedure

The experiments were conducted individually for each participant using customized software built in the MAX/MSP graphical programming environment (version 5.1), running on Mac OS X. The excerpts were presented in a different random order to each participant. Because of the relatively high number and short duration of the musical stimuli, participants were asked to rate perceived (i.e., what the music sounds like) rather than felt emotion. The participants were asked to describe their perceived emotions using six adjective scales (range: 0–100); *sad/melancholic*, *moving/touching*, *tender/warm*, *peaceful/relaxing*, *scary/distressing*, and *happy/joyful* (translated from Finnish by the first author); the scale extremes were “Does not describe the music at all”, and “Describes the music very well.” Participants were also asked to rate how much they liked each excerpt, and how beautiful it sounded. Participants listened to the excerpts through studio quality headphones, and were able to adjust the sound volume according to their own preferences.

## Results

### Descriptive Statistics

The mean ratings of perceived sadness and beauty for the different types of excerpts are displayed in **Table 2**. Note that the excerpts have been categorized according to the mean ratings obtained in the present experiment (three excerpts per category) rather than those used in the stimulus selection process. The mean ratings highlight the difficulty of finding film music excerpts – even from a database of over 400 examples – that would be both highly sad and low in perceived beauty. We further explored the covariance between perceived emotions, liking, and beauty by calculating Pearson correlation coefficients using the raw ratings of participants.

The correlations were again first calculated for each individual participant, and then averaged over participants (see **Table 3** for the correlation matrix). The correlations revealed a strong positive association between liking and beauty ( $r = 0.76$ ), and moderately strong correlations between perceived sadness and movingness ( $r = 0.49$ ), and perceived beauty and movingness ( $r = 0.55$ ). Although every attempt was made to manipulate levels of perceived beauty and sadness as independently as possible, the two concepts were still positively correlated to a limited extent ( $r = 0.25$ ).

### Mediation Analysis

The same multilevel mediation analysis method as used in Experiment 1 was used to analyze the data from the current experiment. First, we tested the hypothesis that perceived movingness would mediate the effect of perceived sadness on liking ratings (Hypothesis A) in a similar manner as being moved mediated the effect of felt sadness on liking in Experiment 1. The total effect of perceived sadness on liking was not quite significant (path c;  $\beta = 0.091$ ,  $t = 2.08$ ). However, the effect of perceived sadness on perceived movingness (path a;  $\beta = 0.47$ ,  $t = 8.16$ ), and the effect of perceived movingness on liking (path b;  $\beta = 0.43$ ,  $t = 6.56$ ) were significant. When perceived movingness was controlled for, the direct effect of perceived sadness on liking became significantly negative (path c';  $\beta = -0.13$ ,  $t = -2.62$ ). The estimated indirect effect of perceived sadness on liking (mediated by perceived movingness) was 0.22 (95% CI [0.14; 0.31]); somewhat smaller than in Experiment 1, but still notable.

Next, we investigated the competing hypothesis (Hypothesis B) that – instead of movingness – perceived beauty would mediate the effect of perceived sadness on liking. As in the first mediation analysis, the total effect of perceived sadness on liking was not quite significant (path c;  $\beta = 0.086$ ,  $t = 2.08$ ). The effect of perceived sadness on beauty (path a;  $\beta = 0.15$ ,  $t = 4.69$ ) and the effect of beauty on liking (path b;  $\beta = 0.75$ ,  $t = 11.35$ ) were both significant. When perceived beauty is controlled for, the direct effect of perceived sadness on liking becomes negative (but non-significant; path c';  $\beta = -0.033$ ,  $t = -0.93$ ). The estimated indirect effect of perceived sadness on liking (mediated by beauty) was 0.12 (95% CI [0.07; 0.17]); half the magnitude of the indirect effect mediated by movingness.

**TABLE 2 | Mean ratings (and standard deviations) of perceived beauty and sadness for the nine different types of excerpts (scale range: 0–100).**

	Low sadness		Medium sadness		High sadness	
	Beauty	Sadness	Beauty	Sadness	Beauty	Sadness
Low beauty	32.2 (23.6)	22.7 (26.6)	50.2 (22.4)	50.8 (30.9)	65.7 (21.7)	68.3 (26.4)
Medium beauty	63.6 (23.7)	15.8 (21.9)	65.4 (24.9)	49.4 (24.6)	72.4 (22.1)	73.5 (18.6)
High beauty	82.3 (18.2)	22.8 (25.6)	76.2 (20.6)	54.7 (29.6)	77.4 (18.6)	73.5 (20.7)

**TABLE 3 | Pearson correlations coefficients between ratings of perceived emotion, beauty, and liking.**

	Liking	Beauty	Sad	Moving	Tender	Peaceful	Scary
Beauty	0.76						
Sad	0.16	0.25					
Moving	0.45	0.55	0.49				
Tender	0.38	0.48	−0.09	0.33			
Peaceful	0.36	0.50	0.04	0.38	0.70		
Scary	−0.36	−0.50	−0.02	−0.32	−0.56	−0.54	
Happy	0.11	0.07	−0.46	−0.18	0.23	0.09	−0.28

The correlations were first calculated within each participant, and then averaged across participants.

Finally, we investigated whether perceived movingness would also mediate the effect of perceived sadness on perceived beauty. The total effect of perceived sadness on beauty was significant (path  $c$ ;  $\beta = 0.16$ ,  $t = 4.99$ ). The effect of perceived sadness on movingness (path  $a$ ;  $\beta = 0.47$ ,  $t = 8.07$ ) and the effect of movingness on beauty (path  $b$ ;  $\beta = 0.50$ ,  $t = 8.37$ ) were also significant. Again, when perceived movingness is controlled for, the direct effect of perceived sadness on beauty becomes negative (but non-significant; path  $c'$ ;  $\beta = -0.068$ ,  $t = -1.30$ ). The estimated indirect effect of perceived sadness on beauty was 0.24 (95% CI [0.16; 0.33]).

## Discussion

The musical stimuli selected for Experiment 2 were intended to represent independently varying levels of sadness and beauty (high, medium, and low levels of both in a factorial design). However, the two concepts were still positively correlated ( $r = 0.25$ ), albeit the correlation was considerably smaller than that reported in a previous study where the selection of stimuli was based on other criteria ( $r = 0.59$ ; Eerola and Vuoskoski, 2011). The mean ratings displayed in **Table 2** indicate that the correlation may have been driven by the highly sad excerpts, as they exhibited less variance in terms of beauty compared to the other types of excerpts. Indeed, finding examples that would be highly sad but low in beauty proved to be especially difficult – at least when using Western film music as the stimulus material. This difficulty may be related to an inherent association between perceived sadness and beauty, or to stylistic conventions that are used in the Western music tradition to convey sadness. Future studies on the topic should strive to use musical materials from more varied cultural settings in order to further elucidate this issue.

In order to use a sufficiently large set of stimuli where levels of sadness and beauty could be varied systematically,

the decision was taken to measure perceived rather than felt emotions. Perceived emotions can be reliably measured using very short excerpts (even as short as 1–5 s; Al'tman et al., 2000; Bigand et al., 2005), while a review by Eerola and Vuoskoski (2012) found that the median stimulus duration in studies investigating music-induced (felt) emotions was 90 s. This suggests that the induction of emotion (and sufficiently accurate introspection) is typically thought to require considerably more time than emotion recognition/perception. Thus, instead of investigating *feelings* of sadness and being moved, Experiment 2 measured the sad and moving qualities perceived in the music stimuli. This might partly explain why the mediation effect was slightly weaker than in Experiment 1 (although the selection of musical stimuli is likely to have played a role as well). However, it should be noted that perceived emotional qualities can directly lead to emotion induction (e.g., through emotional contagion; Juslin and Västfjäll, 2008). Indeed, significant overlap has been shown to exist between perceived and induced emotions (e.g., Kallinen and Ravaja, 2006; Schubert, 2013), and felt emotions that diverge from the perceived emotional expression of the music are typically driven by mechanisms such as episodic memories and evaluative conditioning (Juslin and Västfjäll, 2008; Juslin, 2013), which are typically associated with familiar music. As the present experiment used unfamiliar, experimenter-selected stimuli (familiarity ratings were collected by Eerola and Vuoskoski, 2011), it could be expected that the ratings of perceived emotion would not greatly diverge from hypothetical ratings of felt emotion.

We set out to test two competing hypotheses: (A) that movingness would mediate the effect of perceived sadness on liking, or (B) that perceived beauty would mediate the effect of sadness on liking. The results of the multilevel mediation analyses suggested that both perceived movingness and beauty may mediate the effect of perceived sadness on liking. However, the indirect effect via movingness was twice

the magnitude of that via beauty, suggesting that perceived movingness provides a better account for the link between perceived sadness and liking. This interpretation is further supported by the fact that the positive association between perceived sadness and beauty was entirely mediated by perceived movingness.

## GENERAL DISCUSSION

Our findings suggest that the pleasure drawn from sad music – similarly to the enjoyment of sad films – is mediated by being moved. In both experiments, the initial positive relationships between felt and perceived sadness and liking were entirely mediated by feelings and percepts of movingness. Our findings are in line with previous studies of music-induced sadness that have associated more intensely felt sadness with greater enjoyment (e.g., Vuoskoski et al., 2012; Eerola et al., 2016). However, we have provided new evidence of the significant role of being moved in this relationship. Our findings regarding the contribution of trait empathy to the feelings of sadness and being moved evoked by sad music are highly compatible with the notion of ‘being moved’ as a socially significant emotion that activates the value of social bonds and prosocial behavior (see Menninghaus et al., 2015).

But is there a musical equivalent for the pro-social features that characterize non-musical instances of being moved (e.g., Menninghaus et al., 2015; Wassiliwizky et al., 2015)? A growing body of research has established the significance of social cognition not only in music-making (e.g., Phillips-Silver and Keller, 2012; Moran, 2014), but also in music perception (Aucouturier and Canonne, 2017). A recent study by Aucouturier and Canonne (2017) showed that listeners could accurately decode social intentions (varying in the degree of affiliation and control) from improvised musical interactions, demonstrating that music can be perceived in terms of social relations between real and/or virtual agents. Moreover, certain musical features, namely harmonic and temporal coordination, were causally associated with the affiliation and control dimensions of social behavior (respectively). In addition to establishing that music can directly communicate social relational intentions, these findings are compatible with theoretical accounts proposing that music can be perceived and experienced as narratives of virtual persons ‘inhabiting’ the musical environment (e.g., Levinson, 2006). Thus, it is plausible that those pieces of sad music that listeners find particularly moving are perceived and experienced as communicating pro-social intentions, and may even afford a form of (para)social engagement – empathy – for the listener.

This interpretation is congruent with the finding that the particular subscales of trait empathy (namely Fantasy and Empathic Concern) that were associated with feelings of sadness and being moved in the present study, are specifically those that tap into the tendencies to imaginatively transpose oneself into the feelings of fictitious characters, and to engage in other-oriented, compassionate empathy and helping behavior

(e.g., Davis, 1983; Eisenberg and Fabes, 1990). Furthermore, the association between trait empathy and being moved provides further support and explanation for previous findings that have linked trait empathy with the enjoyment of sad music (e.g., Garrido and Schubert, 2011; Vuoskoski et al., 2012; Taruffi and Koelsch, 2014; Eerola et al., 2016). The findings of the present study suggest that trait empathy contributes directly to the intensity of felt sadness and movingness (and thus enjoyment), but does not modulate the relationships between felt sadness, being moved, and liking.

Our findings and conclusions are subject to certain limitations. As the majority of the participants in Experiment 1 were non-native English speakers, it is possible that there were differences in the participants’ understanding of the rating scale labels. Furthermore, since the majority of the participants in both experiments were female, our findings may not be equally generalizable to men (although most studies that have investigated the role of gender in music-related emotional processing have failed to find significant differences; see Eerola and Vuoskoski, 2012). By default, online experiments have less control in terms of audio quality and participant concentration, but a number of studies have also suggested that the inherent variability present in online studies is linked to clear advantages (such as a wider sample pool; Gosling et al., 2004; Kraut et al., 2004). However, the close similarities between the findings of Experiments 1 and 2 (with Experiment 2 using a more controlled setting albeit a smaller sample of participants), suggest that the pattern of mediation remains consistent despite the differences in language, setting, and musical material.

The present study has not exhaustively explored feelings of ‘being moved’ in the context of music-induced sadness, or its interactions with aesthetic appreciation or liking. It has, however, highlighted the significance of the phenomenon, and opened new avenues for further investigation. Future studies on music-induced feelings of being moved should investigate the physiological responses commonly associated with feelings of being moved (e.g., chills and skin conductance; Benedek and Kaernbach, 2011; Wassiliwizky et al., 2015) as well as the musical aspects that are conducive to being moved. The latter could entail exploring the perceived social intentions in music and their underlying acoustic features (following the work of Aucouturier and Canonne, 2017), or manipulating the aesthetically pleasing qualities of music (sad music in particular). This could help to clarify whether ‘being moved’ in the context of music listening would be better conceptualized as a social emotion (cf. Menninghaus et al., 2015), or as an aesthetic response (as proposed by Konecni, 2005).

Our findings have certainly highlighted the difficulty of disentangling perceived sadness and perceived beauty when using existing musical material. Although we allow that the conceptual distinction between liking and perceived beauty may be somewhat artificial (and that significant overlap most likely exists between the two), we nevertheless showed that perceived movingness mediated the effect of perceived sadness on both liking and perceived beauty.

Furthermore, when two alternative paths from perceived sadness to liking were considered – one via movingness and the other via beauty – the estimated indirect effect via movingness was double the magnitude of that via beauty. Thus, we argue that being moved may provide a more persuasive account of the paradox of “pleasurable sadness” than aesthetic appreciation. Contrary to Juslin’s (2013) statement, we argue that felt sadness does in fact contribute to the enjoyment of sadness-inducing music by directly intensifying feelings of being moved, and that the sadness *per se* can thus be considered a source of pleasure.

## AUTHOR CONTRIBUTIONS

JV conceived the study idea, and was responsible for the study design, carrying out the experiments, analyzing the data, and

writing the article. TE contributed significantly to the study design, data analysis, and writing the article.

## FUNDING

This work was financially supported by the Academy of Finland Grant 270220 (Surun Suloisuus).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.00439/full#supplementary-material>

## REFERENCES

- Al’tman, Y. A., Alyanchikova, Y. O., Guzikov, B. M., and Zakharova, L. E. (2000). Estimation of short musical fragments in normal subjects and patients with chronic depression. *Hum. Physiol.* 26, 553–557. doi: 10.1007/BF02760371
- Aucouturier, J.-J., and Canonne, C. (2017). Musical friends and foes: the social cognition of affiliation and control in improvised interactions. *Cognition* 161, 94–108. doi: 10.1016/j.cognition.2017.01.019
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bauer, D. J., Preacher, K. J., and Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. *Psychol. Methods* 11, 142–163. doi: 10.1037/1082-989X.11.2.142
- Benedek, M., and Kaernbach, C. (2011). Physiological correlates and emotional specificity of human piloerection. *Biol. Psychol.* 86, 320–329. doi: 10.1016/j.biopsycho.2010.12.012
- Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., and Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: the effect of musical expertise and of the duration of the excerpts. *Cogn. Emot.* 19, 1113–1139. doi: 10.1080/02699930500204250
- 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
- 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
- Brattico, E., Bogert, B., and Jacobsen, T. (2013). Toward a neural chronometry for the aesthetic experience of music. *Front. Psychol.* 4:206. doi: 10.3389/fpsyg.2013.00206
- Brattico, E., and Pearce, M. (2013). The neuroaesthetics of music. *Psychol. Aesthet. Creat. Arts* 7, 48–61. doi: 10.1037/a0031624
- Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44, 113–126. doi: 10.1037/0022-3514.44.1.113
- Doherty, R. W. (1997). The emotional contagion scale: a measure of individual differences. *J. Nonverbal Behav.* 21, 131–154. doi: 10.1023/A:1024956003661
- Eerola, T., and Peltola, H.-R. (2016). Memorable experiences with sad music – reasons, reactions and mechanisms of three types of experiences. *PLoS ONE* 11:e0157444. doi: 10.1371/journal.pone.0157444
- Eerola, T., and Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychol. Music* 39, 18–49. doi: 10.1177/0305735610362821
- Eerola, T., and Vuoskoski, J. K. (2012). A review of music and emotion studies: approaches, emotion models and stimuli. *Music Percept.* 30, 307–340. doi: 10.1525/mp.2012.30.3.307
- Eerola, T., Vuoskoski, J. K., and Kautiainen, H. (2016). Being moved by unfamiliar sad music is associated with high empathy. *Front. Psychol.* 7:1176. doi: 10.3389/fpsyg.2016.01176
- Eisenberg, N., and Fabes, R. A. (1990). Empathy: conceptualization, measurement, and relation to prosocial behavior. *Motiv. Emot.* 14, 131–149.
- Gabrielsson, A., and Lindström, E. (1993). On strong experiences of music. *Musikpsychologie* 10, 118–139.
- Garrido, S., and Schubert, E. (2011). Individual differences in the enjoyment of negative emotion in music: a literature review and experiment. *Music Percept.* 28, 279–296. doi: 10.1525/mp.2011.28.3.279
- Gosling, S. D., Vazire, S., Srivastava, S., and John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *Am. Psychol.* 59, 93–104. doi: 10.1037/0003-066X.59.2.93
- Hanich, J., Wagner, V., Shah, M., Jacobsen, T., and Menninghaus, W. (2014). Why we like to watch sad films. The pleasure of being moved in aesthetic experiences. *Psychol. Aesthet. Creat. Arts* 8, 130–143. doi: 10.1037/a0035690
- Huron, D. (2011). Why is sad music pleasurable? A possible role for prolactin. *Music. Sci.* 15, 146–158. doi: 10.3389/fnhum.2015.00404
- Istók, E., Brattico, E., Jacobsen, T., Krohn, K., Müller, M., and Tervaniemi, M. (2009). Aesthetic responses to music: a questionnaire study. *Music. Sci.* 13, 183–206.
- Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: toward a unified theory of musical emotions. *Phys. Life Rev.* 10, 235–266. doi: 10.1016/j.plrev.2013.05.008
- Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 559–575. doi: 10.1017/S0140525X08005293
- Kallinen, K., and Ravaja, N. (2006). Emotion perceived and emotion felt: same and different. *Music. Sci.* 10, 191–213.
- Kawakami, A., Furukawa, K., Katahira, K., and Okanoya, K. (2013). Sad music induces pleasant emotion. *Front. Psychol.* 4:311. doi: 10.3389/fpsyg.2013.00311
- Kawakami, A., and Katahira, K. (2015). Influence of trait empathy on the emotion evoked by sad music and on the preference for it. *Front. Psychol.* 6:1541. doi: 10.3389/fpsyg.2015.01541
- Konecni, V. J. (2005). The aesthetic trinity: awe, being moved, thrills. *Bull. Psychol. Arts* 5, 27–44.
- Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., and Couper, M. (2004). Psychological research online: Report of Board of Scientific Affairs’ advisory group on the conduct of research on the internet. *Am. Psychol.* 59, 105–117.
- Kuehnast, M., Wagner, V., Wassiliwizky, E., Jacobsen, T., and Menninghaus, W. (2014). Being moved: linguistic representation and conceptual structure. *Front. Psychol.* 5:1242. doi: 10.3389/fpsyg.2014.01242

- Levinson, J. (2006). "Musical expressiveness as hearability-as-expression," in *Contemporary Debates in Aesthetics and the Philosophy of Art*, ed. M. Kieran, M (Oxford: Blackwell Publishing), 192–206.
- Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Kuehnast, M., and Jacobsen, T. (2015). Towards a psychological construct of being moved. *PLoS ONE* 10:e0128451. doi: 10.1371/journal.pone.0128451
- Moran, N. (2014). Social implications arise in embodied music cognition research which can counter musicological "individualism". *Front. Psychol.* 5:676. doi: 10.3389/fpsyg.2014.00676
- Panksepp, J. (1995). The emotional sources of "chills" induced by music. *Music Percept.* 13, 171–207.
- Phillips-Silver, J., and Keller, P. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Front. Hum. Neurosci.* 6:26. doi: 10.3389/fnhum.2012.00026
- Preacher, K. J., and Selig, J. P. (2010). *Monte Carlo Method for Assessing Multilevel Mediation: An Interactive Tool for Creating Confidence Intervals for Indirect Effects in 1-1-1 Multilevel Models*. Available at: <http://quantpsy.org/>
- 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
- 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
- Schubert, E. (1996). Enjoyment of negative emotions in music: an associative network explanation. *Psychol. Music* 24, 18–28.
- Schubert, E. (2013). Emotion felt by the listener and expressed by the music: literature review and theoretical perspectives. *Front. Psychol.* 4:837. doi: 10.3389/fpsyg.2013.00837
- 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
- Vuoskoski, J. K., and Eerola, T. (2012). Can sad music really make you sad? Indirect measures of affective states induced by music and autobiographical memories. *Psychol. Aesthet. Creat. Arts* 6, 204–213. doi: 10.1037/a0026937
- Vuoskoski, J. K., and Eerola, T. (2015). Extramusical information contributes to emotions induced by music. *Psychol. Music* 43, 262–274. doi: 10.1177/0305735613502373
- Vuoskoski, J. K., Thompson, B., McIlwain, D., and Eerola, T. (2012). Who enjoys listening to sad music and why? *Music Percept.* 29, 311–317. doi: 10.1525/mp.2012.29.3.311
- Wassiliwizky, E., Wagner, V., Jacobsen, T., and Menninghaus, W. (2015). Art-elicited chills indicate states of being moved. *Psychol. Aesthet. Creat. Arts* 9, 405–416. doi: 10.1037/aca0000023

**Conflict of Interest Statement:** 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.

Copyright © 2017 Vuoskoski and Eerola. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Neurodynamic Perspective on Musical Enjoyment: The Role of Emotional Granularity

Nathaniel F. Barrett<sup>1\*</sup> and Jay Schulkin<sup>2</sup>

<sup>1</sup> Institute for Culture and Society, University of Navarra, Pamplona, Spain, <sup>2</sup> Department of Neuroscience, Georgetown University, Washington, DC, United States

**Keywords:** music cognition, music and emotion, granularity of emotion, neurodynamical models, musical enjoyment

## INTRODUCTION

Musical enjoyment is a nearly universal experience, and yet from a neurocognitive and evolutionary standpoint it presents a conundrum. Why do we respond so powerfully to something apparently without any survival value? A variety of explanations for the evolution of music cognition have been offered (e.g., Wallin et al., 2000; Morley, 2013), nevertheless most current neurocognitive theories of its specifically affective aspects do not posit any specially adapted emotional circuitry (but see Peretz, 2006). Rather, it is assumed that whatever processes are responsible for pleasure and emotion in general—be they subcortical, cortical, or both—are also responsible for the thrills of music. Accordingly, the problem of musical enjoyment is to explain how and why these processes are engaged so effectively by musical stimuli.

While this is a perfectly sensible approach, a major obstacle lies in its path: the paradox of enjoyable sadness in music (Davies, 1994; Levinson, 1997; Garrido and Schubert, 2011; Huron, 2011; Vuoskoski et al., 2011; Kawakami et al., 2013; Taruffi and Koelsch, 2014; Sachs et al., 2015). Music frequently elicits experiences of negative emotions, especially sadness, which we nevertheless find deeply gratifying. For neurocognitive perspectives, the implication of this phenomenon is that normal processes for generating emotional responses are not sufficient to explain musical enjoyment, as something special to music must allow us to enjoy negative emotion. How is musically induced sadness different from “normal” sadness such that the former can be enjoyed while the latter cannot?

To make progress on such questions, in this article we propose a theory of musical enjoyment based on implications of a *neurodynamic* approach to emotion (Pessoa, 2008; Flaig and Large, 2014), which highlights the role of transient patterns of coordinated neural activity spanning multiple regions of the brain. The key advantage of this perspective is its ability to register the possibility that emotional experiences differ not only in kind (happy vs. sad) but also in “granularity,” complexity, or differentiation (Lindquist and Barrett, 2008). Studies indicate that increased emotional granularity functions as a kind of positivity that can meliorate experiences of negative emotions (Smidt and Sudak, 2015). Accordingly, perhaps we can account for enjoyment of negative emotions in music if we can show that these emotions are more finely differentiated than “normal” negative feelings.

Based on this approach, we also propose a general distinction between *pleasure*, defined here as bursts of positively categorized feeling, and *enjoyment*, defined as sustained flows of finely differentiated feeling regardless of emotional categorization (cf. Frederickson, 2002). For this perspective, the categorical meaning of an emotional experience (e.g., happy or sad) is closely related to but separable from its positive or negative affective tone, at least insofar as this tone is influenced by granularity.

## OPEN ACCESS

### Edited by:

Mireille Besson,  
Durham University, United Kingdom

### Reviewed by:

Mireille Besson,  
Institut de Neurosciences Cognitives  
de la Méditerranée (INCM), France  
Mark Reybrouck,  
KU Leuven, Belgium  
Andrea Schiavio,  
University of Graz, Austria

### \*Correspondence:

Nathaniel F. Barrett  
nbarrett@unav.es

### Specialty section:

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

**Received:** 19 October 2016

**Accepted:** 30 November 2017

**Published:** 13 December 2017

### Citation:

Barrett NF and Schulkin J (2017) A  
Neurodynamic Perspective on Musical  
Enjoyment: The Role of Emotional  
Granularity. *Front. Psychol.* 8:2187.  
doi: 10.3389/fpsyg.2017.02187

## NEURODYNAMICS, EMOTION, AND MUSIC

Here “neurodynamic approach” refers to a family of neurocognitive theories that regard transient, large-scale patterns of rhythmically coordinated neural activity as the main vehicles of cognition/emotion (e.g., Freeman, 1997; Bressler and Kelso, 2001, 2016; Varela et al., 2001; Cosmelli et al., 2007; Breakspear and McIntosh, 2011; Sporns, 2011). An especially pertinent feature of this approach is its divergence from traditional notions of functional specialization and localization (Pessoa, 2008, 2014). Broadly speaking, insofar as neurodynamic theories understand cognitive functions as supported by transient, task-specific coalitions rather than stable processing pathways, they tend to affirm the multifunctionality of neural structures at multiple scales (Anderson et al., 2013; see also Hagoort, 2014; Friederici and Singer, 2015). According to this view, which contrasts with the modular approach commonly adopted by computational theories of neural function, the precise functional role of any given neural structure changes according to the context of coordinated neural activity (McIntosh, 2000, 2004; Bressler and McIntosh, 2007). On the other hand, this approach does not herald a return of “equipotentiality,” as it allows for characterizations of the functional “dispositions” of neural structures (Anderson, 2014).

Similarly, with regard to affect and emotion, the neurodynamic perspective can register the importance of distinct structures—e.g., subcortical structures and hedonic “hotspots” (Berridge and Kringelbach, 2015). But it holds that the full range of emotional experience must be understood in terms of continually evolving patterns of globally coordinated neural activity. Thus, while localized structures may have consistent roles in the production of emotional responses, they do not by themselves constitute emotion nor can they be said to govern emotion in any simple way (Flaig and Large, 2014). The point is not just that emotion is the product of the continual interplay of cortical and subcortical dynamics (Panksepp, 2012). Rather the key implication of neurodynamics is that this interplay is constituted by transient patterns of coordinated activity whose dynamic features—especially complexity and continuity—are relevant to the categorization of emotion and affect (cf. Spivey, 2007 on dynamical categorization). Here, we are mainly interested in the possibility that neurodynamic categorizations of emotional content can range in complexity, corresponding to differences of emotional granularity in experience (analogous to the difference between simple and rich color palettes).

It should be noted that this approach encompasses both categorical theories of “basic emotions” (e.g., Panksepp, 1998, 2007) and dimensional theories of “core affect” (Russell, 2003; Barrett et al., 2007). What is essential for present purposes is the way in which the dynamic interplay between subcortical responses and the complex cortical elaboration of emotion (Reybrouck and Eerola, 2017) gives rise to both categorical distinctions (happy vs. sad) and differences of granularity (fine vs. coarse).

Neurodynamic approaches are well-established in the field of music cognition (for reviews see Large, 2010; Flaig and Large, 2014). Among the advantages of a neurodynamic approach is its capacity to register the relationship between bodily movement and music (e.g., Large et al., 2015). This relationship has been indicated by numerous studies of sensorimotor involvement in music perception (e.g., Chen et al., 2008) and must be taken into account by any theory of musical experience, as we briefly indicate below.

However, few attempts have been made to understand musical enjoyment from a neurodynamic perspective (see Chapin et al., 2010; Flaig and Large, 2014). A notable exception is William Benzon’s groundbreaking treatise (Benzon, 2001), which anticipates the perspective offered here. One reason for this neglect is the challenge of empirical verification: like neurodynamic theories of consciousness (Seth et al., 2006), neurodynamic theories of musical experience are in need of high temporal-resolution data (e.g., from EEG or MEG) that show how relevant characteristics of neural dynamics change during musical experience (see Garrett et al., 2013). For the current proposal, the key challenge is to find and measure just those variations that correspond to differences of emotional granularity or complexity. Until such methods are developed, studies of emotional differentiation in musical experience must turn to the refinement of self-reporting methods (Juslin and Sloboda, 2010).

## THE PARADOX OF ENJOYABLE SADNESS IN MUSIC

In this section, we consider the challenge posed by deeply gratifying experiences of musically induced sadness (Sachs et al., 2015). Sadness is being used here as representative case, as it seems that other negative emotions—despair, terror, dread—can also be induced and enjoyed through music (Gabrielsson, 2011). Also, we do not mean to claim that music is the only source of enjoyable sadness. Rather, our purpose is to use the case of enjoyable sadness in music to set up a distinction between *pleasure*, defined as bursts of positively categorized feeling, and *enjoyment*, defined as any sustained flow of high-dimensional feeling, regardless of emotional categorization.

First, let us briefly consider other ways of handling the paradox of enjoyable sadness in music. Some have suggested that music can induce both negative and positive affect at once (Larsen and Stastny, 2011). Others have theorized that sadness might be *perceived* in music but is not actually *felt* (Kivy, 1990; Garrido and Schubert, 2011; Kawakami et al., 2013). A third possibility is that sadness is not enjoyed but is compensated for by other positive emotions or by the positive value of the overall experience (Davies, 1994; Huron, 2011). How to decide among these theories?

Part of the problem is that the phenomenological questions raised by enjoyable sadness are subtle and difficult to settle conclusively. Data from self-reporting shows a variety of responses to sad music, including complex emotions that are difficult to put into words (Taruffi and Koelsch, 2014) as

well as instances in which sadness is perceived but not felt. Nevertheless, there is ample evidence that music is capable of inducing powerful emotions (Gabrielsson, 2011), including sadness (Vuoskoski and Eerola, 2012), and there is little evidence in support of the idea that musically induced emotions are “less real” than normal emotions (Scherer, 2004). From a physiological standpoint they seem to be identical, evoking the same autonomic responses—chills, elevated heart rate, etc. (Hodges, 2010).

Indeed, musically induced emotions can sometimes feel *more* real insofar as they are more precisely specified than normal emotions. Felix Mendelssohn famously observed that our experience of emotion in music is “too precise for words,” suggesting that music is used not only to induce but also to expand and enrich emotion (Krueger, 2014). In light of this possibility, we believe that the simplest explanation for the popularity of sad music (e.g., Adele’s “Someone Like You”) is that people are drawn to the enjoyment of musically enriched negative emotion for its own sake.

While far from exhaustive, we hope this discussion suffices to indicate that the paradox of enjoyable sadness in music is not yet resolved. The emotion of sadness in musical experience can be vividly real in both physiological and subjective senses and yet also thrilling in a way that normal sadness is not. Moreover, the question of enjoyable sadness is only partly explained by musical features that commonly evoke sadness (Guhn et al., 2007) or by evidence that strong experiences of musical emotion are accompanied by the release of dopamine (Salimpoor et al., 2011), as neither explains how an experience can feel sad and enjoyable at once.

## PLEASURE AND ENJOYMENT

Philosophers and psychologists have long distinguished between sensory pleasures and more fulfilling experiences of enjoyment and happiness (see Berridge and Kringelbach, 2011; Katz, 2016). For example, a distinction between pleasure and enjoyment is widely affirmed in positive psychology (Csikszentmihalyi, 1990; Frederickson, 2002). However, to our knowledge this distinction has not been verified experimentally.

We believe that a neurodynamic approach can help to refine this distinction and to develop testable hypotheses concerning its neural basis. For instance, based on the neurodynamic approach sketched above, it can be theorized that pleasure pertains to positively categorized feelings that are constituted by the momentary, stereotypical effects of subcortical responses on cortical dynamics. Meanwhile enjoyment might be associated with temporally extended patterns of cortical dynamics that are marked by sustained high dimensionality and that can vary independently of subcortical input. This hypothesis is consistent with studies that suggest that rich sensorimotor engagement can give rise to enjoyment without stimulating any of the drives or appetites normally associated with pleasure (Nakamura and Csikszentmihalyi, 2002), but it requires more elaboration in both phenomenological and neurological terms. In short, what is needed is a detailed analysis of interrelated but distinct aspects

of positive affect that are frequently lumped together as “positive emotion” (Gruber and Moskowitz, 2014), as granularity seems to be an affective component that is separable from categorical meaning.

Even if we grant this distinction between enjoyment and pleasure, it remains to be demonstrated that emotional granularity can explain enjoyable sadness. The plausibility of our theory is suggested, however, by evidence that finely differentiated negative emotions are experienced as less “unpleasant” (Barrett et al., 2001; Kashdan et al., 2015; Smidt and Sudak, 2015). Although this evidence pertains only to verbal discriminations of emotions, it supports our suggestion that differentiation alters the experience of negative emotion. If Mendelssohn was right about the “preciseness” of musical emotion, then perhaps music can render negative emotions not just endurable but enjoyable. This possibility is supported by evidence that elicitation of a “multifaceted emotional experience” is correlated with the enjoyment of sad music (Taruffi and Koelsch, 2014).

## PLEASURE AND ENJOYMENT IN MUSIC

According to our theory it should be possible to discriminate musical pleasure from enjoyment, although the two are typically mixed together. There are a number of ways for music to give pleasure in the narrow sense; in fact much of what is studied under the rubric of musically induced emotion fits into this category: soothing textures and harmonies, rhythmic coherence and “groove,” and expressive contours or gestures (e.g., Juslin and Västfjäll, 2008). There are also diverse means for the production of musical *displeasure*—dissonance, noise, rhythmic incoherence, etc. However, because the reception of musical feelings depends on how they are embedded within the overall musical experience, reports of isolated feelings of musical displeasure may be rare. For instance, dissonance is often ingredient in enjoyable music and where it is reported as unpleasant it is usually part of a thoroughly unenjoyable experience of “bad” music (Gabrielsson, 2011).

Musical enjoyment as defined here has been theorized elsewhere (Benzon, 2001) but awaits the formulation of a more precise and testable model of its underlying dynamics. We suggest that resources for constructing such a model are emerging from studies of the musical entrainment—i.e., rhythmic synchronization—of sensorimotor dynamics (e.g., Clayton et al., 2005; Janata et al., 2012; Merchant et al., 2015) and the relationship between music perception and movement (e.g., Maes et al., 2014). Especially pertinent are arguments from ecological psychology that the phenomenon of musical motion—the experience that someone or something is moving in music—is not a metaphorical mapping but rather a direct perceptual experience of various kinds of “virtual motion” (gestures, movement within an environment, etc.) specified by dynamic features of the music itself (Clarke, 2001, 2005; Bharucha et al., 2006; Eitan and Granot, 2006). Together, these viewpoints suggest that musical stimuli can be coupled with sensorimotor processes of the brain in a manner that (1) drives widespread rhythmic coordination of neural activity and (2)

overlaps with perceptual experiences of motion in a highly structured environment.

For the present thesis, the key implication is that feelings of negative emotion, when induced by music, are embedded within richly structured experiences of motion which serve to “perceptualize” the experience of emotion. Thus, for instance, a sigh-like phrasing is not just an icon of emotion but also a directly perceived manifestation of emotion. Such manifestations can be categorized in multiple ways—to say that emotions are “perceptualized” does not mean that they are simply “read off” the music. But however they are categorized, musical emotions are experienced through movement and are therefore more concretely formed than “normal” emotion. Musical motion, therefore, is what gives musical emotion its high granularity. What makes negative emotions in music enjoyable is the special way in which music moves us in a quite literal way, as indicated by the close relationship between music and dance (Schulkin, 2013; Fitch, 2016).

In short, what distinguishes the neurodynamic approach is its promise to explain how emotions are co-constituted and enriched by the perceptual experience of music (Krueger, 2014), whereas other approaches usually aim only to explain how emotions are *triggered* by music. But emotions are not just triggered by music; they are vividly rendered in animate and highly “granular” form by the rhythmic entrainment of experience.

## PROSPECTS FOR FURTHER RESEARCH

The central claim of our proposal is that musical enjoyment of negative emotion is distinguished by high emotional granularity in comparison with “normal” (i.e., unenjoyable) experiences of negative emotion. We see two possible ways to test this claim.

One way is to refine methods of self-reporting (Juslin and Sloboda, 2010) in an attempt to gather data about the emotional granularity of musical experiences. It should be emphasized that most published studies of emotional granularity have aimed only to measure subjects’ verbal capacity to discriminate emotions (Smidt and Sudak, 2015); this capacity is, at best, an indirect measure of the actual emotional granularity of experience (see Lindquist and Barrett, 2008). Here, we are interested only in *experienced* emotional granularity; moreover we are interested

in how this granularity *varies* in response to music. Self-reporting is notoriously unreliable, and continuous variations of granularity are likely to be even harder to report than categorical distinctions (e.g., happy/sad). Nevertheless, while acknowledging these limitations, we believe that self-reporting methods could be used to test our theory—as suggested by at least one study (Taruffi and Koelsch, 2014).

A second possible test would be to measure variations of emotional granularity using high temporal-resolution techniques such as EEG and MEG. In recent years, neuroscientists who adopt a neurodynamic approach have begun using these techniques to measure “moment-to-moment brain signal variability” associated with cognitive functioning (Heisz et al., 2012; Garrett et al., 2013; Miskovic et al., 2016) and at least one study has attempted to measure correlations between individual emotional granularity and brain activity during “affective processing” (Lee et al., 2017). To test our proposal what is needed is a way to test variations of granularity within the same individual in response to music, and for this it is necessary to develop a measure of emotional complexity (cf. Seth et al., 2006). If such a measure could be developed, our expectation is that experiences of musically induced negative emotion would be found to correlate with higher levels of complexity than “normally” induced (e.g., IAPS-induced) negative emotion.

At present, however, there is no evidence that directly supports our theory. Even so, we believe that the idea of emotional granularity as a factor in musical enjoyment is plausible and worthy of further investigation. Moreover, it should be noted that the thesis presented here is not just about musical enjoyment: it is also about the role of granularity in enjoyment, pleasure, and emotional experience in general. As such, its various phenomenological and neurological implications need to be further articulated and tested against a very wide array of empirical evidence. The key insight of this perspective, which fits well with a neurodynamic approach, is that emotional experiences can vary in dimensionality (coarse vs. fine) as well as categorical meaning (happy vs. sad).

## AUTHOR CONTRIBUTIONS

NB and JS made contributions to the drafting and revision of this work at all stages and share responsibility for its content.

## REFERENCES

- Anderson, M. L. (2014). *After Phrenology: Neural Reuse and the Interactive Brain*. Cambridge, MA: MIT Press.
- Anderson, M. L., Kinnison, J., and Pessoa, L. (2013). Describing functional diversity of brain regions and brain networks. *Neuroimage* 73, 50–58. doi: 10.1016/j.neuroimage.2013.01.071
- Barrett, L. F., Gross, J., Christensen, T. C., and Benvenuto, M. (2001). Knowing what you’re feeling and knowing what to do about it: mapping the relation between emotion differentiation and emotion regulation. *Cogn. Emot.* 15, 713–724. doi: 10.1080/02699930143000239
- Barrett, L. F., Mesquita, B., Ochsner, K. N., and Gross, J. J. (2007). The experience of emotion. *Annu. Rev. Psychol.* 58, 373–403. doi: 10.1146/annurev.psych.58.110405.085709
- Benzon, W. (2001). *Beethoven’s Anvil: Music in Mind and Culture*. New York, NY: Basic Books.
- Berridge, K., and Kringelbach, M. L. (2011). Building a neuroscience of pleasure and well-being. *Psychol. Well Being* 1:3. doi: 10.1186/2211-1522-1-3
- Berridge, K., and Kringelbach, M. L. (2015). Pleasure systems in the brain. *Neuron* 86, 646–664. doi: 10.1016/j.neuron.2015.02.018
- Bharucha, J. J., Curtis, M., and Paroo, K. (2006). Varieties of musical experience. *Cognition* 100, 131–172. doi: 10.1016/j.cognition.2005.11.008
- Breakspear, M., and McIntosh, A. R. (2011). Networks, noise, and models: reconceptualizing the brain as a complex, distributed system. *Neuroimage* 58, 293–295. doi: 10.1016/j.neuroimage.2011.03.056
- Bressler, S. L., and Kelso, J. A. S. (2001). Cortical coordination dynamics and cognition. *Trends Cogn. Sci. (Regul. Ed.)* 5, 26–36. doi: 10.1016/S1364-6613(00)01564-3

- Bressler, S. L., and Kelso, J. A. S. (2016). Coordination dynamics in cognitive neuroscience. *Front. Neurosci.* 10:397. doi: 10.3389/fnins.2016.00397
- Bressler, S. L., and McIntosh, A. R. (2007). "The role of neural context in large-scale neurocognitive network operations," in *Handbook of Brain Connectivity*, eds V. K. Jirsa and A. R. McIntosh (Berlin: Springer), 403–419.
- Chapin, H., Jantzen, K. J., Kelso, J. A. S., Steinberg, F., and Large, E. W. (2010). Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLoS ONE* 5:e13812. doi: 10.1371/journal.pone.0013812
- Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cereb. Cortex* 18, 2844–2854. doi: 10.1093/cercor/bhn042
- Clarke, E. (2001). Meaning and the specification of motion in music. *Music. Sci.* 5, 213–234. doi: 10.1177/102986490100500205
- Clarke, E. (2005). *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195151947.001.0001
- Clayton, M., Sager, R., and Will, U. (2005). In time with the music: the concept of entrainment and its significance for ethnomusicology. *Eur. Meet. Ethnomusicol.* 11, 3–75.
- Cosmelli, D., Lachaux, J.-P., and Thompson, E. (2007). "Neurodynamical approaches to consciousness," in *The Cambridge Handbook of Consciousness*, eds P. D. Zelazo, M. Moscovitch, and E. Thompson (Cambridge: Cambridge University Press), 731–752.
- Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York, NY: Harper and Row.
- Davies, S. (1994). *Musical Meaning and Expression*. Ithaca, NY: Cornell University Press.
- Eitan, Z., and Granot, R. Y. (2006). How music moves. *Music Percept.* 23, 221–248. doi: 10.1525/mp.2006.23.3.221
- Fitch, W. T. (2016). Dance, music, meter and groove: a forgotten partnership. *Front. Hum. Neurosci.* 10:64. doi: 10.3389/fnhum.2016.00064
- Flaig, N. K., and Large, E. W. (2014). Dynamical musical communication of core affect. *Front. Psychol.* 5:72. doi: 10.3389/fpsyg.2014.00072
- Frederickson, B. L. (2002). "Positive emotions," in *Handbook of Positive Psychology*, eds C. R. Snyder and S. J. Lopez (Oxford: Oxford University Press), 120–134.
- Freeman, W. J. (1997). Nonlinear neurodynamics and intentionality. *J. Mind Behav.* 18, 291–304.
- Friederici, A. D., and Singer, W. (2015). Grounding language processing on basic neurophysiological principles. *Trends Cogn. Sci. (Regul. Ed.)* 19, 329–338. doi: 10.1016/j.tics.2015.03.012
- Gabrielsson, A. (2011). *Strong Experiences with Music* [Transl. by R. Bradbury]. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199695225.001.0001
- Garrett, D. D., Samanez-Larkin, G. R., MacDonald, S. W. S., Lindenberger, U., McIntosh, A. R., and Grady, C. L. (2013). Moment-to-moment brain signal variability: a next frontier in human brain mapping? *Neurosci. Biobehav. Rev.* 37, 610–624. doi: 10.1016/j.neubiorev.2013.02.015
- Garrido, S., and Schubert, E. (2011). Negative emotion in music: what is the attraction? A qualitative study. *Empiric. Musicol. Rev.* 6, 214–230. doi: 10.18061/1811/52950
- Gruber, J., and Moskowitz, J. T. (eds.). (2014). *Positive Emotion: Integrating the Light Sides and Dark Sides*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199926725.001.0001
- Guhn, M., Hamm, A., and Zentner, M. (2007). Physiological and musico-acoustic correlates of the chill response. *Music Perception* 24, 473–484. doi: 10.1525/mp.2007.24.5.473
- Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Curr. Opin. Neurobiol.* 28, 136–141. doi: 10.1016/j.conb.2014.07.013
- Heisz, J. J., Shedden, J. M., and McIntosh, A. R. (2012). Relating brain signal variability to knowledge representation. *Neuroimage* 63, 1384–1392. doi: 10.1016/j.neuroimage.2012.08.018
- Hodges, D. A. (2010). "Psychophysiological measures," in *Handbook of Music and Emotion: Theory, Research, Applications*, eds P. N. Juslin and J. A. Sloboda (Oxford: Oxford University Press), 279–311.
- Huron, D. (2011). Why is sad music pleasurable? A possible role for prolactin. *Music. Sci.* 15, 146–158. doi: 10.1177/1029864911401171
- Janata, P., Tomic, S. T., and Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *J. Exp. Psychol. Gen.* 141, 54–75. doi: 10.1037/a0024208
- Juslin, P. N., and Sloboda, J. A. (eds.). (2010). *Handbook of Music and Emotion: Theory, Research, Applications*. Oxford: Oxford University Press.
- Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 559–621. doi: 10.1017/S0140525X08005293
- Kashdan, T. B., Barrett, L. F., and McKnight, P. E. (2015). Unpacking emotion differentiation: transforming unpleasant experience by perceiving distinctions in negativity. *Curr. Dir. Psychol. Sci.* 24, 10–16. doi: 10.1177/0963721414550708
- Katz, L. D. (2016). "Pleasure," in *The Stanford Encyclopedia of Philosophy*, ed E. N. Zalta. Available online at: <https://plato.stanford.edu/archives/win2016/entries/pleasure/>
- Kawakami, A., Furukawa, K., Katahira, K., and Okanoya, K. (2013). Sad music induces pleasant emotion. *Front. Psychol.* 4:311. doi: 10.3389/fpsyg.2013.00311
- Kivy, P. (1990). *Music Alone: Philosophical Reflections on the Purely Musical Experience*. Ithaca, NY: Cornell University Press.
- Krueger, J. (2014). Affordances and the musically extended mind. *Front. Psychol.* 4:1003. doi: 10.3389/fpsyg.2013.01003
- Large, E. W. (2010). "Neurodynamics of music," in *Music Perception*, ed M. Reiss Jones, R. R. Fay, and A. N. Popper (New York, NY: Springer), 201–231. doi: 10.1007/978-1-4419-6114-3\_7
- 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
- Larsen, J. T., and Stastny, B. J. (2011). It's a bittersweet symphony: simultaneously mixed emotional responses to music with conflicting cues. *Emotion* 11, 1469–1473. doi: 10.1037/a0024081
- Lee, J. Y., Lindquist, K. A., and Nam, C. S. (2017). Emotional granularity effects on event-related brain potentials during affective picture processing. *Front. Hum. Neurosci.* 11:133. doi: 10.3389/fnhum.2017.00133
- Levinson, J. (1997). "Music and negative emotion," in *Music and Meaning*, ed. J. Robinson (Ithaca, NY: Cornell University Press), 215–241.
- Lindquist, K. A., and Barrett, L. F. (2008). "Emotional complexity," in *Handbook of Emotions, 3rd Edn.*, eds M. Lewis, J. M. Haviland-Jones, and L. F. Barrett (New York, NY: Guilford Press), 513–530.
- Maes, P.-J., Leman, M., Palmer, C., and Wanderley, M. M. (2014). Action-based effects on music perception. *Front. Psychol.* 4:1008. doi: 10.3389/fpsyg.2013.01008
- McIntosh, A. R. (2000). Towards a network theory of cognition. *Neural Networks* 13, 861–870. doi: 10.1016/S0893-6080(00)00059-9
- McIntosh, A. R. (2004). Contexts and catalysts: a resolution of the localization and integration of function in the brain. *Neuroinformatics* 4, 175–182. doi: 10.1385/NI:2:2:175
- Merchant, H., Grah, J., Trainor, L., Rohrmeier, M., and Fitch, W. T. (2015). Finding the beat: a neural perspective across humans and non-human primates. *Philos. Trans. R. Soc. B* 370:20140093. doi: 10.1098/rstb.2014.0093
- Miskovic, V., Owens, M., Kuntzleman, K., and Gibb, B. E. (2016). Charting moment-to-moment brain signal variability from early to late childhood. *Cortex* 83, 51–61. doi: 10.1016/j.cortex.2016.07.006
- Morley, I. (2013). *The Prehistory of Music: Human Evolution, Archaeology and the Origins of Musicality*. Oxford: Oxford University Press. doi: 10.1093/acprof:osobl/9780199234080.001.0001
- Nakamura, J., and Csikszentmihalyi, M. (2002). "The concept of flow," in *Handbook of Positive Psychology*, eds C. R. Snyder and S. J. Lopez (Oxford: Oxford University Press), 89–105.
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford University Press.
- Panksepp, J. (2007). Neurologizing the psychology of affects: how appraisal-based constructivism and basic emotion theory can coexist. *Perspect. Psychol. Sci.* 2, 281–296. doi: 10.1111/j.1745-6916.2007.00045.x
- Panksepp, J. (2012). What is an emotional feeling? Lessons about affective origins from cross-species neuroscience. *Motiv. Emot.* 36, 4–15. doi: 10.1007/s11031-011-9232-y
- Peretz, I. (2006). The nature of music from a biological perspective. *Cognition* 100, 1–32. doi: 10.1016/j.cognition.2005.11.004

- Pessoa, L. (2008). On the relationship between emotion and cognition. *Nat. Neurosci. Rev.* 9, 148–158. doi: 10.1038/nrn2317
- Pessoa, L. (2014). Understanding brain networks and brain organization. *Phys. Life Rev.* 11, 400–435. doi: 10.1016/j.plrev.2014.03.005
- Reybrouck, M., and Eerola, T. (2017). Music and its inductive power: a psychobiological and evolutionary approach to musical emotions. *Front. Psychol.* 8:494. doi: 10.3389/fpsyg.2017.00494
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev.* 100, 145–172. doi: 10.1037/0033-295X.110.1.145
- Sachs, M. E., Damasio, A., and Habibi, A. (2015). The pleasures of sad music: a systematic review. *Front. Hum. Neurosci.* 9:404. doi: 10.3389/fnhum.2015.00404
- 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, 257–264. doi: 10.1038/nn.2726
- Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *J. New Music Res.* 33, 239–251. doi: 10.1080/0929821042000317822
- Schulkin, J. (2013). *Reflections on the Musical Mind: An Evolutionary Perspective*. Princeton, NJ: Princeton University Press. doi: 10.1515/9781400849031
- Seth, A. K., Izhikevich, E., Reeke, G. N., and Edelman, G. M. (2006). Theories and measures of consciousness: an extended framework. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10799–10804. doi: 10.1073/pnas.0604347103
- Smidt, K. E., and Sudak, M. K. (2015). A brief, but nuanced, review of emotional granularity and emotion differentiation research. *Curr. Opin. Psychol.* 3, 48–51. doi: 10.1016/j.copsyc.2015.02.007
- Spivey, M. J. (2007). *The Continuity of Mind*. Oxford: Oxford University Press.
- Sporns, O. (2011). *Networks of the Brain*. Cambridge, MA: MIT Press.
- 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
- Varela, F., Lachaux, J.-P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Neurosci. Rev.* 2, 229–239. doi: 10.1038/35067550
- Vuoskoski, J. K., and Eerola, T. (2012). Can sad music really make you sad? Indirect measures of affective states induced by music and autobiographical memories. *Psychol. Aesthet. Creat. Arts* 6, 204–213. doi: 10.1037/a0026937
- Vuoskoski, J. K., Thompson, W. F., McIlwain, D., and Eerola, T. (2011). Who enjoys listening to sad music and why? *Music Percept.* 29, 311–318. doi: 10.1525/mp.2012.29.3.311
- Wallin, N. L., Merker, B., and Brown, S. (eds.). (2000). *The Origins of Music*. Cambridge, MA: MIT Press.

**Conflict of Interest Statement:** 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.

Copyright © 2017 Barrett and Schulkin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Musicians Are Better than Non-musicians in Frequency Change Detection: Behavioral and Electrophysiological Evidence

Chun Liang<sup>1</sup>, Brian Earl<sup>1</sup>, Ivy Thompson<sup>1</sup>, Kayla Whitaker<sup>1</sup>, Steven Cahn<sup>2</sup>, Jing Xiang<sup>3</sup>, Qian-Jie Fu<sup>4</sup> and Fawen Zhang<sup>1\*</sup>

<sup>1</sup> Department of Communication Sciences and Disorders, University of Cincinnati, Cincinnati, OH, USA, <sup>2</sup> Department of Composition, Musicology, and Theory, College-Conservatory of Music, University of Cincinnati, Cincinnati, OH, USA,

<sup>3</sup> Department of Pediatrics and Neurology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA, <sup>4</sup> Department of Head and Neck Surgery, University of California, Los Angeles, Los Angeles, CA, USA

## OPEN ACCESS

### Edited by:

Tuomas Eerola,  
University of Durham, UK

### Reviewed by:

Jed A. Meltzer,  
Baycrest Hospital, Canada  
Lan Shuai,  
Haskins Laboratories, USA

### \*Correspondence:

Fawen Zhang  
fawen.zhang@uc.edu

### Specialty section:

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

**Received:** 15 April 2016

**Accepted:** 27 September 2016

**Published:** 25 October 2016

### Citation:

Liang C, Earl B, Thompson I,  
Whitaker K, Cahn S, Xiang J, Fu Q-J  
and Zhang F (2016) Musicians Are  
Better than Non-musicians in  
Frequency Change Detection:  
Behavioral and Electrophysiological  
Evidence. *Front. Neurosci.* 10:464.  
doi: 10.3389/fnins.2016.00464

**Objective:** The objectives of this study were: (1) to determine if musicians have a better ability to detect frequency changes under quiet and noisy conditions; (2) to use the acoustic change complex (ACC), a type of electroencephalographic (EEG) response, to understand the neural substrates of musician vs. non-musician difference in frequency change detection abilities.

**Methods:** Twenty-four young normal hearing listeners (12 musicians and 12 non-musicians) participated. All participants underwent psychoacoustic frequency detection tests with three types of stimuli: tones (base frequency at 160 Hz) containing frequency changes (Stim 1), tones containing frequency changes masked by low-level noise (Stim 2), and tones containing frequency changes masked by high-level noise (Stim 3). The EEG data were recorded using tones (base frequency at 160 and 1200 Hz, respectively) containing different magnitudes of frequency changes (0, 5, and 50% changes, respectively). The late-latency evoked potential evoked by the onset of the tones (onset LAEP or N1-P2 complex) and that evoked by the frequency change contained in the tone (the acoustic change complex or ACC or N1'-P2' complex) were analyzed.

**Results:** Musicians significantly outperformed non-musicians in all stimulus conditions. The ACC and onset LAEP showed similarities and differences. Increasing the magnitude of frequency change resulted in increased ACC amplitudes. ACC measures were found to be significantly different between musicians (larger P2' amplitude) and non-musicians for the base frequency of 160 Hz but not 1200 Hz. Although the peak amplitude in the onset LAEP appeared to be larger and latency shorter in musicians than in non-musicians, the difference did not reach statistical significance. The amplitude of the onset LAEP is significantly correlated with that of the ACC for the base frequency of 160 Hz.

**Conclusion:** The present study demonstrated that musicians do perform better than non-musicians in detecting frequency changes in quiet and noisy conditions.

The ACC and onset LAEP may involve different but overlapping neural mechanisms.

**Significance:** This is the first study using the ACC to examine music-training effects. The ACC measures provide an objective tool for documenting musical training effects on frequency detection.

**Keywords:** frequency change detection, auditory evoked potentials, acoustic change complex, electrophysiology, cortex

## INTRODUCTION

Frequency information is important for speech and music perception. The fundamental frequency (F0) is the lowest frequency of a periodic sound waveform. The F0 plays a critical role in conveying linguistic and non-linguistic information that is important for perceiving music and tone languages, differentiating vocal emotions, identifying a talker's gender, and extracting speech signals from background noise or competing talkers. Unfortunately, for hearing impaired listeners such as cochlear implant (CI) users, these frequency-based tasks are tremendously challenging due to the limitations of current CI technology (Kong and Zeng, 2004; Fu and Nogaki, 2005; Stickney et al., 2007; Zeng et al., 2014).

Considerable evidence has shown that hearing-impaired listeners achieve maximal benefit from brain plasticity as a result of auditory training, such that the auditory system can be more sensitive to the poorer neural representation of acoustic information at the peripheral auditory system (Fu et al., 2004; Galvin et al., 2009; Lo et al., 2015). The potential benefit of auditory training with music stimuli has drawn increasing attention from researchers in recent years (Looi et al., 2012; Gfeller et al., 2015; Hutter et al., 2015) because of the following reasons: (1) music training, such as the training experienced by musicians, may positively enhance speech perception due to the physical features (e.g., frequency, rhythm, intensity, and duration) and overlapping neural networks for processing between the two stimuli (Patel, 2003; Besson et al., 2007; Kraus et al., 2009; Parbery-Clark et al., 2009; Petersen et al., 2009; Itoh et al., 2012), (2) music training enhances cognitive functions, which are required for both language and music perception (Strait et al., 2010, 2012; Strait and Kraus, 2011; Kraus, 2012), and (3) some behavioral data showed that the performance in hearing impaired listeners significantly improves with music training (Gfeller et al., 2015; Hutter et al., 2015). These findings suggest that music training may be integrated into cross-cultural nonlinguistic training regimens to alleviate perceptual deficits in frequency-based tasks for hearing impaired patients. Therefore, further understanding of the effects of music training has significant implications in pointing to the direction of auditory rehabilitation for hearing impaired listeners.

Numerous studies have examined music training effect through musician vs. non-musician comparisons, because

musicians' brains serve as excellent models to show brain plasticity as a result of routine music training. In the auditory domain, musicians have superior auditory perceptual skills and their brains can better encode frequency information (Koelsch et al., 1999; Parbery-Clark et al., 2009). Most sounds in our environment including speech and music contain frequency changes or transitions, which are important cues for identifying and differentiating these sounds. Most previous studies examining pitch perception in musicians vs. non-musicians used a frequency discrimination task that focuses on the detection of one frequency that is different from the reference frequency (Tervaniemi et al., 2005; Micheyl et al., 2006; Bidelman et al., 2013) rather than the detection of the frequency change contained in an ongoing stimulus. In such a discrimination task, the auditory system needs to detect the individual sounds at different frequencies, thereby the neural mechanism may involve the detection of the onset of the sounds at different frequencies rather than the frequency change *per-se*. Therefore, the frequency discrimination task may not be optimal for the understanding of how the auditory system responds to frequency changes in a context. A frequency change detection task using the stimuli containing frequency changes may provide better insights about the underlying mechanisms of the auditory system responding to frequency changes in a context.

Auditory evoked potentials recorded using EEG techniques have been used to understand the neural substrates of frequency change detection. The late auditory evoked potential (LAEP) is an event-related potential reflecting central processing of the sound. The N1 peak of the LAEP occurs at a latency of ~100 ms and the P2 peak at a latency of 200 ms. The acoustic change complex (ACC) is a type of LAEP evoked by the acoustic change in an ongoing stimulus (Ostroff et al., 1998; Small and Werker, 2012). The ACC can be evoked by the consonant-vowel transition in an ongoing syllable (Ostroff et al., 1998; Friesen and Tremblay, 2006), the change of acoustic feature (e.g., frequency or amplitude, Martin and Boothroyd, 2000; Harris et al., 2007; Dimitrijevic et al., 2008), and the change in place of stimulation within the cochlea (i.e., in CI users, Brown et al., 2008). The minimal acoustic change that can evoke the ACC is similar to the threshold for auditory discrimination threshold (Harris et al., 2007; He et al., 2012). The ACC recording does not require participants' active participation and it provides an objective measure of stimulus differentiation capacity that can be used in difficult-to-measure subjects.

The ACC has been recorded reliably in normal hearing (NH) adults, young infants, hearing aid users, and CI users (Friesen and Tremblay, 2006; Tremblay et al., 2006; Martin, 2007; Kim

**Abbreviations:** ACC, Acoustic Change Complex; ANOVA, Analysis of Variance; CI, Cochlear Implant; EEG, Electroencephalography; LAEP, Late-latency Auditory Evoked Potential; NH, Normal-Hearing.

et al., 2009; Small and Werker, 2012). However, the exact nature of the ACC has not been well understood. One unanswered question is: what are the differences between the ACC evoked by acoustic changes and the conventional LAEP evoked by stimulus onset? The previous studies used stimuli to evoke the ACC containing both onset of new stimulus compared to the base stimulus and acoustic changes (Itoh et al., 2012; Small and Werker, 2012) or stimuli containing acoustic changes in more than one dimension, e.g., the change in stimulation electrode in CI users and the change in perceived frequency due to the change in stimulation electrode (Brown et al., 2008) and the changes in spectral envelope, amplitude, and periodicity at the transition in consonant-vowel syllables (Ostroff et al., 1998; Friesen and Tremblay, 2006; Tremblay et al., 2006).

The current study will examine the musician benefit using frequency changes in the simplest tone, with the onset cues removed, for both behavioral tests of frequency change detection and EEG recordings. Through the combination of behavioral and EEG measures, the neural substrates underlying musician benefit in frequency change detection would be better understood. This information is critical for the design of efficient training strategies. The practical outcome of such research study would be that, if the music training effects can be reflected in EEG measures, the EEG measurement can be used for objective evaluation of music training effects. The objectives of this study were: (1) to determine if musicians have better ability to detect frequency changes in quiet and noisy conditions; (2) to use the ACC measure to understand the neural substrates of musicians vs. non-musicians in frequency change detection. To our knowledge, this is the first study using the ACC to examine music training effects in musicians.

## MATERIALS AND METHODS

### Subjects

Twenty-four healthy young NH individuals (13 males and 11 females; age range: 20–30 years) including 12 musicians and 12 non-musicians participated in the study. All participants had audiometric hearing thresholds  $\leq 20$  dB HL at octave test frequencies from 250 to 8000 Hz, normal type A tympanometry, and normal acoustic reflex thresholds at 0.5, 1, and 2 kHz. All participants were right-handed and did not have neurological or hearing-related disorders. The criteria for musicians were: (a) having at least 10 years of continuous training in Western classical music on their principal instruments, (b) having begun music training at or before the age of 7, (c) having received music training within the last 3 years on a regular basis. All of the 12 musicians were students from the College of Conservatory of Music at the University of Cincinnati; the instruments played by these musicians include piano, guitar saxophone, cello, trumpet, horn, and double bass. The criteria for non-musicians were: (a) having no more than 3 years of formal music training on any combination of instruments throughout their lifetime, (b) having no formal music training within the past 5 years. The above criteria for musicians and non-musicians are similar to the criteria used in previous studies (Bidelman et al., 2013; Fuller et al., 2014). All of the 12 non-musicians were college students

with non-music majors. All participants gave informed written consent prior to their participation. This research was approved by the Institutional Review Board of the University of Cincinnati.

### Stimuli

#### Stimuli for Behavioral Tests

The stimuli were tones generated using Audacity 1.2.5 (<http://audacity.sourceforge.net>) at a sample rate of 44.1 kHz. A tone of 1 s duration at a base frequency of 160 Hz, which is in the frequency range of the F0 of the human voice, was used as the standard tone. To avoid an abrupt onset and offset, the amplitude was reduced to zero over 10 ms using the fade in and fade out function. The target tones were the same as the standard tone except that the target tones contained upward frequency changes at 500 ms after the tone onset, with the magnitude of frequency change varying from 0.05 to 65% (the large range was created so that the same stimuli could be used for CI users in a future study). The frequency change occurred for an integer number of cycles of the base frequency and the change occurred at 0 phase (zero crossing). If the number of base frequency cycles was not an integer at 500 ms, the number of cycles was rounded up to an integer number which leads to a slightly delayed point (not exactly at 500 ms) for the start of the frequency change. Therefore, the onset cue was removed and it did not produce audible transients (Dimitrijevic et al., 2008).

The above stimuli were mixed with broad band noise of the same duration to create two more sets of stimuli: tones containing frequency changes masked by low-level noise, and tones containing frequency changes masked by high-level noise. The onset- and offset-amplitude of the broad band noise was reduced to zero over 10 ms, the same as how the tones were treated. For the low-level noise, the root mean square (RMS) amplitude of the noise was 10 dB lower than that of the tone (SNR = 10 dB); for the high-level noise, the RMS amplitude of the noise was the same as that of the tone stimulus (SNR = 0 dB); The amplitudes of all stimuli were normalized. The stimuli were calibrated using a Brüel and Kjær (Investigator 2260) sound level meter set on linear frequency and slow time weighting with a 2 cc coupler.

For convenience, the stimuli for frequency detection tasks were renamed numerically: tone stimuli containing frequency changes (Stim 1), tones containing frequency changes masked by low-level noise (Stim 2), and tones containing frequency changes masked by high-level noise (Stim 3).

#### Stimuli for EEG Recording

Tones of 160 and 1200 Hz with 1 s duration that contained upward frequency changes were used as stimuli for EEG recordings. These two different base frequencies were used for EEG recording for the following reasons. First, while 160 Hz is in the frequency range of the F0 of the human voice, 1200 Hz is in the frequency range of the 2nd formant of vowels. Examining the ACC at these two base frequencies would help understand how the auditory system processes the frequency change near the F0 and the 2nd formant of vowels; second, this would help us better understand the differences of the ACC and the onset LAEP. Specifically, the ACC is evoked by frequency change

from the base frequency. But is the ACC evoked by a frequency change (e.g., a small change from 160 to 168 Hz for a 5% change) predictable using the onset LAEP evoked by the onset of different frequencies (160 vs. 1200 Hz)? The amount of the frequency change was manipulated at 0% (no change), 5, and 50%, respectively. Note, that the stimuli used for EEG recordings were presented in quiet conditions. Therefore, the six stimuli (3 types of frequency changes  $\times$  2 base frequencies) were presented with 200 trials for each, with a randomized order. The inter-stimulus interval was 800 ms.

## Procedure

### Behavioral Tests of Frequency Change Detection

The participants were comfortably seated in a sound-treated booth. Stimuli were delivered in the sound field via a single loudspeaker placed at ear level, 50 cm in front of the participant at the most comfortable level (7 on a 0–10 loudness scale). Such a presentation approach, which has been commonly used in CI users, was used so that the current data can be compared with those from CI users in a future study. The stimuli were presented using APEX (Francart et al., 2008). An adaptive, 2-alternative forced-choice procedure with an up-down stepping rule was employed to measure the minimum frequency change the participant was able to detect. In each trial, a target stimulus and a standard stimulus were included. The standard stimulus was the tone without frequency change and the target stimulus was the tone with a frequency change. The order of standard and target stimulus was randomized and the interval between the stimuli in a trial was 0.5 s. The participant was instructed to choose the target signal by pressing the button on the computer screen and was given a visual feedback regarding the correct response. Each run generated a total of five reversals. The asymptotic amount of frequency change (the average of the last three trials) then became an estimate of the threshold for frequency change detection. Each participant was required to do the frequency change detection task with the three types of stimuli (Stim 1, 2, and 3). The order of the three stimulus type conditions was randomized and counterbalanced across participants.

### EEG Recording

Participants were fitted with a 40-channel Neuroscan quick-cap (NuAmps, Compumedics Neuroscan, Inc., Charlotte, NC). The cap was placed according to the International 10–20 system, with the linked ear as the reference. Electro-ocular activity (EOG) was monitored so that eye movement artifacts could be identified and rejected during the offline analysis. Electrode impedances for the remaining electrodes were kept at or below 5 k $\Omega$ . EEG recordings were collected using the SCAN software (version 4.3, Compumedics Neuroscan, Inc., Charlotte, NC) with a band-pass filter setting from 0.1 to 100 Hz and an analog-to-digital converter (ADC) sampling rate of 1000 Hz. During testing, participants were instructed to avoid excessive eye and body movements. Participants read self-selected magazines to keep alert and were asked to ignore the acoustic stimuli. Participants were periodically given short breaks in order to shift body position and to maximize alertness during the experiment.

## Data Processing

For the behavioral test, the frequency change detection threshold was measured for each of the three stimuli in each participant. For EEG results, continuous EEG data collected from each participant were digitally filtered using a band-pass filter (0.1–30 Hz). Then the data were segmented into epochs over a window of 1500 ms (including a 100 ms pre-stimulus duration). Following segmentation, baseline was corrected by the mean amplitude of the 100 ms pre-stimulus time window and epochs in which voltages exceeded  $\pm 150 \mu\text{V}$  were rejected from further analysis. Then EEG data were averaged separately for each of the six types of stimuli (2 base frequencies  $\times$  3 types of frequency changes) in each participant. Then, MATLAB (Mathworks, Natick, MA) was used to objectively identify peak components, which were confirmed by visual evaluation of the experimenters. Because the LAEP was largest at electrode Cz, we restricted the later analysis to data from Cz.

The onset LAEP response peaks were labeled using standard nomenclature of N1 and P2. The ACC response peaks were labeled using N1' and P2'. The N1 and P2 peaks of the onset LAEP were identified in a latency range 70–180 and 150–250 ms, respectively, after the onset of the tone; The N1' and P2' peaks of the ACC were identified in a latency range 70–180 and 150–250 ms, respectively, after the onset of the frequency change. The measures used for statistical analysis include: N1 and P2 amplitude and latency, N1-P2 peak-to-peak amplitude for the onset LAEP and the corresponding measures for the ACC.

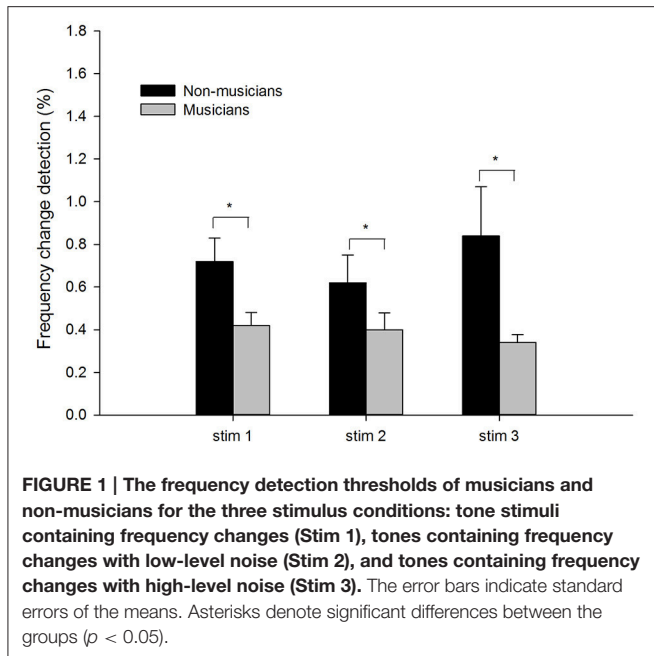
The series of mixed-design repeated analysis of variance (ANOVA) were performed to examine the difference in behavioral and EEG measures between the musician and non-musician groups under different stimulus conditions. Pearson correlation analysis was performed to determine if ACC measures correlate to onset LAEP measures, and if behavioral frequency detection thresholds correlate to ACC measures. A  $p$ -value of 0.05 was used as the significance level for all analyses.

## RESULTS

### Psychoacoustic Performance

Figure 1 shows the means and standard errors of the frequency change detection thresholds in musician and non-musician groups under three different stimulus conditions. The mean frequency thresholds were higher (poorer performance) in non-musicians (Stim 1:  $M = 0.72\%$ ; Stim 2:  $M = 0.62\%$ ; Stim 3:  $M = 0.84\%$ ) than in musicians (Stim 1:  $M = 0.42\%$ ; Stim 2:  $M = 0.40\%$ ; Stim 3:  $M = 0.34\%$ ).

A two-way mixed ANOVA was used to determine the effects of the Subject Group (between-subject factor) and the Stimulus Condition (within-subject factor). There was a significant effects of Subject Group, with musicians showing a lower threshold than non-musicians [ $F_{(1, 21)} = 12.64$ ,  $p < 0.05$ ,  $\eta^2 = 0.38$ ]. There was no significant effect of Stimulus Condition [ $F_{(2, 42)} = 0.63$ ,  $p > 0.05$ ,  $\eta^2 = 0.03$ ] nor a significant interaction between Stimulus Condition and Subject Group [ $F_{(2, 42)} = 0.11$ ,  $p > 0.05$ ,  $\eta^2 = 0.01$ ].



## EEG Results

**Figure 2** shows the mean waveforms between musicians (black traces) and non-musicians (red traces) for 160 Hz (left panel) and 1200 Hz (right panel) with a frequency change of 0% (top), 5% (middle), and 50% (bottom). Two types of LAEP responses were observed: one with a latency of ~100–250 ms after the stimulus onset and the other occurring 100–250 ms after the acoustic change, respectively, with the former being the onset LAEP or N1-P2 complex and the latter being the ACC or N1'-P2' complex.

### Onset LAEP

**Figure 3** shows the onset LAEP measures (N1 latency, P2 latency, and N1-P2 amplitude) for musicians and non-musicians. The error bars indicate standard errors of the means. As shown in the figure, the onset LAEPs appear to be similar for the tones with three frequency changes (top, middle, and bottom) at each base frequency, because they are evoked by the onset of the same tone regardless of the acoustic change inserted in the middle of the tone. This is an indication of the high repeatability of the LAEP. Compared to non-musicians, musicians have shorter latencies for N1 and P2 for base frequency of 160 Hz and shorter N1 latency for base frequency of 1200 Hz as well as a larger N1-P2 amplitude. The onset LAEP peak latencies tend to be shorter and amplitudes greater for base frequency of 1200 Hz than 160 Hz.

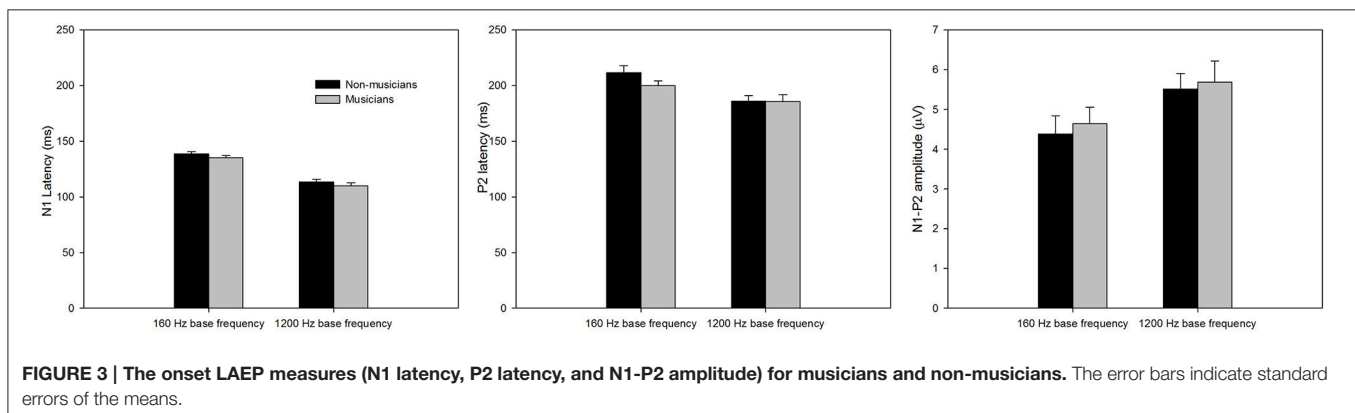
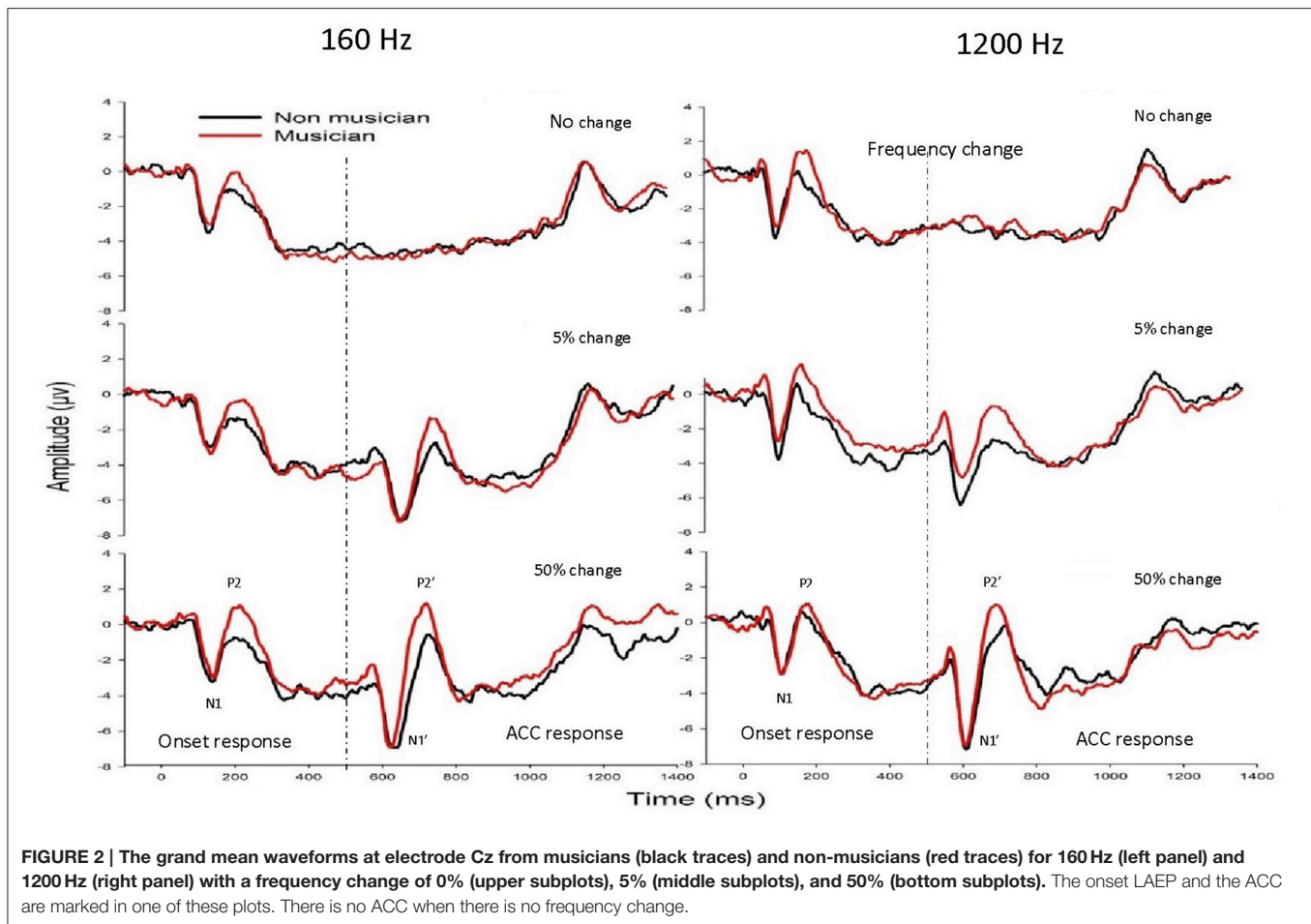
A mixed  $2 \times 2$  repeated ANOVA (Subject Group as the between-subject factor and Base Frequency as the within-subject factor) was performed. The data for different frequency change magnitude (0, 5, and 50%) for each base frequency were averaged since there was no difference in the onset LAEP evoked by these stimulus conditions. Although musicians do show a shorter latency of N1 and P2 and a larger N1-P2 amplitude, the difference did not reach statistical significance ( $p > 0.05$ ). There was a significant main effect of Base Frequency for N1, P2 latency, and

N1-P2 amplitude ( $p < 0.05$ ). The 1200 Hz base frequency evoked an onset LAEP with a shorter N1 latency, shorter P2 latency, and larger N1-P2 amplitude.

### Acoustic Change Complex

The general morphologies of the ACC were similar to those of the onset LAEP, but the amplitude of the ACC appeared to be bigger than the onset LAEP. The ACC occurs only when there is a frequency change in the tone but not when there is no frequency change (**Figure 2**). **Table 1** shows the means and standard deviations of the ACC measures. Musicians have shorter N1' latency, larger P2' amplitude, and larger N1'-P2' amplitude for both base frequencies with both 5% and 50% changes. The ACC amplitudes are bigger and peak latencies shorter for 50% frequency change than for 5% change. The frequency changes at the base frequency 1200 Hz evoked shorter latencies than those at 160 Hz. **Figure 4** shows the ACC measures (N1' amplitude and latency, P2' amplitude and latency, and N1'-P2' amplitude) for musicians and non-musicians. The error bars indicate standard errors of the means.

To explore the effects of Base Frequency (within-subject factor), Frequency Change (within-subject factor), and Subject Group (between-subject factor) on the ACC measures, a  $2 \times 2 \times 2$  mixed-model repeated ANOVA was conducted separately for N1' amplitude and latency, P2' amplitude and latency, and N1'-P2' peak-to-peak amplitude. Statistical significance was found for N1' latency and P2' amplitude. For N1' latency, there was a main effect of Base Frequency [ $F_{(1, 22)} = 116.01, p < 0.01, \eta^2 = 0.84$ ], Frequency Change [ $F_{(1, 22)} = 84.88, p < 0.01, \eta^2 = 0.79$ ] and significant interaction between Base Frequency and Subject Group [ $F_{(1, 22)} = 5.26, p < 0.05, \eta^2 = 0.19$ ]. No statistical significance was found in Subject Group [ $F_{(1, 22)} = 2.84, p > 0.05, \eta^2 = 0.12$ ]. For P2' amplitude, there was a main effect of Base Frequency [ $F_{(1, 22)} = 6.00, p < 0.05, \eta^2 = 0.21$ ], Frequency Change [ $F_{(1, 22)} = 22.61, p < 0.01, \eta^2 = 0.51$ ]. No statistical significance was found in Subject Group [ $F_{(1, 22)} = 3.07, p > 0.05, \eta^2 = 0.12$ ]. Further,  $2 \times 2$  mixed-model repeated ANOVA tests were conducted to examine the effects of Base Frequency and Subject Group on the P2' amplitude and N1' latency separately for 5 and 50% change, respectively. For P2' amplitude, there was a main effect of Base Frequency [ $F_{(1, 22)} = 11.64, p < 0.05, \eta^2 = 0.35$ ] and Subject Group [ $F_{(1, 22)} = 6.86, p < 0.05, \eta^2 = 0.24$ ] for 160 Hz 5% change. No statistical significance was found in P2' amplitude for 160 Hz 50% change ( $p > 0.05$ ). For N1' latency, there was a main effect of Base Frequency [ $F_{(1, 22)} = 43.00, p < 0.05, \eta^2 = 0.66$ ] and Subject Group [ $F_{(1, 22)} = 4.82, p < 0.05, \eta^2 = 0.18$ ] as well as significant interaction between Base Frequency and Subject Group [ $F_{(1, 22)} = 4.73, p < 0.05, \eta^2 = 0.18$ ] for the 160 Hz, 50% change condition. No statistical significance was found in N1' latency for the 160 Hz, 5% change condition. In summary, musicians have shorter N1' latency for 160 Hz with 50% change and larger P2' amplitude for 160 Hz 5% change; ACCs for 1200 Hz have a shorter N1' peak latency and larger P2' amplitude than for 160 Hz. After adjusting the significance level for conducting multiple ANOVAs, the P2' amplitude was significantly greater in musicians than non-musicians for 160 Hz 5% change.



### Comparison between the Onset LAEP and the ACC

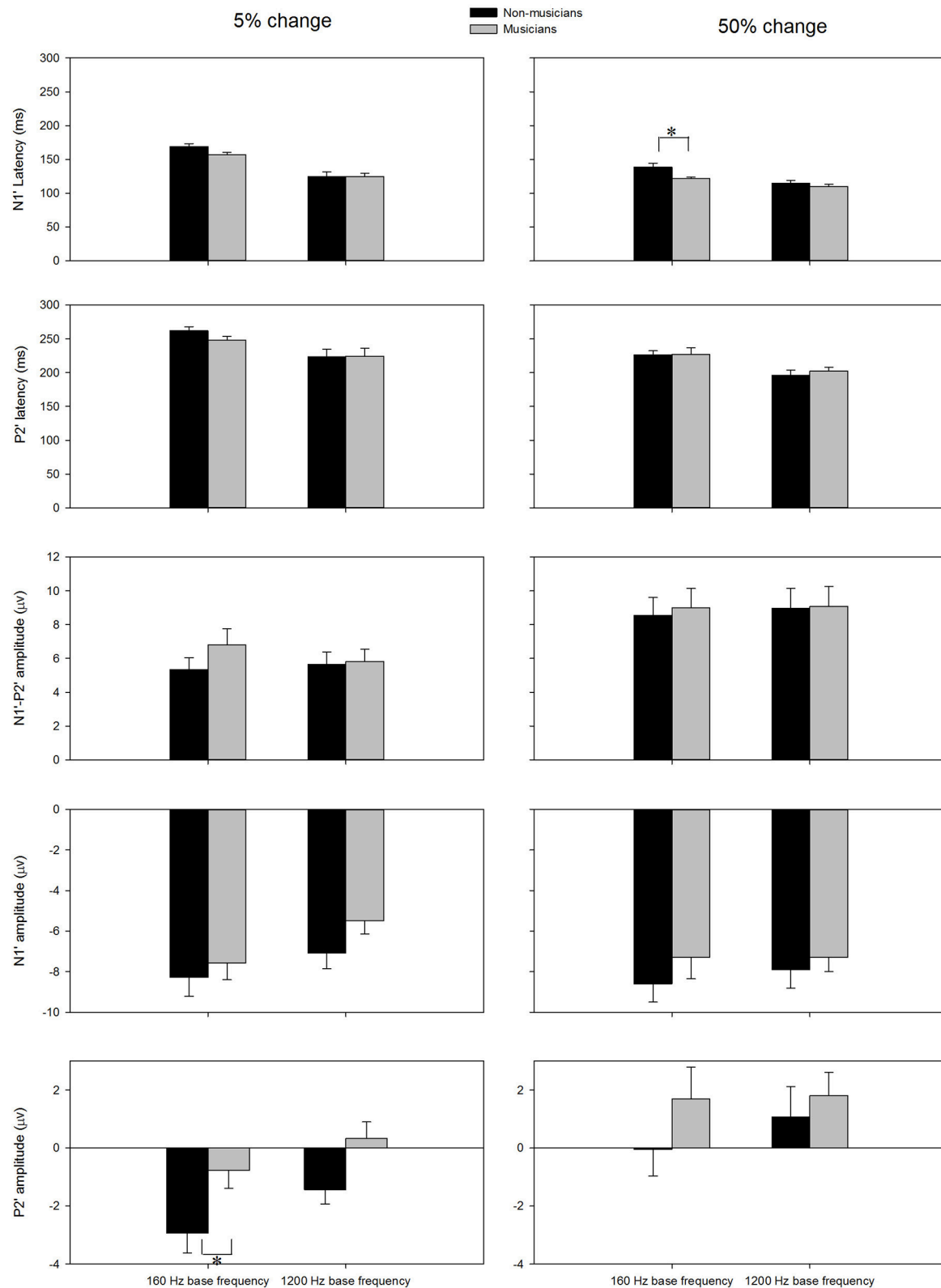
Pearson correlation analyses were performed to determine the correlations between the onset LAEP and the ACC measures. There were significant correlations between the onset LAEP N1-P2 amplitude and the ACC N1'-P2' amplitude for 160 Hz base frequency with 5% ( $r = 0.77$ ,  $p < 0.01$ ) and 50% change ( $r = 0.71$ ,  $p < 0.01$ ). However, there was no such correlation for 1200 Hz base frequency. **Figure 5** shows scatter plots of ACC amplitude vs. onset LAEP amplitude for the 160 Hz base

frequency with 5 and 50% frequency changes, respectively. Data from participants in both musician and non-musician groups were included. This finding indicates that participants who display a larger onset LAEP tend to display a larger ACC for the 160 Hz.

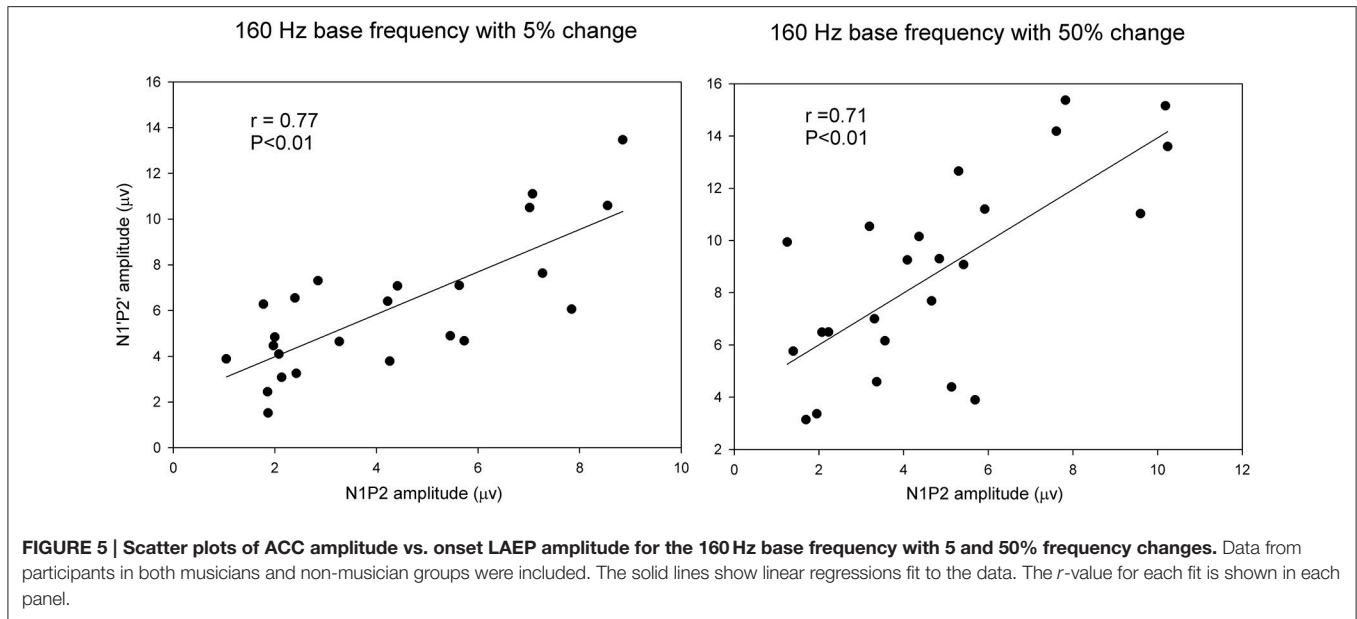
### Comparison between ACC and Behavioral Measures

The correlation between frequency detection thresholds and the ACC measures were examined. Pearson product moment





**FIGURE 4 | The ACC measures (N1' latency, P2' latency, N1' amplitude, P2' amplitude, and N1'-P2' amplitude) for musicians and non-musicians. The error bars indicate standard errors of the means. Asterisk denote significant differences between the groups ( $p < 0.05$ ).**



musicians can detect a 0.15% of frequency difference while the non-musicians can detect 0.5% of frequency difference using 330 Hz as a standard frequency. The authors also reported that the musician advantage of detecting frequency differences is even larger for complex harmonic tones than for pure tone. Kishon-Rabin et al. (2001) reported that the frequency differentiation threshold was  $\sim 1.8\%$  for musicians and  $3.4\%$  for non-musicians with a standard frequency of 250 Hz (see Figure 1 in Kishon-Rabin et al., 2001).

In the current study, the amount of frequency change at base frequency of 160 Hz can be detected by musicians and non-musicians was 0.42 and 0.72%, respectively, for Stim 1 which is comparable to stimulus condition used in other studies. The difference in the thresholds for frequency change detection between the current study and previous studies may be mainly related to the differences in the specific stimuli or stimulus paradigms used. Most previous studies used trials including standard tone at one frequency and the target tone that has a different frequency. The current study used trials including one standard tone at 160 Hz and one target tone of the same base frequency that contained a frequency change in the middle of the tone. The major difference between these stimulus paradigms is that the behavioral response with the conventional paradigm may reflect that the detection ability of the auditory system for a different frequency plus onset of the different frequency and the response with the current stimulus paradigm may reflect the detection of frequency change only. Additionally, other factors related to stimulus parameters could affect the performance of frequency change detection. Such factors include, but are not limited to, the duration, frequency, whether the stimulus is a pure tone or complex tone, intensity of the stimuli, and the interval between the stimuli in the trials. Despite the differences in the values of thresholds for frequency change detection/frequency discrimination among studies, the current finding that musicians outperform non-musicians in frequency change detection is

consistent with that in prior studies (Kishon-Rabin et al., 2001; Michey et al., 2006).

This musician benefit extends to the noisy conditions. Although the amount of musician benefits (differences between musicians vs. non-musicians, **Figure 1**) appears to be greater for Stim 3 compared to other conditions, the difference did not show statistical significance (no significant effect of interaction effect between Stimulus Condition and Subject Group). The lack of group difference in the 3 stimulus conditions suggests that the degree of musicians' benefits in pitch detection is similar in quiet and noisy conditions. The finding that musician benefits exist not only in quiet but also noisy conditions has been reported previously. Fuller et al. (2014) compared musicians and non-musicians in the performance of different types of tasks. The degree of musician effects varied greatly across stimulus conditions. In the conditions involving melodic/pitch pattern identification, there was a significant musician benefit when the melodic/pitch patterns were presented in quiet and noisy conditions. The authors suggested that musicians may be better overall listeners due to better high-level auditory cognitive functioning, not only in noise, but also in general. It would be worthwhile to examine if musician benefits also exist in pitch-based speech and music tasks in noisy conditions in laboratories and real life, which is still controversial with the results from the literature (Parbery-Clark et al., 2009; Strait et al., 2012; Fuller et al., 2014; Ruggles et al., 2014).

### Is Musician Effect Reflected by the ACC?

This is the first study examining the ACC in musician vs. non-musician comparisons. Although the training effects on the ACC have not been reported in previous studies, the training effects of the LAEP have been reported (Shahin et al., 2003; Tremblay et al., 2006). In general, previous studies reported that P2 was enlarged, although there have been some controversies on whether N1 is enhanced in musicians (Shahin et al., 2003, 2005). The previous

findings suggest that the P2 is the main component that is susceptible to training. However, the enhanced P2 in musicians does not necessarily reflect the long-term training in musicians only. A short-term auditory training may result in an enlarged P2. Tremblay et al. (2006) examined how a 10-day voice-onset-time (VOT) training changes the LAEP evoked by consonant-vowel speech tokens with various VOTs. The results showed that P2 is the dominant component that is enhanced after this training.

The current study found that the superiority of frequency change detection in musicians can be reflected by ACC measures (P2' amplitude). N1' latency was also shorter in musicians than non-musicians, although the difference was not statistically significant. Shorter latencies are thought to reflect faster and more efficient neural transmission and larger amplitudes reflect increased neural synchrony. The shorter N1' latency and larger P2' amplitude in musicians may suggest that musicians have a more efficient central processing of pitch changes than non-musicians.

There are some questions that need to be addressed in future studies. For example, what is the difference between the P2/P2' enhancements after the long-term vs. short-term training? If the P2/P2' enhancement in the waveform is the same after long-term and short-term training, would the neural generators for the P2/P2' be the same? Additionally, it remains a question how much contribution to the enhanced P2/P2' is from repeated exposures of stimulus trials during the testing (Seppänen et al., 2013; Tremblay et al., 2014)?

## Is Musician Effect Bottom-Up or Top-Down: Behavioral and EEG Evidence

Sound perception involves "bottom-up" and "top-down" processes that may be disrupted in hearing impaired individuals (Pisoni and Cleary, 2003). "Bottom-up" refers to the early automatic mechanisms (from the peripheral to auditory cortex) that encode the physical properties of sensory inputs (Noesselt et al., 2003). "Top-down" refers to processing (working memory, auditory attention, semantic, syntactical processing, etc.) after passively receiving and automatically detecting sounds (Noesselt et al., 2003).

Previous research provided evidence that the bottom-up process in musicians is enhanced (Jeon and Fricke, 1997; Tervaniemi et al., 2005; Micheyl et al., 2006). For instance, there is a strong correlation between neural responses (e.g., the Frequency Following Response/FFR) from subcortical level and the behavioral measure of frequency perception (Krishnan et al., 2012). Numerous studies have shown enhanced top-down processes in musicians. Behavioral data showed that musicians have better working memory (George and Coch, 2011; Bidelman et al., 2013). The EEG data showed a shorter latency in the P3 response of musicians, which is regarded as the effect of musical experience on cognitive abilities (George and Coch, 2011; Marie et al., 2011). Musicians show a larger gamma-band response (GBR), which has been associated with attentional, expectation, memory retrieval, and integration of top-down, bottom-up, and multisensory processes (Trainor et al., 2009; Ott et al., 2013). Tervaniemi et al. (2005) examined ERPs under passive and

active listening conditions. Under passive listening condition, the MMN and P3a, which reflect automatic sound differentiation, did not show difference between musicians and non-musicians. Under active listening conditions during which participants were required to pay attention to the stimuli and identify the deviant stimuli embedded in standard stimuli, the N2b and P3 were larger in musicians than in non-musicians. The authors suggested that musical expertise facilitates effects selectively for cognitive processes under attentional control.

Some researchers have examined the top-down control over bottom-up processes. Using the speech-evoked auditory brainstem response (ABR; Wong et al., 2007; Musacchia et al., 2008; Parbery-Clark et al., 2009; Strait et al., 2010, 2012; Strait and Kraus, 2011; Skoe and Kraus, 2013), Kraus and her colleagues reported enhanced encoding of fundamental frequencies and harmonics in musicians compared with non-musicians; there is a significant correlation between auditory working memory and attention and the ABR properties. Based on these findings, Kraus' group proposed that musicians' perceptual and neural enhancement are driven in a corticofugal or top-down manner. The top-down influence on cortical sensory processing in musicians can also be seen in the stronger efferent fibers linking cortical to subcortical auditory structures and even more peripheral stages of the auditory pathway (Perrot and Collet, 2014). Taken together, a rich amount of evidence suggested that musicians' auditory function is enhanced in a corticofugal top-down driven fashion. In the current study, musicians show a significantly larger P2' amplitude in the ACC. The N1' latency in musicians is shorter than non-musicians, although this difference did not reach statistical significance. The shorter N1 and larger P2 was also observed in the onset LAEP of musicians, although the musician vs. non-musician difference did not reach a statistical significance. The N1 has been considered the obligatory response that reflects the sound registration in the auditory cortex and the P2 is not simply an obligatory part of N1-P2 complex; evidence showed that the P2 is a more cognitive component reflecting attention-modulated process required for the performance of auditory discrimination tasks (Crowley and Colrain, 2004). This shortened N1/N1' and enlarged P2/P2' in musicians may be the result of a stronger interaction of bottom-up and top-down mechanisms in musicians. Future research with more comprehensive testing (e.g., the use of passive and active listening conditions for EEG and behavioral measures of cognitive function) can be designed to examine or disentangle the role of bottom-up and top-down processing in musician effects on the ACC. The interaction between behavioral outcomes (Hit/Miss) and the features of the ACC will be better revealed in the active listening condition in which the EEG is recorded while the participant performs the behavioral task.

## The ACC vs. the Onset LAEP

In the current study, the ACC was evoked by frequency changes contained in pure tones and the onset LAEP was evoked by the onset of the pure tone. The morphologies of onset LAEP and the ACC are very similar. However, several differences between them described below may suggest that the onset LAEP and the ACC involve different neural mechanisms.

First, the onset LAEP is evoked by the stimulus onset. The ACC is evoked by frequency changes not the onset of a new sound, which was removed when the frequency change occurred. Second, the ACC has longer peak latencies, especially for the 5% change. Specifically, ACC measures in **Table 1** show that the N1' and P2' latencies for the 5% change of base frequency 160 Hz are ~160 and 250 ms, respectively. Onset LAEP measures in **Figure 3** show that the N1 and P2 latencies for 160 Hz are ~140 and 210 ms, respectively. The same trend of longer peak latencies for the ACC than for the onset LAEP can be seen for base frequency of 1200 Hz. Third, the ACC has a larger amplitude than the onset LAEP. The amplitude difference of the ACC and the onset LAEP does not look like the result of a higher frequency contained in the tone relative to the base frequency, because the difference of the onset LAEPs evoked by 160 and 1200 Hz is not as dramatic as the amplitude difference between the ACC and onset LAEP. Finally, the amplitude of the ACC is significantly greater when the frequency change is perceptually greater (50 vs. 5%). The ACC amplitude difference between 50 vs. 5% cannot be explained by the difference of the onset LAEP evoked by 160 and 1200 Hz. Hence, the present data suggested that the ACC is evoked by acoustic changes in ongoing stimuli rather than the onset of a new frequency. This supports a previous speculation by some researchers that the ACC is more than a simple onset response (Ostroff et al., 1998).

The distinctions between the ACC and onset LAEP may be caused by different groups of neurons that are responsible for these two different responses, respectively. Animal studies have provided evidence that different groups of neurons in the auditory cortex have functional differences. For instance, in cats' primary auditory cortex, the tonic cells encode information of static auditory signals (e.g., tonal stimulus) with a significant firing increase throughout the stimulus period after a long latency; the phasic-tonic cells encode information of the change of auditory signal during the stimulus period after a medium latency; and the phasic cells (short latency) encode information of rapid change of the auditory signal at onset and offset after a short latency (Chimoto et al., 2002). We speculate that, the onset LAEP is dominantly contributed by the neurons that are sensitive to stimulus onset (e.g., the phasic cells that have shorter response latencies) and the ACC by neurons sensitive to acoustic changes (e.g., the phasic-tonic cells that have longer response latencies). If the onset LAEP and the ACC involves the activation of different groups of cortical neurons, this may suggest that it is important to use stimuli that have acoustic changes with removed onset cues in order to evoke the ACC.

It should be noted that, however, the ACC and onset LAEP may have shared neural mechanisms based on the following findings. Individuals displaying larger onset LAEPs tend to have larger ACCs evoked by frequency changes in 160 Hz base frequency. This finding is consistent with the finding in a previous ACC study in CI users (Brown et al., 2008). Additionally, musician vs. non-musician comparisons in the onset LAEP and the ACC show that musicians have a larger P2/P2', although the group difference is not significant for the onset LAEP. One possible shared mechanism for these two responses is novelty detection, which may be activated by the

stimulus onset that is different from the pre-stimulus quiet period and the frequency change that is different from the previous base frequency. This explanation can be further examined using source mapping in future studies.

It is noted, that the correlation between the onset LAEP and ACC exists only for base frequency 160 not 1200 Hz. This may suggest that the auditory system treats the frequency changes differently for different base frequencies.

## ACC Reflects Musician Benefits

ACC has been suggested by recent studies as a promising tool to show training effects. The current study supports this conclusion with the following findings: (1) there was no ACC when there was no frequency change in the tone, while there was an ACC when there were perceivable frequency changes; (2) the ACC is bigger when the frequency change is perceptually greater (50 vs. 5%); and (3) Musicians had significantly more robust P2' amplitude compared with non-musicians for 160 Hz base frequency. The lack of correlation between ACC and behavioral measures in the current finding does not exclude the correlation between ACC and behavioral measures. Note, that the ACC in this study was evoked by supra-threshold frequency changes (5 and 50%) rather than the threshold. The use of supra-threshold frequency changes for ACC recordings may be the reason for the failure of observing a significant relationship between the behavioral and ACC measures in the current study. Previous studies have reported that the minimal acoustic change that can evoke the ACC is similar to the threshold for auditory discrimination threshold (He et al., 2012; Kim, 2015). A refined EEG stimulus paradigm (e.g., tones containing a wider range of magnitude of frequency changes), should be used to determine if the perceptual threshold is corresponding to the minimum frequency change that can evoke an ACC in musicians and non-musicians in the future studies.

## ACC vs. MMN

The ACC has the following advantages over another EEG tool that can potentially be used to reflect training effects on perceptual discrimination ability, the mismatch negativity (MMN, Tremblay et al., 1997; Koelsch et al., 1999; Tervaniemi et al., 2005; Itoh et al., 2012). First, the ACC is a more time-efficient measure compared to the MMN (Martin and Boothroyd, 1999). While the MMN requires a large number of trials of stimuli to ensure there are enough trials (e.g., 100–200) of rarely presented stimulus (deviant stimuli) embedded in the frequently presented stimulus (standard stimuli) of the oddball paradigm, ACC is contributed by every trial in the stimulus paradigm. Second, ACC is a more sensitive and efficient evaluation tool than the MMN (Martin and Boothroyd, 1999) due to relatively larger and more stable amplitude. Repeated recording of the ACC has revealed that the ACC is stable and repeatable (Friesen and Tremblay, 2006).

Although the ACC and the MMN are thought to reflect automatic differentiation of sounds at the pre-attentive stage of auditory processing, these responses may involve different neural generators/neurons due to the differences in the stimulus paradigms used to evoke them. Specifically, the ACC is evoked

by the acoustic change in the ongoing stimuli. The neurons activated may be the ones that are sensitive to the acoustic change rather than the acoustic onset. In contrast, the MMN reflects the discrepancy between the neural response to the deviant stimuli and the response to the standard stimuli, the neurons activated for the MMN may be the ones that respond to the onset of the deviant stimuli rather than the acoustic change *per-se*. The above speculation can be further confirmed using source mapping in future studies. This speculated difference between the ACC and the MMN may be the reason why there is a significant difference in the ACC between musicians and non-musicians in the current study while there was no difference in the MMN between these two groups in a previous study (Tervaniemi et al., 2005).

## IMPLICATIONS AND FUTURE WORK

This study has important implications. First, our findings suggest that long-term music training in individuals with normal auditory systems provides advantages in the frequency tasks that are challenging for hearing impaired patients; the musician benefit is persistent in noisy conditions. However, future studies are still needed to determine if the short-term training in hearing impaired patients, who have different degrees of neural deficits, would result in neurological changes and perceptual improvement in pitch change detection.

Second, the current results also have implications in other populations who have problems in frequency-based perception. For instance, dyslexic children are found to have difficulties discriminating frequency changes that are easily discriminated by normal readers (Besson et al., 2007). Music training would be beneficial to facilitate the brain plasticity toward improving frequency perception and further language perception.

Finally, the ACC can be evoked by frequency changes. The ACC in musicians showed a larger P2' amplitude than in non-musicians. Because the ACC is recorded without participant's voluntary response, it provides an objective tool to estimate

frequency change detection ability and to document training effects.

## CONCLUSION

To summarize, musicians outperform non-musicians in pitch change detection in quiet and noisy conditions. This musician benefit can be reflected in the ACC measures: the ACC evoked by the frequency change from a base frequency 160 Hz showed a greater P2' amplitude in musicians than in non-musicians. The ACC displays differences and similarities compared to the onset LAEP, which may suggest these two responses involve different but overlapping neural mechanisms.

## AUTHOR CONTRIBUTIONS

FZ has substantial contributions to the conception and design of the work, the acquisition, analysis, interpretation of data, manuscript preparation, final approval of the version to be submitted, and be accountable for all aspects of the work. CL has substantial contributions to the design of the work, the acquisition, analysis, interpretation of data, drafting the manuscript, final approval of the version to be submitted, and be accountable for all aspects of the work related to data collection and analysis. IT, BE, KW, SC, JX, and QF all have significant contributions to the design of the work, the acquisition, manuscript preparation, final approval of the version to be submitted, and be accountable for their portions of work.

## ACKNOWLEDGMENTS

This research was supported by the University Research Council at the University of Cincinnati. We would like to thank all participants for their contribution to this research. The authors also thank Mr. Cody Curry, Ms. Sarah Colligan, and Ms. Anna Herrmann for their efforts in recruiting participants.

## REFERENCES

- Besson, M., Schön, D., Moreno, S., Santos, A., and Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restor. Neurol. Neurosci.* 25, 399–410.
- Bidelman, G. M., Hutka, S., and Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PLoS ONE* 8:e60676. doi: 10.1371/journal.pone.0060676
- Brown, C. J., Etler, C., He, S., O'Brien, S., Erenberg, S., Kim, J. R., et al. (2008). The electrically evoked auditory change complex: preliminary results from nucleus cochlear implant users. *Ear Hear.* 29, 704–717. doi: 10.1097/AUD.0b013e31817a98af
- Chimoto, S., Kitama, T., Qin, L., Sakayori, S., and Sato, Y. (2002). Tonal response patterns of primary auditory cortex neurons in alert cats. *Brain Res.* 934, 34–42. doi: 10.1016/S0006-8993(02)02316-8
- Crowley, K. E., and Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clin. Neurophysiol.* 115, 732–744. doi: 10.1016/j.clinph.2003.11.021
- Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., et al. (2012). Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: behavioral, electrophysiological and psychoacoustic study. *Brain Res.* 1455, 75–89. doi: 10.1016/j.brainres.2012.03.034
- Dimitrijevic, A., Michalewski, H. J., Zeng, F.-G., Pratt, H., and Starr, A. (2008). Frequency changes in a continuous tone: auditory cortical potentials. *Clin. Neurophysiol.* 119, 2111–2124. doi: 10.1016/j.clinph.2008.06.002
- Francart, T., van Wieringen, A., and Wouters, J. (2008). APEX 3: a multi-purpose test platform for auditory psychophysical experiments. *J. Neurosci. Methods* 172, 283–293. doi: 10.1016/j.jneumeth.2008.04.020
- Friesen, L. M., and Tremblay, K. L. (2006). Acoustic change complexes recorded in adult cochlear implant listeners. *Ear Hear.* 27, 678–685. doi: 10.1097/01.aud.0000240620.63453.c3
- Fu, Q. J., Galvin, J., Wang, X., and Nogaki, G. (2004). Effects of auditory training on adult cochlear implant patients: a preliminary report. *Cochlear Implants Int.* 5(Suppl. 1), 84–90. doi: 10.1002/cii.181
- Fu, Q.-J., and Nogaki, G. (2005). Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *J. Assoc. Res. Otolaryngol.* 6, 19–27. doi: 10.1007/s10162-004-5024-3
- Fuller, C. D., Galvin, J. J. III, Maat, B., Free, R. H., and Başkent, D. (2014). The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? *Front. Neurosci.* 8:179. doi: 10.3389/fnins.2014.00179

- Galvin, J. J. III, Fu, Q. J., and Shannon, R. V. (2009). Melodic contour identification and music perception by cochlear implant users. *Ann. N.Y. Acad. Sci.* 1169, 518–533. doi: 10.1111/j.1749-6632.2009.04551.x
- Gaser, C., and Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *J. Neurosci.* 23, 9240–9245.
- George, E. M., and Coch, D. (2011). Music training and working memory: an ERP study. *Neuropsychologia* 49, 1083–1094. doi: 10.1016/j.neuropsychologia.2011.02.001
- Gfeller, K., Guthe, E., Driscoll, V., and Brown, C. J. (2015). A preliminary report of music-based training for adult cochlear implant users: rationales and development. *Cochlear Implants Int.* 16(Suppl. 3), S22–S31. doi: 10.1179/1467010015Z.000000000269
- Harris, K. C., Mills, J. H., and Dubno, J. R. (2007). Electrophysiologic correlates of intensity discrimination in cortical evoked potentials of younger and older adults. *Hear. Res.* 228, 58–68. doi: 10.1016/j.heares.2007.01.021
- He, S., Grose, J. H., and Buchman, C. A. (2012). Auditory discrimination: The relationship between psychophysical and electrophysiological measures. *Int. J. Audiol.* 51, 771–782. doi: 10.3109/14992027.2012.699198
- Hutter, E., Argstatter, H., Grapp, M., and Plinkert, P. K. (2015). Music therapy as specific and complementary training for adults after cochlear implantation: A pilot study. *Cochlear Implants Int.* 16(Suppl. 3), S13–S21. doi: 10.1179/1467010015Z.000000000261
- Itoh, K., Okumiya-Kanke, Y., Nakayama, Y., Kwee, I. L., and Nakada, T. (2012). Effects of musical training on the early auditory cortical representation of pitch transitions as indexed by change-N1. *Eur. J. Neurosci.* 36, 3580–3592. doi: 10.1111/j.1460-9568.2012.08278.x
- James, C. E., Oechslin, M. S., Van De Ville, D., Hauert, C. A., Descoux, C., and Lazeyras, F. (2014). Musical training intensity yields opposite effects on grey matter density in cognitive versus sensorimotor networks. *Brain Struct. Funct.* 219, 353–366. doi: 10.1007/s00429-013-0504-z
- Jentschke, S., and Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *Neuroimage* 47, 735–744. doi: 10.1016/j.neuroimage.2009.04.090
- Jeon, J. Y., and Fricke, F. R. (1997). Duration of perceived and performed sounds. *Psychol. Music* 25, 70–83. doi: 10.1177/0305735697251006
- Kim, J. R. (2015). Acoustic change complex: clinical implications. *J. Audiol. Otol.* 19, 120–124. doi: 10.7874/jao.2015.19.3.120
- Kim, J. R., Brown, C. J., Abbas, P. J., Etlar, C. P., and O'Brien, S. (2009). The effect of changes in stimulus level on electrically evoked cortical auditory potentials. *Ear Hear.* 30, 320–329. doi: 10.1097/AUD.0b013e31819c42b7
- Kishon-Rabin, L., Amir, O., Vexler, Y., and Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *J. Basic Clin. Physiol. Pharmacol.* 12, 125–144. doi: 10.1515/JBCPP.2001.12.2.125
- Koelsch, S., Gunter, T. C., Wittfoth, M., and Sammler, D. (2005). Interaction between syntax processing in language and in music: an ERP study. *J. Cogn. Neurosci.* 17, 1565–1577. doi: 10.1162/089892905774597290
- Koelsch, S., Schröger, E., and Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport* 10, 1309–1313. doi: 10.1097/0001756-199904260-00029
- Koelsch, S., and Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends Cogn. Sci.* 9, 578–584. doi: 10.1016/j.tics.2005.10.001
- Kong, Y. Y., and Zeng, F. G. (2004). Mandarin tone recognition in acoustic and electric hearing. *Cochlear Implants Int.* 5(Suppl. 1), 175–177. doi: 10.1002/cii.218
- Kraus, N. (2012). Biological impact of music and software-based auditory training. *J. Commun. Disord.* 45, 403–410. doi: 10.1016/j.jcomdis.2012.06.005
- Kraus, N., Skoe, E., Parbery-Clark, A., and Ashley, R. (2009). Experience-induced malleability in neural encoding of pitch, timbre, and timing. *Ann. N.Y. Acad. Sci.* 1169, 543–557. doi: 10.1111/j.1749-6632.2009.04549.x
- Kraus, N., and Strait, D. L. (2015). Emergence of biological markers of musicianship with school-based music instruction. *Ann. N.Y. Acad. Sci.* 1337, 163–169. doi: 10.1111/nyas.12631
- Krishnan, A., Bidelman, G. M., Smalt, C. J., Ananthakrishnan, S., and Gandour, J. T. (2012). Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia* 50, 2849–2859. doi: 10.1016/j.neuropsychologia.2012.08.013
- Lo, C. Y., McMahon, C. M., Looi, V., and Thompson, W. F. (2015). Melodic contour training and its effect on speech in noise, consonant discrimination, and prosody perception for cochlear implant recipients. *Behav. Neurol.* 2015:352869. doi: 10.1155/2015/352869
- Looi, V., Gfeller, K., and Driscoll, V. (2012). Music appreciation and training for cochlear implant recipients: a review. *Semin. Hear.* 33, 307–334. doi: 10.1055/s-0032-1329222
- Magne, C., Schön, D., and Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *J. Cogn. Neurosci.* 18, 199–211. doi: 10.1162/jocn.2006.18.2.199
- Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., and Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in mandarin Chinese. *J. Cogn. Neurosci.* 23, 2701–2715. doi: 10.1162/jocn.2010.21585
- Marques, C., Moreno, S., Castro, S. L., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463. doi: 10.1162/jocn.2007.19.9.1453
- Martin, B. A. (2007). Can the acoustic change complex be recorded in an individual with a cochlear implant? separating neural responses from cochlear implant artifact. *J. Am. Acad. Audiol.* 18, 126–140. doi: 10.3766/jaaa.18.2.5
- Martin, B. A., and Boothroyd, A. (1999). Cortical, auditory, event-related potentials in response to periodic and aperiodic stimuli with the same spectral envelope. *Ear Hear.* 20, 33–44. doi: 10.1097/00003446-199902000-00004
- Martin, B. A., and Boothroyd, A. (2000). Cortical, auditory, evoked potentials in response to changes of spectrum and amplitude. *J. Acoust. Soc. Am.* 107, 2155–2161. doi: 10.1121/1.428556
- Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hear. Res.* 219, 36–47. doi: 10.1016/j.heares.2006.05.004
- Musacchia, G., Strait, D., and Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear. Res.* 241, 34–42. doi: 10.1016/j.heares.2008.04.013
- Noesselt, T., Shah, N. J., and Jäncke, L. (2003). Top-down and bottom-up modulation of language related areas—an fMRI study. *BMC Neurosci.* 4:13. doi: 10.1186/1471-2202-4-13
- Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., et al. (2001). Functional anatomy of musical perception in musicians. *Cereb. Cortex* 11, 754–760. doi: 10.1093/cercor/11.8.754
- Ostroff, J. M., Martin, B. A., and Boothroyd, A. (1998). Cortical evoked response to acoustic change within a syllable. *Ear Hear.* 19, 290–297. doi: 10.1097/00003446-199808000-00004
- Ott, C. G., Stier, C., Herrmann, C. S., and Jäncke, L. (2013). Musical expertise affects attention as reflected by auditory-evoked gamma-band activity in human EEG. *Neuroreport* 24, 445–450. doi: 10.1097/WNR.0b013e328360abdb
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature* 392, 811–814. doi: 10.1038/33918
- Pantev, C., Ross, B., Fujioka, T., Trainor, L. J., Schulte, M., and Schulz, M. (2003). Music and learning-induced cortical plasticity. *Ann. N.Y. Acad. Sci.* 999, 438–450. doi: 10.1196/annals.1284.054
- Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear Hear.* 30, 653–661. doi: 10.1097/AUD.0b013e3181b412e9
- Patel, A. D. (2003). Rhythm in language and music: parallels and differences. *Ann. N.Y. Acad. Sci.* 999, 140–143. doi: 10.1196/annals.1284.015
- Perrot, X., and Collet, L. (2014). Function and plasticity of the medial olivocochlear system in musicians: A review. *Hear. Res.* 308, 27–40. doi: 10.1016/j.heares.2013.08.010
- Petersen, B., Mortensen, M. V., Gjedde, A., and Vuust, P. (2009). Reestablishing speech understanding through musical ear training after cochlear implantation: a study of the potential cortical plasticity in the brain. *Ann. N.Y. Acad. Sci.* 1169, 437–440. doi: 10.1111/j.1749-6632.2009.04796.x
- Pisoni, D. B., and Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear Hear.* 24(1 Suppl.), 106S–120S. doi: 10.1097/01.AUD.0000051692.05140.8E
- Ruggles, D. R., Freyman, R. L., and Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PLoS ONE* 9:e86980. doi: 10.1371/journal.pone.0086980

- Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). Morphology of heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* 5, 688–694. doi: 10.1038/nn871
- Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349. doi: 10.1111/1469-8986.00172.x
- Seppänen, M., Hämäläinen, J., Pesonen, A. K., and Tervaniemi, M. (2013). Passive sound exposure induces rapid perceptual learning in musicians: event-related potential evidence. *Biol. Psychol.* 94, 341–353. doi: 10.1016/j.biopsycho.2013.07.004
- Seung, Y., Kyong, J. S., Woo, S. H., Lee, B. T., and Lee, K. M. (2005). Brain activation during music listening in individuals with or without prior music training. *Neurosci. Res.* 52, 323–329. doi: 10.1016/j.neures.2005.04.011
- Shahin, A., Bosnyak, D. J., Trainor, L. J., and Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J. Neurosci.* 23, 5545–5552.
- Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., and Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport* 16, 1781–1785. doi: 10.1097/01.wnr.0000185017.29316.63
- Skoe, E., and Kraus, N. (2013). Musical training heightens auditory brainstem function during sensitive periods in development. *Front. Psychol.* 4:622. doi: 10.3389/fpsyg.2013.00622
- Small, S. A., and Werker, J. F. (2012). Does the ACC have potential as an index of early speech discrimination ability? A preliminary study in 4-month-old infants with normal hearing. *Ear Hear.* 33, e59–e69. doi: 10.1097/AUD.0b013e31825f29be
- Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F. G. (2007). Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. *J. Acoust. Soc. Am.* 122, 1069–1078. doi: 10.1121/1.2750159
- Strait, D. L., and Kraus, N. (2011). Can you hear me now? musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Front. Psychol.* 2:113. doi: 10.3389/fpsyg.2011.00113
- Strait, D. L., Kraus, N., Parbery-Clark, A., and Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear. Res.* 261, 22–29. doi: 10.1016/j.heares.2009.12.021
- Strait, D. L., Parbery-Clark, A., Hittner, E., and Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain Lang.* 123, 191–201. doi: 10.1016/j.bandl.2012.09.001
- Strait, D. L., Slater, J., O'Connell, S., and Kraus, N. (2015). Music training relates to the development of neural mechanisms of selective auditory attention. *Dev. Cogn. Neurosci.* 12, 94–104. doi: 10.1016/j.dcn.2015.01.001
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., and Schröger, E. (2005). Pitch discrimination accuracy in musicians vs. nonmusicians: an event-related potential and behavioral study. *Exp. Brain Res.* 161, 1–10. doi: 10.1007/s00221-004-2044-5
- Thompson, W. F., Schellenberg, E. G., and Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emotion* 4, 46–64. doi: 10.1037/1528-3542.4.1.46
- Trainor, L. J., Shahin, A. J., and Roberts, L. E. (2009). Understanding the benefits of musical training: effects on oscillatory brain activity. *Ann. N.Y. Acad. Sci.* 1169, 133–142. doi: 10.1111/j.1749-6632.2009.04589.x
- Tremblay, K., Kraus, N., Carrell, T. D., and McGee, T. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *J. Acoust. Soc. Am.* 102, 3762–3773. doi: 10.1121/1.420139
- Tremblay, K. L., Kalstein, L., Billings, C. J., and Souza, P. E. (2006). The neural representation of consonant-vowel transitions in adults who wear hearing AIDS. *Trends Amplif.* 10, 155–162. doi: 10.1177/1084713806292655
- Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K., and Collet, G. (2014). Is the auditory evoked P2 response a biomarker of learning? *Front. Syst. Neurosci.* 8:28. doi: 10.3389/fnsys.2014.00028
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422. doi: 10.1038/nn1872
- Zeng, F. G., Tang, Q., and Lu, T. (2014). Abnormal pitch perception produced by cochlear implant stimulation. *PLoS ONE* 9:e88662. doi: 10.1371/journal.pone.0088662

**Conflict of Interest Statement:** 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.

Copyright © 2016 Liang, Earl, Thompson, Whitaker, Cahn, Xiang, Fu and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# High-Resolution Audio with Inaudible High-Frequency Components Induces a Relaxed Attentional State without Conscious Awareness

Ryuma Kuribayashi and Hiroshi Nittono\*

Graduate School of Human Sciences, Osaka University, Osaka, Japan

## OPEN ACCESS

### Edited by:

Mark Reybrouck,  
KU Leuven, Belgium

### Reviewed by:

Lutz Jäncke,  
University of Zurich, Switzerland  
Robert J. Barry,  
University of Wollongong, Australia

### \*Correspondence:

Hiroshi Nittono  
nittono@hus.osaka-u.ac.jp

### Specialty section:

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

**Received:** 01 October 2016

**Accepted:** 13 January 2017

**Published:** 01 February 2017

### Citation:

Kuribayashi R and Nittono H (2017)  
High-Resolution Audio with Inaudible  
High-Frequency Components  
Induces a Relaxed Attentional State  
without Conscious Awareness.  
Front. Psychol. 8:93.  
doi: 10.3389/fpsyg.2017.00093

High-resolution audio has a higher sampling frequency and a greater bit depth than conventional low-resolution audio such as compact disks. The higher sampling frequency enables inaudible sound components (above 20 kHz) that are cut off in low-resolution audio to be reproduced. Previous studies of high-resolution audio have mainly focused on the effect of such high-frequency components. It is known that alpha-band power in a human electroencephalogram (EEG) is larger when the inaudible high-frequency components are present than when they are absent. Traditionally, alpha-band EEG activity has been associated with arousal level. However, no previous studies have explored whether sound sources with high-frequency components affect the arousal level of listeners. The present study examined this possibility by having 22 participants listen to two types of a 400-s musical excerpt of *French Suite No. 5* by J. S. Bach (on cembalo, 24-bit quantization, 192 kHz A/D sampling), with or without inaudible high-frequency components, while performing a visual vigilance task. High-alpha (10.5–13 Hz) and low-beta (13–20 Hz) EEG powers were larger for the excerpt with high-frequency components than for the excerpt without them. Reaction times and error rates did not change during the task and were not different between the excerpts. The amplitude of the P3 component elicited by target stimuli in the vigilance task increased in the second half of the listening period for the excerpt with high-frequency components, whereas no such P3 amplitude change was observed for the other excerpt without them. The participants did not distinguish between these excerpts in terms of sound quality. Only a subjective rating of inactive pleasantness after listening was higher for the excerpt with high-frequency components than for the other excerpt. The present study shows that high-resolution audio that retains high-frequency components has an advantage over similar and indistinguishable digital sound sources in which such components are artificially cut off, suggesting that high-resolution audio with inaudible high-frequency components induces a relaxed attentional state without conscious awareness.

**Keywords:** high-resolution audio, electroencephalogram, alpha power, event-related potential, vigilance task, attention, conscious awareness, hypersonic effect

## INTRODUCTION

High-resolution audio has recently emerged in the digital music market due to recent advances in information and communications technologies. Because of a higher sampling frequency and a greater bit depth than conventional low-resolution audio such as compact disks (CDs), it provides a closer replication of the real analog sound waves. Sampling frequency means the number of samples per second taken from a sound source through analog-to-digital conversion. Bit depth is the number of possible values in each sample and expressed as a power of two. A higher sampling frequency makes the digitization of sound more accurate in the time-frequency domain, whereas a greater bit depth increases the resolution of the sound. What kind of advantage does the latest digital audio have for human beings? This question has not been sufficiently discussed. The present investigation used physiological, behavioral, and subjective measures to provide evidence that high-resolution audio affects human psychophysiological state without conscious awareness.

The higher sampling frequency enables higher frequency sound components to be reproduced, because one-half of the sampling frequency defines the upper limit of reproducible frequencies (as dictated by the Nyquist-Shannon sampling theorem). However, in conventional digital audio, sampling frequency is usually restrained so that sounds above 20 kHz are cut off in order to reduce file sizes for convenience. This reduction is based on the knowledge that sounds above 20 kHz do not influence sound quality ratings (Muraoka et al., 1981) and do not appear to produce evoked brain magnetic field responses (Fujioka et al., 2002).

In contrast to this conventional digital recording process in which inaudible high-frequency components are cut off, high-resolution music that retains such components has been repeatedly shown to affect human electroencephalographic (EEG) activity (Oohashi et al., 2000, 2006; Yagi et al., 2003a; Fukushima et al., 2014; Kuribayashi et al., 2014; Ito et al., 2016). This effect is often called “hypersonic” effect. In these studies, only the presence or absence of inaudible high-frequency components is manipulated while the sampling frequency and the bit depth are held constant. Interestingly, this effect appears with a considerable delay (i.e., 100–200 s after the onset of music). However, it remains unclear what kind of psychological and cognitive states are associated with this effect. These studies also suggest that it is difficult to distinguish in a conscious sense between sounds with and without inaudible high-frequency components (full-range vs. high-cut). Some studies have shown that full-range audio is rated as better sound quality (e.g., a softer tone, more comfortable to the ears) than high-cut audio (Oohashi et al., 2000; Yagi et al., 2003a). Another study has shown that participants are not able to distinguish between the two types of digital audio, with no significant differences found for subjective ratings of sound qualities (Kuribayashi et al., 2014). The feasibility of discrimination seems to depend on the kinds of audio sources and individuals (Nishiguchi et al., 2009). Regarding

behavioral aspects, it has been shown that people listen to full-range sounds at a higher level of sound volume than high-cut sounds (Yagi et al., 2003b, 2006; Oohashi et al., 2006).

Previous studies have examined the effect of inaudible high-frequency components on EEG activity while listening to music under resting conditions. It has been shown that EEG alpha-band (8–13 Hz) frequency power is greater for high-resolution music with high-frequency components than for the same sound sources without them (Oohashi et al., 2000, 2006; Yagi et al., 2003a). The effect appears more clearly in a higher part of the conventional alpha-band frequency of 8–13 Hz (10.5–13 Hz: Kuribayashi et al., 2014; 11–13 Hz: Ito et al., 2016). Ito et al. (2016) reported that low beta-band (14–20 Hz) EEG power also showed the same tendency to increase as high alpha-band EEG power.

A study using positron emission tomography (PET) revealed that the brainstem and thalamus areas were more activated when hearing full-range as compared with high-cut sounds (Oohashi et al., 2000). Because such activation may support a role of the thalamus in emotional experience (LeDoux, 1993; Vogt and Gabriel, 1993; Blood and Zatorre, 2001; Jeffries et al., 2003; Brown et al., 2004; Lupien et al., 2009) and also in filtering or gating sensory input (Andreasen et al., 1994), Oohashi et al. (2000) speculated that the presence of inaudible high-frequency components may affect the perception of sounds and some aspects of human behavior.

Another line of research suggests a link between cognitive function and alpha-band as well as beta-band EEG activities. Alpha-band EEG activity is thought to be associated not only with arousal and vigilance levels (Barry et al., 2007) but also with cognitive tasks involving perception, working memory, long-term memory, and attention (e.g., Başar, 1999; Klimesch, 1999; Ward, 2003; Klimesch et al., 2005). Higher alpha-band activity is considered to inhibit task-irrelevant brain regions so as to serve effective disengagement for optimal processing (Jensen and Mazaheri, 2010; Foxe and Snyder, 2011; Weisz et al., 2011; Klimesch, 2012; De Blasio et al., 2013).

Beta-band power is broadly thought to be associated with motor function when it is derived from motor areas (Hari and Salmelin, 1997; Crone et al., 1998; Pfurtscheller and Lopes da Silva, 1999; Pfurtscheller et al., 2003). Moreover, beta power has been shown to increase with corresponding increases in arousal and vigilance levels, which indicates that participants get engaged in a task (e.g., Sebastiani et al., 2003; Aftanas et al., 2006; Gola et al., 2012; Kamiński et al., 2012).

What kind of advantage does high-resolution audio with inaudible high-frequency components have for human beings? What remains unclear is what kind of psychophysiological states high-resolution audio induces, along with the corresponding increase in alpha- and beta-band EEG activities. To monitor listeners' arousal level, we asked participants to listen to a musical piece while performing a visual vigilance task that required sustained attention in order to continuously respond to specific stimuli. Two types of high-resolution audio of the same musical piece were presented using a double-blind method: With or without inaudible high-frequency components.

EEG was recorded along with other psychophysiological measures: Heart rate (HR), heart rate variability (HRV), and facial electromyograms (EMGs). The former two measures index autonomic nervous system activities. HRV contains two components with different frequency bands: High frequency (HF; 0.15–0.4 Hz), and low frequency (LF; 0.04–0.15 Hz). HF and LF activities are mediated by vagal and vagosympathetic activations, respectively (Malliani et al., 1991; Malliani and Montano, 2002). The LF/HF power ratio is sometimes used as an index parameter that shows the sympathetic activities. Facial EMGs in the regions of the corrugator supercilii and the zygomaticus major muscles have been used as indices of negative and positive affects, respectively (Larsen et al., 2003). Decrements in vigilance task performance such as longer reaction times (RTs) and higher error rates are interpreted as reflecting the decrease in arousal level, which is also reflected in the electrical activity of the brain (Fruhstorfer and Bergström, 1969). Besides ongoing EEG activity, event-related potentials (ERPs) are associated with vigilance task performance. When vigilance task performance decreases, the amplitude of P3, a positive ERP component observed dominantly at parietal recording sites between 300 and 600 ms after stimulus onset, decreases and its latency increases (Fruhstorfer and Bergström, 1969; Davies and Parasuraman, 1977; Parasuraman, 1983). The P3 outcomes are thought to be modulated not only by overall arousal level but also by attentional resource allocation (Polich, 2007). P3 amplitude has been shown to be larger when greater attentional resources are allocated to the eliciting stimulus. It is thus thought that P3 amplitude can serve as a measure of processing capacity and mental workload (Kok, 1997, 2001).

In the present study, physiological, behavioral, and subjective measures were recorded to examine what kind of advantage high-resolution audio with inaudible high-frequency components has. Specifically, we were interested in how the increase in alpha- and beta-band EEG activities is associated with listeners' arousal and vigilance level. Using a double-blind method, two types of high-resolution audio of the same musical piece (with or without inaudible high-frequency components) were presented while participants performed a vigilance task in the visual modality.

## MATERIALS AND METHODS

### Participants

Twenty-six student volunteers at Hiroshima University gave their informed consent and participated in the study. Four participants had to be excluded due to technical problems. The remaining 22 participants (14 women, 18–24 years,  $M = 20.6$  years) did not report any known neurological dysfunction or hearing deficit. They were right-handed according to the Edinburgh Inventory ( $M = 84.1 \pm 12.8$ ). All reported to have correct or corrected-to-normal vision. Eight participants had the experience of learning musical instruments for a few years, but none of them were professional musicians. The Research Ethics Committee of the Graduate School of Integrated Arts and Sciences in Hiroshima University approved the experimental protocol.

### Stimuli and Task

The present study used the same materials that were used in Kuribayashi et al. (2014). The first 200-s portion of *French Suite No. 5* by J. S. Bach (on cembalo, 24-bit quantization, 192 kHz A/D sampling) was selected. In the present study, this portion was played twice to produce a 400-s excerpt. The original (full-range) excerpt is rich in high-frequency components. A high-cut version of the excerpt was produced by removing such components using a low-pass finite impulse response digital filter with a very steep slope (cutoff = 20 kHz, slope =  $-1,673$  dB/oct). This linear-phase filter does not cause any phase distortion. Although the filter produces very small ripples ( $1.04 \times 10^{-2}$  dB), they are negligible and it is unlikely to affect auditory perception. Sounds were amplified using AI-501DA (TEAC Corporation, Tokyo, Japan) controlled by dedicated software on a laptop PC. Two loudspeakers with high-frequency tweeters (PM1; Bowers & Wilkins, Worthing, England) were located 1.5 m diagonally forward from the listening position. The sound pressure level was set at approximately 70 dB (A). Calibration measurements at the listening position ensured that the full-range excerpt contained abundant high-frequency components and that the high-frequency power of the high-cut excerpt (i.e., components over 20 kHz) did not differ from that of background noise. The average power spectra of the excerpts are available at <http://links.lww.com/WNR/A279> as Supplemental Digital Content of Kuribayashi et al. (2014).

An equiprobable visual Go/NoGo task was conducted using a cathode ray tube (CRT) computer monitor (refresh rate = 100 Hz) in front of participants. A block consisted of 120 visual stimuli: 60 targets (either 'T' or 'V', 30 each) and 60 non-targets ('O') in a randomized order. The visual stimuli were 200 ms in duration and presented with a mean stimulus onset asynchrony (SOA) of 5 s (range = 3–7 s). Button-press responses with the left and right index fingers were required to 'T' and 'V' (or 'V' and 'T') respectively, as quickly and accurately as possible.

### Procedure

The study was conducted using a double-blind method. Participants listened to two versions of the 400-s musical excerpt (with or without high-frequency components) while performing the Go/NoGo task. Participants also performed the task under silent conditions for 100 s before and after music presentation (pre- and post-music periods). The presentation order of the two excerpts was counterbalanced across the participants. EEG, HR, and facial EMGs were recorded during task performance. After listening to each excerpt, participants completed a sound quality questionnaire consisting of 10 pairs of adjectives and then reported their mood states on the Affect Grid (Russell et al., 1989) and multiple mood scales (Terasaki et al., 1992). At the end of the experiment, participants judged which excerpt contained high-frequency components by making a binary choice between them.

### Physiological Recording

Psychophysiological measures were recorded with a sampling rate of 1000 Hz using QuickAmp (Brain Products, Gilching,

Germany). Filter bandpass was DC to 200 Hz. EEG was recorded from 34 scalp electrodes (Fp1/2, Fz, F3/4, F7/8, FC1/2, FC5/6, FT9/10, Cz, T7/8, C3/4, CP1/2, CP5/6, TP9/10, Pz, P3/4, P7/8, PO9/10, Oz, O1/2) according to the extended 10–20 system. Four additional electrodes (supra-orbital and infra-orbital ridges of the right eye and outer canthi) were used to monitor eye movements and blinks. EEG data were recorded using the average reference online and re-referenced to the digitally linked earlobes (A1–A2) offline. EEG data were resampled at 250 Hz and were filtered offline (1–60 Hz band pass, 24 dB/oct for EEG analysis; 0.1–60 Hz band pass, 24 dB/oct for ERP analysis). Ocular artifacts were corrected using a semi-automatic independent component analysis method implemented on Brain Vision Analyzer 2.04 (Brain Products). The components that were easily identifiable as artifacts related to blinks and eye movements were removed.

Heart rate was measured by recording electrocardiograms from the left ankle and the right hand. The R–R intervals were calculated and converted into HR in bpm. For facial EMGs, electrical activities over the zygomaticus major and corrugator supercilii regions were recorded using bipolar electrodes affixed above the left brow and on the left cheek, respectively (Fridlund and Cacioppo, 1986). The EMG data were filtered offline (15 Hz high-pass, 12 dB/oct) and fully rectified (Larsen et al., 2003).

## Data Reduction and Statistical Analysis

A total of 600 s (including silent periods) was divided into six 100-s epochs. For EEG analysis, each 100-s epoch was divided into 97 2.048-s segments with 1.024 s overlap. Power spectrum was calculated by Fast Fourier Transform with a Hanning window. The total powers ( $\mu V^2$ ) of the following frequency bands were calculated: Delta (1–4 Hz), theta (4–8 Hz), low-alpha (8–10.5 Hz), high-alpha (10.5–13 Hz), low-beta (13–20 Hz), high-beta (20–30 Hz), and gamma (36–44 Hz). The square root of the total power ( $\mu V$ ) was used for statistical analysis, following the procedure of previous studies (Oohashi et al., 2000, 2006; Kuribayashi et al., 2014). The scalp electrode sites were grouped into four regions: Anterior Left (AL: Fp1, F3, F7, FC1, FC5, FT9), Anterior Right (AR: Fp2, F4, F8, FC2, FC6, FT10), Posterior Left (PL: CP1, CP5, TP9, P3, P7, PO9, O1), and Posterior Right (PR: CP2, CP6, TP10, P4, P8, PO10, O2). For RT, EMG, and HR, the mean values of each 100-s epoch were calculated. Mean RT and EMG values were log-transformed before statistical analysis.

Heart rate variability analysis was done by using Kubios HRV 2.2 (Tarvainen et al., 2014). The last 300-s (5-min) epoch of the 400-s listening period was selected according to previously established guidelines (Berntson et al., 1997). Prior to spectrum estimation, the R–R interval series is converted to equidistantly sampled series via piecewise cubic spline interpolation. The spectrum is estimated using an autoregressive modeling based method. The total powers ( $ms^2$ ) were calculated for LF (0.04–0.15 Hz) and HF (0.15–0.4 Hz) bands, and LF/HF power ratio was obtained. The square roots of LF and HF ( $ms$ ) were used for statistical analysis.

For ERP analysis, the total 400-s listening period was divided into two 200-s epochs, to secure a reasonable number of Go trials (around 20). Silent periods (pre- and post-music epoch) were not included in the calculation. Those trials found to have Go

omissions, Go misses (incorrect hand response to ‘T’ or ‘V’), or NoGo responses (commission errors) were excluded from further processing steps. Go and NoGo responses were separately averaged to produce ERPs. Epochs (200 ms before stimulus presentation until 1000 ms after the presentation) were baseline corrected (–200 ms until 0 ms). The peak of a P3 wave was identified within a latency range of 350–500 ms at Pz where P3 amplitude is dominant topographically.

Each measure was subjected to a repeated measures analysis of variance (ANOVA) with sound type (full-range vs. high-cut) and epoch (pre-music, 0–100, 100–200, 200–300, 300–400 s, and post-music for EEG data; 0–200 and 200–400 s for ERP data) as factors. To compensate for possible type I error inflation by the violation of sphericity, multivariate ANOVA solutions are reported (Vasey and Thayer, 1987). The significance level was set at 0.05. For *post hoc* multiple comparisons of means, the comparison-wise level of significance was determined by the Bonferroni method.

## RESULTS

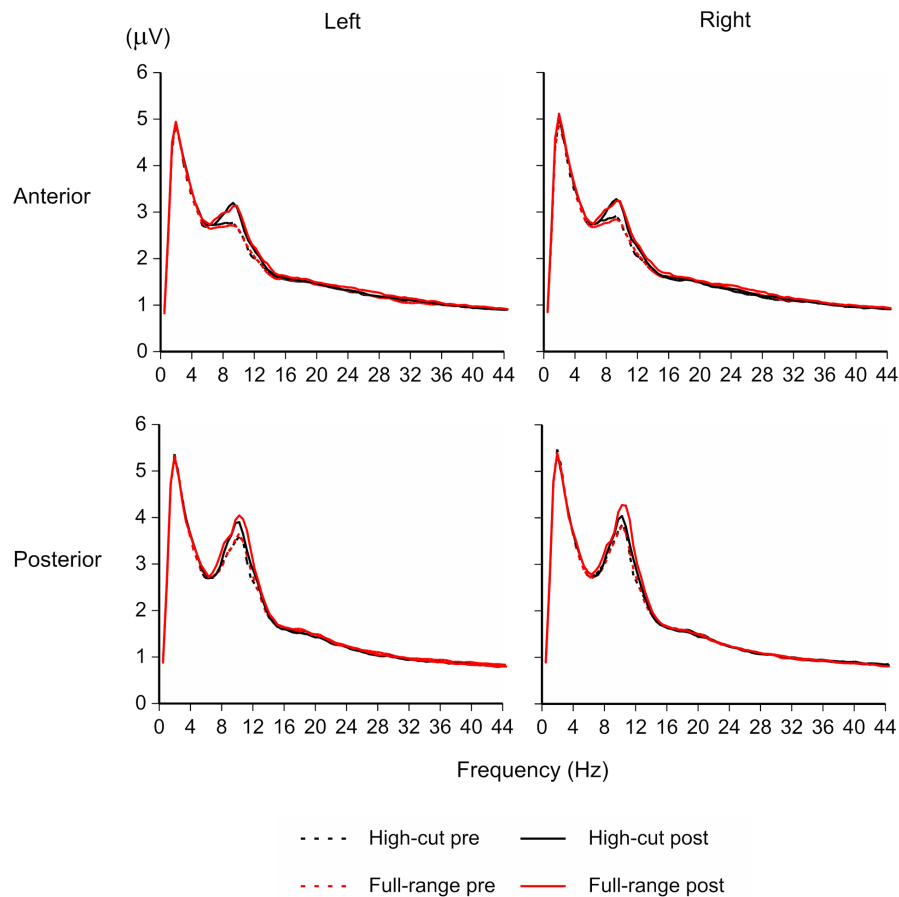
### EEG Measures

Figure 1 shows the EEG amplitude spectrogram for the four regions in the silent conditions (pre- and post-music periods). Although participants were performing a visual vigilance task with eyes opened, a peak around 10 Hz appears clearly. The amplitude of the peak appears to be increased after listening to music, in particular after listening to the full-range version of the musical piece.

Figure 2 shows the time course and scalp topography of high-alpha EEG (10.5–13 Hz) and low-beta EEG (13–20 Hz) bands. For EEG measures, a Sound Type  $\times$  Epoch  $\times$  Anterior–Posterior  $\times$  Hemisphere ANOVA was conducted for each frequency band. Significant effects of sound type were found for both bands. For other frequency bands, only the theta EEG band (4–8 Hz) power showed a significant Sound Type  $\times$  Anterior–Posterior  $\times$  Hemisphere interaction,  $F(1,21) = 5.37$ ,  $p = 0.031$ ,  $\eta_p^2 = 0.20$ . However, no significant simple main effects were found.

For high-alpha EEG band, the Sound Type  $\times$  Epoch  $\times$  Hemisphere interaction was significant,  $F(5,17) = 7.06$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.67$ . Separate ANOVAs for each epoch revealed a significant Sound Type  $\times$  Hemisphere interaction at the 200–300-s epoch,  $F(1,21) = 12.63$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.38$ , and a significant effect of sound type at the post-music period,  $F(1,21) = 6.99$ ,  $p = 0.015$ ,  $\eta_p^2 = 0.25$ . *Post hoc* tests revealed that high-alpha EEG power was greater for the full-range excerpt than for the high-cut excerpt and that the sound type effect was found for the left but not right hemisphere at the 200–300-s epoch. No effects of sound type were obtained at the epochs before 200 s. The main effect of anterior–posterior was also significant,  $F(1,21) = 15.67$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.43$ , showing that the high-alpha EEG was dominant over posterior scalp sites.

For low-beta EEG band, the Sound Type  $\times$  Anterior–Posterior  $\times$  Hemisphere interaction and the main effect of sound



**FIGURE 1 | EEG amplitude spectrogram for the full-range and high-cut conditions in the silent periods (100-s epochs before and after listening to music).** The amplitude of the peak around 10 Hz was increased after listening to music, in particular after listening to the full-range version of the sound source that contains high-frequency components.

type effect were significant,  $F(1,21) = 4.49$ ,  $p = 0.046$ ,  $\eta_p^2 = 0.18$ ;  $F(1,21) = 5.43$ ,  $p = 0.030$ ,  $\eta_p^2 = 0.21$ . Low-beta EEG power was greater in the full-range condition than in the high-cut condition. Separate ANOVAs for anterior-posterior and hemisphere also revealed significant effects of sound type, for posterior region:  $F(1,21) = 7.07$ ,  $p = 0.015$ ,  $\eta_p^2 = 0.25$ ; for left hemisphere:  $F(1,21) = 5.26$ ,  $p = 0.032$ ,  $\eta_p^2 = 0.20$ ; for right hemisphere:  $F(1,21) = 5.27$ ,  $p = 0.032$ ,  $\eta_p^2 = 0.20$ ; except for anterior region:  $F(1,21) = 3.94$ ,  $p = 0.060$ ,  $\eta_p^2 = 0.16$ . Although there were no significant interaction effects including epoch, **Figure 2** shows that the difference between the full-range and high-cut excerpts seems to be more prominent at later epochs. Two-tailed  $t$ -tests revealed significant differences between the two excerpts at the 200–300-s, 300–400-s, and post epochs,  $ts(21) > 2.37$ ,  $ps < 0.027$ ;  $p > 0.114$  at the epochs before 200 s.

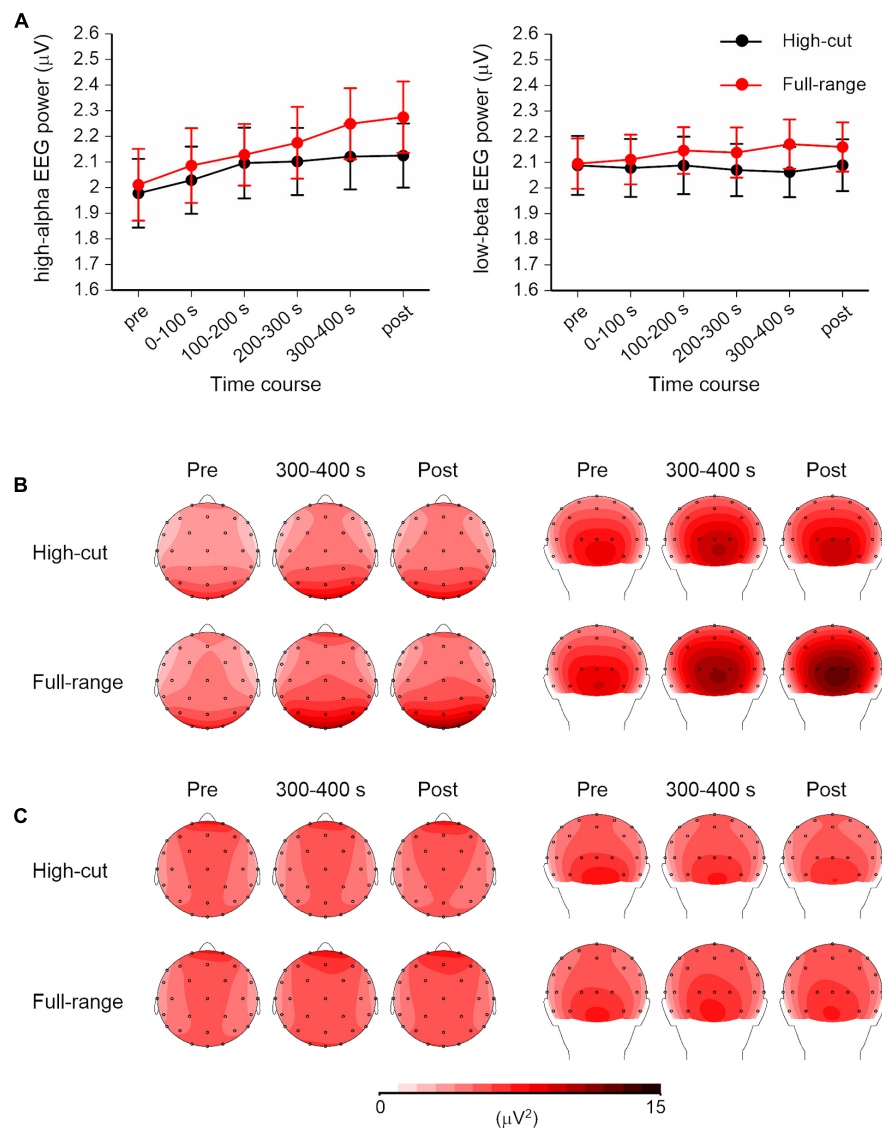
## Grand Mean ERPs for the Visual Vigilance Task

**Figure 3** shows grand mean ERP waveforms and the scalp topography of the Go and NoGo P3 amplitudes. The mean

number of averaged trials was 18.6 (range = 13–20). Although this is less than an optimal number of averages for P3 (Cohen and Polich, 1997), P3 peaks can be detected for all individual ERP waveforms. **Table 1** shows the mean amplitudes and latencies of the P3 peaks.

For the P3 amplitude, a Sound Type  $\times$  Epoch ANOVA was conducted for Go and NoGo stimulus conditions separately. A significant interaction was found for the Go condition,  $F(1,21) = 4.39$ ,  $p = 0.049$ ,  $\eta_p^2 = 0.17$ , but not for the NoGo condition,  $F(1,21) = 2.64$ ,  $p = 0.119$ ,  $\eta_p^2 = 0.11$ . *Post hoc* tests revealed that Go P3 amplitude increased from the 0–200 s to the 200–400 s epoch for the full-range excerpt, whereas Go P3 amplitude did not change for the high-cut excerpt. The main effect of epoch was significant for the NoGo condition,  $F(1,21) = 13.39$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.39$ , showing that NoGo P3 amplitude decreased during the task for both musical excerpts.

Similar ANOVAs were conducted for latencies. No significant main or interaction effects of sound type were found. The main effect of epoch was significant for the Go stimulus condition,  $F(1,21) = 5.01$ ,  $p = 0.036$ ,  $\eta_p^2 = 0.19$ , showing that Go P3 latency increased through the task.



**FIGURE 2 | High-alpha (10.5–13 Hz) and low-beta (13–20 Hz) EEG powers for the full-range and high-cut conditions. (A)** Time course of the square root of EEG powers over four scalp regions. Error bars show SEs. No music was played in the pre- and post-music epochs. **(B,C)** Scalp topography of the high-alpha and low-beta EEG powers in the pre-music, 300–400 s (i.e., last quarter of the listening period), and post-music epochs. Left: top view. Right: back view.

One of the reviewers questioned about the effects of sound type on the Nogo N2 (Falkenstein et al., 1999). We conducted a Sound Type  $\times$  Epoch ANOVA on the amplitude of the Nogo N2 (Nogo minus Go in the 200–300 ms period at Fz and Cz). No significant main or interaction effects were found.

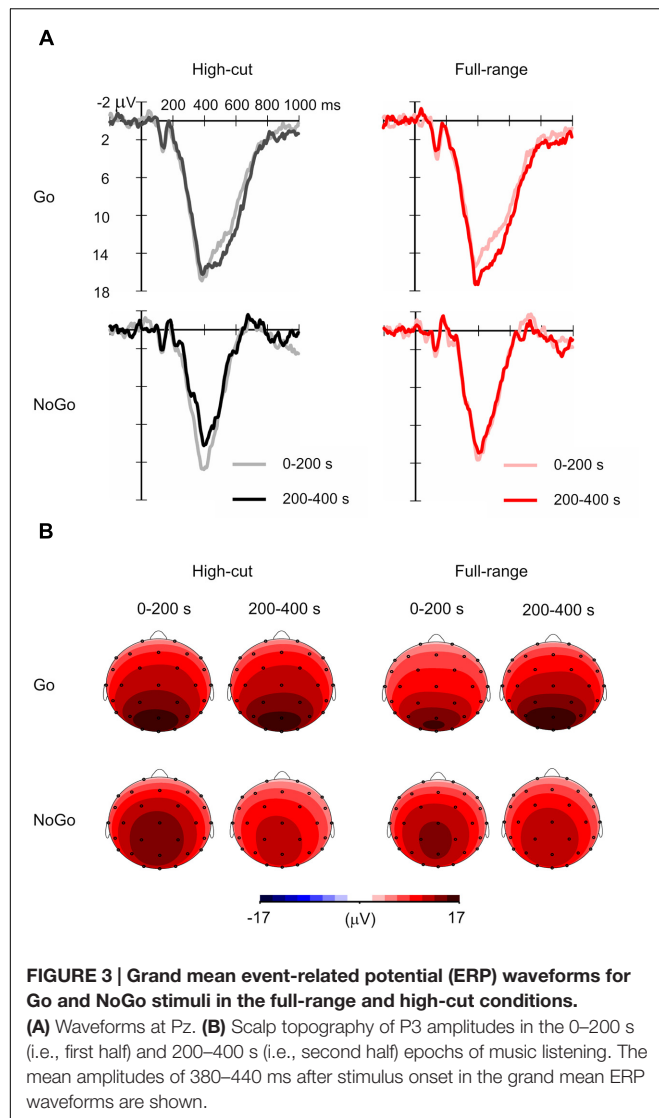
## Behavioral and Other Physiological Measures

Participants performed the vigilance task with considerable accuracy (high-cut:  $M = 98.6\%$ , 95.8–100%; full-range:  $M = 97.9\%$ , 95.0–99.2%). **Figure 4** shows the time course of mean Go reaction times, HR, and facial EMGs (corrugator supercilii, zygomaticus major), and the HRV components for the last 300-s epoch of the musical excerpts. For the corrugator

supercilii, a Sound Type  $\times$  Epoch ANOVA showed a significant main effect of epoch,  $F(5,17) = 5.69$ ,  $p = 0.003$ ,  $\eta_p^2 = 0.63$ . Corrugator activity increased over the course of the task. No significant main or interaction effects of sound type were found for RT or other physiological measures.

## Subjective Ratings

**Table 2** shows mean scores for participants' mood states. A significant difference between the two types of musical excerpt was found only for inactive pleasantness scores,  $t(21) = 3.13$ ,  $p = 0.005$ . Participants provided higher inactive pleasantness scores under the full-range than under the high-cut excerpt. **Figure 5** shows the mean sound quality ratings for the full-range and high-cut musical excerpts. No significant differences were



found between the two types of audio source for any adjective pairs,  $t(21) < 1.92$ ,  $ps > 0.069$ . The correct rate of the forced choices was 41.0%, which did not exceed chance level ( $p = 0.523$ , binomial test).

## DISCUSSION

High-resolution audio with inaudible high-frequency components is a closer replication of real sounds than similar and indistinguishable sounds in which these components are artificially cut off. It remains unclear what kind of advantages high-resolution audio might have for human beings. Previous research in which participants listened to high-resolution music under resting conditions have shown that alpha and low-beta EEG powers were larger for an excerpt with high-frequency components as compared with an excerpt without them (Oohashi et al., 2000, 2006; Yagi et al., 2003a; Fukushima et al., 2014; Kuribayashi et al., 2014; Ito et al., 2016). The present

**TABLE 1 | The peak amplitudes and latencies of the P3 component at Pz.**

		High-cut		Full-range	
		0–200 s	200–400 s	0–200 s	200–400 s
Go	Amplitude ( $\mu V$ )	18.3 (6.4)	18.0 (5.7)	16.7 (6.3)	18.6 (6.3)
	Latency (ms)	404.9 (37.2)	419.5 (42.2)	405.1 (34.8)	420.2 (41.7)
No Go	Amplitude ( $\mu V$ )	16.3 (6.6)	13.1 (5.3)	15.1 (6.9)	14.0 (7.1)
	Latency (ms)	408.0 (35.0)	412.2 (35.7)	417.8 (41.2)	413.5 (27.9)

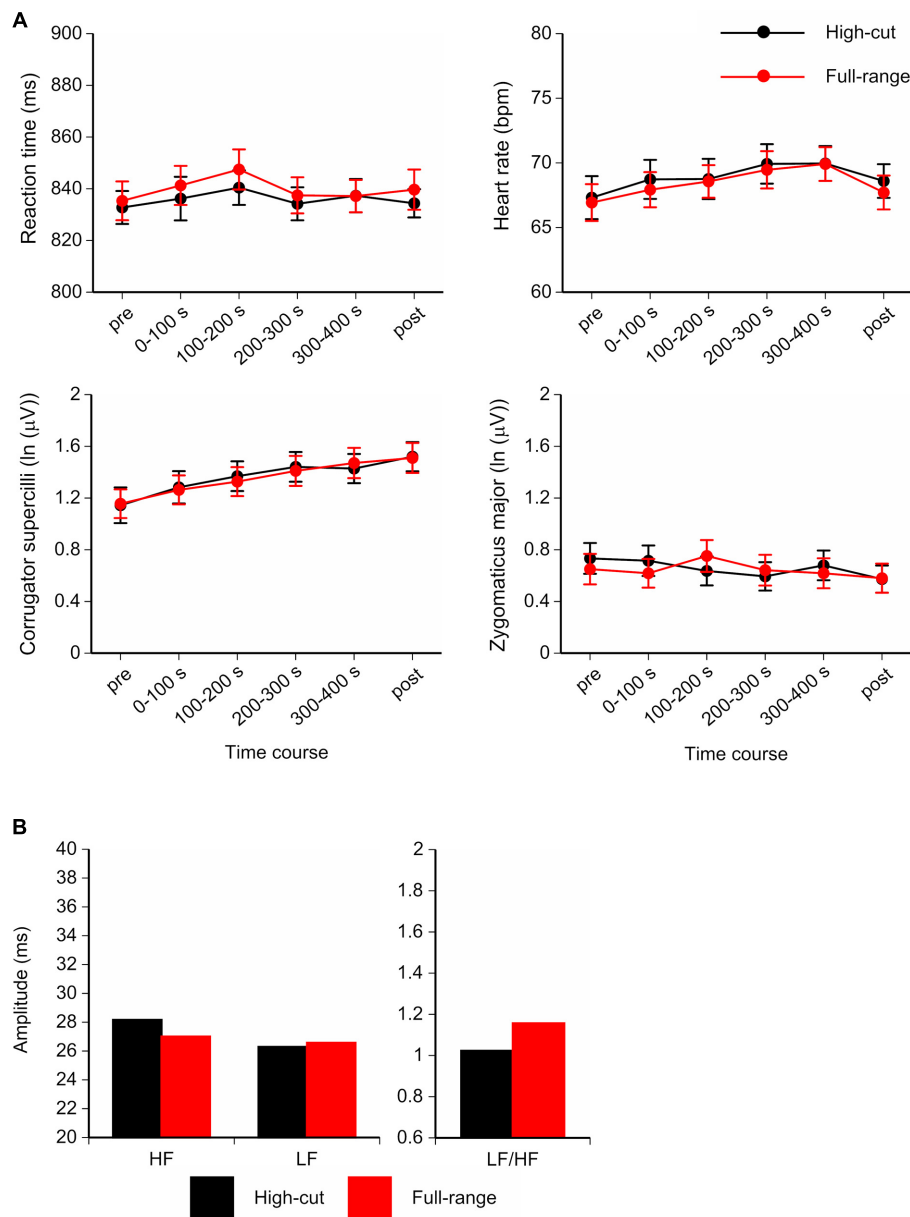
Values in brackets are SD.

study asked participants to listen to two types of high-resolution audio of the same musical piece (with or without inaudible high-frequency components) while performing a vigilance task in the visual modality. Although the effect size is small, the overall results support the view that the effect of high-resolution audio with inaudible high-frequency components on brain activity reflects a relaxed attentional state without conscious awareness.

We found greater high-alpha (10.5–13 Hz) and low-beta (13–20 Hz) EEG powers for the excerpt with high-frequency components as compared with the excerpt without them. The effect appeared in the latter half of the listening period (200–400 s) and during the 100-s period after music presentation (post-music epoch). Furthermore, for full-range sounds compared with high-cut sounds, Go trial P3 amplitude increased, and subjective relaxation scores were greater. Because task performance did not change across musical excerpts, with no difference in self-reported arousal, the effects of high-resolution audio with inaudible high-frequency components on brain activities should not reflect a decrease of listeners' arousal level. These findings show that listeners seem to experience a relaxed attentional state when listening to high-resolution audio with inaudible high-frequency components compared to similar sounds without these components.

It has been shown that listening to musical pieces increases EEG powers of theta, alpha, and beta bands (Pavlygina et al., 2004; Jäncke et al., 2015), and that the enhanced alpha-band power holds for approximately 100 s after listening (Sanyal et al., 2013). Therefore, high-resolution audio with inaudible high-frequency components would be advantageous compared to a similar digital audio in which these components are removed, in terms of the enhanced brain activity. Kuribayashi and Nittono (2014) have localized the intracerebral sources of this alpha EEG effect using standardized low-resolution brain electromagnetic tomography (sLORETA). The analysis revealed that the difference between full-range and high-cut sounds appeared in the right inferior temporal cortex, whereas the main source of the alpha-band activity was located in the parietal-occipital region. The finding that the alpha-band activity difference was obtained in specific but not whole regions is suggestive that this increase may reflect an activity related to task performance rather than a global arousal effect (Barry et al., 2007).

The present study shows that not only high-alpha and low-beta EEG powers but also P3 amplitude increased in the last half of the listening period (200–400 s). Alpha-band EEG activity and



**FIGURE 4 | (A)** Time course of the mean Go reaction times, Heart rate (HR), and facial electromyograms (EMGs; corrugator supercilii and zygomaticus major) for the full-range and high-cut conditions. **(B)** Amplitudes and ratio of heart rate variability (HRV) components for the last 300-s epoch of musical excerpts. Error bars show SEs.

P3 amplitude have been shown to be positively correlated, in such a way that prestimulus alpha directly modulates positive potential amplitude in an auditory equiprobable Go/NoGo task (Barry et al., 2000; De Blasio et al., 2013). P3 amplitude is larger when greater attentional resources are allocated to the eliciting stimulus (Kok, 1997, 2001; Polich, 2007). Alpha power is increased in tasks requiring a relaxed attentional state such as mindfulness and imagination of music (Cooper et al., 2006; Schaefer et al., 2011; Lomas et al., 2015). Increased alpha power is thought to be a signifier of enhanced processing, with attention focused on

internally generated stimuli (Lomas et al., 2015). Beta power has been shown to increase when arousal and vigilance level increase (e.g., Sebastiani et al., 2003; Aftanas et al., 2006; Gola et al., 2012; Kamiński et al., 2012). Taken together, the EEG and ERP results support the idea that listening to high-resolution audio with inaudible high-frequency components enhances the cortical activity related to the attention allocated to task-relevant stimuli. Although the effect was not observed in behavior, the gap between behavioral and EEG and ERP results is probably due to the ceiling effect of the vigilance task performance. Such a gap is often

**TABLE 2 | Mean scores and the results of two-tailed *t*-tests for participants' mood states.**

	High-cut	Full-range	<i>t</i> (21)	<i>p</i>
Affect Grid (9-point scale, 1–9)				
Pleasantness	6.3 (1.4)	6.4 (1.0)	0.32	0.754
Arousal	4.9 (2.3)	4.4 (2.1)	1.15	0.264
Multiple mood scale (4-point scale, 1–4)				
Depression	1.5 (0.5)	1.5 (0.5)	0.00	1.00
Aggression	1.1 (0.3)	1.1 (0.3)	0.40	0.690
Fatigue	1.9 (0.6)	1.8 (0.5)	1.27	0.219
Active	1.8 (0.6)	1.8 (0.5)	0.09	0.929
pleasantness				
Inactive	2.7 (0.6)	3.0 (0.6)	3.13	0.005
pleasantness				
Affinity	1.7 (0.6)	1.7 (0.7)	0.66	0.518
Concentration	2.1 (0.5)	2.2 (0.6)	1.50	0.148
Startle	1.5 (0.5)	1.4 (0.5)	0.86	0.400

Values in brackets are SD.

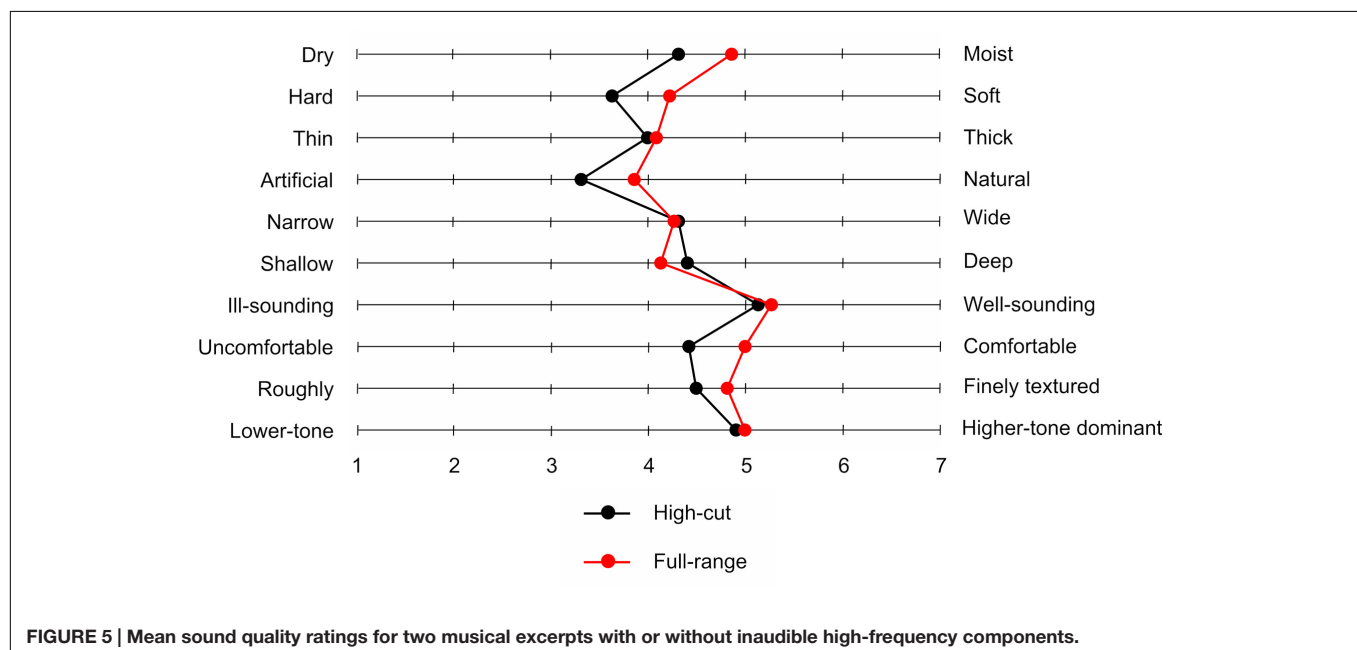
observed in other studies. For example, Okamoto and Nakagawa (2016) similarly reported that event-related synchronization in the alpha band during working memory task was increased 20–30 min after the onset of the exposure to blue (short-wavelength) light, as compared with green (middle-wavelength) light, while task performance was high irrespective of light colors.

As a mechanism underlying the effect of inaudible high-frequency sound components, we speculate that the brain may subconsciously recognize high-resolution audio that retains high-frequency components as being more natural, as compared with similar sounds in which such components

are artificially removed. A link between alpha power and ratings of 'naturalness' of music has been reported. When listening to the same musical piece with different tempos, alpha-band EEG power increased for excerpts that were rated to be more natural, the ratings of which were not directly related to subjective arousal (Ma et al., 2012; Tian et al., 2013). As high-resolution audio replicates real sound waves more closely, it may sound more natural (at least on a subconscious level) and facilitate music-related psychophysiological responses.

Our findings have some limitations. First, because we used only a visual vigilance task, it is unclear whether high-resolution audio can improve performance on tasks that involve working memory and long-term memory. Because a vigilance task is relatively easy, our participants were able to sustain high performance. Other research using an n-back task requiring memory has shown that high-resolution audio also enhances task performance (Suzuki, 2013). Future research will benefit from using other tasks requiring various cognitive domains and processes.

Second, the underlying mechanism of how inaudible high-frequency components affect EEG activities cannot be revealed by the current data. It is noteworthy that presenting high-frequency components above 20 kHz alone did not produce any change in EEG activities (Oohashi et al., 2000). Therefore, the combination of inaudible high-frequency components and audible low-frequency components should be a key factor that causes this phenomenon. A possible clue was obtained by a recent study of Kuribayashi and Nittono (2015). Recording sound spectra of various musical instruments, they found that high-frequency components above 20 kHz appear abundantly during the rising period of a sound wave (i.e., from the silence to the maximal intensity, usually less than 0.1 s), but occur much less after that. Artificially cutting off the high-frequency components

**FIGURE 5 | Mean sound quality ratings for two musical excerpts with or without inaudible high-frequency components.**

may cause a subtle distortion in this short period. It will take some time to accumulate these small, short-lasting differences until they produce discernible psychophysiological effects. This explanation is consistent with the fact that the effect of high-frequency components on EEG activities appears only after a 100–200-s exposure to the music (Oohashi et al., 2000, 2006; Yagi et al., 2003a; Fukushima et al., 2014; Kuribayashi et al., 2014; Ito et al., 2016).

Third, it remains unclear why there was a time lag until the effects of high-resolution audio on brain activity show up, and why this effect was maintained for 100 s after music stopped. A possible reason is that, as mentioned above, sufficiently long exposure is needed for the effects of inaudible high-frequency components. Another possibility is that listening to music has psychophysiological impact through the engagement of various neurochemical systems (Chanda and Levitin, 2013). Humoral effects are characterized by slow and durable responses, which might be underlying the lagged effect of high-resolution audio with inaudible high-frequency components. Although the present study did not reveal this effect on autonomic nervous system (HR and HRV) indices during music listening, participants reported greater relaxation scores after listening to high-resolution music with inaudible high-frequency components. It is a task for future research to determine the time course of the effect more precisely.

Fourth, the present study did not manipulate the sampling frequency and the bit depth of digital audio. High-resolution audio is characterized not only by the capability of reproducing inaudible high-frequency components but also by more accurate sampling and quantization (i.e., a higher sampling frequency and a greater bit depth) as compared with low-resolution audio. If the naturalness derived by a closer replication of real sounds affects EEG activities, the sampling frequency and the bit depth would do too regardless of whether the real sounds feature high-frequency components. This idea would be worth examining in future research.

In summary, high-resolution audio with inaudible high-frequency components has some advantages over similar

and indistinguishable sounds in which these components are artificially cut off, such that the former type of digital audio induces a relaxed attentional state. Even without conscious awareness, a closer replication of real sounds in terms of frequency structure appears to bring out greater potential effects of music on human psychophysiological state and behavior.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of The Research Ethics Committee of the Graduate School of Integrated Arts and Sciences in Hiroshima University. All participants gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by The Research Ethics Committee of the Graduate School of Integrated Arts and Sciences in Hiroshima University.

## AUTHOR CONTRIBUTIONS

RK and HN planned the experiment, interpreted the data, and wrote the paper. RK collected and analyzed the data.

## FUNDING

This work was supported by JSPS KAKENHI Grant Number 15J06118.

## ACKNOWLEDGMENTS

The authors thank Ryuta Yamamoto, Katsuyuki Niyada, Kazushi Uemura, and Fujio Iwaki for their support as research coordinators. Hiroshima Innovation Center for Biomedical Engineering and Advanced Medicine offered the sound equipment.

## REFERENCES

- Aftanas, L. I., Reva, N. V., Savotina, L. N., and Makhnev, V. P. (2006). Neurophysiological correlates of induced discrete emotions in humans: an individually oriented analysis. *Neurosci. Behav. Physiol.* 36, 119–130. doi: 10.1007/s11055-005-0170-6
- Andreasen, N. C., Arndt, S., Swayze, V., Cizadlo, T., Flaum, M., O'Leary, D., et al. (1994). Thalamic abnormalities in schizophrenia visualized through magnetic resonance image averaging. *Science* 266, 294–298. doi: 10.1126/science.7939669
- Barry, R. J., Clarke, A. R., Johnstone, S. J., Magee, C. A., and Rushby, J. A. (2007). EEG differences between eyes-closed and eyes-open resting conditions. *Clin. Neurophysiol.* 118, 2765–2773. doi: 10.1016/j.clinph.2007.07.028
- Barry, R. J., Kirkaikul, S., and Hodder, D. (2000). EEG alpha activity and the ERP to target stimuli in an auditory oddball paradigm. *Int. J. Psychophysiol.* 39, 39–50. doi: 10.1016/S0167-8760(00)00114-8
- Başar, E. (1999). *Brain Function and Oscillations: Integrative Brain Function. Neurophysiology and Cognitive Processes*, Vol. II. Berlin: Springer, doi: 10.1007/978-3-642-59893-7
- Berntson, G. G., Bigger, J. T., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., et al. (1997). Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology* 34, 623–648. doi: 10.1111/j.1469-8986.1997.tb02140.x
- 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
- Brown, S., Martinez, M. J., and Parsons, L. M. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. *Neuroreport* 15, 2033–2037. doi: 10.1097/00001756-200409150-00008
- Chanda, M. L., and Levitin, D. J. (2013). The neurochemistry of music. *Trends Cogn. Sci.* 17, 179–193. doi: 10.1016/j.tics.2013.02.007
- Cohen, J., and Polich, J. (1997). On the number of trials needed for P300. *Int. J. Psychophysiol.* 25, 249–255. doi: 10.1016/S0167-8760(96)00743-X
- Cooper, N. R., Burgess, A. P., Croft, R. J., and Gruzelier, J. H. (2006). Investigating evoked and induced electroencephalogram activity in task-related alpha power increases during an internally directed attention task. *Neuroreport* 17, 205–208. doi: 10.1097/01.wnr.0000198433.29389.54

- Crone, N. E., Miglioretti, D. L., Gordon, B., Sieracki, J. M., Wilson, M. T., Uematsu, S., et al. (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. I. Alpha and beta event-related desynchronization. *Brain* 121, 2271–2299. doi: 10.1093/brain/121.12.2271
- Davies, D. R., and Parasuraman, R. (1977). "Cortical evoked potentials and vigilance: a decision theory analysis," in *NATO Conference Series. Vigilance*, Vol. 3, ed. R. Mackie (New York, NY: Plenum Press), 285–306. doi: 10.1007/978-1-4684-2529-1
- De Blasio, F. M., Barry, R. J., and Steiner, G. Z. (2013). Prestimulus EEG amplitude determinants of ERP responses in a habituation paradigm. *Int. J. Psychophysiol.* 89, 444–450. doi: 10.1016/j.ijpsycho.2013.05.015
- Falkenstein, M., Hoormann, J., and Hohnsbein, J. (1999). ERP components in Go/Nogo tasks and their relation to inhibition. *Acta Psychol. (Amst.)* 101, 267–291. doi: 10.1016/S0001-6918(99)00008-6
- Foxe, J. J., and Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front. Psychol.* 2:154. doi: 10.3389/fpsyg.2011.00154
- Fridlund, A. J., and Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology* 23, 567–589. doi: 10.1111/j.1469-8986.1986.tb00676.x
- Fruhstorfer, H., and Bergström, R. M. (1969). Human vigilance and auditory evoked responses. *Electroencephalogr. Clin. Neurophysiol.* 27, 346–355. doi: 10.1016/0013-4694(69)91443-6
- Fujioka, T., Kakigi, R., Gunji, A., and Takeshima, Y. (2002). The auditory evoked magnetic fields to very high frequency tones. *Neuroscience* 112, 367–381. doi: 10.1016/S0306-4522(02)00086-6
- Fukushima, A., Yagi, R., Kawai, N., Honda, M., Nishina, E., and Oohashi, T. (2014). Frequencies of inaudible high-frequency sounds differentially affect brain activity: positive and negative hypersonic effects. *PLoS ONE* 9:e95464. doi: 10.1371/journal.pone.0095464
- Gola, M., Kamiński, J., Brzezicka, A., and Wróbel, A. (2012). Beta band oscillations as a correlate of alertness — Changes in aging. *Int. J. Psychophysiol.* 85, 62–67. doi: 10.1016/j.ijpsycho.2011.09.00
- Hari, R., and Salmelin, R. (1997). Human cortical oscillations: a neuromagnetic view through the skull. *Trends. Neurosci.* 20, 44–49. doi: 10.1016/S0166-2236(96)10065-5
- Ito, S., Harada, T., Miyaguchi, M., Ishizaki, F., Chikamura, C., Kodama, Y., et al. (2016). Effect of high-resolution audio music box sound on EEG. *Int. Med. J.* 23, 1–3.
- Jäncke, L., Kühnis, J., Rogenmoser, L., and Elmer, S. (2015). Time course of EEG oscillations during repeated listening of a well-known aria. *Front. Hum. Neurosci.* 9:401. doi: 10.3389/fnhum.2015.00401
- Jeffries, K. J., Fritz, J. B., and Braun, A. R. (2003). Words in melody: an h(2)15o pet study of brain activation during singing and speaking. *Neuroreport* 14, 749–754. doi: 10.1097/01.wnr.0000066198.94941.a4
- Jensen, O., and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* 4:186. doi: 10.3389/fnhum.2010.00186
- Kamiński, J., Brzezicka, A., Gola, M., and Wróbel, A. (2012). Beta band oscillations engagement in human alertness process. *Int. J. Psychophysiol.* 85, 125–128. doi: 10.1016/j.ijpsycho.2011.11.006
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195. doi: 10.1016/S0165-0173(98)00056-3
- Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617. doi: 10.1016/j.tics.2012.10.007
- Klimesch, W., Schack, B., and Sauseng, P. (2005). The functional significance of theta and upper alpha oscillations. *Exp. Psychol.* 52, 99–108. doi: 10.1027/1618-3169.52.2.99
- Kok, A. (1997). Event-related-potential (ERP) reflections of mental resources: a review and synthesis. *Biol. Psychol.* 45, 19–56. doi: 10.1016/S0301-0511(96)05221-0
- Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology* 38, 557–577. doi: 10.1017/S004857720190559
- Kuribayashi, R., and Nittono, H. (2014). Source localization of brain electrical activity while listening to high-resolution digital sounds with inaudible high-frequency components (Abstract). *Int. J. Psychophysiol.* 94:192. doi: 10.1016/j.ijpsycho.2014.08.796
- Kuribayashi, R., and Nittono, H. (2015). Music instruments that produce sounds with inaudible high-frequency components (in Japanese). *Stud. Hum. Sci.* 10, 35–41. doi: 10.15027/39146
- Kuribayashi, R., Yamamoto, R., and Nittono, H. (2014). High-resolution music with inaudible high-frequency components produces a lagged effect on human electroencephalographic activities. *Neuroreport* 25, 651–655. doi: 10.1097/wnr.0000000000000151
- Larsen, J. T., Norris, C. J., and Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology* 40, 776–785. doi: 10.1111/1469-8986.00078
- LeDoux, J. E. (1993). Emotional memory systems in the brain. *Behav. Brain Res.* 58, 69–79. doi: 10.1016/0166-4328(93)90091-4
- Lomas, T., Ivtzan, I., and Fu, C. H. (2015). A systematic review of the neurophysiology of mindfulness on EEG oscillations. *Neurosci. Biobehav. Rev.* 57, 401–410. doi: 10.1016/j.neubiorev.2015.09.018
- Lupien, S. J., McEwen, B. S., Gunnar, M. R., and Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat. Rev. Neurosci.* 10, 434–445. doi: 10.1038/nrn2639
- Ma, W., Lai, Y., Yuan, Y., Wu, D., and Yao, D. (2012). Electroencephalogram variations in the  $\alpha$  band during tempo-specific perception. *Neuroreport* 23, 125–128. doi: 10.1097/WNR.0b013e32834e7eac
- Malliani, A., and Montano, N. (2002). Heart rate variability as a clinical tool. *Ital. Heart J.* 3, 439–445.
- Malliani, A. I., Pagani, M., Lombardi, F., and Cerutti, S. (1991). Cardiovascular neural regulation explored in the frequency domain. *Circulation* 84, 482–489. doi: 10.1161/01.CIR.84.2.482
- Muraoka, T., Iwahara, M., and Yamada, Y. (1981). Examination of audio-bandwidth requirements for optimum sound signal transmission. *J. Audio Eng. Soc.* 29, 2–9.
- Nishiguchi, T., Hamasaki, K., Ono, K., Iwaki, M., and Ando, A. (2009). Perceptual discrimination of very high frequency components in wide frequency range musical sound. *Appl. Acoust.* 70, 921–934. doi: 10.1016/j.apacoust.2009.01.002
- Okamoto, Y., and Nakagawa, S. (2016). Effects of light wavelength on MEG ERD/ERS during a working memory task. *Int. J. Psychophysiol.* 104, 10–16. doi: 10.1016/j.ijpsycho.2016.03.008
- Oohashi, T., Kawai, N., Nishina, E., Honda, M., Yagi, R., Nakamura, S., et al. (2006). The role of biological system other than auditory air-conduction in the emergence of the hypersonic effect. *Brain Res.* 107, 339–347. doi: 10.1016/j.brainres.2005.12.096
- Oohashi, T., Nishina, E., Honda, M., Yonekura, Y., Fuwamoto, Y., Kawai, N., et al. (2000). Inaudible high-frequency sounds affect brain activity: hypersonic effect. *J. Neurophysiol.* 83, 3548–3558.
- Parasuraman, R. (1983). "Vigilance, arousal, and the brain," in *Physiological Correlates of Human Behavior*, eds A. Gale and J. A. Edwards (London: Academic Press), 35–55.
- Pavlygina, R. A., Sakharov, D. S., and Davydov, V. I. (2004). Spectral analysis of the human EEG during listening to musical compositions. *Hum. Physiol.* 30, 54–60. doi: 10.1023/B:HUMP.0000013765.64276.e6
- Pfurtscheller, G., and Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/S1388-2457(99)00141-8
- Pfurtscheller, G., Woertz, M., Supp, G., and Lopes da Silva, F. H. (2003). Early onset of post-movement beta electroencephalogram synchronization in the supplementary motor area during self-paced finger movement in man. *Neurosci. Lett.* 339, 111–114. doi: 10.1016/S0304-3940(02)01479-9
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019
- Russell, J. A., Weiss, A., and Mendelsohn, G. A. (1989). Affect grid – a single-item scale of pleasure and arousal. *J. Pers. Soc. Psychol.* 57, 493–502. doi: 10.1037/0022-3514.57.3.493
- Sanyal, S., Banerjee, A., Guhathakurta, T., Sengupta, R., Ghosh, D., and Ghose, P. (2013). "EEG study on the neural patterns of brain with music stimuli: an

- evidence of Hysteresis?," in *Proceedings of the International Seminar on 'Creating and Teaching Music Patterns'*, (Kolkata: Rabindra Bharati University), 51–61.
- Schaefer, R. S., Vlek, R. J., and Desain, P. (2011). Music perception and imagery in EEG: alpha band effects of task and stimulus. *Int. J. Psychophysiol.* 82, 254–259. doi: 10.1016/j.ijpsycho.2011.09.007
- Sebastiani, L., Simoni, A., Gemignani, A., Ghelarducci, B., and Santarcangelo, E. L. (2003). Autonomic and EEG correlates of emotional imagery in subjects with different hypnotic susceptibility. *Brain Res. Bull.* 60, 151–160. doi: 10.1016/S0361-9230(03)00025-X
- Suzuki, K. (2013). Hypersonic effect and performance of recognition tests (in Japanese). *Kagaku* 83, 343–345.
- Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-Aho, P. O., and Karjalainen, P. A. (2014). Kubios HRV—heart rate variability analysis software. *Comput. Methods Programs Biomed.* 113, 210–220. doi: 10.1016/j.cmpb.2013.07.024
- Terasaki, M., Kishimoto, Y., and Koga, A. (1992). Construction of a multiple mood scale (In Japanese). *Shinrigaku Kenkyu* 62, 350–356. doi: 10.4992/jjpsy.62.350
- Tian, Y., Ma, W., Tian, C., Xu, P., and Yao, D. (2013). Brain oscillations and electroencephalography scalp networks during tempo perception. *Neurosci. Bull.* 29, 731–736. doi: 10.1007/s12264-013-1352-9
- Vasey, M. W., and Thayer, J. F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: a multivariate solution. *Psychophysiology* 24, 479–486. doi: 10.1111/j.1469-8986.1987.tb00324.x
- Vogt, B. A., and Gabriel, M. (1993). *Neurobiology of Cingulate Cortex and Limbic Thalamus. A Comprehensive Handbook*. Boston, MA: Birkhauser.
- Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends Cogn. Sci.* 7, 553–559. doi: 10.1016/j.tics.2003.10.012
- Weisz, N., Hartmann, T., Muller, N., Lorenz, I., and Obleser, J. (2011). Alpha rhythms in audition: cognitive and clinical perspectives. *Front. Psychol.* 2:73. doi: 10.3389/fpsyg.2011.00073
- Yagi, R., Nishina, E., Honda, M., and Oohashi, T. (2003a). Modulatory effect of inaudible high-frequency sounds on human acoustic perception. *Neurosci. Lett.* 351, 191–195. doi: 10.1016/j.neulet.2003.07.020
- Yagi, R., Nishina, E., and Oohashi, T. (2003b). A method for behavioral evaluation of the "hypersonic effect". *Acoust. Sci. Technol.* 24, 197–200. doi: 10.1250/ast.24.197
- Yagi, T., Kawai, N., Nishina, E., Honda, M., Yagi, R., Nakamura, S., et al. (2006). The role of biological system other than auditory air-conduction in the emergence of the hypersonic effect. *Brain Res.* 1073–1074, 339–347. doi: 10.1016/j.brainres.2005.12.096

**Conflict of Interest Statement:** 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.

Copyright © 2017 Kuribayashi and Nittono. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Emotional Responses to Music: Shifts in Frontal Brain Asymmetry Mark Periods of Musical Change

Hussain-Abdulah Arjmand<sup>1</sup>, Jesper Hohagen<sup>2</sup>, Bryan Paton<sup>3</sup> and Nikki S. Rickard<sup>1,4\*</sup>

<sup>1</sup> School of Psychological Sciences, Monash University, Melbourne, VIC, Australia, <sup>2</sup> Institute for Systematic Musicology, University of Hamburg, Hamburg, Germany, <sup>3</sup> Monash Biomedical Imaging, Monash University, University of Newcastle, Newcastle, NSW, Australia, <sup>4</sup> Centre for Positive Psychology, Graduate School of Education, University of Melbourne, Melbourne, VIC, Australia

## OPEN ACCESS

### Edited by:

Mark Reybrouck,  
KU Leuven, Belgium

### Reviewed by:

Tuomas Eerola,  
Durham University, United Kingdom  
Elvira Brattico,  
Aarhus University, Denmark  
Jan Wikgren,  
University of Jyväskylä, Finland

### \*Correspondence:

Nikki S. Rickard  
nikki.rickard@monash.edu

### Specialty section:

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

**Received:** 08 November 2016

**Accepted:** 08 November 2017

**Published:** 04 December 2017

### Citation:

Arjmand H-A, Hohagen J, Paton B  
and Rickard NS (2017) Emotional  
Responses to Music: Shifts in Frontal  
Brain Asymmetry Mark Periods  
of Musical Change.  
Front. Psychol. 8:2044.  
doi: 10.3389/fpsyg.2017.02044

Recent studies have demonstrated increased activity in brain regions associated with emotion and reward when listening to pleasurable music. Unexpected change in musical features intensity and tempo – and thereby enhanced tension and anticipation – is proposed to be one of the primary mechanisms by which music induces a strong emotional response in listeners. Whether such musical features coincide with central measures of emotional response has not, however, been extensively examined. In this study, subjective and physiological measures of experienced emotion were obtained continuously from 18 participants (12 females, 6 males; 18–38 years) who listened to four stimuli—pleasant music, unpleasant music (dissonant manipulations of their own music), neutral music, and no music, in a counter-balanced order. Each stimulus was presented twice: electroencephalograph (EEG) data were collected during the first, while participants continuously subjectively rated the stimuli during the second presentation. Frontal asymmetry (FA) indices from frontal and temporal sites were calculated, and peak periods of bias toward the left (indicating a shift toward positive affect) were identified across the sample. The music pieces were also examined to define the temporal onset of key musical features. Subjective reports of emotional experience averaged across the condition confirmed participants rated their music selection as very positive, the scrambled music as negative, and the neutral music and silence as neither positive nor negative. Significant effects in FA were observed in the frontal electrode pair FC3–FC4, and the greatest increase in left bias from baseline was observed in response to pleasurable music. These results are consistent with findings from previous research. Peak FA responses at this site were also found to co-occur with key musical events relating to change, for instance, the introduction of a new motif, or an instrument change, or a change in low level acoustic factors such as pitch, dynamics or texture. These findings provide empirical support for the proposal that change in basic musical features is a fundamental trigger of emotional responses in listeners.

**Keywords:** frontal asymmetry, subjective emotions, pleasurable music, musicology, positive and negative affect

## INTRODUCTION

One of the most intriguing debates in music psychology research is whether the emotions people report when listening to music are ‘real.’ Various authorities have argued that music is one of the most powerful means of inducing emotions, from Tolstoy’s mantra that “music is the shorthand of emotion,” to the deeply researched and influential reference texts of Leonard Meyer (“Emotion and meaning in music”; Meyer, 1956) and Juslin and Sloboda (“The Handbook of music and emotion”; Juslin and Sloboda, 2010). Emotions evolved as a response to events in the environment which are potentially significant for the organism’s survival. Key features of these ‘utilitarian’ emotions include goal relevance, action readiness and multicomponentiality (Frijda and Scherer, 2009). Emotions are therefore triggered by events that are appraised as relevant to one’s survival, and help prepare us to respond, for instance via fight or flight. In addition to the cognitive appraisal, emotions are also widely acknowledged to be multidimensional, yielding changes in subjective feeling, physiological arousal, and behavioral response (Scherer, 2009). The absence of clear goal implications of music listening, or any need to become ‘action ready,’ however, challenges the claim that music-induced emotions are real (Kivy, 1990; Konecni, 2013).

A growing body of ‘emotivist’ music psychology research has nonetheless demonstrated that music does elicit a response in multiple components, as observed with non-aesthetic (or ‘utilitarian’) emotions. The generation of an emotion in subcortical regions of the brain (such as the amygdala) lead to hypothalamic and autonomic nervous system activation and release of arousal hormones, such as noradrenaline and cortisol. Sympathetic nervous system changes associated with physiological arousal, such as increased heart rate and reduced skin conductance, are most commonly measured as peripheral indices of emotion. A large body of work now illustrates, under a range of conditions and with a variety of music genres, that emotionally exciting or powerful music impacts on these autonomic measures of emotion (see Bartlett, 1996; Panksepp and Bernatzky, 2002; Hodges, 2010; Rickard, 2012 for reviews). For example, Krumhansl (1997) recorded physiological (heart rate, blood pressure, transit time and amplitude, respiration, skin conductance, and skin temperature) and subjective measures of emotion in real time while participants listened to music. The observed changes in these measures differed according to the emotion category of the music, and was similar (although not identical) to that observed for non-musical stimuli. Rickard (2004) also observed coherent subjective and physiological (chills and skin conductance) responses to music selected by participants as emotionally powerful, which was interpreted as support for the emotivist perspective on music-induced emotions.

It appears then that the evidence supporting music evoked emotions being ‘real’ is substantive, despite no obvious goal implications, or need for action, of this primarily aesthetic stimulus. Scherer and Coutinho (2013) have argued that music may induce a particular ‘kind’ of emotion – aesthetic emotions – that are triggered by novelty and complexity, rather than direct relevance to one’s survival. Novelty and complexity are

nonetheless features of goal relevant stimuli, even though in the case of music, there is no significance to the listener’s survival. In the same way that secondary reinforcers appropriate the physiological systems of primary reinforcers via association, it is possible then that music may also hijack the emotion system by sharing some key features of goal relevant stimuli.

Multiple mechanisms have been proposed to explain how music is capable of inducing emotions (e.g., Juslin et al., 2010; Scherer and Coutinho, 2013). Common to most theories is an almost primal response elicited by psychoacoustic features of music (but also shared by other auditory stimuli). Juslin et al. (2010) describe how the ‘brain stem reflex’ (from their ‘BRECHEMA’ theory) is activated by changes in basic acoustic events – such as sudden loudness or fast rhythms – by tapping into an evolutionarily ancient survival system. This is because these acoustic events are associated with events that do in fact signal relevance for survival for real events (such as a nearby loud noise, or a rapidly approaching predator). Any unexpected *change* in acoustic feature, whether it be in pitch, timbre, loudness, or tempo, in music could therefore fundamentally be worthy of special attention, and therefore trigger an arousal response (Gabrielsson and Lindstrom, 2010; Juslin et al., 2010). Huron (2006) has elaborated on how music exploits this response by using extended anticipation and violation of expectations to intensify an emotional response. Higher level music events – such as motifs, or instrumental changes – may therefore also induce emotions via expectancy. In seminal work in this field, Sloboda (1991) asked participants to identify music passages which evoked strong, physical emotional responses in them, such as tears or chills. The most frequent musical events coded within these passages were new or unexpected harmonies, or appoggiaturas (which delay an expected principal note), supporting the proposal that unexpected musical events, or substantial changes in music features, were associated with physiological responses. Interestingly, a survey by Scherer et al. (2002) rated musical structure and acoustic features as more important in determining emotional reactions than the listener’s mood, affective involvement, personality or contextual factors. Importantly, because music events can elicit emotions via both expectation of an upcoming event and experience of that event, physiological markers of peak emotional responses may occur prior to, during or after a music event.

This proposal has received some empirical support via research demonstrating physiological peak responses to psychoacoustic ‘events’ in music (see **Table 1**). On the whole, changes in physiological arousal – primarily, chills, heart rate or skin conductance changes – coincided with sudden *changes* in acoustic features (such as changes in volume or tempo), or novel musical events (such as entry of new voices, or harmonic changes).

Supporting evidence for the similarity between music-evoked emotions and ‘real’ emotions has also emerged from research using central measures of emotional response. Importantly, brain regions associated with emotion and reward have been shown to also respond to emotionally powerful music. For instance, Blood and Zatorre (2001) found that pleasant music activated the dorsal amygdala (which connects to the ‘positive emotion’ network

**TABLE 1 |** Music features identified in the literature to be associated with various physiological markers of emotion.

Study	Physiological markers of emotion	Associated musical feature
Egermann et al. (2013)	Subjective and physiological arousal (skin conductance and heart rate)	Passages which violated expectations generated by a computational model which analyzed pitch features
Gomez and Danuser (2007)	Faster and higher minute ventilation, skin conductance, and heart rate	Tempo, accentuation, and rhythmic articulation Fast, accentuated, and staccato
Grewe et al. (2007a)	Chills	Entry of voice and changes in volume
Grewe et al. (2007b)	Skin conductance and facial muscle activity	The first entrance of a solo voice or choir and the beginning of new sections
Guhn et al. (2007)	Chills coinciding with distinct patterns of skin conductance increases	Passages which evoked chills had a number of similar characteristics: (1) they were from slow movements (adagio or larghetto), (2) they were characterized by alternation, or contrast, of the solo instrument and the orchestra, (3) a sudden or gradual volume increase from soft to loud, and (4) chill passages were characterized by an expansion in its frequency range in the high or low register, (5) all chill passages were characterized by harmonic peculiar progressions that potentially elicited ambiguity in the listener; that is, it deviated from what was expected based on the previous section
Coutinho and Cangelosi (2011)	Skin conductance and heart rate	Six low level music structural parameters: loudness, pitch level, pitch contour, tempo, texture, and sharpness
Steinbeis et al. (2006)	Electrodermal activity	Increases in harmonic unexpectedness

comprising the ventral striatum and orbitofrontal cortex), while reducing activity in central regions of the amygdala (which appear to be associated with unpleasant or aversive stimuli). Listening to pleasant music was also found to release dopamine in the striatum (Salimpoor et al., 2011, 2013). Further, the release was higher in the dorsal striatum during the anticipation of the peak emotional period of the music, but higher in the ventral striatum during the actual peak experience of the music. This is entirely consistent with the differentiated pattern of dopamine release during craving and consummation of other rewarding stimuli, e.g., amphetamines. Only one group to date has, however, attempted to identify musical features associated with central measures of emotional response. Koelsch et al. (2008a) performed a functional MRI study with musicians and non-musicians. While musicians tended to perceive syntactically irregular music events (single irregular chords) as slightly more pleasant than non-musicians, these generally perceived unpleasant events induced increased blood oxygen levels in the emotion-related brain region, the amygdala. Unexpected chords were also found to elicit specific event related potentials (ERAN and N5) as well as changes in skin conductance (Koelsch et al., 2008b). Specific music events associated with *pleasurable* emotions have not yet been examined using central measures of emotion.

Davidson and Irwin (1999), Davidson (2000, 2004), and Davidson et al. (2000), have demonstrated that a left bias in frontal cortical activity is associated with positive affect. Broadly, a left bias frontal asymmetry (FA) in the alpha band (8–13 Hz) has been associated with a positive affective style, higher levels of wellbeing and effective emotion regulation (Tomarken et al., 1992; Jackson et al., 2000). Interventions have been demonstrated to shift frontal electroencephalograph (EEG) activity to the left. An 8-week meditation training program significantly increased left sided FA when compared to wait list controls (Davidson et al., 2003). Blood et al. (1999) observed that left frontal brain areas were more likely to be activated by pleasant music than by

unpleasant music. The amygdala appears to demonstrate valence-specific lateralization with pleasant music increasing responses in the left amygdala and unpleasant music increasing responses in the right amygdala (Brattico, 2015; Bogert et al., 2016). Positively valenced music has also been found to elicit greater frontal EEG activity in the left hemisphere, while negatively valenced music elicits greater frontal activity in the right hemisphere (Schmidt and Trainor, 2001; Altenmüller et al., 2002; Flores-Gutierrez et al., 2007). The pattern of data in these studies suggests that this frontal lateralization is mediated by the emotions *induced* by the music, rather than just the emotional valence they *perceive* in the music. Hausmann et al. (2013) provided support for this conclusion via mood induction through a musical procedure using happy or sad music, which reduced the right lateralization bias typically observed for emotional faces and visual tasks, and increased the left lateralization bias typically observed for language tasks. A right FA pattern associated with depression was found to be shifted by a music intervention (listening to 15 min of ‘uplifting’ popular music previously selected by another group of adolescents) in a group of adolescents (Jones and Field, 1999). This measure therefore provides a useful objective marker of emotional response to further identify whether specific music events are associated with physiological measures of emotion.

The aim in this study was to examine whether: (1) music perceived as ‘emotionally powerful’ and pleasant by listeners also elicited a response in a central marker of emotional response (frontal alpha asymmetry), as found in previous research; and (2) peaks in frontal alpha asymmetry were associated with changes in key musical or psychoacoustic events associated with emotion. To optimize the likelihood that emotions were induced (that is, felt rather than just perceived), participants listened to their own selections of highly pleasurable music. Two validation hypotheses were proposed to confirm the methodology was consistent with previous research. It was hypothesized that: (1)

emotionally powerful and pleasant music selected by participants would be rated as more positive than silence, neutral music or a dissonant (unpleasant) version of their music; and (2) emotionally powerful pleasant music would elicit greater shifts in frontal alpha asymmetry than control auditory stimuli or silence. The primary novel hypothesis was that peak alpha periods would coincide with changes in basic psychoacoustic features, reflecting unexpected or anticipatory musical events. Since music-induced emotions can occur both before and after key music events, FA peaks were considered associated with music events if the music event occurred within 5 s before to 5 s after the FA event. Music background and affective style were also taken into account as potential confounds.

## MATERIALS AND METHODS

### Participants

The sample for this study consisted of 18 participants (6 males, 12 females) recruited from tertiary institutions located in Melbourne, Australia. Participants' ages ranged between 18 and 38 years ( $M = 22.22$ ,  $SD = 5.00$ ). Participants were excluded if they were younger than 17 years of age, had an uncorrected hearing loss, were taking medication that may impact on mood or concentration, were left-handed, or had a history of severe head injuries or seizure-related disorder. Despite clearly stated exclusion criteria, two left handed participants attended the lab, although as the pattern of their hemispheric activity did not appear to differ to right-handed participants, their data were retained. Informed consent was obtained through an online questionnaire that participants completed prior to the laboratory session.

### Materials

#### Online Survey

The online survey consisted of questions pertaining to demographic information (gender, age, a left-handedness question, education, employment status and income), music background (MUSE questionnaire; Chin and Rickard, 2012) and affective style (PANAS; Watson and Tellegen, 1988). The survey also provided an anonymous code to allow matching with laboratory data, instructions for attending the laboratory and music choices, and explanatory information about the study and a consent form.

#### Peak Frontal Asymmetry in Alpha EEG Frequency Band

The physiological index of emotion was measured using electroencephalography (EEG). EEG data were recorded using a 64-electrode silver-silver chloride (Ag-AgCl) EEG elastic Quik-cap (Compumedics) in accordance with the international 10–20 system. Data are, however, analyzed and reported from midfrontal sites (F3/F4 and FC3/FC4) only, as hemispheric asymmetry associated with positive and negative affect has been observed primarily in frontal cortex (Davidson et al., 1990; Tomarken et al., 1992; Dennis and Solomon, 2010). Further spatial exploration of data for structural mapping purposes was

beyond of the scope of this paper. In addition, analyses were performed for the P3–P4 sites as a negative control (Schmidt and Trainor, 2001; Dennis and Solomon, 2010). All channels were referenced to the mastoid electrodes (M1–M2). The ground electrode was situated between FPZ and FZ and impedances were kept below 10 kOhms. Data were collected and analyzed offline using the Compumedics Neuroscan 4.5 software.

### Subjective Emotional Response

The subjective feeling component of emotion was measured using 'EmuJoy' software (Nagel et al., 2007). This software allows participants to indicate how they feel in real time as they listen to the stimulus by moving the cursor along the screen. The Emujoy program utilizes the circumplex model of affect (Russell, 1980) where emotion is measured in a two dimensional affective space, with axes of arousal and valence. Previous studies have shown that valence and arousal account for a large portion of the variation observed in the emotional labeling of musical (e.g., Thayer, 1986), as well as linguistic (Russell, 1980) and picture-oriented (Bradley and Lang, 1994) experimental stimuli. The sampling rate was 20 Hz (one sample every 50 ms), which is consistent with recommendations for continuous monitoring of subjective ratings of emotion (Schubert, 2010). Consistent with Nagel et al. (2007), the visual scale was quantified as an interval scale from  $-10$  to  $+10$ .

### Music Stimuli

Four music stimuli—practice, pleasant, unpleasant, and neutral—were presented throughout the experiment. Each stimulus lasted between 3 and 5 min in duration. The practice stimulus was presented to familiarize participants with the Emujoy program and to acclimatize participants to the sound and the onset and offset of the music stimulus (fading in at the start and fading out at the end). The song was selected on the basis that it was likely to be familiar to participants, positive in affective valence, and containing segments of both arousing and calming music—The Lion King musical theme song, "*The circle of life*."

The pleasant music stimulus was participant-selected. This option was preferred over experimenter-selected music as participant-selected music was considered more likely to induce robust emotions (Thaut and Davis, 1993; Panksepp, 1995; Blood and Zatorre, 2001; Rickard, 2004). Participants were instructed to select a music piece that made them, "experience positive emotions (happy, joyful, excited, etc.) – like those songs you absolutely love or make you get goose bumps." This song selection also had to be one that would be considered a happy song by the general public. That is, it could not be sad music that participants enjoyed. While previous research has used both positively and negatively valenced music to elicit strong experiences with music, in the current study, we limited the music choices to those that expressed positive emotions. This decision was made to reduce variability in EEG responses arising from perception of negative emotions and experience of positive emotions, as EEG can be sensitive to differences in both perception and experience of emotional valence. The music

also had to be alyrical<sup>1</sup>—music with unintelligible words, for example in another language or skat singing, were permitted—as language processing might conceivably elicit different patterns of hemisphere activation solely as a function of the processing of vocabulary included in the song. [It should be noted that there are numerous mechanisms by which a piece of music might induce an emotion (see Juslin and Västfjäll, 2008), including associations with autobiographical events, visual imagery and brain stem reflexes. Differentiating between these various causes of emotion was, however, beyond the scope of the current study.]

The unpleasant music stimulus was intended to induce negative emotions. This was a dissonant piece produced by manipulating the participant's pleasant music stimulus and was achieved using Sony Sound Forge® 8 software. This stimulus consisted of three versions of the song played simultaneously—one shifted a tritone down, one pitch shifted a whole tone up, and one played in reverse (adapted from Koelsch et al., 2006). The neutral condition was an operatic track, La Traviata, chosen based upon its neutrality observed in previous research (Mitterschiffthaler et al., 2007).

The presentation of music stimuli was controlled by the experimenter via the EmuJoy program. The music volume was set to a comfortable listening level, and participants listened to all stimuli via bud earphones (to avoid interference with the EEG cap).

## Procedure

Prior to attending the laboratory session, participants completed the anonymously coded online survey. Within 2 weeks, participants attended the EEG laboratory at the Monash Biomedical Imaging Centre. Participants were tested individually during a 3 h session. An identification code was requested in order to match questionnaire data with laboratory session data.

Participants were seated in a comfortable chair and were prepared for fitting of the EEG cap. The participant's forehead was cleaned using medical grade alcohol swabs and exfoliated using NuPrep exfoliant gel. Participants were fitted with the EEG cap according to the standardized international 10/20 system (Jasper, 1958). Blinks and vertical/horizontal movements were recorded by attaching loose electrodes from the cap above and below the left eye, as well as laterally on the outer canthi of each eye. The structure of the testing was explained to participants and was as follows (see **Figure 1**):

The testing comprised four within-subjects conditions: pleasant, unpleasant, neutral, and control. Differing only in the type of auditory stimulus presented, each condition consisted of:

- (a) Baseline recording (B)—no stimulus was presented during the baseline recordings. These lasted 3 min in duration and participants were asked to close their eyes and relax.
- (b) Physiological recording (P)—the stimulus (depending on what condition) was played and participants were asked to have their eyes closed and to just experience the sounds.

- (c) Subjective rating (S)—the stimulus was repeated, however, this time participants were asked to indicate, with eyes open, how they felt as they listened to the same music on the computer screen using the cursor and the EmuJoy software.

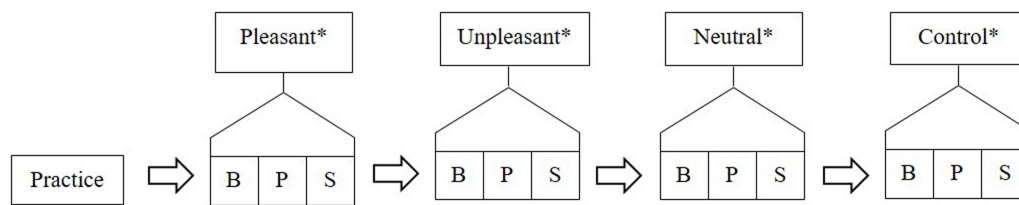
At every step of each condition, participants were guided by the experimenter (e.g., “I’m going to present a stimulus to you now, it may be music, something that sounds like music, or it could be nothing at all. All I would like you to do is to close your eyes and just experience the sounds”). Before the official testing began, the participant was asked to practice using the EmuJoy program in response to the practice stimulus. Participants were asked about their level of comfort and understanding with regards to using the EmuJoy software; experimentation did not begin until participants felt comfortable and understood the use of EmuJoy. Participants were reminded of the distinction between rating emotions *felt* vs. emotions *perceived* in the music; the former was encouraged throughout relevant sections of the experiment. After this, the experimental procedure began with each condition being presented to participants in a counterbalanced fashion. All procedures in this study were approved by the Monash University Human Research Ethics Committee.

## EEG Data Analysis for Frontal Asymmetry

Electroencephalograph data from each participant was visually inspected for artifacts (eye movements and muscle artifacts were manually removed prior to any analyses). EEG data were also digitally filtered with a low-pass zero phase-shift filter set to 30 Hz and 96 dB/oct. All data were re-referenced to mastoid processes. The sampling rate was 1250 Hz and eye movements were controlled for with automatic artifact rejection of  $> 50 \mu V$  in reference to VEO. Data were baseline corrected to 100 ms pre-stimulus period. EEG data were aggregated for all artifact-free periods within a condition to form a set of data for the positive music, negative music, neutral, and the control.

Chunks of 1024 ms were extracted for analyses using a Cosine window. A Fast Fourier Transform (FFT) was applied to each chunk of EEG permitting the computation of the amount of power at different frequencies. Power values from all chunks within an epoch were averaged (see Dumermuth and Molinari, 1987). The dependent measure that was extracted from this analysis was power density ( $\mu V^2/Hz$ ) in the alpha band (8–13 Hz). The data were log transformed to normalize their distribution because power values are positively skewed (Davidson, 1988). Power in the alpha band is inversely related to activation (e.g., Lindsley and Wicke, 1974) and has been the measure most consistently obtained in studies of EEG asymmetry (Davidson, 1988). Cortical asymmetry [ $\ln(\text{right}) - \ln(\text{left})$ ] was computed for the alpha band. This FA score provides a simple unidimensional scale representing relative activity of the right and left hemispheres for an electrode pair (e.g., F3 [left]/F4 [right]). FA scores of 0 indicate no asymmetry, while scores greater than 0 putatively are indicative of greater left frontal activity (positive affective response) and scores below 0 are

<sup>1</sup>One participant only chose music with lyrical content; the experimenter confirmed with this participant that the language (Italian) was unknown to them.



**FIGURE 1** | Example of testing structure with conditions ordered; pleasant, unpleasant, neutral, and control. B, baseline; P, physiological recording; and S, subjective rating. \*These stimuli were presented to participants in a counter balanced order.

indicative of greater right frontal activity (negative affective response), assuming that alpha is inversely related to activity (Allen et al., 2004). Peak FA periods at the FC3/FC4 site were also identified across each participant's pleasant music piece for purposes of music event analysis. FA (difference between left and right power densities) values were ranked from highest (most asymmetric, left biased) to lowest using spectrograms (see **Figure 2** for an example). Due to considerable inter-individual variability in asymmetry ranges, descriptive ranking was used as a selection criterion instead of an absolute threshold or statistical difference criterion. The ranked FA differences were inspected and those that were clearly separated from the others (on average six peaks were clearly more asymmetric than the rest of the record) were selected for each individual as their greatest moments of FA. This process was performed by two raters (authors H-AA and NR), with 100% interrater reliability, so no further analysis was performed/considered necessary required to rank the FA peaks.

## Music Event Data Coding

A subjective method of annotating each pleasant music piece with temporal onsets and types of all notable changes in musical features was utilized in this study. Coding was performed by a music performer and producer with postgraduate qualifications in systematic musicology. A decision was made to use subjective coding as it has been successfully used previously to identify significant changes in a broad range of music features associated with emotional induction by music (Sloboda, 1991). This method was framed within a hierarchical category model which contained both low-level and high-level factors of important changes. First, each participant's music piece was described by annotating the audiogram, noting the types of music changes at respective times. Secondly, the low-level factor model utilized by Coutinho and Cangelosi (2011) was applied to assign the identified music features deductively to changes within six low-level factors: loudness, pitch level, pitch contour, tempo, texture, and sharpness. Each low-level factor change was coded as a change toward one of the two anchors of the feature. For example, if a modification was marked in terms of loudness with 'loud,' it described an increase in loudness of the current part compared to the part before (see **Table 2**).

Due to the high variability of the analyzed musical pieces from a musicological perspective – including the genre, which ranged from classical and jazz to pop and electronica – every song had

a different frequency of changes in terms of these six factors. Hence, we applied a third step of categorization which led to a more abstract layer of changes in musical features that included two higher-level factors: motif changes and instrument changes. A time point in the music is marked with 'motif change' if the theme, movement or motif of the leading melody change from one part to the next one. The factor 'instrument change' can be defined as an increase or decrease of the number of playing instruments or as a change of instruments used within the current part.

## RESULTS

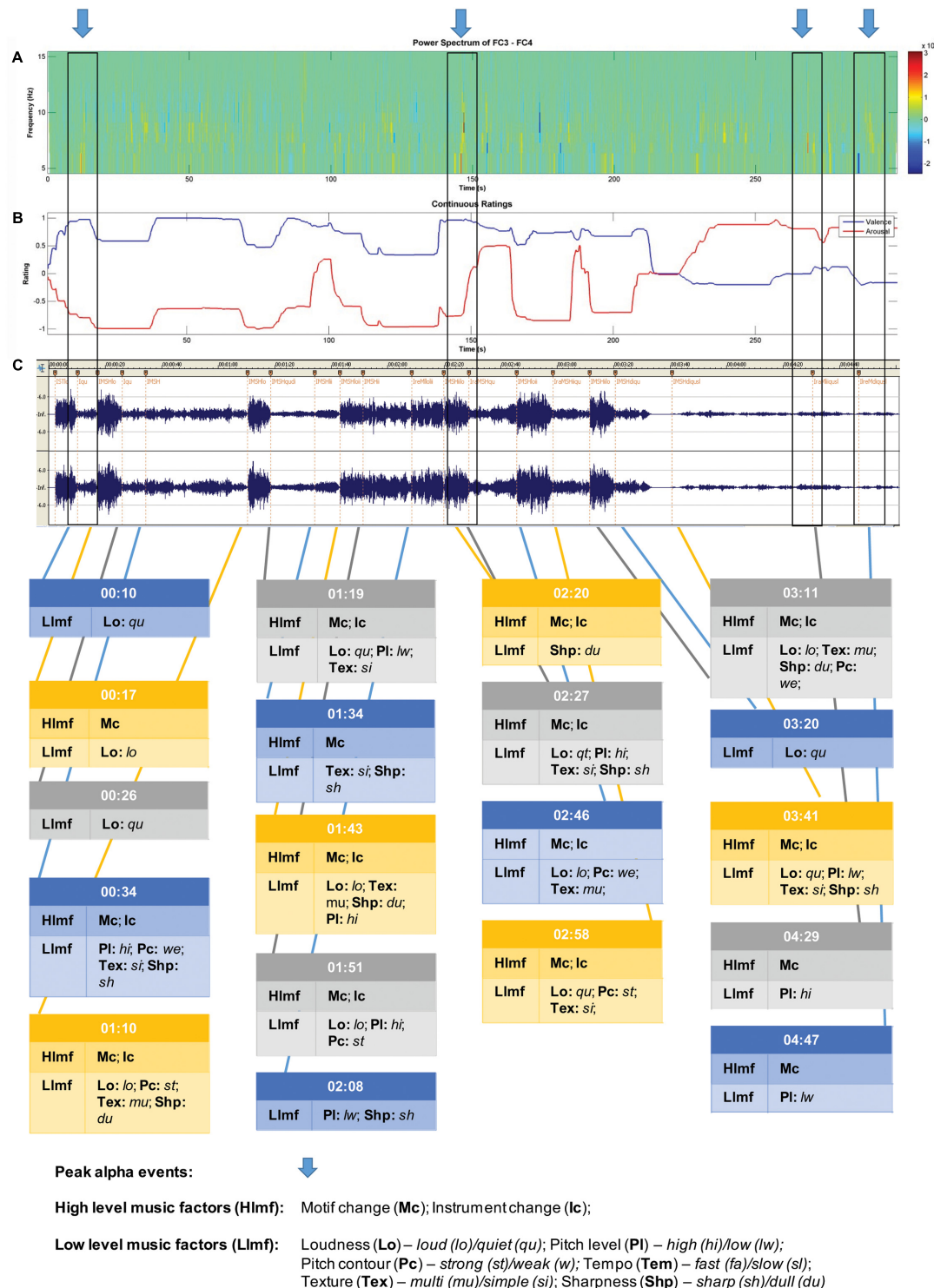
Data were scored and entered into PASW Statistics 18 for analyses. No missing data or outliers were observed in the survey data. Bivariate correlations were run between potential confounding variables – Positive affect negative affect schedule (PANAS), and the Music use questionnaire (MUSE) – and FA to determine if they were potential confounds, but no correlations were observed.

A sample of data obtained for each participant is shown in **Figure 2**. For this participant, five peak alpha periods were identified (shown in blue arrows at top). Changes in subjective valence and arousal across the piece are shown in the second panel, and then the musicological analysis in the final section of the figure.

## Subjective Ratings of Emotion – Averaged Emotional Responses

A one-way analysis of variance (ANOVA) was conducted to compare mean subjective ratings of emotional valence. Kolmogorov–Smirnov tests of normality indicated that distributions were normal for each condition except the subjective ratings of the control condition  $D(9) = 0.35$ ,  $p < 0.001$ . Nonetheless, as ANOVAs are robust to violations of normality when group sizes are equal (Howell, 2002), parametric tests were retained. No missing data or outliers were observed in the subjective rating data. **Figure 3** below shows the mean ratings of each condition.

**Figure 3** shows that both the direction and magnitude of subjective emotional valence differed across conditions, with the pleasant condition rated very positively, the unpleasant condition rated negatively, and the control and neutral conditions rated as neutral. Arousal ratings appeared to be reduced in response



**FIGURE 2 |** Sample data for participant 4 – music selection: The Four Seasons: Spring; Antonio Vivaldi. Recording: Karoly Botvay, Budapest Strings, Cobra Entertainment). (A) EEG alpha band spectrogram; (B) subjective valence and arousal ratings; and (C) music feature analysis.

to unpleasant and pleasant music. (Anecdotal reports from participants indicated that in addition to being very familiar with their own music, participants recognized the unpleasant piece as

a dissonant manipulation of their own music selection, and were therefore familiar with it also. Several participants noted that this made the piece even more unpleasant to listen to for them.)

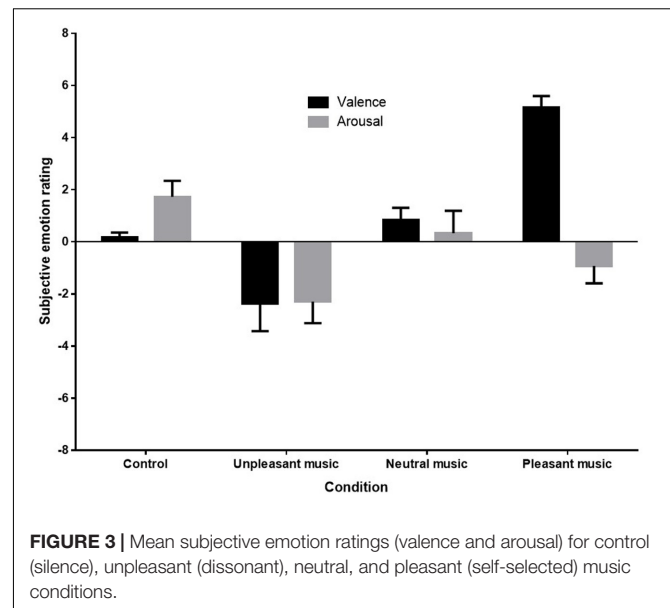
**TABLE 2 |** Operational definitions of high and low level musical features investigated in the current study.

Music feature	Operational definition
<b>High level factors</b>	
Motif changes	A theme/movement/motif of the leading melody changes from one part to another (e.g., from one motif to another/from first to second movement/from verse to bridge to chorus)
Instrumentation changes	the number and/or type of instruments played in one part is different compared to the part before
<b>Low level factors</b>	
Loudness	Is <b>loud</b> , if the general sound of one part is louder compared to the part before
Pitch level	Is <b>high</b> , if the pitch level of the leading melody is higher compared to the part before
Pitch contour	Is <b>strong</b> , if there are more pitch changes in the leading melody of one part compared to the number of changes in the part before
Tempo	Is <b>fast</b> , if the perceived tempo of one part is faster compared to the part before
Texture	Is <b>multi</b> , if there are more musical moments in which a high number of tones are playing simultaneously in one part compared to the part before
Sharpness	Is <b>sharp</b> , if there are more single differences in loudness and timbre from the different instruments or tones in one part compared to the part before

Sphericity was met for the arousal ratings, but not for valence ratings, so a Greenhouse-Geisser correction was made for analyses on valence ratings. A one-way repeated measures ANOVA revealed a significant effect of stimulus condition on valence ratings,  $F(1.6, 27.07) = 23.442$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.58$ . *Post hoc* contrasts revealed that the mean subjective valence rating for the unpleasant music was significantly lower than for the control  $F(1, 17) = 5.59$ ,  $p = 0.030$ ,  $\eta_p^2 = 0.25$ , and the mean subjective valence rating for the pleasant music was significantly higher than for the control condition,  $F(1, 17) = 112.42$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.87$ . The one-way repeated measures ANOVA for arousal ratings also showed a significant effect for stimulus condition,  $F(3, 51) = 5.20$ ,  $p = 0.003$ ,  $\eta_p^2 = 0.23$ . *Post hoc* contrasts revealed that arousal ratings were significantly reduced by both unpleasant,  $F(1, 17) = 10.11$ ,  $p = 0.005$ ,  $\eta_p^2 = 0.37$ , and pleasant music,  $F(1, 17) = 6.88$ ,  $p = 0.018$ ,  $\eta_p^2 = 0.29$ , when compared with ratings for the control.

## Aim 1: Can Emotionally Pleasant Music Be Detected by a Central Marker of Emotion (FA)?

Two-way repeated measures ANOVAs were conducted on the FA scores (averaged across baseline period, and averaged across condition) for each of the two frontal electrode pairs, and the control parietal site pair. The within-subjects factor included the music condition (positive, negative, neutral, and control) and time (baseline and stimulus). Despite the robustness of ANOVA to assumptions, caution should be taken in interpreting results as both the normality and sphericity assumptions were violated across each electrode pair. Where sphericity was violated, a

**FIGURE 3 |** Mean subjective emotion ratings (valence and arousal) for control (silence), unpleasant (dissonant), neutral, and pleasant (self-selected) music conditions.

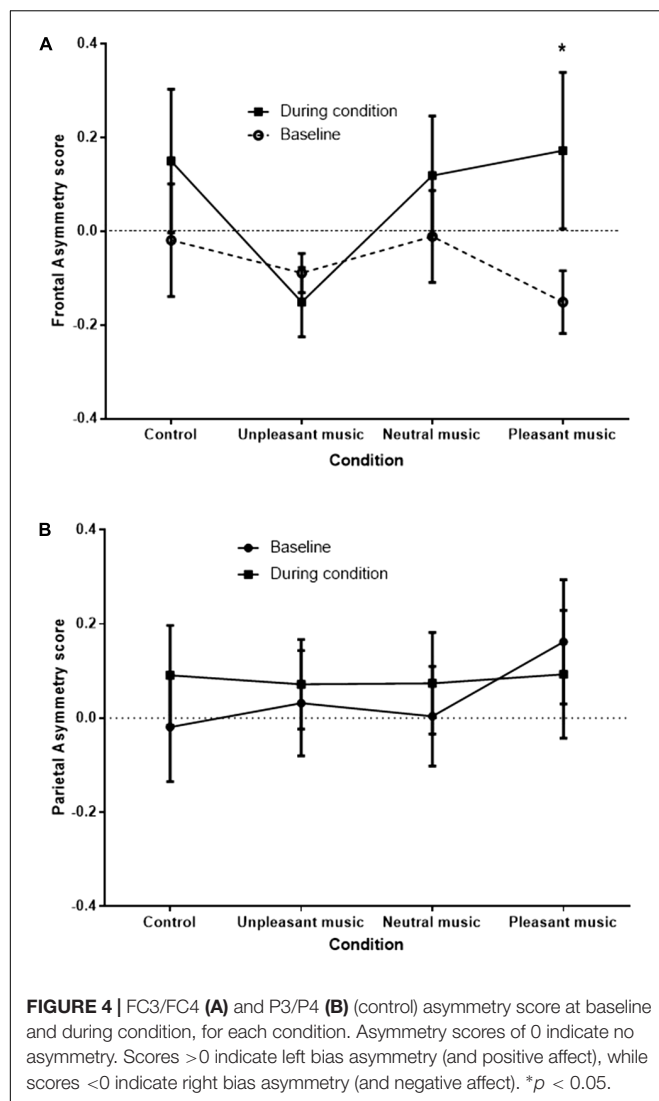
Greenhouse-Geisser correction was applied. Asymmetry scores above two were considered likely a result of noisy or damaged electrodes (62 points out of 864) and were omitted as missing data which were excluded pairwise. Two outliers were identified in the data and were replaced with a score  $\pm 3.29$  standard deviations from the mean.

A significant time by condition interaction effect was observed at the FC3/FC4 site,  $F(2.09, 27.17) = 3.45$ ,  $p = 0.045$ ,  $\eta_p^2 = 0.210$ , and a significant condition main effect was observed at the F3/F4 site,  $F(2.58, 21.59) = 3.22$ ,  $p = 0.039$ ,  $\eta_p^2 = 0.168$ . No significant effects were observed at the P3/P4 site [time by condition effect,  $F(1.98, 23.76) = 2.27$ ,  $p = 0.126$ ]. The significant interaction at FC3/FC4 is shown in Figure 4.

The greatest difference between baseline and during condition FA scores was for the pleasant music, representative of a positive shift in asymmetry from the right hemisphere to the left when comparing the baseline period to the stimulus period. Planned simple contrasts revealed that when compared with the unpleasant music condition, only the pleasant music condition showed a significant positive shift in FA score,  $F(1, 13) = 6.27$ ,  $p = 0.026$ . Positive shifts in FA were also apparent for control and neutral music conditions, although not significantly greater than for the unpleasant music condition [ $F(1, 13) = 2.60$ ,  $p = 0.131$ , and  $F(1, 13) = 3.28$ ,  $p = 0.093$ ], respectively.

## Aim 2: Are Peak FA Periods Associated with Particular Musical Events?

Peak periods of FA were identified for each participant, and the sum varied between 2 and 9 ( $M = 6.5$ ,  $SD = 2.0$ ). The music event description was then examined for presence or absence of coded musical events within a 10 s time window of (5 s before to 5 s after) the peak FA time-points. Across all participants, 106 peak alpha periods were identified, 78 of which (74%) were associated with particular music events. The type of music event coinciding with peak



alpha periods is shown in **Table 3**. A two-step cluster analysis was also performed to explore natural groupings of peak alpha asymmetry events that coincided with distinct *combinations* (2 or more) of musical features. A musical feature was to be deemed a salient characteristic of a cluster if present in at least 70% of the peak alpha events within the same cluster.

**Table 3** shows that, considered independently, the most frequent music features associated with peak alpha periods were primarily high level factors (changes in motif and instruments), with the addition of one low level factor (pitch). In exploring the data for clusters of peak alpha events associated with combinations of musical features, a four cluster solution was found to successfully group approximately half (53%) of the events into groups with identifiable patterns. This equated to 3 separate clusters characterized by distinct combinations of musical features, with the remaining half (47%) deemed unclassifiable as higher factor solutions provided no further differentiation.

**TABLE 3 |** Frequency and percentages of musical features associated with a physiological marker of emotion (peak alpha FA). High level, low level, and clusters of music features are distinguished.

Music-structural features	Frequency during peak alpha periods	Percentage during peak alpha periods
<b>High level music factors</b>		
Motif change	51	65.4
Instrument change	38	48.7
<b>Low level music factors</b>		
Pitch (high)	35	44.9
Loudness (loud)	20	25.6
Texture (simple)	11	14.1
Pitch contour (strong)	18	23.1
Sharpness (dull)	15	19.2
Sharpness (sharp)	14	18.0
Pitch (low)	14	18.0
Loudness (soft)	14	18.0
Texture (multi)	19	24.3
Pitch contour (weak)	7	9.0
Tempo (slow)	2	2.6
Tempo (fast)	1	1.2
<b>Feature cluster 1 (14/78 alpha periods)</b>		
Motif change	10	71
Instrument change	14	100
Texture (multi)	12	86
Sharpness (dull)	13	93
<b>Feature cluster 2 (11 of 78 alpha periods)</b>		
Motif change	8	73
Loudness (high level factor)	11	100
Loudness (loud; low level factor)	11	100
<b>Feature cluster 3 (14 of 78 alpha periods)</b>		
Motif change	10	71
Instrument change	11	79
Softness (high level factor)	14	100
Loudness (soft, low level factor)	12	86

## DISCUSSION

In the current study, a central physiological marker (alpha FA) was used to investigate the emotional response of music selected by participants to be ‘emotionally powerful’ and pleasant. Musical features of these pieces were also examined to explore associations between key musical events and central physiological markers of emotional responding. The first aim of this study was to examine whether pleasant music elicited physiological reactions in this central marker of emotional responding. As hypothesized, pleasant musical stimuli elicited greater shifts in FA than did the control auditory stimulus, silence or an unpleasant dissonant version of each participant’s music. This finding confirmed previous research findings and demonstrated that the methodology was robust and appropriate for further investigation. The second aim was to examine associations between key musical features (affiliated with emotion), contained within participant-selected musical pieces, and peaks in FA. FA peaks were commonly associated with changes in both high and

low level music features, including changes in motif, instrument, loudness and pitch, supporting the hypothesis that key events in music are marked by significant physiological changes in the listener. Further, specific combinations of individual musical features were identified that tended to predict FA peaks.

## Central Physiological Measures of Responding to Musical Stimuli

Participants' subjective valence ratings of music were consistent with expectations; control and neutral music were both rated neutrally, while unpleasant music was rated negatively and pleasant music was rated very positively. These findings are consistent with previous research indicating that music is capable of eliciting strong *felt* positive affective reports (Panksepp, 1995; Rickard, 2004; Juslin et al., 2008; Zenter et al., 2008; Eerola and Vuoskoski, 2011). The current findings were also consistent with previous negative subjective ratings (unpleasantness) by participants listening to the dissonant manipulation of musical stimuli (Koelsch et al., 2006). It is not entirely clear why arousal ratings were reduced by both the unpleasant and pleasant music. The variety of pieces selected by participants means that both relaxing and stimulating pieces were present in these conditions, although overall, the findings suggest that listening to music (regardless of whether pleasant or unpleasant) was more calming than silence for this sample. In addition, as both pieces were likely to be familiar (as participants reported that they recognized the dissonant manipulations of their own piece), familiarity could have reduced the arousal response expected for unpleasant music.

As hypothesized, FA responses from the FC3/FC4 site were consistent with subjective valence ratings, with the largest shift to the left hemisphere observed for the pleasant music condition. While not statistically significant, the small shifts to the left hemisphere during both control and neutral music conditions, and the small shift to the right hemisphere during the unpleasant music condition, indicate the trends in FA were also consistent with subjective valence reports. These findings support previous research findings on the involvement of the left frontal lobe in positive emotional experiences, and the right frontal lobe in negative emotional experiences (Davidson et al., 1979, 1990; Fox and Davidson, 1986; Davidson and Fox, 1989; Tomarken et al., 1990). The demonstration of these effects in the FC3/FC4 site is consistent with previous findings (Davidson et al., 1990; Jackson et al., 2003; Travis and Arenander, 2006; Kline and Allen, 2008; Dennis and Solomon, 2010), although meaningful findings are also commonly obtained from data collected from the F3/F4 site (see Schmidt and Trainor, 2001; Thibodeau et al., 2006), which was not observed in the current study. The asymmetry findings also verify findings observed in response to positive and negative emotion induction by music (Schmidt and Trainor, 2001; Altenmüller et al., 2002; Flores-Gutierrez et al., 2007; Hausmann et al., 2013). Importantly, no significant FA effect was observed in the control P3/P4 sites, which is an area not implicated in emotional responding.

## Associations between Musical Features and Peak Periods of Frontal Asymmetry Individual Musical Features

Several individual musical features coincided with peak FA events. Each of these musical features occurred in over 40% of the total peak alpha asymmetry events identified throughout the sample and appear to be closely related to *changes* in musical structure. These included changes in motif and instruments (high level factors), as well as pitch (low level factor). Such findings are in line with previous studies measuring non-central physiological measures of affective responding. For example, high level factor musical features such as instrument change, specifically changes and alternations between orchestra and solo piece instruments have been cited to coincide with chill responses (Grewe et al., 2007b; Guhn et al., 2007). Similarly, pitch events have been observed in previous research to coincide with various physiological measures of emotional responding including skin conductance and heart rate (Coutinho and Cangelosi, 2011; Egermann et al., 2013). In the current study, instances of *high* pitch were most closely associated with physiological reactions. These findings can be explained through Juslin and Sloboda's (2010) description of the activation of a 'brain stem reflex' in response to changes in basic acoustic events. Changes in loudness and high pitch levels may trigger physiological reactions on account of being psychoacoustic features of music that are shared with more primitive auditory stimuli that signal relevance for survival to real events.

Changes in instruments and motif, however, may be less related to primitive auditory stimuli and stimulate physiological reactions differently. Motif changes have not been observed in previous studies yet appeared most frequently throughout the peak alpha asymmetry events identified in the sample. In music, motif has been described as "...the smallest structural unit possessing thematic identity" (White, 1976, p. 26–27) and exists as a salient and recurring characteristic musical fragment throughout a musical piece. Within the descriptive analysis of the current study, however, a *motif* can be understood in a much broader sense (see definitions in **Table 2**). Due to the broad musical diversity of the songs selected by participants, the term *motif change* emerged as most appropriate description to cover high level structural changes in all the different musical pieces (e.g., changes from one small unit to another in a classic piece and changes from a long repetitive pattern to a chorus in an electronic dance piece). Changes in such a fundamental musical feature, as well as changes in instrument, are likely to stimulate a sense of novelty and add complexity, and possibly unexpectedness (i.e., features of goal oriented stimuli), to a musical piece. This may therefore also recruit the same neural system which has evolved to yield an emotional response, which in this study, is manifest in the greater activation in the left frontal hemisphere compared to the right frontal hemisphere. Many of the other musical features identified, however, did not coincide frequently with peak FA events. While peripheral markers of emotion, such as skin conductance and heart rate changes, are likely to respond strongly to basic psychoacoustic events associated with arousal, it is

likely that central markers such as FA are more sensitive to higher level musical events associated with positive affect. This may explain why motif changes were a particularly frequent event associated with FA peaks. Alternatively, some musical features may evoke emotional and physiological reactions only when present in conjunction with other musical features. It is recognized that an objective method of low level music feature identification would also be useful in future research to validate the current findings relating to low level psychoacoustic events. A limitation of the current study, however, was that the coding of both peak FA events and music events was subjective, which limits both replicability and objectivity. It is recommended future research utilize more objective coding techniques including statistical identification of peak FA events, and formal psychoacoustic analysis (such as achieved using software tools such as MIR Toolbox or PsySound). While an objective method of detecting FA events occurring within a specific time period after a music event is also appealing, the current methodology operationalized synchrony of FA and music events within a 10 s time window to include mechanisms of anticipation as well as experience of the event. Future research may be able to provide further distinction between these emotion induction mechanisms by applying different time windows to such analyses.

### Feature Clusters of Musical Feature Combinations

Several clusters comprising combinations of musical features were identified in the current study. A number of musical events which on their own did not coincide with FA peaks did nonetheless appear in music event clusters that were associated with FA peaks. For example, feature cluster 1 consists of motif and instrument changes—both individually considered to coincide frequently with peak alpha asymmetry events—as well as texture (multi) and sharpness (dull). Changes in texture and sharpness, individually, were observed to occur in only 24.3 and 19.2% of the total peak alpha asymmetry events, respectively. After exploring the data for natural groupings of musical events that occurred during peak alpha asymmetry scores, however, texture and sharpness changes appeared to occur frequently in conjunction with motif changes and instrument changes. Within feature cluster 1, texture and sharpness occurred in 86 and 93% of the peak alpha asymmetry periods. This suggests that certain musical features, like texture and sharpness, may lead to stronger emotional responses in central markers of physiological functioning when presented concurrently with specific musical events as compared to instances where they are present in isolation.

An interesting related observation is the specificity with which these musical events can combine to form a cluster. While motif and instrument changes occurred often in conjunction with texture (multi) and sharpness (dull) during peak alpha asymmetry events, both also occurred distinctly in conjunction with dynamic changes in volume (high level factor) and softness (low level factor) in a separate feature cluster. While both the texture/sharpness and loudness change/softness combinations frequently occur with motif and instrument changes, they appear

to do so in a mutually exclusive manner. This suggests a high level of complexity and specificity with which musical features may complement one another to stimulate physiological reactions during musical pieces.

The current findings extend previous research which has demonstrated that emotionally powerful music elicits changes in physiological, as well as subjective, measures of emotion. This study provides further empirical support for the emotivist theory of music and emotion which proposes that if emotional responses to music are ‘real,’ then they should be observable in physiological indices of emotion (Krumhansl, 1997; Rickard, 2004). The pattern of FA observed in this study is consistent with that observed in previous research in response to positive and negative music (Blood et al., 1999; Schmidt and Trainor, 2001), and non-musical stimuli (Fox, 1991; Davidson, 1993, 2000). However, the current study utilized music which expressed and induced positive emotions only, whereas previous research has also included powerful emotions induced by music expressing negative emotions. It would be of interest to replicate the current study with a broader range of powerful music to determine whether FA is indeed a marker of emotional experience, or a mixture of emotion perception and experience.

The findings also extend those obtained in studies which have examined musical features associated with strong emotional responses. Consistent with the broad consensus in this research, strong emotional responses often coincide with music events that signal change, novelty or violated expectations (Sloboda, 1991; Huron, 2006; Steinbeis et al., 2006; Eggermann et al., 2013). In particular, FA peaks were found to be associated in the current sample's music selections with motif changes, instrument changes, dynamic changes in volume, and pitch, or specific clusters of music events. Importantly, however, these conclusions are limited by the modest sample size, and consequently by the music pieces selected. Further research utilizing a different set of music pieces may identify a quite distinct pattern of music features associated with FA peaks. In sum, these findings provide empirical support for anticipation/expectation as a fundamental mechanism underlying music's capacity to evoke strong emotional responses in listeners.

### ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the National Statement on Ethical Conduct in Human Research, National Health and Medical Research Council, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Monash University Standing Committee for Ethical Research on Humans.

### AUTHOR CONTRIBUTIONS

H-AA conducted the experiments, contributed to the design and methods of the study, analysis of data and preparation of all

sections of the manuscript. NR contributed to the design and methods of the study, analysis of data and preparation of all sections the manuscript, and provided oversight of this study. JH conducted the musicological analyses of the music selections,

and contributed to the methods and results sections of the manuscript. BP performed the analyses of the EEG recordings and contributed to the methods and results sections of the manuscript.

## REFERENCES

- Allen, J., Coan, J., and Nazarian, M. (2004). Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion. *Biol. Psychol.* 67, 183–218. doi: 10.1016/j.biopsycho.2004.03.007
- Altenmüller, E., Schürmann, K., Lim, V. K., and Parlitz, D. (2002). Hits to the left, flops to the right: different emotions during listening to music are reflected in cortical lateralisation patterns. *Neuropsychologia* 40, 2242–2256. doi: 10.1016/S0028-3932(02)00107-0
- Bartlett, D. L. (1996). “Physiological reactions to music and acoustic stimuli,” in *Handbook of Music Psychology*, 2nd Edn, ed. D. A. Hodges (San Antonio, TX: IMR Press), 343–385.
- 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
- Blood, A. J., Zatorre, R. J., Bermudez, P., and Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nat. Neurosci.* 2, 382–387. doi: 10.1038/7299
- Bogert, B., Numminen-Kontti, T., Gold, B., Sams, M., Numminen, J., Burunat, I., et al. (2016). Hidden sources of joy, fear, and sadness: explicit versus implicit neural processing of musical emotions. *Neuropsychologia* 89, 393–402. doi: 10.1016/j.neuropsychologia.2016.07.005
- Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9
- Brattico, E. (2015). “From pleasure to liking and back: bottom-up and top-down neural routes to the aesthetic enjoyment of music,” in *Art, Aesthetics and the Brain*, eds M. Nadal, J. P. Houston, L. Agnati, F. Mora, and C. J. CelaConde (Oxford, NY: Oxford University Press), 303–318. doi: 10.1093/acprof:oso/9780199670000.003.0015
- Chin, T. C., and Rickard, N. S. (2012). The Music USE (MUSE) questionnaire: an instrument to measure engagement in music. *Music Percept.* 29, 429–446. doi: 10.1525/mp.2012.29.4.429
- Coutinho, E., and Cangelosi, A. (2011). Musical emotions: predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. *Emotion* 11, 921–937. doi: 10.1037/a0024700
- Davidson, R. J. (1988). EEG measures of cerebral asymmetry: conceptual and methodological issues. *Int. J. Neurosci.* 39, 71–89. doi: 10.3109/00207458808985694
- Davidson, R. J. (1993). “The neuropsychology of emotion and affective style,” in *Handbook of Emotion*, eds M. Lewis and J. M. Haviland (New York, NY: The Guilford Press), 143–154.
- Davidson, R. J. (2000). Affective style, psychopathology, and resilience. Brain mechanisms and plasticity. *Am. Psychol.* 55, 1196–1214. doi: 10.1037/0003-066X.55.11.1196
- Davidson, R. J. (2004). Well-being and affective style: neural substrates and biobehavioural correlates. *Philos. Trans. R. Soc.* 359, 1395–1411. doi: 10.1098/rstb.2004.1510
- Davidson, R. J., Ekman, P., Saron, C. D., Senulis, J. A., and Friesen, W. V. (1990). Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology. *I. J. Pers. Soc. Psychol.* 58, 330–341. doi: 10.1037/0022-3514.58.2.330
- Davidson, R. J., and Fox, N. A. (1989). Frontal brain asymmetry predicts infants’ response to maternal separation. *J. Abnorm. Psychol.* 98, 127–131. doi: 10.1037/0021-843X.98.2.127
- Davidson, R. J., and Irwin, W. (1999). The functional neuroanatomy of emotion and affective style. *Trends Cogn. Sci.* 3, 11–21. doi: 10.1016/S1364-6613(98)01265-0
- Davidson, R. J., Jackson, D. C., and Kalin, N. H. (2000). Emotion, plasticity, context, and regulation: perspectives from affective neuroscience. *Psychol. Bull.* 126, 890–909. doi: 10.1037/0033-2909.126.6.890
- Davidson, R. J., Kabat-Zinn, J., Schumacher, J., Rosenkranz, M., Muller, D., Santorelli, S. F., et al. (2003). Alterations in brain and immune function produced by mindfulness meditation. *Psychosom. Med.* 65, 564–570. doi: 10.1097/01.PSY.0000077505.67574.E3
- Davidson, R. J., Schwartz, G. E., Saron, C., Bennett, J., and Goleman, D. J. (1979). Frontal versus parietal EEG asymmetry during positive and negative affect. *Psychophysiology* 16, 202–203.
- Dennis, T. A., and Solomon, B. (2010). Frontal EEG and emotion regulation: electrocortical activity in response to emotional film clips is associated with reduced mood induction and attention interference effects. *Biol. Psychol.* 85, 456–464. doi: 10.1016/j.biopsycho.2010.09.008
- Dumermuth, G., and Molinari, L. (1987). “Spectral analysis of EEG background activity,” in *Handbook of Electroencephalography and Clinical Neurophysiology: Methods of Analysis of Brain Electrical and Magnetic Signals*, Vol. 1, eds A. S. Gevins and A. Remond (Amsterdam: Elsevier), 85–130.
- Eerola, T., and Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychol. Music* 39, 18–49. doi: 10.1093/scan/nsv032
- 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
- Flores-Gutierrez, E. O., Diaz, J.-L., Barrios, F. A., Favila-Humara, R., Guevara, M. A., del Rio-Portilla, Y., et al. (2007). Metabolic and electric brain patterns during pleasant and unpleasant emotions induced by music masterpieces. *Int. J. Psychophysiol.* 65, 69–84. doi: 10.1016/j.ijpsycho.2007.03.004
- Fox, N. A. (1991). If it’s not left, it’s right: electroencephalogram asymmetry and the development of emotion. *Am. Psychol.* 46, 863–872. doi: 10.1037/0003-066X.46.8.863
- Fox, N. A., and Davidson, R. J. (1986). Taste-elicited changes in facial signs of emotion and the asymmetry of brain electrical activity in human newborns. *Neuropsychologia* 24, 417–422. doi: 10.1016/0028-3932(86)90028-X
- Frijda, N. H., and Scherer, K. R. (2009). “Emotion definition (psychological perspectives),” in *Oxford Companion to Emotion and the Affective Sciences*, eds D. Sander and K. R. Scherer (Oxford: Oxford University Press), 142–143.
- Gabrielsson, A., and Lindstrom, E. (2010). “The role of structure in the musical expression of emotions,” in *Handbook of Music and Emotion: Theory, Research, Applications*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 367–400.
- Gomez, P., and Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion* 7, 377–387. doi: 10.1037/1528-3542.7.2.377
- Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2007a). Emotions over time: synchronicity and development of subjective, physiological, and facial affective reactions to music. *Emotion* 7, 774–788.
- Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2007b). Listening to music as a re-creative process: physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Percept.* 24, 297–314. doi: 10.1525/mp.2007.24.3.297
- Guhn, M., Hamm, A., and Zentner, M. (2007). Physiological and musico-acoustic correlates of the chill response. *Music Percept.* 24, 473–484. doi: 10.1525/mp.2007.24.5.473
- Hausmann, M., Hodgetts, S., and Eerola, T. (2013). Music-induced changes in functional cerebral asymmetries. *Brain Cogn.* 104, 58–71. doi: 10.1016/j.bandc.2016.03.001
- Hodges, D. (2010). “Psychophysiological measures,” in *Handbook of Music and Emotion: Theory, Research and Applications*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 279–312.
- Howell, D. C. (2002). *Statistical Methods for Psychology*, 5th Edn. Belmont, CA: Duxbury.

- Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.
- Jackson, D. C., Malmstadt, J. R., Larson, C. L., and Davidson, R. J. (2000). Suppression and enhancement of emotional responses to unpleasant pictures. *Psychophysiology* 37, 515–522. doi: 10.1111/1469-8986.3740515
- Jackson, D. C., Mueller, C. J., Dolski, I., Dalton, K. M., Nitschke, J. B., Urry, H. L., et al. (2003). Now you feel it now you don't: frontal brain electrical asymmetry and individual differences in emotion regulation. *Psychol. Sci.* 14, 612–617. doi: 10.1046/j.0956-7976.2003.psci.1473.x
- Jasper, H. H. (1958). Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalogr. Clin. Neurophysiol.* 10, 370–375. doi: 10.1016/0013-4694(58)90053-1
- Jones, N. A., and Field, T. (1999). Massage and music therapies attenuate frontal EEG asymmetry in depressed adolescents. *Adolescence* 34, 529–534.
- Juslin, P. N., Liljestrom, S., Västfjäll, D., Barradas, G., and Silva, A. (2008). An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion* 8, 668–683. doi: 10.1037/a0013505
- Juslin, P. N., Liljestrom, S., Västfjäll, D., and Lundqvist, L. (2010). "How does music evoke emotions? Exploring the underlying mechanisms," in *Music and Emotion: Theory, Research and Applications*, eds P. N. Juslin and J. A. Sloboda (Oxford: Oxford University Press), 605–642.
- Juslin, P. N., and Sloboda, J. A. (eds) (2010). *Handbook of Music and Emotion: Theory, Research and Applications*. New York, NY: Oxford University Press.
- Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 559–621. doi: 10.1017/S0140525X08005293
- Kivy, P. (1990). *Music Alone: Philosophical Reflections on the Purely Musical Experience*. London: Cornell University Press.
- Kline, J. P., and Allen, S. (2008). The failed repressor: EEG asymmetry as a moderator of the relation between defensiveness and depressive symptoms. *Int. J. Psychophysiol.* 68, 228–234. doi: 10.1016/j.ijpsycho.2008.02.002
- Koelsch, S., Fritz, T., and Schlaug, G. (2008a). Amygdala activity can be modulated by unexpected chord functions during music listening. *Neuroreport* 19, 1815–1819. doi: 10.1097/WNR.0b013e32831a8722
- Koelsch, S., Fritz, T., von Cramon, Y., Muller, K., and Friederici, A. D. (2006). Investigating emotion with music: an fMRI study. *Hum. Brain Mapp.* 27, 239–250. doi: 10.1002/hbm.20180
- Koelsch, S., Kilches, S., Steinbeis, N., and Schelinski, S. (2008b). Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *PLOS ONE* 3:e2631. doi: 10.1371/journal.pone.0002631
- Konecni, V. (2013). Music, affect, method, data: reflections on the Carroll versus Kivy debate. *Am. J. Psychol.* 126, 179–195. doi: 10.5406/amerjpsyc.126.2.0179
- Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Can. J. Exp. Psychol.* 51, 336–352. doi: 10.1037/1196-1961.51.4.336
- Lindsley, D. B., and Wicke, J. D. (1974). "The electroencephalogram: autonomous electrical activity in man and animals," in *Bioelectric Recording Techniques*, eds R. Thompson and M. N. Patterson (New York, NY: Academic Press), 3–79.
- Meyer, L. B. (1956). "Emotion and meaning in music," in *Handbook of Music and Emotion: Theory, Research and Applications*, eds P. N. Juslin and J. A. Sloboda (Oxford: Oxford University Press), 279–312.
- Mitterschiffthaler, M. T., Fu, C. H. Y., Dalton, J. A., Andrew, C. M., and Williams, S. C. R. (2007). A functional MRI study of happy and sad affective states induced by classical music. *Hum. Brain Mapp.* 28, 1150–1162. doi: 10.1002/hbm.20337
- Nagel, F., Kopiez, R., Grewe, O., and Altenmuller, E. (2007). EMUJoy: software for continuous measurement of perceived emotions in music. *Behav. Res. Methods* 39, 283–290. doi: 10.3758/BF03193159
- Panksepp, J. (1995). The emotional sources of 'chills' induced by music. *Music Percept.* 13, 171–207. doi: 10.2307/40285693
- Panksepp, J., and Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behav. Process.* 60, 133–155. doi: 10.1016/S0376-6357(02)00080-3
- Rickard, N. S. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychol. Music* 32, 371–388. doi: 10.1177/0305735604046096
- Rickard, N. S. (2012). "Music listening and emotional well-being," in *Lifelong Engagement with Music: Benefits for Mental Health and Well-Being*, eds N. S. Rickard and K. McFerran (Hauppauge, NY: de Sitter), 207–238.
- Russell, J. A. (1980). A circumplex model of affect. *J. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714
- 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, 257–264. doi: 10.1038/nn.2726
- Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., and Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science* 340, 216–219. doi: 10.1126/science.1231059
- Scherer, K. R. (2009). Emotions are emergent processes: they require a dynamic computational architecture. *Philos. Trans. R. Soc. Ser. B* 364, 3459–3474. doi: 10.1098/rstb.2009.0141
- Scherer, K. R., and Coutinho, E. (2013). "How music creates emotion: a multifactorial process approach," in *The Emotional Power of Music*, eds T. Cochrane, B. Fantini, and K. R. Scherer (Oxford: Oxford University Press). doi: 10.1093/acprof:oso/9780199654888.003.0010
- Scherer, K. R., Zentner, M. R., and Schacht, A. (2002). Emotional states generated by music: an exploratory study of music experts. *Musica Sci.* 5, 149–171. doi: 10.1177/102986490200505106
- Schmidt, L. A., and Trainor, L. J. (2001). Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cogn. Emot.* 15, 487–500. doi: 10.1080/02699930126048
- Schubert, E. (2010). "Continuous self-report methods," in *Handbook of Music and Emotion: Theory, Research and Applications*, eds P. N. Juslin and J. A. Sloboda (Oxford: Oxford University Press), 223–224.
- Sloboda, J. (1991). Music structure and emotional response: some empirical findings. *Psychol. Music* 19, 110–120. doi: 10.1177/0305735691192002
- Steinbeis, N., Koelsch, S., and Sloboda, J. (2006). The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological, and neural responses. *J. Cogn. Neurosci.* 18, 1380–1393. doi: 10.1162/jocn.2006.18.8.1380
- Thaut, M. H., and Davis, W. B. (1993). The influence of subject-selected versus experimenter-chosen music on affect, anxiety, and relaxation. *J. Music Ther.* 30, 210–233. doi: 10.1093/jmt/30.4.210
- Thayer, J. F. (1986). *Multiple Indicators of Affective Response to Music*. Doctoral Dissertation, New York University, New York, NY.
- Thibodeau, R., Jorgensen, R. S., and Kim, S. (2006). Depression, anxiety, and resting frontal EEG asymmetry: a meta-analytic review. *J. Abnorm. Psychol.* 115, 715–729. doi: 10.1037/0021-843X.115.4.715
- Tomarken, A. J., Davidson, R. J., and Henriques, J. B. (1990). Resting frontal brain asymmetry predicts affective responses to films. *J. Pers. Soc. Psychol.* 59, 791–801. doi: 10.1037/0022-3514.59.4.791
- Tomarken, A. J., Davidson, R. J., Wheeler, R. E., and Doss, R. C. (1992). Individual differences in anterior brain asymmetry and fundamental dimensions of emotion. *J. Pers. Soc. Psychol.* 62, 676–687. doi: 10.1037/0022-3514.62.4.676
- Travis, F., and Arenander, A. (2006). Cross-sectional and longitudinal study of effects of transcendental meditation practice on interhemispheric frontal asymmetry and frontal coherence. *Int. J. Neurosci.* 116, 1519–1538. doi: 10.1080/00207450600575482
- Watson, D., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063
- White, J. D. (1976). *The Analysis of Music*. Duke, NC: Duke University Press.
- Zenter, M., Grandjean, D., and Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* 8, 494–521. doi: 10.1037/1528-3542.8.4.494

**Conflict of Interest Statement:** 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.

Copyright © 2017 Arjmand, Hohagen, Paton and Rickard. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Reviewing the Effectiveness of Music Interventions in Treating Depression

Daniel Leubner\* and Thilo Hinterberger

Department of Psychosomatic Medicine, Research Section of Applied Consciousness Sciences, University Clinic Regensburg, Regensburg, Germany

## OPEN ACCESS

### Edited by:

Mark Reybrouck,  
KU Leuven, Belgium

### Reviewed by:

Jeanette Tamplin,  
University of Melbourne, Australia  
Amanda E. Krause,  
University of Melbourne, Australia

### \*Correspondence:

Daniel Leubner  
leubner@ymail.com

### Specialty section:

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

**Received:** 13 October 2016

**Accepted:** 15 June 2017

**Published:** 07 July 2017

### Citation:

Leubner D and Hinterberger T (2017)  
Reviewing the Effectiveness of Music  
Interventions in Treating Depression.  
Front. Psychol. 8:1109.  
doi: 10.3389/fpsyg.2017.01109

Depression is a very common mood disorder, resulting in a loss of social function, reduced quality of life and increased mortality. Music interventions have been shown to be a potential alternative for depression therapy but the number of up-to-date research literature is quite limited. We present a review of original research trials which utilize music or music therapy as intervention to treat participants with depressive symptoms. Our goal was to differentiate the impact of certain therapeutic uses of music used in the various experiments. Randomized controlled study designs were preferred but also longitudinal studies were chosen to be included. 28 studies with a total number of 1,810 participants met our inclusion criteria and were finally selected. We distinguished between passive listening to music (record from a CD or live music) (79%), and active singing, playing, or improvising with instruments (46%). Within certain boundaries of variance an analysis of similar studies was attempted. Critical parameters were for example length of trial, number of sessions, participants' age, kind of music, active or passive participation and single- or group setting. In 26 studies, a statistically significant reduction in depression levels was found over time in the experimental (music intervention) group compared to a control ( $n = 25$ ) or comparison group ( $n = 2$ ). In particular, elderly participants showed impressive improvements when they listened to music or participated in music therapy projects. Researchers used group settings more often than individual sessions and our results indicated a slightly better outcome for those cases. Additional questionnaires about participants confidence, self-esteem or motivation, confirmed further improvements after music treatment. Consequently, the present review offers an extensive set of comparable data, observations about the range of treatment options these papers addressed, and thus might represent a valuable aid for future projects for the use of music-based interventions to improve symptoms of depression.

**Keywords:** depression, music therapy, meta-analysis, neuropsychology, psychosomatic medicine, neurophysiology, anxiety, stress

## INTRODUCTION

*"If I were not a physicist, I would probably be a musician. I often think in music. I live my daydreams in music. I see my life in terms of music."*

—Einstein, 1929.

Depression is one of the most serious and frequent mental disorders worldwide. International studies predict that approximately 322 million (WHO, 2017) of the world's population suffer from a clinical depression. This disorder can occur from infancy to old age, with women being affected more often than men (WHO, 2017). Thus, depression is one of the most common chronic diseases. Depressive suffering is associated with psychological, physical, emotional, and social impairments. This can influence the whole human being in a fundamental way. Without clinical treatment, it has the tendency to recur or to take a chronic course that can lead to loneliness (Alpass and Neville, 2003) and an increasing social isolation (Teo, 2012). Depression can have many causes that range from genetic, over psychological factors (negative self-concept, pessimism, anxiety and compulsive states, etc.) to psychological trauma. In addition, substance abuse (Neighbors et al., 1992) or chronic diseases (Moussavi et al., 2007) can also trigger depression. The colloquial use of the term "depressed" has nothing to do with the depression in the clinical sense. The ICD-10 (WHO, 1992) and the DSM-V (APA, 2013) provide a classification based on symptoms, considering the patient's history and its severity, duration, course and frequency. Within the last two decades, research on the use of music medicine or music therapy to treat depression, showed a growing popularity and several publications have appeared that documented this movement (e.g., Lee, 2000; Loewy, 2004; Esfandiari and Mansouri, 2014; Verrusio et al., 2014; Chen et al., 2016; Fancourt et al., 2016). However, most researchers used a very specific experimental setup (Hillecke et al., 2005) and thus, for example, focused only on one music genre (i.e., classical, modern; instrumental, vocal), used a predefined experimental setup (group or individual) (e.g., Kim et al., 2006; Chen et al., 2016), or specified precisely the age range (i.e., adolescents, elderly) of participants (e.g., Koelsch et al., 2010; Verrusio et al., 2014). A recent meta-analysis (Hole et al., 2015) reviewed 72 randomized controlled trials and concluded that music was a notable aid for reducing postoperative symptoms of anxiety and pain.

Dementia patients showed significant cognitive and emotional benefits when they sang, or listened to familiar songs (Särkämö et al., 2008, 2014). Beneficial effects were also described for CNMP (Chronic Non-Malignant Pain) patients with depression (Siedliecki and Good, 2006)<sup>1</sup>. Cardiology is an area where music interventions are commonly used for intervention purposes. Various explanations were postulated and the broad range of effects on the cardiovascular system was investigated (Trappe,

2010; Hanser, 2014). Music as a therapeutic approach was evaluated (Gold et al., 2004), and found to have positive effects before heart surgery (Twiss et al., 2006), used to increase relaxation during angiography (Bally et al., 2003), or decrease anxiety (Doğan and Senturan, 2012; Yinger and Gooding, 2015). A systematic review (Jespersen et al., 2015) concluded that music improved subjective sleep quality in adults with insomnia, verbal memory in children (Chan et al., 1998; Ho et al., 2003), and episodic long-term memory (Eschrich et al., 2008). Music conveyed a certain mood or atmosphere (Husain et al., 2002), allowed composers to trigger emotions (Bodner et al., 2007; Droit-Volet et al., 2013), based on the cultural background (Balkwill and Thompson, 1999), or ethnic group (Werner et al., 2009) someone belonged to. In contrast, the emotional state itself plays a role (Al'tman et al., 2004) on how music is interpreted (Al'tman et al., 2000), and durations are evaluated (Schäfer et al., 2013). Subjective impressions embedded in a composition caused physiological body reactions (Grewe et al., 2007; Jäncke, 2008) and even strengthened the immune system (McCraty et al., 1996; Bittman et al., 2001). The pace of (background) music (Oakes, 2003), has also been used as an essential element of many marketing concepts (North and Hargreaves, 1999), to create a relaxed atmosphere. An in-depth, detailed illustration described the wide variety of conscious, as well as subconscious influences music can have (Panksepp and Bernatzky, 2002), and endorsed future research on this subject.

## Distinction between the Terms "Music Therapy [MT]" and "Music Medicine [MM]"

Most of us know what kind of music or song "can cheer us up." To treat someone else is something completely different though. Therefore, evidence-based procedures were created for a more pragmatic approach. It is important to differentiate between music therapy and the therapeutic use of music. Music used for patient treatment can be divided into two major categories, namely [MT] and [MM], although the distinction is not always that clear.

### Music Therapy [MT]

Term used primarily for a setting, where sessions are provided by a board-certified music therapist. Music therapy [MT] (Maratos et al., 2008; Bradt et al., 2015) stands for the "...clinical and evidence-based use of music interventions to accomplish individualized goals within a therapeutic relationship by a credentialed professional who has completed an approved music therapy program" (AMTA)<sup>2</sup>. Many different fields of practice, mostly in the health care system, show an increasing amount of interest in [MT]. Mandatory is a systematic constructed therapy process that was created by a board-certified music therapist and requires an individual-specific music selection that is developed uniquely for and together with the patient in one or more sessions. Therapy settings are not limited to listening, but may also include playing, composing, or interacting with music. Presentations can be pre-recorded or live. In other cases

<sup>1</sup>Participants in the two music groups (standard or patterning music) showed an increased belief in their personal power as well as a reduction in pain, depression and disability, compared to the relevant control group. The two experimental groups listened to 1 h of music each day for 7 days in a row.

<sup>2</sup>Official definition of the American Music Therapy Association [AMTA] <http://www.musictherapy.org/about/quotes/>

(basic) instruments are built together. The process to create these tailor-made selections requires specific knowledge on how to select, then construct and combine the most suitable stimuli or hardware. It must also be noted that music therapy is offered as a profession-qualifying course of study.

### Music Medicine [MM] (i.e., Functional Music, Music in Medicine)

Carried out independently by professionals, who are not qualified music therapists, like relaxation therapists, physicians or (natural) scientists. A previous consultation, or collaboration, with a certified music therapist can be helpful (Register, 2002). In recent years, significant progress has been made in both the research and clinical application of music as a form of treatment. It has valuable therapeutic properties, suitable for the treatment of several diseases. The term “music medicine” is used as a term for the therapeutic use of music in medicine (Bradt et al., 2015, 2016), to be able to differentiate it from “music therapy.” [MM] stands for a medical, physiological and physical evaluation of the use of music. If someone listens to his or her favorite music, this is sometimes also considered as a form of music medicine. [MM] deliberately differs from music therapy as part of psychiatric care or psychotherapy. It is important to stress out that the term “Music Therapy [MT]” should not be used for any kind of treatment involving music, although there is without doubt a relationship between [MT] and [MM]. What all of them have in common is the focus on a scientifically, artistically or clinically based approach to music.

### “Seamless Transitions” between Music Therapy [MT] and Music Medicine [MM]

Activity used for treatment is ambiguous or not clearly labeled as “Music Therapy” or “Music Medicine.” It should not be forgotten that the definition of “Music Therapy” is not always clearly distinguishable from “Music Medicine.” One possible scenario would be a physician (i.e., “non-professional”), who is not officially certified by the AMTA (or comparable institutions), but still acts according to the mandatory rules. In addition, depending on one’s home country, uniform standards or eligibility requirements might be substantially different. We think that every effort should be recognized and therefore postulate one definition that can describe the main principle of [MT], [MM], and everything in between, in one sentence: *“Implementation of acoustic stimuli (“music”) as a medium for the purpose of improving symptoms in a defined group of participants (patients) suffering from depression.”*

## MATERIALS AND METHODS

### Literature Search

Search strategy and selection process was performed according to the recommended guidelines of the Cochrane Centre on systematic literature search (Higgins and Green, 2008). Our approach (Figure 2) was according to their scientific relevance, supplemented by the analysis of relevant journals, conferences and workshops of recent years. We obtained 60,795 articles from various search engines as initial result. Retrieved data was

collected and processed on an existing personal computer with the latest Windows operating system.

### Search, Collection, Selection, and Review Strategies

We used a combination of words defining three search-categories (Music-, Treatment-, and Depression associated) as well as several words (e.g., Sound, Unhappy, and Treatment) assigned to each category as described in the collection process below. If synonyms of those keywords were identified, they were added as well. Theme-categories<sup>3</sup> were created next, then related keywords identified and added into a table. “Boolean Operators<sup>4</sup>” were used as logical connectives to broaden and/or narrow our search results within many databases (mostly search engines as described below).

This way the systematic variation of keyword-based queries and search terms could be performed with much more efficiency. To find the most relevant literature on the subject, keywords were entered into various scientific search engines, namely PubMed, MEDLINE, and Google Scholar. After the collection process, several different steps were used to reduce the number of retrieved results. Selection out of the collected material included to narrow down search results to a limited period of time. We decided to choose a period between 1990 and 2016 (i.e., not exceeding 26 years), because within these years several very interesting works of research were published, but often not mentioned explicitly, discussed in detail, or the main target of a comparative review. After several papers were excluded, a systematic key phrases search was conducted once more to retrieve results, limited to original research articles<sup>5</sup>. We also removed search results that quoted book chapters, as well as reports from international congresses and conferences. Research papers that remained were distinguished from duplicates (or miss-matches not dismissed yet). Based on our predefined criteria for in- and ex-clusion, relevant publications were then selected for an intensified review process. Our plan was to apply the following inclusion criteria: Original research article, published at time of selection, music and/or instruments were used intentionally to improve the emotional status of participants (i.e., intended or officially confirmed as music therapy). The following exclusion criteria were used: No original research, article was not published (e.g., project phase, in review), unverified data or literature was used, participants did neither receive nor interact with music. Not relevant for in- or exclusion was the kind of questionnaire used to measure depression, additional diagnostic measures for pathologies other than depression, spatial and temporal implementation of treatment, demographics (i.e., number, age,

<sup>3</sup>Clinical speciality areas; Diagnostic, Treatment, and Therapeutic procedures, approaches, tools; Disorders; Age groups; Scientific; Country-specific; Musical Aspects; Recording hardware and equipment; Literature Genre; Publication type or medium; Year of publication; Number of authors.

<sup>4</sup>Boolean Operators for searching databases: Concept explained by the Massachusetts Institute of Technology [MIT].

<sup>5</sup>Our preference was an experimental-control setting, but unfortunately three authors (Ashida, 2000; Guétin et al., 2009b; Schwantes and McKinney, 2010) did not use a control sample.

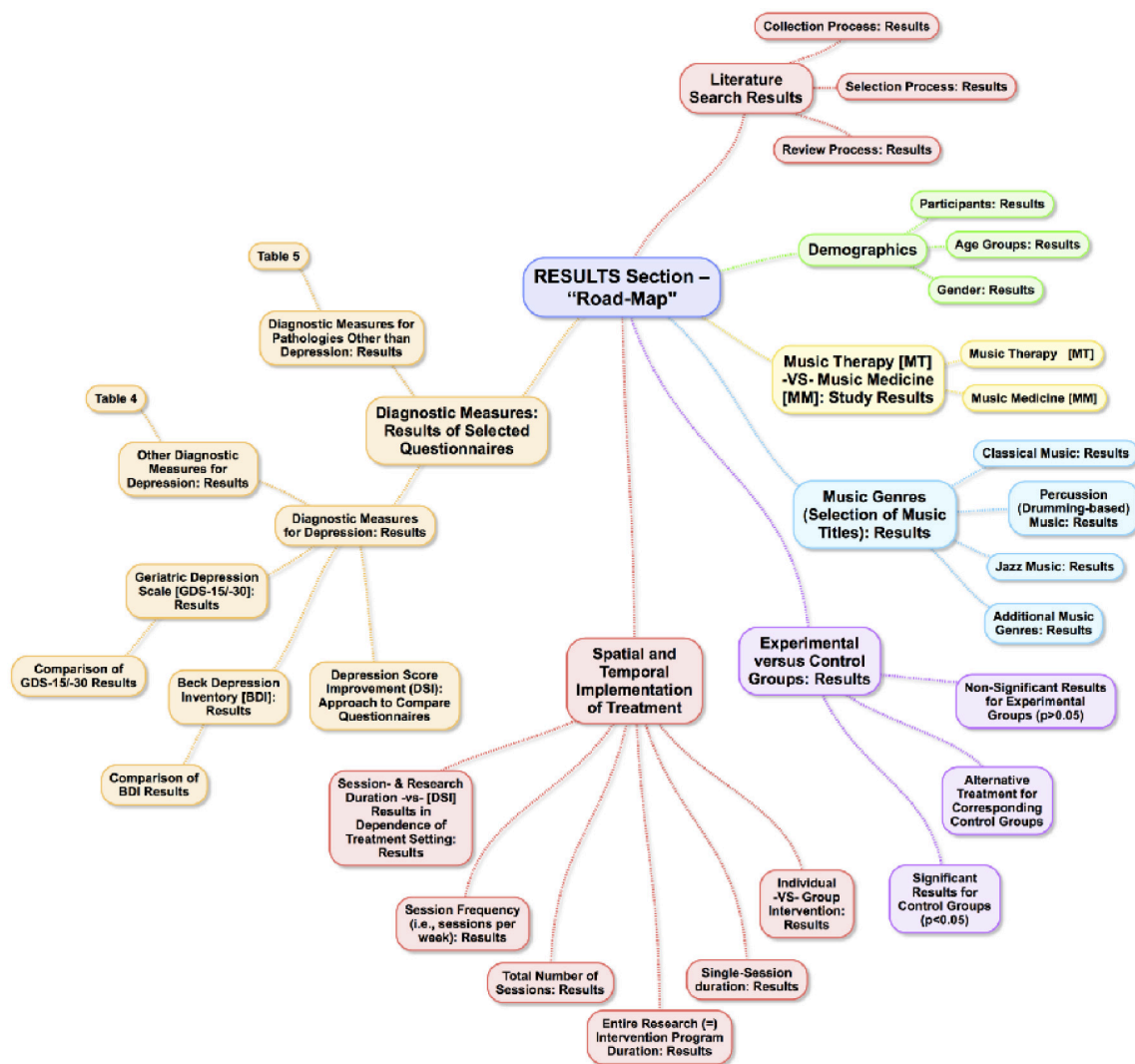


FIGURE 1 | "Road-Map" Outline of the following results section (idea, concept and creation of this Figure by Leubner).

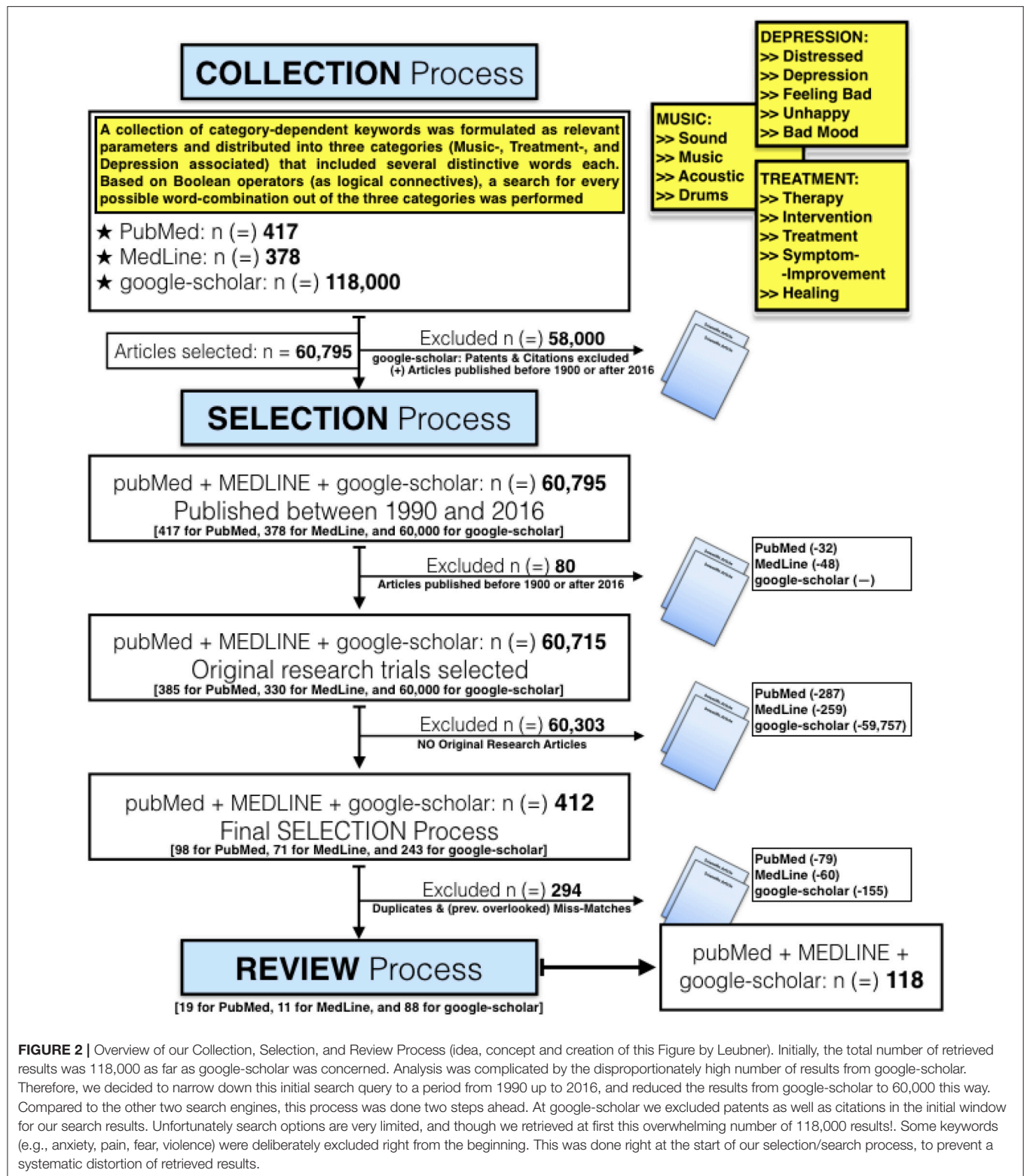
and gender) participants had, or distinctive features (like setting, duration, speakers, live version, and recorded) of stimuli. After the initial number of results, the remaining articles were manually checked for completeness and accuracy of information. Our final selection of articles included 28 research papers.

## General Information – (Figure 1) Evaluating the Methodological Quality of Our Meta-Review

During the review process, we used a very strict self-monitoring procedure to ensure that the quality of scientific research was met to the best of our knowledge and stood in accordance with the standards of good scientific practice. Every effort has been made to provide the accuracy of contents as well as completeness of data published within our meta-review. Inspired by another author's meta-review (Kamioka et al.,

2014), we evaluated our work by the AMSTAR checklist (Shea et al., 2007)<sup>6</sup> and found no reasons for objection regarding our selection of reviews. AMSTAR (acronym for "Assessment of Multiple SysTemAtic Reviews"), a questionnaire for assessing systemic reviews, is based on a rating scale with

<sup>6</sup>AMSTAR (Shea et al., 2007)–Further Info & AMSTAR online calculator: [https://amstar.ca/Amstar\\_Checklist.php](https://amstar.ca/Amstar_Checklist.php); National Collaborating Centre for Methods and Tools (NCCMT): <http://www.nccmt.ca/resources/search/97> Questions included in the AMSTAR-Checklist (Shea et al., 2007) are: (I) Was an "a priori" design provided? (II) Was there duplicate study selection and data extraction? (III) Was a comprehensive literature search performed? (IV) Was the status of publication (i.e. gray literature) used as an inclusion criterion? (V) Was a list of studies (included and excluded) provided? (VI) Were the characteristics of the included studies provided? (VII) Was the scientific quality of the included studies assessed and documented? (VIII) Was the scientific quality of the included studies used appropriately in formulating conclusions? (IX) Were the methods used to combine the findings of studies appropriate? (X) Was the likelihood of publication bias assessed? (XI) Was any conflict of interest included?



11 items (i.e., questions). AMSTAR allows authors to determine and graduate the methodological quality of their systematic review.

### Effect Size

We investigated a wide variety of scholarly papers within our review. There were many different approaches and several

procedures. As far as intervention approaches and procedures were concerned, we found (very) similar trends in several papers. To ensure that those different tendencies were not only based on our pure assumption as well as biased interpretation, we also calculated the effect-size correlation by using the mean scores as well as standard deviations for each of the treatment and control groups, if this setup was used by the respective researcher. Most trials showed a small difference in between the experimental and control group at baseline, what almost always turned into a large effect size regarding post-measurement.

### Depression Score Improvement (DSI) – Approach to Compare Questionnaires

As mentioned above, we selected 28 scholarly articles that used different questionnaires to measure symptoms of depression for experimental and control groups. According to common statistical standards we used a formula to evaluate and compare the relative standing of each mean to every other mean. To avoid confusion, we decided to refer to it as “Depression Score Improvement (DSI).” Mathematically speaking it stands for the mean difference between the pre-test and post-test results (i.e., score changes) in percent.  $DSI_{ind}$  stands for an individual and  $DSI_{Gr}$  for a group setting. Please refer to the Supplementary Materials (Table: “Complete Display of Statistical Data”)<sup>7</sup> for additional information.

## RESULTS

The results will review the works in terms of demographics, treatment implementation, and diagnostic measures.

### Literature Search Results – (Figure 2) Collection Process – Results

A large list of keywords, based on several questions we had, was created initially. They were combined into search-terms and finally put into search-categories as category-dependent keywords. In addition, we discussed several parameters and agreed on three categories (associated to music, treatment, and depression). By querying scientific databases, using the above-mentioned category-dependent keywords as input criteria, we retrieved a very large number of results. We then searched for a combination of the following words and/or phrases (e.g., “*music AND therapy AND depression*”; “*acoustic AND intervention AND unhappy*”), narrowed down the retrieved results according to a combination of several keywords (e.g., “*music therapy*”; “*acoustic intervention*”), and sorted this data according to relevance.

### Selection Process – Results

In step two we applied the above-mentioned approach and narrowed down our search query to a limited period of time, then systematically searched for key phrases, and excluded duplicates as well as previously overlooked miss-matches. Our inclusion criteria can be summarized as follows: Original research

article, already published at time of selection, music and/or instruments were used intentionally to improve the emotional status of participants. Our exclusion criteria were: No original research, article was not published (e.g., project phase, in review), unverified data or literature was used, participants did neither receive nor interact with music.

### Review Process – Results

Based on our predefined criteria for inclusion and exclusion, relevant publications were then selected and used for our intensified review process. After reducing the initial number of results, we obtained the remaining articles, conducted a hand-search in selected scientific journals, and manually checked for completeness as well as accuracy of the contained information. The final selection of articles, according to our selection criteria, included 28 papers.

### Demographics<sup>8</sup>

To begin with, the number of participants as well as age and gender related basic demographics were analyzed.

### Participants – Results

Our final selection of 28 studies included 1,810 participants, with group sizes between five and 236 persons ( $n_{av} = 64.64$ ;  $SD = 56.13$ ). For experimental groups, we counted 954 individuals ( $n_{min} = 5$ ;  $n_{max} = 116$ ;  $n_{av} = 34.07$ ;  $SD = 27.78$ ), and 856 ( $n_{min} = 10$ ;  $n_{max} = 120$ ;  $n_{av} = 30.57$ ;  $SD = 29.10$ ) for the control respectively. Although three authors (Ashida, 2000; Guétin et al., 2009b; Schwantes and Mckinney, 2010) did not use a control sample, those articles were nevertheless considered for calculating accurate and up-to-date data. Depending on each review, sample groups differed profoundly in number of participants. The smallest one had five participants (Schwantes and Mckinney, 2010), followed by three authors (Hendricks et al., 1999; Ashida, 2000; Guétin et al., 2009b) who used between 10 and 20 individuals in their clinical trials. Medium sized groups of up to 100 participants were found in six articles (Gupta and Gupta, 2005; Castillo-Pérez et al., 2010; Erkkilä et al., 2011; Wang et al., 2011; Lu et al., 2013). Large groups with more than 100 (Koelsch et al., 2010; Silverman, 2011), or 200 (Chen et al., 2016) participants were the exception, and 236 participants (Chang et al., 2008) presented the upper end in our selection.

### Age Groups – Results

Within our selected articles, the youngest participant was 14 (Hendricks et al., 1999), and the oldest 95 years of age (Guétin et al., 2009a). We then separated relevant groups, according to their age, into three categories, namely “young,” “medium,” and “elderly.”

### Young

Participants were defined as “young,” if their mean age was below or equal to 30 years ( $\leq 30$ ). Young individuals did show minimal better (i.e., higher) depression score improvements (DSI) (mean difference between the pre-test and post-test results

<sup>7</sup>In our Supplementary Table (“Complete Display of Statistical Data”), DSI was referred to as “Change [%].”

<sup>8</sup>A much More Detailed Representation of Demographics is Available in the Supplementary Table (Appendix-B).

TABLE 1 | Music-Therapy interventions—music types and results.

Author and year	Intervention		Test	Depression measurement methods							
	Music type	[MT] or [MM]		Experimental [mean (SD)]				Control [mean (SD)]			
				Pre (SD)	Post (SD)	Change (%)	p-Level	PRE (SD)	POST (SD)	Change [%]	p-Level
Choi et al., 2008	Drumming, Relax-Music, Singing	[MT]	BDI	49.30 (3.10)	25.50 (2.20)	48.28	$p < 0.001$	47.40 (2.80)	44.80 (3.80)	5.49	$p > 0.05$
Erkkilä et al., 2011	Drumming (Djembe Drum)	[MT]	MADRS	24.60 (6.40)	14.10 (8.77)	42.68	$p < 0.05$	23.00 (7.60)	16.43 (9.33)	28.57	$p > 0.05$
Han et al., 2011	Drumming Singing, Dancing, Improvisation with various Instruments	[MT]	RMBPC <sub>D</sub>	20.50 (23.5)	11.70 (15.9)	42.93	$p < 0.05$	13.10 (21.0)	24.60 (34.70)	−87.79	$p > 0.05$
Hanser and Thompson, 1994	Relax-Music, Improvisational Harp Music, PMR (spoken cues)	[MT] Visit	AES	18.20 (6.40)	19.0 (4.80)	−4.40	$p > 0.05$	17.10 (4.30)	16.60 (5.10)	2.92	$p > 0.05$
			GDS-30	17.30 (5.85)	7.70 (3.66)	55.49	$p < 0.05$	15.30 (5.85)	16.20 (6.13)	−5.88	$p > 0.05$
			GDS-30	17.60 (7.89)	12.30 (8.65)	30.11	$p < 0.05$				
Schwantes and McKinney, 2010	Drumming, Guitar, Piano	[MT]	CES-D	21.60 (3.22)	15.60 (2.66)	27.78	$p < 0.05$	–	–	–	–
Silverman, 2011	12-bar BLUES, Blues Songwriting	[MT]	BDI	n/a	18.79 (9.14)	(−?−)	$p > 0.05$	n/a	20.28 (9.53)	(−?−)	$p > 0.05$

was calculated in percent), if they attended group (mean  $DSI_{Gr} = 53.83\%$ )<sup>9</sup>, rather than individual ( $DSI_{Ind} = 40.47\%$ ) music intervention sessions. These results may be due to the beneficial consequences of social interactions within groups, and thus confirm previous study results (Garber et al., 2009; Tartakovsky, 2015).

### Medium

We used the term “medium” for groups of participants, whose mean ages ranged between 31 (>30) and 59 years (<60). Medium-aged participants presented much better results (i.e., higher depression score improvements), if they attended a group (mean  $DSI_{Gr} = 48.37\%$ ), rather than an individual (mean  $DSI_{Ind} = 24.79\%$ ) intervention setting. However, it should be stressed that our findings only show a positive trend and thus should not be evidence.

### Elderly

The third and final group was defined by us as “elderly” and included participants with a mean age of 60 years or above ( $\geq 60$ ). Noticeable results were found for the age group we defined as elderly, as participants showed slightly better (i.e., higher) score improvements (mean  $DSI_{Ind} = 48.96\%$ ), if they attended an individual setting. Considering the music selection that had been used for elderly participants, a strong tendency toward classical compositions was found (e.g., Chan et al., 2010; Han et al., 2011). Because a relevant number of participants came from Asian countries (e.g., China, Korea), elderly people from those research articles received, in addition to classical music, quite often Asian oriented compositions as well. Despite our extensive investigations, the influence this combination had on results, remained uncertain. Positive tendencies within those groups might be due to “traditional” and/or “culture related” factors. It is, however, also conceivable that combining Western classical with traditional Asian music is notably suited to produce better results. Concerning this matter, future research on western depression patients treated with a combination of classical Western, and traditional Asian music might be a promising concept to be further explored.

### Gender – Results

As far as gender was concerned, we subdivided each sample group in its female and male participants. Women and men were found in 20 study designs. This was the most frequently used constellation. Within this selection, we did not find any significant differences, and so no further analysis was done. Only women took part in two studies (Chang et al., 2008; Esfandiari and Mansouri, 2014)<sup>10</sup>. Interestingly the same stimuli setup was used in both cases. It consisted of instrumental music without vocals, stored on a digital record, and was presented via loudspeakers from a CD (Chang et al., 2008) or MP3 player (Esfandiari and Mansouri, 2014). Only men

were seen in four research papers (Gupta and Gupta, 2005; Schwantes and McKinney, 2010; Albornoz, 2011; Chen et al., 2016). A significant improvement of depression scores was reported for every experimental group, and once (Albornoz, 2011) for a corresponding control setting (received only standard and no alternative treatment). Three articles (Schwantes and McKinney, 2010; Albornoz, 2011; Chen et al., 2016) shared several similarities, as percussion instruments (e.g., drums, tambourines) were part of each genre selection, all participants received music interventions in a group setting, and stimuli were actively produced within a live performance. In addition, the BDI questionnaire has also been used in three cases (Gupta and Gupta, 2005; Albornoz, 2011; Chen et al., 2016), and thus we were able to perform a search for similarities or tendencies. The average duration for one music intervention was 80 ( $SD = 45$ ) min and the total number of sessions was 17 ( $SD = 5$ ) in average. Two publications (Hsu and Lai, 2004; Wang et al., 2011) did not offer any information about gender related distribution of participants.

## Music Therapy [MT] vs. Music Medicine [MM] – Study Results

### Music-Therapy [MT]

Within our selection of 28 articles, six explicitly mentioned a certified music therapist (Hanser and Thompson, 1994; Choi et al., 2008; Schwantes and McKinney, 2010; Erkkilä et al., 2011; Han et al., 2011; Silverman, 2011)<sup>11</sup>. For five articles with available data, a combined average depression score improvement (DSI) of 40.87% ( $SD = 7.70\%$ ) was calculated for the experimental groups. As far as the relevant control groups were concerned, only twice depression scores decreased at all (Choi et al., 2008; Erkkilä et al., 2011; **Table 1**).

Regarding the kind of music provided by a board-certified music therapist, we found some similarities that stood out and appeared more frequently, when compared to music medicine interventions. Percussion music (mainly drumming) was used by four researchers (Choi et al., 2008; Schwantes and McKinney, 2010; Erkkilä et al., 2011; Han et al., 2011). One author (Choi et al., 2008) used music based on instruments that were selected according to participant's preferences. Included were, for example, egg shakes, base-, ocean-, and paddle-drums. Participants actively played and passively listened to instruments or sounds, complemented by singing together. Another researcher (Erkkilä et al., 2011) preferred the African Djembe<sup>12</sup> drum as well as a selection of several percussion sounds created digitally by an external MIDI (*Musical Instrument Digital Interface*) synthesizer. Percussion-oriented improvisation that included rhythmic drumming and vocal patterns was another approach one scholar (Han et al., 2011) used for his stimuli selection. Congas, Cabassas, Ago-Gos, and Claves was the percussion-based selection (in addition a guitar and a Piano was also available) in the fourth music-therapy article

<sup>9</sup>DSI: Depression Score Improvement stands for the mean difference between the pre-test and post-test results (i.e., score changes) in percent. Please refer to the supplementary materials for additional information.

<sup>10</sup>Music interventions: Individual setting (Chang et al., 2008); Group setting (Esfandiari and Mansouri, 2014).

<sup>11</sup>One [MT] music-therapy article (Silverman, 2011) was not used for comparison and calculations because the relevant data was unavailable.

<sup>12</sup>Djembe is based on the expression “anke djé, anke bé” which roughly translates as “everyone should come together in peace and harmony.”

(Schwantes and McKinney, 2010). Twice, music without the use of percussion instruments or drums in general, was selected for the intervention. Once (Hanser and Thompson, 1994) relaxing, slow and rhythmic harp-samples, played from a cassette-player, were used. In addition, each of the participants was invited to bring some samples of her or his favored music titles. The second one (Silverman, 2011) decided to play a “12-bar Blues” (i.e., “blues changes”)<sup>13</sup> progression as an introduction, followed by a Blues songwriting session. The last-mentioned music-therapy project was the only article out of six, where participants within their respective music intervention group did not present a significant reduction of depression. A very interesting “fund” was that none of the music-therapy articles neither concentrated their main music selection on classical, nor on Jazz music. When we looked for other distinctive features it turned out that stimuli were actively produced within a live performance in five articles. There was only one exception (Hanser and Thompson, 1994), where a passive presentation of recorded stimuli was preferred by the scholar.

### Music-Medicine [MM]

The remaining 22 research articles did not explicitly mention a certified music therapist. In those cases, some variant of music medicine was used for intervention. Often the expression music therapy was used, although a more detailed description or specific information was neither published nor available upon our request. With one exception (Castillo-Pérez et al., 2010), we could calculate the (DSI)<sup>9</sup> for 25 articles that used some variant of music-medicine [MM].

When we investigated the kind of music that was used, a broader selection of genres was found. Percussion based tracks and drumming appeared in five scholarly papers (Ashida, 2000; Alborno, 2011; Lu et al., 2013; Chen et al., 2016; Fancourt et al., 2016). Researchers that used drums reported a significant depression score improvement for every experimental group and we calculated an average of 53.71% for those five articles. Regarding the kind of genre used in our selection of music-medicine articles, a wider range of genres was found. One of the biggest differences was that only music-medicine articles used, in addition to percussion stimuli, also classical and Jazz music for their intervention. Please note that for reasons of confusion, we do not mention the Seamless Transitions between Music Therapy [MT] and Music Medicine [MM] from the “Materials and Methods Section.”

### Music Genres (Selection of Music Titles) – Results

Regarding the kind of music used in our selection of research articles, a wide range of genres was found. Mainly three styles,

classical<sup>14</sup> (9x), percussion<sup>15</sup> (9x), and Jazz (5x) music were used more frequently for music intervention. The evaluation took place when specific compositions showed significantly greater improvements in depression compared to other research attempts. Utilizing our comprehensive data analysis, music titles were categorized according to genre or style (e.g., classical music, Jazz), narrowed down (e.g., Jazz), sorted by magnitude of depression score improvements (DSI)<sup>9</sup>, and finally examined for distinctive features (like setting, duration, speakers, live version, recorded). Similarities that stood out and appeared more frequently among one selected music genre were compared with the 28 scholarly articles we selected for our meta-review.

### Classical Music – Results

In nine articles, classical music (Classical or Baroque period)<sup>22</sup> was used. Several well-known composers such as W.A. Mozart (Castillo-Pérez et al., 2010), L. v. Beethoven (Chang et al., 2008; Chan et al., 2009) and J. S. Bach (Castillo-Pérez et al., 2010; Koelsch et al., 2010) have been among the selected samples. If classical music was used as intervention, our calculations revealed that four studies out of eight<sup>16</sup> were among those with depression score improvements (DSI)<sup>31</sup> that were above the average<sup>17</sup> of 39.98% ( $SD = 12$ ). When we looked for similarities between these, three of the four studies (Harmat et al., 2008; Chan et al., 2009; Guétin et al., 2009a) used individual sessions, rather than a group setting (Koelsch et al., 2010). For all four articles mentioned above, we calculated an average of 11 ( $SD = 10$ ) for the total number of sessions that included classical music. The remaining five articles on the other hand, presenting results not as good as the aforementioned, showed an average of 30 ( $SD = 21$ ) music interventions. One plausible hypothesis might be “saturation effect” caused by too many interventions in total. Too little variety within the selection of music titles has probably played an important role as well. A general tendency that less intervention sessions in total would lead to better results for every case where classical compositions were included could not be confirmed for our selection.

### Percussion (Drumming-based) Music – Results

Percussion music (mainly drumming) was used by nine<sup>18</sup> researchers, and among those, two ways of integration were

<sup>14</sup> Ambiguity of the term “classical” music: In our review, this term refers to “Western Art Music” and thus includes, but is not limited to the “Classical” music period. Most of the time we used this term for music from the Baroque (1600–1750), Classical (1750–1820), and Romantic (1804–1910) period.

<sup>15</sup> Within percussion groups various types of drums presented the instrument of choice most of the time.

<sup>16</sup> Eight out of nine articles because in one case (Castillo-Pérez et al., 2010) scores were missing. The remaining were: Hsu and Lai, 2004; Chang et al., 2008; Harmat et al., 2008; Chan et al., 2009, 2010; Guétin et al., 2009a; Koelsch et al., 2010; Verrusio et al., 2014.

<sup>17</sup> Average: Arithmetic mean of all score-changes in [%] for a defined selection (e.g., classical music). Example: We calculated the score-change in [%] for each of the eight experimental groups that received classical music as intervention. In this case the arithmetic mean ( $DSI_{Clas}$ ) was 39.98% (i.e., average). Then every individual score can be compared to this average. If it was above, we called it “above average”.

<sup>18</sup> Percussion music (drumming): Ashida, 2000; Choi et al., 2008; Schwantes and McKinney, 2010; Alborno, 2011; Erkkilä et al., 2011; Han et al., 2011; Lu et al., 2013; Chen et al., 2016; Fancourt et al., 2016.

<sup>13</sup> 12-bar Blues: Traditional Blues pattern that is 12 measures long. This chord progression is also used for many other music genres and quite popular in pop-music.

found. On the one hand, rhythmic percussion compositions were included as part of the music title selection used for intervention. On the other hand, and this was the case in nine articles, various forms of drums had been offered to those who joined the experimental groups, allowing them to “produce their own” music. Sometimes participants were accompanied by a music therapist (e.g., Albornoz, 2011) or professional artist (Fancourt et al., 2016), who gave instructions on how to use and play these instruments. When we looked for trends or distinctive features percussion music (in particular drumming) had, it turned out that, except one article (Erkkilä et al., 2011), all were carried out within a group, rather than an individual setting. A further search for additional similarities, leading to better outcome scores, did not deliver any new findings as far as improvement of depression was concerned. Participants in altogether 7 out of 9 percussion groups were medium aged, two authors (Ashida, 2000; Han et al., 2011) described elderly participants, whereas none of the percussion groups included young participants.

A wide and even distribution of reduced depression scores across all outcome levels became apparent, when participants received percussion (or drumming) interventions. We calculated an average depression score improvement (DSI) of 47.80% ( $SD = 14$ ). Above-average results regarding depression score improvement (DSI), were achieved in four experiments that had an average percussion session duration of 63 ( $SD = 19$ ) min. In comparison, we calculated for the remaining five articles an average of 93 ( $SD = 26$ ) min. Although a difference of 30 min showed a clear tendency, it was not enough of a difference to draw any definitive conclusions.

### Jazz Music – Results

Finally, five<sup>19</sup> researchers used primarily Jazz<sup>20</sup> as music genre for their intervention. Featured performers (artists) were Vernon Duke (“April in Paris”) (Chan et al., 2009), M. Greger (“Up to Date”), and Louis Armstrong (“St. Louis Blues”) (Koelsch et al., 2010). Unfortunately, available data was quite limited, mainly since most authors did not disclose relevant information and a detailed description was rarely seen. Some interesting points were also found for research articles that used Jazz as a treatment option. All five of them were among those with good outcome scores, as far as depression reduction was concerned. Test scores ranged between a significance level of  $p < 0.01$  (Guétin et al., 2009a; Verrusio et al., 2014; Chen et al., 2016) and sometimes even better than  $p < 0.001$  (e.g., Koelsch et al., 2010; Fancourt et al., 2016). Depression score improvement (DSI) had an average of 43.41% ( $SD = 6$ ). However, there was no clear trend leading toward Jazz as a more effective intervention option, when compared to other music genres. This was assumed because the two studies that showed the best<sup>21</sup> reduction in depression [Chan et al., 2010 (DSI = 48.78%); Koelsch et al., 2010 (DSI = 4 6.58%)] used both classical music in addition to

Jazz as an intervention. Experimental groups received two types of intervention (i.e., classical music and Jazz) which eventually blurred outcome scores or prevented more accurate results. Since it was not possible to differentiate to what extent either classical music or Jazz was responsible for the positive trend in reducing symptoms of depression, further research in this field is needed.

### Additional Music Genres – Results

Numerous other music styles were used in the experiments, ranging from Indian ragas<sup>22</sup> played on a flute (Gupta and Gupta, 2005; Deshmukh et al., 2009), nature sound compositions (Ashida, 2000; Chang et al., 2008), meditative (Chan et al., 2010), or slow rhythm music (Chan et al., 2012), to lullabies (Chang et al., 2008), pop or rock (Kim et al., 2006; Erkkilä et al., 2011), Irish folk, Salsa, and Reggae (Koelsch et al., 2010), only to name a few. As far as we were concerned all those genres mentioned above would present interesting approaches for future research. Due to a relatively small number and simultaneously wide-ranging variety, more thorough investigations are needed, though. These should be examined independently. As far as the above-mentioned music genres, other than classical, percussion, or Jazz were concerned, no indication for a preferable combination was observed.

### Experimental vs. Control Groups – Results

#### Non-significant Results for Experimental Groups ( $p > 0.05$ )

In two (Deshmukh et al., 2009; Silverman, 2011) out of 28 studies within our selection of research papers, no significant reduction in depression scores was reported, after participants participated in music interventions. Within those two cases all relevant statistical observations differed without any obvious similarities indicating reasons for non-significant results. Although the results did not meet statistical significance for symptom improvement, both authors explicitly pointed out that positive changes in the severity of depression became obvious for the respective experimental groups. We declared one article (Guétin et al., 2009b) as significant, although it was marked as non-significant in our complete table. This was due to the overall results of this specific research paper, with significant [HADS-D] test scores for weeks 5, 10, and 15. Only week 20 did not follow this positive trend of improvement. It is also important to mention that after music treatment every one of the additional tests [HADS-A for Anxiety; Face(-Scale) to measure mood] showed significant improvements for the experimental group.

#### Alternative Treatment for Corresponding Control Groups

Control groups, who received an alternative (i.e., non-music) intervention, were found in nine research articles (e.g., Guétin et al., 2009a; Castillo-Pérez et al., 2010)<sup>23</sup>. We investigated whether there were particularly noticeable differences in outcome scores, when relevant control groups, who received an alternative

<sup>19</sup>Jazz: Chan et al., 2009, 2010; Guétin et al., 2009a; Koelsch et al., 2010; Verrusio et al., 2014.

<sup>20</sup>In most cases there was no further categorization between different musical sub-genres of Jazz.

<sup>21</sup>Greatest: Best in terms of depression score improvement (DSI) (i.e., pre-post score reduction in percent) with Jazz as intervention.

<sup>22</sup>Raga: Classification system for music that originated during the eleventh century in Asia (mainly India).

<sup>23</sup>Setting was always: Experimental group received music as intervention, and the corresponding control group received an (non-music) alternative.

treatment, were compared to those who received no additional intervention at all (or only the usual treatment)<sup>24</sup>. As far as these nine articles were concerned, a significant reduction ( $p < 0.05$ ) in depression scores was found in every experimental but only one control setting (Hendricks et al., 1999). In this case, an entirely different result became apparent, when control participants received a Cognitive-Behavioral Therapy [CBT] and a significant reduction ( $p < 0.05$ ) in depression scores was measured compared to the respective baseline score, although music still lead to better results. Another scholar (Chan et al., 2012)<sup>25</sup>, instructed participants in the control group to take a resting period, while simultaneously the experimental attendees joined their music intervention session. This alternative approach did not reduce the [GDS-15] depression score, but even increased it. Interestingly, the same author previously published (Chan et al., 2009) a significant ( $p = 0.007$ ) increase (i.e., worsening of depression) for the relevant control setting. To be complete, a resting period was also conducted in another case (Hsu and Lai, 2004), but results showed also no significant reduction in depression scores. Other attempts to provide an alternative intervention for the control group have been monomorphic tones (Koelsch et al., 2010) that corresponded to the experimental music samples (in pitch-, BPM-, and duration), verbal treatment sessions (Silverman, 2011), antidepressant drugs (Verrusio et al., 2014)<sup>26</sup>, reading sessions (Guétin et al., 2009a) or a “conductive-behavioral” psychotherapy (Castillo-Pérez et al., 2010).

### Significant Results for Control Groups ( $p < 0.05$ ):

Significant reduction of depression ( $p < 0.05$ ) in corresponding control (“non-music treatment”) groups was reported twice (Hendricks et al., 1999; Albornoz, 2011) within our selection of scholarly articles. In one instance (Albornoz, 2011) the relevant participants received only standard care, but in the other case (Hendricks et al., 1999) an already above mentioned alternative treatment (i.e., “Cognitive-Behavioral Activities”) was reported.

## Spatial and Temporal Implementation of Treatment

### Individual vs. Group Intervention – Results

As postulated by previous literature (Wheeler et al., 2003; Maratos et al., 2008), we differentiated mainly two scenarios based on the number of participants who attended music intervention sessions and referred to them as “group” or “individual.” Group sessions can awaken participants’ social interactions and individual sessions often provides motivation (Wheeler et al., 2003). Here, a “group” scenario was specified, if two or more persons ( $n \geq 2$ ) were treated simultaneously, whereas “individual” determined experimental settings where only one single person received music interventions individually

( $n = 1$ ). Among our article selection we could find a well-balanced distribution of 15 trials with participants who received music interventions in a group, while 13 researchers used an individual setting. First, the impact of individual compared to group treatment was evaluated. Here an almost equivalent outcome (for the significance-level of results) across all 13 individual, compared to 15 group settings was found, without any advantage to one over the other. Non-significant improvements were seen once for a group (Silverman, 2011) and once<sup>27</sup> for an individual (Deshmukh et al., 2009) intervention.

### Single-Session Duration – Results

The question whether groups showed different (i.e., more or less) improvements, if the duration of one single session was altered, we decided to use the intervention length as a key metric (Figure 3). Except for two instances (Hendricks et al., 1999; Wang et al., 2011), 26 research papers reported the duration one single treatment had. Among those 20 min (Guétin et al., 2009a) was the shortest, and 120 min (Albornoz, 2011; Han et al., 2011) the longest duration for one session. The average for all 26 articles was 55 min, 70 min for 13<sup>28</sup> group settings, and 40 min as far as the 13 individual intervention setups were concerned.

### Entire Research (=) Intervention Program Duration – Results

Continuing our review process, some interesting diversity was found for the scheduled (i.e., total) treatment duration (Figure 3). It ranged from 1 day in two cases (Koelsch et al., 2010; Silverman, 2011) up to 20 (Guétin et al., 2009b), or even 24 weeks (Verrusio et al., 2014). Out of 26 trials an average duration of 7 weeks was found. In two cases, the data was missing (Wang et al., 2011; Esfandiari and Mansouri, 2014). The scheduled (i.e., total) treatment duration was determined by a variety of factors. Our investigation, whether there was any relationship between the entire duration of experimental projects and relevant outcome scores, delivered the following results. For an individual (Ind) therapy setting, we isolated eight<sup>29</sup> research papers with above average<sup>30</sup> results in depression score improvement ( $DSI_{Ind} > 36.50\%$ ). We then calculated for the entire project an average duration of almost 7 weeks. For the remaining five<sup>31</sup> articles that also used an individual approach, but had below average depression score improvements, an average duration of 6 weeks was found. A different picture became apparent when we selected those four<sup>32</sup> articles that presented better than average ( $DSI_{Gr} > 49.09\%$ ) results in depression score improvement, after participants received music intervention in a group (Gr). Percussion music (mainly drumming) was used

<sup>24</sup>For example, if elderly people lived in a retirement home, a standard daily routine or common everyday activities were seen as usual or regular treatment. If, on the other hand, a resting period (e.g., Chan et al., 2012) was carried out simultaneously, this was interpreted as an (“non-music”) alternative.

<sup>25</sup>In all three of his articles within our selection (Chan et al., 2009, 2010, 2012) participants were instructed to rest.

<sup>26</sup>Pharmacotherapy treatment included SSRI (Paroxetine 20mg/die), NaSSA (Mirtazapine 30 mg/die), and Benzodiazepine (Alprazolam).

<sup>27</sup>As already described above, the other individual setting (Guétin Soua, et al., 2009) with pre-post results of  $p > 0.05$  was still counted as significant.

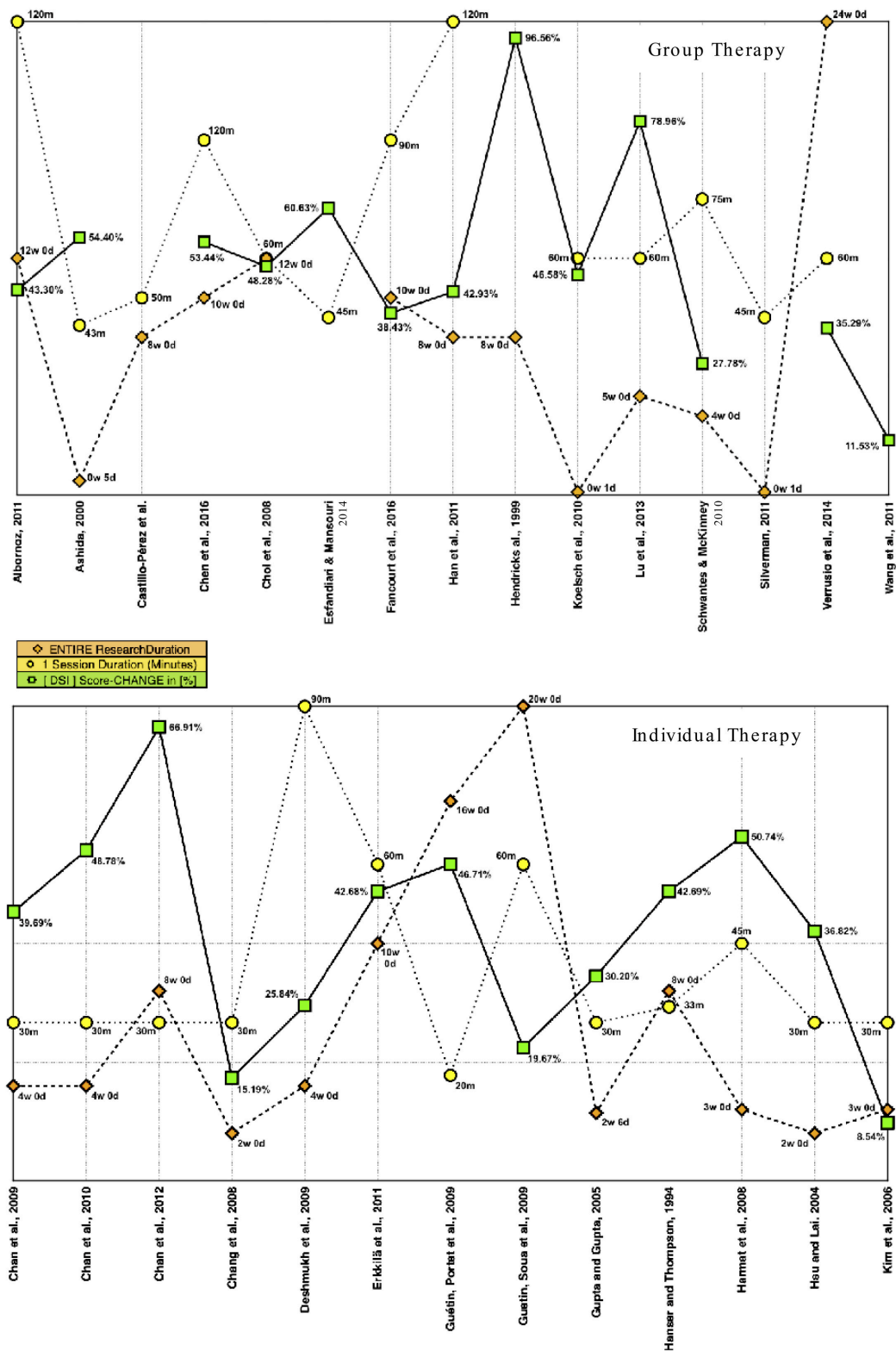
<sup>28</sup>Information regarding the duration for one group session was unavailable in two articles (Hendricks et al., 1999; Wang et al., 2011).

<sup>29</sup>Hanser and Thompson, 1994; Hsu and Lai, 2004; Harmat et al., 2008; Chan et al., 2009, 2010, 2012; Guétin et al., 2009a; Erkkilä et al., 2011.

<sup>30</sup>Average DSI for all 13 articles that used an individual (\*Ind) treatment as intervention was 36.50%.

<sup>31</sup>Gupta and Gupta, 2005; Kim et al., 2006; Chang et al., 2008; Deshmukh et al., 2009; Guétin et al., 2009b.

<sup>32</sup>Once (Esfandiari and Mansouri, 2014) the relevant score was unavailable.



**FIGURE 3 |** Session- and research duration-vs.-[DSI] results in dependence of treatment setting.

by three researchers (Ashida, 2000; Lu et al., 2013; Chen et al., 2016). In comparison, the fourth author (Hendricks et al., 1999) used a selection of relaxing music for treatment. For this setup, a combined duration of six ( $SD = 4$ ) weeks was calculated for the entire project length. On the other hand, a mean close to 10 ( $SD = 7$ ) weeks was found for the remaining 7<sup>33</sup> group intervention projects that were less successful (i.e., below average), as far as depression score reduction was concerned. Based on these results, we concluded that the length for the entire music intervention procedure might be a crucial element for successful results, and seems to be associated with the intervention type. These findings were not enough to draw further conclusions for every project though, but as far as our selection was concerned, a slightly longer intervention duration of 7 weeks led to better results if participants were treated individually. In comparison, for a group setting our calculations revealed a different picture, when we calculated the average entire duration for all relevant research projects. Here it was 6 weeks that produced the most beneficial results within groups. Drums were used for three out of the four projects that presented above average results. Once (Ashida, 2000) a small African drum was used for “drumming activity” at the start of every session. Each time a different participant was asked to perform with this instrument, although nobody in the experimental group was neither a professional drummer nor a musician. African drums were also used by another researcher (Chen et al., 2016). In addition, equipment also included one stereo, one electronic piano, two guitars, one set of hand glockenspiel, and other percussion instruments such as cymbals, tambourines, and xylophones. Finally, percussion instruments used in the third study (Lu et al., 2013) included hand bells, snare drums, a castanet, a tambourine, some claves, a triangle and wood blocks.

### Total Number of Sessions – Results

Continuing the analysis, we evaluated the total number of music intervention sessions. Apparently, this metric was dependent on the duration as well as frequency (“session frequency”) each intervention had. With one exception (Wang et al., 2011), where relevant data was missing, the number of sessions varied considerably. Only a single treatment session was used by three authors (Chan et al., 2010; Koelsch et al., 2010; Silverman, 2011), whereas 56 sessions (Castillo-Pérez et al., 2010) marked the opposite end of the scale. For 27 articles with available data, a combined average of 15 sessions was found. As far as the total number of sessions in an individual type of setting was concerned, above average results had a combined number of 13 ( $SD = 5$ ) sessions, whereas the remaining six research works had 18 ( $SD = 8$ ) interventions. The best results in a group setting showed an average of 17 sessions ( $SD = 15$ ) and they were found in 7 scholarly publications. In comparison, we calculated 14 sessions in total for the remaining 7 articles.

### Session Frequency (i.e., Sessions per Week) – Results

As described previously (Wheeler et al., 2003), the number of sessions can produce different results. Researchers, within

our selection of 28 articles, used various approaches for their experiment, as far as the “session frequency” (i.e., number of sessions within a defined duration) was concerned. Pre-defined intervals ranged from once a week up to one time a day. Once (Choi et al., 2008), the article did mention the total number of sessions ( $n = 15$ ) with a “frequency” of one to two times a week and a total intervention duration of 12 weeks. To be able to present an appropriate comparison of statistical data, a mean of 1.25 sessions per week was calculated. Besides two cases (Wang et al., 2011; Esfandiari and Mansouri, 2014) where no information was provided, the combined average session frequency for the remaining 26 articles was 2.89 ( $SD = 2.50$ ) interventions per week. Usually sessions were held once a week.

### Session- and Research Duration – vs. – [DSI] Results in Dependence of Treatment Setting

We further investigated if there was an association between therapy setting (individual or group), the length of a single session, and trial duration with regard to symptom improvement. Groups (Figure 3) showed better (i.e., above average) improvements in depression, if each session had an average duration of 60 min, and the mean length of treatment was 4–8 weeks.

In comparison, the two variables, session length and trial duration, had different effects for individual treatment approaches (Figure 3). Above average results were found for sessions lasting 30 min combined with a treatment duration between 4 and 8 weeks.

### Diagnostic Measures – Results of Selected Questionnaires

We discovered some distinctive features as well as certain similarities in our selection of 28 articles. They might be a guidance for future research projects and as such are presented in more detail in the subsections below.

### Beck Depression Inventory [BDI]

There are three versions of the BDI. The original [BDI] (Beck et al., 1961), followed by its first [BDI-I/-1A] (Beck et al., 1988) and second [BDI-II] revision (Beck et al., 1996). Beck used a novel approach to develop his inventory by writing down the verbal symptom description of his patients with depression and later sorted his notes according to intensity or severity.

### Beck Depression Inventory [BDI] – Results

The BDI<sup>34</sup> (Beck et al., 1961, 1996) was the most widely used screening tool in our scholarly selection. It was used in eight trials, but we only selected 7<sup>35</sup> studies for evaluating pre-post BDI scores. Once (Harmat et al., 2008), results were only provided for the experimental group, although an experimental control setting was described by the author. Twice (Harmat et al., 2008; Esfandiari and Mansouri, 2014) two experimental groups and

<sup>33</sup>Once (Wang et al., 2011) the relevant score was unavailable.

<sup>34</sup>BDI: Original BDI from 1961; (1st) Revision (=) BDI-I or BDI-1A from 1978; (2nd) Revision (=) BDI-II from 1996.

<sup>35</sup>BDI-scores were measured only once (Silverman, 2011), either at the end (experimental group), or at the beginning (control group) and thus was excluded for this calculation.

one control group were reported. In one case (Esfandiari and Mansouri, 2014) two different music genres were used (“Light Pop & Heavy Rock”), and in another incident (Harmat et al., 2008) the second experimental group listened to an audiobook (“Music & Audiobook”). BDI baseline scores, that indicated a minimal<sup>36</sup> to mild<sup>37</sup> depression, were found in two articles (Gupta and Gupta, 2005; Harmat et al., 2008). Both authors reported for their experimental group a significant improvement of (BDI) depression scores. We calculated an overall average reduction of 2.72 ( $SD = 0.03$ ). Moderate<sup>38</sup> signs of depression, with BDI baseline scores that ranged from 18.66 (Albornoz, 2011) to 24.72 (Chen et al., 2016), were found twice. Music intervention improved BDI scores significantly, with an overall average reduction of 10.65 ( $SD = 3.63$ ) for both articles mentioned above. For the respective control groups one author (Chen et al., 2016) reported non-significant pre-post changes, whereas the other researcher (Albornoz, 2011) described a significant<sup>39</sup> reduction in the standard treatment group as well. The remaining three scholarly papers (Hendricks et al., 1999; Choi et al., 2008; Esfandiari and Mansouri, 2014) described participants with a severe<sup>40</sup> depression, as confirmed by the initial (baseline) BDI results. One article (Esfandiari and Mansouri, 2014), of the three mentioned above, used one control and two experimental groups, who were treated with either “light” or “heavy” music. To be able to compare this work with the other studies one single baseline (31.75), post treatment (12.50), and pre-post difference score of 19.25 ( $SD = 2.47$ )<sup>41</sup> was calculated (according to common statistical standards) for both experimental settings. Interestingly, the corresponding control sample showed a three-point increased BDI score ( $p > 0.05$ ) and no decrease at any time. Continuing with the remaining articles, even bigger initial baseline BDI scores of 39.00 ( $SD = n/a$ ) (Hendricks et al., 1999) and 49.30 ( $SD = 3.10$ ) (Choi et al., 2008) were found. In addition, both authors reported a significant pre-post BDI score reduction<sup>42</sup> for their experimental groups. Based on the published data it became evident that BDI scores improved significantly in each of the cases and this time an overall average reduction of 26.90 ( $SD = 9.59$ ) was calculated. Once (Hendricks et al., 1999) a significantly reduced BDI pre-post score was also reported for the control setting, where participants received a cognitive-behavioral activities program as an alternative (non-music) intervention.

We compared all research projects that used the BDI questionnaire (Table 2). Higher baseline scores almost always led

TABLE 2 | Comparison of BDI results.

Authors [BDI]	Music Type/Genre (focus on)	Test-score [mean] Exp-Gr. Baseline	Test-score [mean] Exp-Gr. End/Post	Pre-Post Means Exp-Gr.	Score-Change in (%) Exp-Gr.	p-Level Exp-Gr.	Pre-Post Means Con-Gr.	p-Level Con-Gr.
Albornoz, 2011	Percussion	18.66	10.58	[-]08.08	43.30	$p < 0.005$	[-]02.25	$p < 0.05$
Chen et al., 2016	Percussion	24.72	11.51	[-]13.21	53.44	$p < 0.01$	[-]03.58	$p > 0.05$
Choi et al., 2008	Korean	49.30	25.50	[-]23.80	48.28	$p < 0.001$	[-]02.90	$p > 0.05$
Esfandiari and Mansouri, 2014	Pop & Rock	31.75	12.50	[-]19.25	60.63	$p < 0.05$	[+]03.00	$p > 0.05$
Gupta and Gupta, 2005	Indian Flute	08.94	06.24	[-]02.70	30.20	$p < 0.001$	[-]00.27	$p > 0.05$
Harmat et al., 2008	Classical	05.40	02.66	[-]02.74	50.74	$p < 0.05$	n/a	$p > 0.05$
Hendricks et al., 1999	Relaxing	39.00	01.34	[-]37.66	96.56	$p < 0.05$	[-]15.30	$p < 0.05$
Silverman, 2011	Songwriting	-	18.79	[-]00.00	-	$p > 0.05$	[-]01.49	$p > 0.05$

<sup>36</sup>Minimal depression: BDI-I (= BDI-1A) score (=) 00–09; BDI-II score (=) 00–13.

<sup>37</sup>Mild depression: BDI-I (= BDI-1A) score (=) 10–18; BDI-II score (=) 14–19.

<sup>38</sup>Moderate depression: BDI-I (= BDI-1A) score (=) 19–29; BDI-II score (=) 20–28.

<sup>39</sup>Albornoz (2011) found in both groups a significant reduction for BDI scores albeit to a significantly greater extent in the experimental (–8.08;  $p < 0.01$ ) than in the control (–2.25;  $p < 0.05$ ) setting.

<sup>40</sup>Severe depression: BDI-I (=BDI-1A) score (=) 30–63; BDI-II score (=) 29–63.

<sup>41</sup>Pre-post difference: experimental (1) “light” music (=) 17.50; experimental (2) “heavy” music (=) 21.00 (both  $p < 0.05$  within groups) (Esfandiari and Mansouri, 2014).

<sup>42</sup>Average pre-post BDI reduction of –30.73 ( $SD = 9.80$ ) combined (Hendricks et al., 1999; Choi et al., 2008).

to comparatively bigger score reductions in those experimental groups, who received music intervention. Except for two articles (Hendricks et al., 1999; Alborno, 2011), no significant improvements were found for control samples. For one of the above-mentioned exceptions (Hendricks et al., 1999) an alternative treatment (“Cognitive-Behavioral” activities) was provided, which might be a plausible explanation why those relatively young participants (all 14 or 15 years old) showed such reductions in BDI values. Nevertheless, it is also important to mention that the relevant experimental group improved to a greater extent ( $BDI_{PRE} - BDI_{POST} = 37.66$ ) after treatment. As far as the other case (Alborno, 2011) was concerned, no alternatives (i.e., other than basic or usual care) were offered, and thus no explanation had been established as to how the results could be explained.

### Geriatric Depression scale [GDS-15/-30]

The original *Geriatric Depression Scale* [GDS-30] (Yesavage et al., 1983) includes 30 questions (Hanser and Thompson, 1994; Chan et al., 2009; Guétin et al., 2009a) and its shorter equivalent [GDS-15] (Yesavage and Sheikh, 1986) contains 15 items (Chan et al., 2010, 2012; Verrusio et al., 2014).

### Geriatric Depression Scale [GDS-15/-30] – Results

A more precise analysis of results was also done for the Geriatric Depression Scale (GDS-15/-30) scores. As already suggested by its name, all 223 participants were elderly. Because both GDS versions are based on the same questionnaire, we combined scores of the long (i.e., GDS-30) with the short (i.e., GDS-15) test version and found a total of 223 participants in six articles (e.g., Chan et al., 2009; Verrusio et al., 2014). A possible bias could be prevented because tests were evenly distributed in number, and with respect to higher GDS-30 as well as lower GDS-15 scores, calculations were adapted accordingly. Taking a closer look at the GDS-15/-30 results (Table 3), some similarities could be found for the most successful (all  $p \leq 0.01$ ) four research articles (Chan et al., 2009, 2010; Guétin et al., 2009a; Verrusio et al., 2014). All of them used and mainly focused on classical compositions as far as their music title selection was concerned. The average reduction in depression as measured by the GDS-15/-30 depression scores was 43% ( $-42.62\%$ ;  $SD = 6.24\%$ ). In comparison, every one of the remaining four research projects (Hanser and Thompson, 1994; Ashida, 2000; Han et al., 2011; Chan et al., 2012) also presented significant results, albeit not as good as the above-mentioned (all  $p \leq 0.05$ ). Interestingly, as far as music genres were concerned, the focus of these less successful projects was rhythmic drumming in two cases (Ashida, 2000; Han et al., 2011). For the remaining two (Hanser and Thompson, 1994; Chan et al., 2012) primarily relaxing, slow paced titles<sup>43</sup> were selected as intervention.

<sup>43</sup>One author (Chan et al., 2012) limited his selection to slow music (60–80 beats per minute). The other researcher (Hanser and Thompson, 1994) also used some “energetic” or “empowering” titles, but mainly concentrated on relaxing compositions.

TABLE 3 | Comparison of GDS-15/-30 Results (\*)GDS-15, (\*\*)GDS-30.

Authors [BDI]	Music Type/Genre (focus on)	Test-score [mean] Exp-Gr. Baseline	Test-score [mean] Exp-Gr. End/Post	Pre-Post Means Exp-Gr.	Score-change in [%] Exp-Gr.	p-Level Exp-Gr.	Pre-Post Means Con-Gr.	p-Level Con-Gr.
Chan et al., 2010*	Relaxing	04.10	02.10	[−]02.00	48.78	$p < 0.001$	[+]00.20	$p > 0.05$
Chan et al., 2012*	Relaxing	04.17	01.38	[−]02.79	66.91	$p < 0.05$	[−]00.08	$p > 0.05$
Verrusio et al., 2014*	Jazz/Classical	08.50	05.50	[−]03.00	35.29	$p < 0.01$	[−]00.40	$p > 0.05$
Chan et al., 2009**	Jazz/Classical	13.10	07.90	[−]05.20	39.69	$p < 0.005$	[+]02.40	$p = 0.007$
Guétin et al., 2009a**	Jazz/Classical	16.70	08.90	[−]07.80	46.71	$p < 0.01$	[−]00.60	$p > 0.05$
Hanser and Thompson, 1994**	Harp & PMR	17.45	10.00	[−]07.45	42.69	$p < 0.05$	[+]00.90	$p > 0.05$

## Other Diagnostic Measures for Depression<sup>44</sup> – Results<sup>45</sup>

Several times, additional questionnaires were used to measure changes in the severity of depression.

Researchers performed those surveys (Table 4) in addition to their “main” depression questionnaire. Please refer to our Supplementary Material for a more comprehensive test description.

## Diagnostic Measures for Pathologies Other than Depression – Results

In many instances, additional questionnaires were used (Table 5)<sup>49</sup> to measure symptoms other than depression (e.g., Anxiety is known to be one of the most common depression comorbidities, Sartorius et al., 1996; Bradt et al., 2013; Tiller, 2013). Eight<sup>46</sup> researchers concentrated their investigation entirely on depression, and thus only performed questionnaires related to this pathology. In comparison, most of the remaining studies measured additional pathologies, with some of them known to be often associated comorbidities with depressive symptoms. However, because these topics were not the focus of this review, we won’t discuss them here in detail. A much more detailed representation is available in the Supplementary Table. Please refer to the original studies for a more comprehensive test description.

## DISCUSSION, CONCLUSION AND FURTHER THOUGHTS

Depression often reduces participation in social activities. It also has an impact on reliability or stamina at daily work and may even result in a greater susceptibility to diseases. Music can be considered an emerging treatment option for mood disorders that has not yet been explored to its full potential. To the best of our knowledge, there were only very few meta-analyses, or systematic reviews of randomized controlled trials available that generated the amount of statistical data, which we presented here.

Certain individual-specific attributes of music are recognizable, when the medium of music is decomposed (Durkin, 2014)<sup>47</sup> into its components. Numerous researchers reported the beneficial effects of music, such as strengthening awareness and sensitiveness for positive emotions (Croom, 2012), or improvement of psychiatric symptoms (Nizamie and Tikka, 2014). Group drumming, for example, helped soldiers to deal with their traumatic experiences, while they were in the process of recovery (Bensimon et al., 2008). However, we have concentrated our focus of interest on patients diagnosed with

clinical depression, one of the most serious and frequent mental disorders worldwide.

In this review we examined whether, and to what extent, music intervention could significantly affect the emotional state of people living with depression. Our primary objective was to accurately identify, select, and analyze up-to-date research literature, which utilized music as intervention to treat participants with depressive symptoms. After a multi-stage review process, a total of 1.810 participants in 28 scholarly papers met our inclusion criteria and were finally selected for further investigations about the effectiveness music had to treat their depression. Both, quantitative as well as qualitative empirical approaches were performed to interpret the data obtained from those original research papers. To consider the different methods researchers used, we presented a detailed illustration of approaches and evaluated them during our investigation process.

Interventions included, for example, various instrumental or vocal versions of classical compositions, Jazz, world music, and meditative songs to name just a few genres. Classical music (Classical or Baroque period) for treatment was used in nine articles. Notable composers were W.A. Mozart, L. v. Beethoven and J. S. Bach. Jazz was used five times for intervention. Vernon Duke (Title: “April in Paris”), M. Greger (Title: “Up to Date”), or Louis Armstrong (Title: “St. Louis Blues”) are some of the featured artists. The third major genre researchers used for their experimental groups was percussion and drumming-based music.

Significant criteria were complete trial duration, amount of intervention sessions, age distribution within participants, and individual or group setting. We compared passive listening to recorded music (e.g., CD), with active experiencing of live music (e.g., singing, improvising with instruments). Furthermore, the analysis of similar studies has enhanced and complemented our work. Previous studies indicated positive effects of music on emotions and anxiety, what we tried to confirm in more detail. The length of an entire music treatment procedure was suspected to be an important element for reducing symptoms of depression. A longer treatment duration of 7 weeks for an individual, compared to nearly 6 weeks in a group setting led to better (i.e., above average) outcomes. Although a difference was discovered, 1 week was not enough to draw further conclusions for each and every project. As far as intervals between sessions were concerned, we found no differences between those research articles that were among the best, compared to the remaining experimental designs. Consequently no trend was becoming apparent, favoring one over the others. We further investigated if there was any association between an individual or a group setting, if the length of a single session and trial duration were compared with regard to symptom improvement. Groups showed better improvements in depression, if each session had an average duration of 60 min, and a treatment between 4 and 8 weeks long. In comparison, the two variables, session length and trial duration, had different effects for individual treatment approaches. Above average results were found for sessions lasting 30 min combined with a treatment duration between 4 and 8 weeks. Furthermore, results were compared according to age groups (“young,” “medium,” and “elderly”). Overall, elderly

<sup>44</sup>For a reference “Intervention Review” about Music Therapy for Depression see: Maratos et al. (2008).

<sup>45</sup>Every available test-result (Pre-/Post-Scores for experimental/control) can be found in our Supplementary Table 12.

<sup>46</sup>Hendricks et al., 1999; Ashida, 2000; Hsu and Lai, 2004; Kim et al., 2006; Chan et al., 2009, 2012; Castillo-Pérez et al., 2010; Albornoz, 2011.

<sup>47</sup>We used the metaphor “decomposed” based on the inspiring book by Andrew Durkin (“Decomposition: A Music Manifesto”), who refers to it “as a way...to demythologize music without demeaning it” (Review by Madison Heying).

**TABLE 4 |** Additional tests, conducted by researchers within our article selection for investigating changes in depression.

Test-Name	Abbreviations	Creator(s)	Used by
Calgary Depression (rating) Scale for Schizophrenia	CDSS	Addington et al., 1990	Lu et al., 2013
Center for Epidemiological Studies Depression Scale	CES-D	Radloff, 1977	Schwantes and Mckinney, 2010
Cornell Scale for Depression in Dementia	CSDD	Alexopoulos et al., 1988	Ashida, 2000
Edinburgh Postnatal Depression Scale	EPDS	Cox et al., 1987	Chang et al., 2008
Hospital Anxiety and Depression Scale - Depression-Subscale	HAD(S)-D	Zigmond and Snaith, 1983	Guétin et al., 2009b; Castillo-Pérez et al., 2010; Fancourt et al., 2016
Hamilton Rating Scale for Depression	HAM-D (=) HRSD	Hamilton, 1960	Albornoz, 2011
Montgomery-Åsberg Depression Rating Scale	MADRS	Montgomery and Asberg, 1979	Deshmukh et al., 2009; Erkkilä et al., 2011
Perception ("Profile") of Mood States (short 35-item version)-Depression Sub-Scale	POMS(-SF) <sub>D</sub>	Curran et al., 1995	Koelsch et al., 2010
Revised Memory and Behavioral Problems Checklist-Depression	RMBPC <sub>D</sub>	Johnson et al., 2001	Han et al., 2011
Self-Rating Depression Scale	SDS	Zung, 1965	Hsu and Lai, 2004; Kim et al., 2006; Castillo-Pérez et al., 2010; Wang et al., 2011

**TABLE 5 |** Additional tests, conducted by researchers within our selection for investigating changes in other pathologies.

Area	Questionnaire-Name	Abbreviations	Creator(s)	Authors
Anxiety	Apparent Emotion Scale-Anxiety	AESA	Lawton et al., 1999	Han et al., 2011
Anxiety	Four Factor Anxiety Inventory	FFAI	Gupta and Gupta, 1998	Gupta and Gupta, 2005
Anxiety	Hamilton's Anxiety Rating Scale	HAM-A (=) HAS	Hamilton, 1959	Guétin et al., 2009b; Verrusio et al., 2014
Anxiety	Hospital Anxiety and Depression Scale-Anxiety-Subscale	HAD(S)-A	Zigmond and Snaith, 1983	Guétin et al., 2009b; Erkkilä et al., 2011; Fancourt et al., 2016
Anxiety	State-Trait Anxiety Inventory	STAI	Spielberger et al., 1983	Gupta and Gupta, 2005; Chang et al., 2008; Choi et al., 2008; Chen et al., 2016
Confidence and Self-esteem	Rosenberg Self-Esteem Inventory	RSI (=) SEI	Rosenberg, 1965	Hanser and Thompson, 1994; Chen et al., 2016
Psychopathology	Symptom Checklist-90	SCL-90	Derogatis, 1975	Wang et al., 2011
Sleep	Epworth Sleepiness Scale	ESS	Johns, 1991	Harmat et al., 2008
Sleep	Pittsburgh Sleep Quality Index	PSQI	Buysse et al., 1989	Harmat et al., 2008; Deshmukh et al., 2009; Chan et al., 2010
Likert-Scale	Self-Assessment Manikins	SAMs	Bradley and Lang, 1994	Koelsch et al., 2010
Likert-Scale	Seven-point Likert scale	—	Silverman, 2011	Silverman, 2011
Likert-Scale	Toronto Alexithymia Scale	TAS-26	Kupfer et al., 2001	Koelsch et al., 2010
Symptom	Brief Symptom Inventory-General Severity Index	BSI-GSI	Derogatis and Spencer, 1982; Derogatis and Melisaratos, 1983	Hanser and Thompson, 1994
Symptom	Brief Symptom Inventory-18	BSI-18	Derogatis, 2000	Schwantes and Mckinney, 2010
Symptom	Positive and Negative Syndrome Scale	PANSS	Kay et al., 1987	Lu et al., 2013
Comorbidity	Comorbidity Index	CInd	Charlson et al., 1987	Verrusio et al., 2014
Illness	Cumulative Illness Rating Scale	CIRS	Linn et al., 1968	Verrusio et al., 2014

people benefitted in particular from this kind of non-invasive treatment. During, but mainly after completion of music-driven interventions, positive effects became apparent. Those included primarily social aspects of life (e.g., an increased motivation to participate in life again), as well as concerned participants' psychological status (e.g., a strengthened self-confidence, an improved resilience to withstand stress).

We described similarities, the integration of different music intervention approaches had on participants in experimental vs. control groups, who received an alternative, or no additional treatment at all. Additional questionnaires confirmed further improvements regarding confidence, self-esteem and motivation. Trends in the improvement of frequently occurring comorbidities (e.g., anxiety, sleeping disorders, confidence and

self-esteem)<sup>48</sup>, associated with depression, were also discussed briefly, and showed promising outcomes after intervention as well. Particularly anxiety (Sartorius et al., 1996; Tiller, 2013) is known to be a common burden, many patients with mood disorders are additionally affected with. Interpreted as manifestation of fear, anxiety is a basic feeling in situations that are regarded as threatening. Triggers can be expected threats such as physical integrity, self-esteem or self-image. Unfortunately, researchers merely distinguished between “anxiety disorder” (i.e., mildly exceeded anxiety) and the physiological reaction. Also, the question should be raised if the response to music differs if patients are suffering from both, depression and anxiety. Sleep quality in combination with symptoms of depression (Mayers and Baldwin, 2006) raised the question, whether sleep disturbances lead to depression or, vice versa, depression was responsible for a reduced quantity of sleep instead. Most studies used questionnaires that were based on self-assessment. However, it is unclear whether this approach is sufficiently valid and reliable enough to diagnose changes regarding to symptom improvement. Future approaches should not solely rely on questionnaires, but rather add measurements of physiological body reactions (e.g., skin conductance, heart and respiratory rate, or AEP’s via an EEG) for more objectivity.

The way auditory stimuli were presented, also raised some additional questions. We found that for individual intervention most of the times headphones were used. For a group setting speakers were the number one choice instead. For elderly participants, a different sensitivity for music perception was a concern, when music was presented directly through headphones. Headphones add at least some isolation from background noises (i.e., able to reduce noise disturbances and surround-soundings). Another concern was that most of the time a certified hearing test was not used. Although, a tendency toward a reduction in the ability to hear higher frequencies is quite common with an increased age, there might still be substantial differences between participants.

Two authors (Deshmukh et al., 2009; Silverman, 2011) reported that participants within their respective music intervention group, did not present a significant reduction of depression. Those two had almost nothing in common<sup>49</sup> and were not investigated further.

Control groups, who received an alternative (“non-music”) intervention, were found in nine research articles. Significant reduction of depression in corresponding control (“non-music intervention”) groups was reported by two authors (Hendricks et al., 1999; Albornoz, 2011). In one instance (Albornoz, 2011) the relevant participants received only standard care, but in the other case (Hendricks et al., 1999) an alternative treatment (Cognitive-Behavioral activities) was reported. Medical conceptions are in a constant state of change. To achieve improvements in areas of disease prevention and treatment, psychology is increasingly associated with clinical medicine and general practitioners. Under the guidance of an experienced music therapist, the

patient receives a multimodal and very structured treatment approach. That is the reason why we can find specialists for music therapy in fields other than psychosomatics or psychiatry today. Examples are internal medicine departments and almost all rehabilitation centers. The acoustic and musical environment literally opens a portal to our unconscious mind. Music therapy often comes into play when other forms of treatment are not effective enough or fail completely.

Music connects us to the time when we only had preverbal communication skills (Hwang and Hughes, 2000; Graham, 2004; i.e., communication before a fully functioning language is developed; e.g., infants or children with autism spectrum disorder), without being dependent on language. Although board-certified music therapy is undeniable the most regulated, developed and professional variant, this should not hinder health professionals and researchers from other areas in the execution of their own projects using music-based interventions. The only thing they should be very precise about, is the way they define their work. Within our selection of articles the expression music therapy was used sometimes, although a more detailed description or specific information was neither published nor available upon our request. In those cases, the term “music therapy” should not be used, but instead music medicine or some of the alternatives mentioned in this manuscript (e.g., therapy with music, music for treatment). This way many obstacles as well as misunderstandings can be prevented in the first place, but high-quality research is still produced. Also, it is very important that researchers contemplate and report the details of the music intervention that they use. For example, they should report whether the music is researcher-selected or participant-selected, the specific tracks they used, the delivery method (speakers, headphones), and any other relevant details.

Encouraged by the promising potential of music as an intervention (Kemper and Danhauer, 2005), we pursued our ambitious goal to contribute knowledge that provides help for the affected individuals, both the patients themselves as well as their nearest relatives. Furthermore, we wanted to provide detailed information about each randomized controlled study, and therefore made all our data available, so others may benefit for their potential upcoming research project. The overall outcome of our analysis, with all significant effects considered, produced highly convincing results that music is a potential treatment option, to improve depression symptoms and quality of life across many age groups. We hope that our results provide some support for future concepts.

## AUTHOR CONTRIBUTIONS

DL (Substantial contributor who meets all four authorship criteria): (1) Project idea, article concept and design, as well as planning the timeline, substantially involved in the data, material, and article acquisition, (2) mainly responsible for drafting, writing, and revising the review article, (3) responsible for selecting and final approving of the scholarly publication, (4) agreed and is accountable for all aspects connected to the work. TH (Substantial contributor who meets all four authorship criteria): (1) Substantial help with the concept and design,

<sup>48</sup>A complete list, with all results we could extract, can be found in the Supplementary Table.

<sup>49</sup>Music Therapy; Duration 90min./session; Session Frequency 7x/week; Raagas Music (Deshmukh et al., 2009).

substantially contributed to the article and material acquisition, (2) substantially contributed to the project by drafting and revising the review article, (3) responsible for final approval of the scholarly publication, (4) agreed and is accountable for all aspects connected to the work.

## REFERENCES

- Addington, D., Addington, J., and Schissel, B. (1990). A depression rating scale for schizophrenics. *Schizophr. Res.* 3, 247–251. doi: 10.1016/0920-9964(90)90005-R
- Al'tman, Y. A., Alyanchikova, Y. O., Guzikov, B. M., and Zakharova, L. E. (2000). Estimation of short musical fragments in normal subjects and patients with chronic depression. *Hum. Physiol.* 26, 553–557. doi: 10.1007/BF02760371
- Albornoz, Y. (2011). The effects of group improvisational music therapy on depression in adolescents and adults with substance abuse: a randomized controlled trial. *Nordic J. Music Ther.* 20, 208–224. doi: 10.1080/08098131.2010.522717
- Alexopoulos, G. S., Abrams, R. C., Young, R. C., and Shamoian, C. A. (1988). Cornell scale for depression in dementia. *Biol. Psychiatry* 23, 271–284. doi: 10.1016/0006-3223(88)90038-8
- Alpass, F. M., and Neville, S. (2003). Loneliness, health and depression in older males. *Aging Mental Health* 7, 212–216. doi: 10.1080/1360786031000101193
- Al'tman, Y. A., Guzikov, B. M., and Golubev, A. A. (2004). The influence of various depressive states on the emotional evaluation of short musical fragments in humans. *Hum. Physiol.* 30, 415–417. doi: 10.1023/B:HUMP.0000036334.57824.46
- American Psychiatric Association (APA) (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.
- Ashida, S. (2000). The effect of reminiscence music therapy sessions on changes in depressive symptoms in elderly persons with dementia. *J. Music Ther.* 37, 170–182. doi: 10.1093/jmt/37.3.170
- Balkwill, L. L., and Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: psychophysical and cultural cues. *Music Percept. Interdisc. J.* 17, 43–64. doi: 10.2307/40285811
- Bally, K., Campbell, D., Chesnick, K., and Tranmer, J. E. (2003). Effects of patient-controlled music therapy during coronary angiography on procedural pain and anxiety distress syndrome. *Crit. Care Nurse* 23, 50–58.
- Beck, A. T., Steer, R. A., Ball, R., and Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients. *J. Person. Assess.* 67, 588–597. doi: 10.1207/s15327752jpa6703\_13
- Beck, A. T., Steer, R. A., and Carbin, M. G. (1988). Psychometric properties of the beck depression inventory: twenty-five years of evaluation. *Clin. Psychol. Rev.* 8, 77–100. doi: 10.1016/0272-7358(88)90050-5
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., and Erbaugh, J. (1961). An inventory for measuring depression. *Arch. Gen. Psychiatry* 4, 561–571. doi: 10.1001/archpsyc.1961.01710120031004
- Bensimon, M., Amir, D., and Wolf, Y. (2008). Drumming through trauma: music therapy with post-traumatic soldiers. *Arts Psychother.* 35, 34–48. doi: 10.1016/j.aip.2007.09.002
- Bittman, B. B., Berk, L. S., Felten, D. L., Westengard, J., Simonton, O. C., Pappas, J., et al. (2001). Composite effects of group drumming music therapy on modulation of neuroendocrine-immune parameters in normal subjects. *Altern. Ther. Health Med.* 7, 38–47.
- Bodner, E., Iancu, I., Gilboa, A., Sarel, A., Mazar, A., and Amir, D. (2007). Finding words for emotions: the reactions of patients with major depressive disorder towards various musical excerpts. *Arts Psychother.* 34, 142–150. doi: 10.1016/j.aip.2006.12.002
- Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exper. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9
- Bradt, J., Dileo, C., Magill, L., and Teague, A. (2016). Music interventions for improving psychological and physical outcomes in cancer patients. *Cochrane Database Syst. Rev.* CD006911. doi: 10.1002/14651858.CD006911.pub3
- Bradt, J., Dileo, C., and Shim, M. (2013). Music interventions for preoperative anxiety. *Cochrane Database Syst. Rev.* CD006908. doi: 10.1002/14651858.CD006908.pub2
- Bradt, J., Potvin, N., Kesslick, A., Shim, M., Radl, D., Schriver, E., et al. (2015). The impact of music therapy versus music medicine on psychological outcomes and pain in cancer patients: a mixed methods study. *Support. Care Cancer* 23, 1261–1271. doi: 10.1007/s00520-014-2478-7
- Buyse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., and Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res.* 28, 193–213. doi: 10.1016/0165-1781(89)90047-4
- Castillo-Pérez, S., Gómez-Pérez, V., Velasco, M. C., Pérez-Campos, E., and Mayoral, M. A. (2010). Effects of music therapy on depression compared with psychotherapy. *Arts Psychother.* 37, 387–390. doi: 10.1016/j.aip.2010.07.001
- Chan, A. S., Ho, Y. C., and Cheung, M. C. (1998). Music training improves verbal memory. *Nature* 396, 128–128. doi: 10.1038/24075
- Chan, M. F., Chan, E. A., and Mok, E. (2010). Effects of music on depression and sleep quality in elderly people: a randomised controlled trial. *Complement. Ther. Med.* 8, 150–159. doi: 10.1016/j.ctim.2010.02.004
- Chan, M. F., Chan, E. A., Mok, E., Tse, K., and Yuk, F. (2009). Effect of music on depression levels and physiological responses in community-based older adults. *Int. J. Mental Health Nurs.* 18, 285–294. doi: 10.1111/j.1447-0349.2009.00614.x
- Chan, M. F., Wong, Z. Y., Onishi, H., and Thayala, N. V. (2012). Effects of music on depression in older people: a randomised controlled trial. *J. Clin. Nurs.* 21, 776–783. doi: 10.1111/j.1365-2702.2011.03954.x
- Chang, M. Y., Chen, C. H., and Huang, K. F. (2008). Effects of music therapy on psychological health of women during pregnancy. *J. Clin. Nurs.* 17, 2580–2587. doi: 10.1111/j.1365-2702.2007.02064.x
- Charlson, M. E., Pompei, P., Ales, K. L., and MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J. Chronic Dis.* 40, 373–383. doi: 10.1016/0021-9681(87)90171-8
- Chen, X. J., Hannibal, N., and Gold, C. (2016). Randomized trial of group music therapy with Chinese prisoners: impact on anxiety, depression, and self-esteem. *Int. J. Offend. Ther. Comp. Criminol.* 60, 1064–1081. doi: 10.1177/0306624X15572795
- Choi, A. N., Lee, M. S., and Lim, H. J. (2008). Effects of group music intervention on depression, anxiety, and relationships in psychiatric patients: a pilot study. *J. Alternat. Complement. Med.* 14, 567–570. doi: 10.1089/acm.2008.0006
- Cox, J. L., Holden, J. M., and Sagovsky, R. (1987). Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *Br. J. Psychiatry* 150, 782–786. doi: 10.1192/bjp.150.6.782
- Croom, A. M. (2012). Music, neuroscience, and the psychology of well-being: a précis. *Front. Psychol.* 2:393. doi: 10.3389/fpsyg.2011.00393
- Curran, S. L., Andrykowski, M. A., and Studts, J. L. (1995). Short form of the Profile of Mood States (POMS-SF): psychometric information. *Psychol. Assess.* 7:80. doi: 10.1037/1040-3590.7.1.80
- Derogatis, L. R. (1975). *How to Use the Symptom Distress Checklist (SCL-90) in Clinical Evaluation: Psychiatric Rating Scale, Vol III, Self-Report Rating Scale*. Nutley, NJ: Hoffmann-La Roche.
- Derogatis, L. R. (2000). *BSI 18, Brief Symptom Inventory 18: Administration, Scoring and Procedures Manual*. Minneapolis, MN: NCS Pearson, Incorporated.
- Derogatis, L. R., and Melisaratos, N. (1983). The brief symptom inventory: an introductory report. *Psychol. Med.* 13, 595–605. doi: 10.1017/S0033291700048017
- Derogatis, L. R., and Spencer, P. M. (1982). *The Brief Symptom Inventory (BSI): Administration, Scoring and Procedures Manual-I*. Baltimore, MD: Johns Hopkins University School of Medicine, Clinical Psychometrics Research Unit.
- Deshmukh, A. D., Sarvaiya, A. A., Seethalakshmi, R., and Nayak, A. S. (2009). Effect of Indian classical music on quality of sleep in depressed patients: a randomized controlled trial. *Nordic J. Music Ther.* 18, 70–78. doi: 10.1080/08098130802697269

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01109/full#supplementary-material>

- Doğan, M., and Senturan, L. (2012). The effect of music therapy on the level of anxiety in the patients undergoing coronary angiography. *Open J. Nurs.* 2, 165–169. doi: 10.4236/ojn.2012.23025
- Droit-Volet, S., Ramos, D., Bueno, J. L., and Bigand, E. (2013). Music, emotion, and time perception: the influence of subjective emotional valence and arousal? *Front. Psychol.* 4:417. doi: 10.3389/fpsyg.2013.00417
- Durkin, A. (2014). *Decomposition: A Music Manifesto, 1st Edn.* New York, NY: Pantheon Books.
- Einstein, A. (1929). *What life means to Einstein – An Interview by G. S. Viereck, The Saturday Evening Post.* Indianapolis, IN: Business & Editorial Offices.
- Erkkilä, J., Punkanen, M., Fachner, J., Ala-Ruona, E., Pöntiö, I., Tervaniemi, M., et al. (2011). Individual music therapy for depression: randomised controlled trial. *Br. J. Psychiatry* 199, 132–139. doi: 10.1192/bjp.bp.110.085431
- Eschrich, S., Münte, T. F., and Altenmüller, E. O. (2008). Unforgettable film music: the role of emotion in episodic long-term memory for music. *BMC Neurosci.* 9:48. doi: 10.1186/1471-2202-9-48
- Esfandiari, N., and Mansouri, S. (2014). The effect of listening to light and heavy music on reducing the symptoms of depression among female students. *Arts Psychother.* 41, 211–213. doi: 10.1016/j.aip.2014.02.001
- Fancourt, D., Perkins, R., Ascenso, S., Carvalho, L. A., Steptoe, A., and Williamson, A. (2016). Effects of group drumming interventions on anxiety, depression, social resilience and inflammatory immune response among mental health service users. *PLoS ONE* 11:e0151136. doi: 10.1371/journal.pone.0151136
- Garber, J., Clarke, G. N., Weersing, V. R., Beardslee, W. R., Brent, D. A., Gladstone, T. R., et al. (2009). Prevention of depression in at-risk adolescents: a randomized controlled trial. *JAMA* 301, 2215–2224. doi: 10.1001/jama.2009.788.
- Gold, C., Voracek, M., and Wigram, T. (2004). Effects of music therapy for children and adolescents with psychopathology: a meta-analysis. *J. Child Psychol. Psychiatry* 45, 1054–1063. doi: 10.1111/j.1469-7610.2004.t01-1-00298.x
- Graham, J. (2004). Communicating with the uncommunicative: music therapy with pre-verbal adults. *Br. J. Learn. Disabil.* 32, 24–29. doi: 10.1111/j.1468-3156.2004.00247.x
- Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2007). Listening to music as a re-creative process: physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Percept.* 24, 297–314. doi: 10.1525/mp.2007.24.3.297
- Guétin, S., Portet, F., Picot, M. C., Pomié, C., Messaoudi, M., Djabelkir, L. et al. (2009a). Effect of music therapy on anxiety and depression in patients with Alzheimer's type dementia: randomised, controlled study. *Dement. Geriatr. Cogn. Disord.* 28, 36–46. doi: 10.1159/000229024
- Guétin, S., Soua, B., Voiriot, G., Picot, M. C., and Herisson, C. (2009b). The effect of music therapy on mood and anxiety-depression: An observational study in institutionalised patients with traumatic brain injury. *Anna. Phys. Rehab. Med.* 52, 30–40. doi: 10.1016/j.annrmp.2008.08.009
- Gupta, B. S., and Gupta, U. (1998). *Manual of the Four Factor Anxiety Inventory.* Varanasi: Pharmacopsychocological Association.
- Gupta, U., and Gupta, B. S. (2005). Psychophysiological responsivity to Indian instrumental music. *Psychol. Music* 33, 363–372. doi: 10.1177/0305735605056144
- Hamilton, M. (1959). The assessment of anxiety states by rating. *Br. J. Med. Psychol.* 32, 50–55. doi: 10.1111/j.2044-8341.1959.tb00467.x
- Hamilton, M. (1960). A rating scale for depression. *J. Neurol Neurosurg. Psychiatry* 23, 56–62. doi: 10.1136/jnnp.23.1.56
- Han, P., Kwan, M., Chen, D., Yusoff, S. Z., Chionh, H. L., Goh, J., et al. (2011). A controlled naturalistic study on a weekly music therapy and activity program on disruptive and depressive behaviors in dementia. *Dement. Geriatr. Cogn. Disord.* 30, 540–546. doi: 10.1159/000321668
- Hanser, S. B. (2014). Music therapy in cardiac health care: current issues in research. *Cardiol. Rev.* 22, 37–42. doi: 10.1097/CRD.0b013e318291c5fc
- Hanser, S. B., and Thompson, L. W. (1994). Effects of a music therapy strategy on depressed older adults. *J. Gerontol.* 49, P265–P269. doi: 10.1093/geronj/49.6.P265
- Harmat, L., Takács, J., and Bodizs, R. (2008). Music improves sleep quality in students. *J. Adv. Nurs.* 62, 327–335. doi: 10.1111/j.1365-2648.2008.04602.x
- Hendricks, C. B., Robinson, B., Bradley, L. J., and Davis, K. (1999). Using music techniques to treat adolescent depression. *J. Human. Counsel.* 38:39. doi: 10.1002/j.2164-490X.1999.tb00160.x
- Higgins, J. P., and Green, S. (eds.). (2008). *Cochrane Handbook for Systematic Reviews of Interventions*, Vol. 5. Chichester: Wiley-Blackwell.
- Hillecke, T., Nickel, A., and Bolay, H. V. (2005). Scientific perspectives on music therapy. *Ann. N.Y. Acad. Sci.* 1060, 271–282. doi: 10.1196/annals.1360.020
- Ho, Y. C., Cheung, M. C., and Chan, A. S. (2003). Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology* 17:439. doi: 10.1037/0894-4105.17.3.439
- Hole, J., Hirsch, M., Ball, E., and Meads, C. (2015). Music as an aid for postoperative recovery in adults: a systematic review and meta-analysis. *Lancet* 386, 1659–1671. doi: 10.1016/S0140-6736(15)60169-61
- Hsu, W. C., and Lai, H. L. (2004). Effects of music on major depression in psychiatric inpatients. *Arch. Psychiat. Nurs.* 18, 193–199. doi: 10.1016/j.apnu.2004.07.007
- Husain, G., Thompson, W. F., and Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Percept. Interdisc. J.* 20, 151–171. doi: 10.1525/mp.2002.20.2.151
- Hwang, B., and Hughes, C. (2000). Increasing early social-communicative skills of preverbal preschool children with autism through social interactive training. *J. Assoc. Persons Severe Hand.* 25, 18–28. doi: 10.2511/rpsd.25.1.18
- Jäncke, L. (2008). Music, memory and emotion. *J. Biol.* 7:1. doi: 10.1186/jbiol82
- Jespersen, K. V., Koenig, J., Jennum, P., and Vuust, P. (2015). *Music for Insomnia in Adults.* London: The Cochrane Library.
- Johns, M. W. (1991). A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 14, 540–545.
- Johnson, M. M., Wackerbarth, S. B., and Schmitt, F. A. (2001). Revised memory and behavior problems checklist. *Clin. Gerontol.* 22, 87–108. doi: 10.1300/J018v22n03\_09
- Kamioka, H., Tsutani, K., Yamada, M., Park, H., Okuizumi, H., Tsuruoka, K., et al. (2014). Effectiveness of music therapy: a summary of systematic reviews based on randomized controlled trials of music interventions. *Patient Prefer. Adher.* 8:727. doi: 10.2147/PPA.S61340
- Kay, S. R., Flszbein, A., and Opfer, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bull.* 13:261. doi: 10.1093/schbul/13.2.261
- Kemper, K. J., and Danhauer, S. C. (2005). Music as therapy. *South Med. J.* 98, 282–288. doi: 10.1097/01.SMJ.0000154773.11986.39
- Kim, K. B., Lee, M. H., and Sok, S. R. (2006). [The effect of music therapy on anxiety and depression in patients undergoing hemodialysis]. *Taehan Kanho Hakhoe Chi* 36, 321–329. doi: 10.4040/jkan.2006.36.2.321
- Koelsch, S., Offermanns, K., and Franzke, P. (2010). Music in the treatment of affective disorders: an exploratory investigation of a new method for music-therapeutic research. *Music Percept. Interdisc. J.* 27, 307–316. doi: 10.1525/mp.2010.27.4.307
- Kupfer, J., Brosig, B., and Brähler, E. (2001). *Toronto-Alexithymie-Skala-26 (TAS-26).* Göttingen: Hogrefe.
- Lawton, M. P., Van Haitsma, K., Perkinson, M., and Ruckdeschel, K. (1999). Observed affect and quality of life in dementia: further affirmations and problems. *J. Mental Health Aging* 5, 69–82.
- Lee, C. (2000). A method of analyzing improvisations in music therapy. *J. Music Ther.* 37, 147–167. doi: 10.1093/jmt/37.2.147
- Linn, B. S., Linn, M. W., and Gurel, L. E. E. (1968). Cumulative illness rating scale. *J. Am. Geriatr. Soc.* 16, 622–626. doi: 10.1111/j.1532-5415.1968.tb02103.x
- Loewy, J. (2004). Integrating music, language and the voice in music therapy. *Voices: A World Forum Music Therapy* 4. doi: 10.15845/voices.v4i1.140
- Lu, S. F., Lo, C. H. K., Sung, H. C., Hsieh, T. C., Yu, S. C., and Chang, S. C. (2013). Effects of group music intervention on psychiatric symptoms and depression in patient with schizophrenia. *Complement. Ther. Med.* 21, 682–688. doi: 10.1016/j.ctim.2013.09.002
- Maratos, A., Gold, C., Wang, X., and Crawford, M. (2008). *Music Therapy for Depression.* London: The Cochrane Library.
- Mayers, A. G., and Baldwin, D. S. (2006). The relationship between sleep disturbance and depression. *Int. J. Psychiatry Clin. Pract.* 10, 2–16. doi: 10.1080/13651500500328087
- McCraty, R., Atkinson, M., Rein, G., and Watkins, A. D. (1996). Music enhances the effect of positive emotional states on salivary IgA. *Stress Med.* 12, 167–175. doi: 10.1002/(SICI)1099-1700(199607)12:3<167::AID-SMI697>3.0.CO;2-2
- Montgomery, S. A., and Asberg, M. (1979). A new depression scale designed to be sensitive to change. *Br. J. Psychiatry*, 134, 382–389. doi: 10.1192/bjp.134.4.382

- Moussavi, S., Chatterji, S., Verdes, E., Tandon, A., Patel, V., and Ustun, B. (2007). Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet* 370, 851–858. doi: 10.1016/S0140-6736(07)61415-9
- Neighbors, B., Kempton, T., and Forehand, R. (1992). Co-occurrence of substance abuse with conduct, anxiety, and depression disorders in juvenile delinquents. *Addict. Behav.* 17, 379–386. doi: 10.1016/0306-4603(92)90043-U
- Nizamie, S. H., and Tikka, S. K. (2014). Psychiatry and music. *Ind. J. Psychiatry* 56, 128–140. doi: 10.4103/0019-5545.130482
- North, A. C., and Hargreaves, D. J. (1999). Can music move people? The effects of musical complexity and silence on waiting time. *Environ. Behav.* 31, 136–149. doi: 10.1177/00139169921972038
- Oakes, S. (2003). Musical tempo and waiting perceptions. *Psychol. Market.* 20, 685–705. doi: 10.1002/mar.10092
- Panksepp, J., and Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behav. Process.* 60, 133–155. doi: 10.1016/S0376-6357(02)00080-3
- Radloff, L. S. (1977). The CES-D scale a self-report depression scale for research in the general population. *Appl. Psychol. Meas.* 1, 385–401. doi: 10.1177/014662167700100306
- Register, D. (2002). Collaboration and consultation: a survey of board certified music therapists. *J. Music Ther.* 39, 305–321. doi: 10.1093/jmt/39.4.305
- Rosenberg, M. (1965). Self esteem and the adolescent (Economics and the social sciences: society and the adolescent self-image). *Science* 148:804. doi: 10.1126/science.148.3671.804
- Särkämö, T., Tervaniemi, M., Laitinen, S., Forsblom, A., Soinila, S., Mikkonen, M., et al. (2008). Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain* 131, 866–876. doi: 10.1093/brain/awn013
- Särkämö, T., Tervaniemi, M., Laitinen, S., Numminen, A., Kurki, M., Johnson, J. K., et al. (2014). Cognitive, emotional, and social benefits of regular musical activities in early dementia: randomized controlled study. *Gerontologist* 54, 634–650. doi: 10.1093/geront/gnt100
- Sartorius, N., Üstün, T. B., Lecrubier, Y., and Wittchen, H. U. (1996). Depression comorbid with anxiety: results from the WHO study on “Psychological disorders in primary health care.” *Br. J. Psychiatry* 168(Suppl. 30), 38–43.
- Schäfer, T., Fachner, J., and Smukalla, M. (2013). Changes in the representation of space and time while listening to music. *Front. Psychol.* 4:508. doi: 10.3389/fpsyg.2013.00508
- Schwantes, M., and McKinney, C. (2010). Music therapy with Mexican migrant farmworkers: a pilot study. *Music Ther. Perspect.* 28, 22–28. doi: 10.1093/mtp/28.1.22
- Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., et al. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med. Res. Methodol.* 7:10. doi: 10.1186/1471-2288-7-10
- Siedliecki, S. L., and Good, M. (2006). Effect of music on power, pain, depression and disability. *J. Adv. Nurs.* 54, 553–562. doi: 10.1111/j.1365-2648.2006.03860.x
- Silverman, M. J. (2011). Effects of music therapy on change and depression on clients in detoxification. *J. Addict. Nurs.* 22, 185–192. doi: 10.3109/10884602.2011.616606
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., and Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Tartakovsky, M. (2015). *5 Benefits of Group Therapy*. Psychological Central. Available online at: <https://psychcentral.com/lib/5-benefits-of-group-therapy/>
- Teo, A. R. (2012). Social isolation associated with depression: a case report of hikikomori. *Int. J. Soc. Psychiatry* 59, 339–341. doi: 10.1177/0020764012437128
- Tiller, J. W. (2013). Depression and anxiety. *Med. J. Aust.* 199, 28–31. doi: 10.5694/mjaol2.10628
- Trappe, H. J. (2010). The effects of music on the cardiovascular system and cardiovascular health. *Heart* 96, 1868–1871. doi: 10.1136/hrt.2010.209858
- Twiss, E., Seaver, J., and McCaffrey, R. (2006). The effect of music listening on older adults undergoing cardiovascular surgery. *Nurs. Crit. Care* 11, 224–231. doi: 10.1111/j.1478-5153.2006.00174.x
- Verrusio, W., Andreozzi, P., Marigliano, B., Renzi, A., Gianturco, V., Pecci, M. T., et al. (2014). Exercise training and music therapy in elderly with depressive syndrome: a pilot study. *Complement. Ther. Med.* 22, 614–620. doi: 10.1016/j.ctim.2014.05.012
- Wang, J., Wang, H., and Zhang, D. (2011). Impact of group music therapy on the depression mood of college students. *Health* 3:151. doi: 10.4236/health.2011.33028
- Werner, P. D., Swope, A. J., and Heide, F. J. (2009). Ethnicity, music experience, and depression. *J. Music Ther.* 46, 339–358. doi: 10.1093/jmt/46.4.339
- Wheeler, B. L., Shiflett, S. C., and Nayak, S. (2003). Effects of number of sessions and group or individual music therapy on the mood and behavior of people who have had strokes or traumatic brain injuries. *Nordic J. Music Ther.* 12, 139–151. doi: 10.1080/08098130309478084
- World Health Organization (WHO) (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: World Health Organization.
- World Health Organization (WHO) (2017). *Depression and Other Common Mental Disorders: Global Health Estimates*. Geneva: World Health Organization.
- Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., et al. (1983). Development and validation of a geriatric depression screening scale: a preliminary report. *J. Psychiatric Res.* 17, 37–49. doi: 10.1016/0022-3956(82)90033-4
- Yesavage, J. A., and Sheikh, J. I. (1986). 9/Geriatric Depression Scale (GDS) recent evidence and development of a shorter version. *Clin. Gerontol.* 5, 165–173. doi: 10.1300/J018v05n01\_09
- Yinger, O. S., and Gooding, L. F. (2015). A systematic review of music-based interventions for procedural support. *J. Music Ther.* 52, 1–77. doi: 10.1093/jmt/thv004
- Zigmond, A. S., and Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatr. Scand.* 67, 361–370. doi: 10.1111/j.1600-0447.1983.tb09716.x
- Zung, W. W. (1965). A self-rating depression scale. *Arch. General Psychiatry* 12, 63–70. doi: 10.1001/archpsyc.1965.01720310065008.

**Conflict of Interest Statement:** 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.

Copyright © 2017 Leubner and Hinterberger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

