

OVERLAP OF NEURAL SYSTEMS FOR PROCESSING LANGUAGE AND MUSIC

EDITED BY: McNeel Gordon Jantzen, Edward W. Large and Cyrille Magne
PUBLISHED IN: Frontiers in Psychology and Frontiers in Neuroscience



frontiers

Frontiers Copyright Statement

© Copyright 2007-2016 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 978-2-88919-911-2

DOI 10.3389/978-2-88919-911-2

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

OVERLAP OF NEURAL SYSTEMS FOR PROCESSING LANGUAGE AND MUSIC

Topic Editors:

McNeel Gordon Jantzen, Western Washington University, USA

Edward W. Large, University of Connecticut, USA

Cyrille Magne, Middle Tennessee State University, USA

The interplay between musical training and speech perception continues to intrigue researchers in the areas of language and music alike. Historically, language function has been attributed to brain regions localized predominately in left hemisphere, whereas music has been attributed to right hemisphere dominant regions. Recent studies demonstrating neural overlap for processing speech and music, and enhanced speech perception and production in musicians suggest that these regions may be inextricably intertwined. The extent of neural overlap between music and speech remains hotly debated, with surprisingly little empirical research exploring specific neural homologies and analogs. Moreover, despite recognition that shared processes likely exist throughout development and depend upon an individual's acoustic experiences, even less research exists on how overlapping neural structures for music and language are affected by developmental trajectories.

Nonetheless, the field is well poised to address key empirical questions, in part because of the recent development of new theories that address the neural and developmental interaction between music and language processing in conjunction with the broad availability of sophisticated tools for quantifying brain activity and dynamics. To understand the overlap of neural structures for language and music processing, research is needed to identify those specific functions of each that influence the other, with areas for enhanced perception of pitch and onset time having already been targeted. Research is also needed to identify the extent to which this overlap is developed in infancy or early childhood and the process by which it affects neural reorganization, plasticity, and trainability in adulthood.

For this research topic, we would like to further explore the relationship between language and music in the brain from two perspectives: 1) understanding the nature of shared neural and cognitive processing for music and language and 2) understanding the developmental trajectory of these neural systems and how they are influenced by experience. We seek to gather technically diverse original research articles that present new empirical findings relevant to understanding:

1. When, in the brain, acoustic information becomes processed specifically as language or music.
2. The shared and independent neural structures for processing music and language.
3. How acoustic experiences such as musical training influence overlap of neural structures for language and music.
4. How the overlap of processing regions changes over time due to experiences at any developmental stage.

Citation: Jantzen, M. G., Large, E. W., Magne, C., eds. (2016). *Overlap of Neural Systems for Processing Language and Music*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-911-2

Table of Contents

04	<i>Editorial: Overlap of Neural Systems for Processing Language and Music</i> McNeel G. Jantzen, Edward W. Large and Cyrille Magne
06	<i>The relationship between the neural computations for speech and music perception is context-dependent: an activation likelihood estimate study</i> Arianna N. LaCroix, Alvaro F. Diaz and Corianne Rogalsky
25	<i>Studying Musical and Linguistic Prediction in Comparable Ways: The Melodic Cloze Probability Method</i> Allison R. Fogel, Jason C. Rosenberg, Frank M. Lehman, Gina R. Kuperberg and Aniruddh D. Patel
39	<i>Rhythmic Effects of Syntax Processing in Music and Language</i> Harim Jung, Samuel Sontag, YeBin S. Park and Psyche Loui
50	<i>Pronunciation difficulty, temporal regularity, and the speech-to-song illusion</i> Elizabeth H. Margulis, Rhimmon Simchy-Gross and Justin L. Black
57	<i>Sex Differences in Music: A Female Advantage at Recognizing Familiar Melodies</i> Scott A. Miles, Robbin A. Miranda and Michael T. Ullman
69	<i>Associations between musical abilities and precursors of reading in preschool aged children</i> Franziska Degé, Claudia Kubicek and Gudrun Schwarzer
79	<i>Does Music Training Enhance Literacy Skills? A Meta-Analysis</i> Reyna L. Gordon, Hilda M. Fehd and Bruce D. McCandliss
95	<i>Sound frequency affects speech emotion perception: results from congenital amusia</i> Sydney L. Lolli, Ari D. Lewenstein, Julian Basurto, Sean Winnik and Psyche Loui
105	<i>Cross-domain processing of musical and vocal emotions in cochlear implant users</i> Alexandre Lehmann and Sébastien Paquette
110	<i>Music and literature: are there shared empathy and predictive mechanisms underlying their affective impact?</i> Diana Omigie



Editorial: Overlap of Neural Systems for Processing Language and Music

McNeel G. Jantzen^{1*}, Edward W. Large² and Cyrille Magne³

¹ Language and Neural Systems Laboratory, Department of Psychology and Behavioral Neuroscience Program, Western Washington University, Bellingham, WA, USA, ² Music Dynamics Laboratory, Department of Psychology, University of Connecticut, Storrs, CT, USA, ³ Brain and Language Laboratory, Department of Psychology and Interdisciplinary Program in Literacy Studies, Middle Tennessee State University, Murfreesboro, TN, USA

Keywords: musical training, language, speech, music, auditory perception, neural plasticity, development, neural overlap

The Editorial on the Research Topic

Overlap of Neural Systems for Processing Language and Music

The relationship between musical training and speech perception has intrigued researchers in language and music for decades, from Bever and Chiarello's (1974) work emphasizing hemispheric specialization to Tallal and Gaab's (2006) findings of shared neural circuitry. Recent studies demonstrating neural overlap for processing speech and music, and enhanced speech perception and production in musicians, suggest that these regions may be inextricably intertwined (Sammler et al., 2007; Wong P.C. et al., 2007; Wong P. et al., 2007; Rogalsky et al., 2011; Schulze et al., 2011). Patel's OPERA hypothesis and Hickok and Poeppel's (2000, 2007) neuroanatomical models continue to evolve and guide this field of research. However, the extent of neural overlap between music and speech remains hotly debated (Norman-Haignere et al., 2015; Peretz et al., 2015), with surprisingly little empirical research exploring specific neural homologs and analogs. Emerging evidence suggests that shared processes likely exist throughout development, depend upon an individual's acoustic experiences, and are affected by developmental trajectories. Moreover, developing theories that address the neural and developmental interaction between music and language processing in conjunction with the broad availability of sophisticated tools for quantifying brain activity and dynamics offer the perfect opportunity for researchers to address these key empirical questions. Taken together, this field of research has begun to elucidate the complex dynamics of overlapping neural areas for processing language and music. This special issue highlights the development of this overlap in early childhood and explores how the interaction between language and musical training enhances cognitive functioning in adults.

This E-Book comprises 10 opinion, perspective, and research papers that focus on the overlap of neural systems for processing language and music. Eight of these papers report original research and new findings that support overlapping neural systems for processing language and music. LaCroix et al. performed a meta-analysis of 171 neuroimaging studies to examine the role of context in processing music and language. Their findings suggest that observed neural overlaps for speech and music might be task-dependent. Fogel et al. developed a novel method for studying and quantifying predictions in musical tasks that is consistent with language tasks. Their melodic cloze probability task can be used to test computational models of melodic expectation and allows for a more precise examination of the relationship between predictive mechanisms in music and language. Using a garden-path design, Jung et al. demonstrated that rhythmic expectancy is crucial to the interaction of processing musical and linguistic syntax. Additionally, their findings support the incorporation of dynamic models of attentional entrainment into existing theories of musical and linguistic syntactical processing. Margulis et al. used the speech-to-song illusion

OPEN ACCESS

Edited and reviewed by:

Isabelle Peretz,
Université de Montréal, Canada

*Correspondence:

McNeel G. Jantzen
mcneel.jantzen@www.edu

Specialty section:

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

Received: 03 May 2016

Accepted: 27 May 2016

Published: 14 June 2016

Citation:

Jantzen MG, Large EW and Magne C
(2016) Editorial: Overlap of Neural
Systems for Processing Language
and Music. *Front. Psychol.* 7:876.
doi: 10.3389/fpsyg.2016.00876

to examine the role of pronunciation difficulty and temporal regularity. Their finding—that difficult to pronounce languages, not differing temporal intervals, elicited a stronger speech-to-song illusion—suggests a stronger speech representation for native and easy to pronounce languages. Miles et al. demonstrated that females have an advantage for recognizing familiar musical melodies. They believe this advantage is related to superior declarative memory, which may underlie the storage and knowledge of both the mental lexicon in language (e.g., Ullman, 2001) and some aspects of familiar melodies in music (Miranda and Ullman, 2007). Two papers report finding that musical training during development enhances literacy skills, including phonological awareness and reading fluency, via neural mechanisms for both language and music (Degé et al.; Gordon et al.). Moreover, Degé and colleagues provide evidence that music production and music perception are associated with multiple precursors of reading. Finally, Lolli et al. examined the effect of sound frequency on judgments of emotion in speech by congenital amusics. Using both high and low-pass filtered speech in a pitch discrimination and emotion identification task, their findings demonstrate the important role of low frequency information in identifying the emotional content of speech.

In addition to these eight research papers there are two perspective and opinion papers that emphasize the affective

and emotive commonalities between music and language (Lehmann and Paquette; Omigie). Lehmann and Paquette provide a neurobehavioral approach for examining cross-domain processing of musical and vocal emotions, suggesting that studying cochlear implant users may allow for a richer understanding of neural overlap between music and language. Omigie (2015) provides evolutionary evidence for shared underlying neural mechanisms for our emotive responses to music and literature.

This E-Book provides a comprehensive snapshot of the research examining the complex overlap of neural systems for processing language and music. Both musical experience and training enhance the development of linguistic representations, emotion perception, and other cognitive skills. Furthermore, the research presented here contributes to current knowledge of neuroplastic reorganization and repair in clinical populations, and may aid in the design of new and more effective rehabilitative protocols.

AUTHOR CONTRIBUTIONS

MJ prepared and wrote the editorial soliciting feedback from EL and CM. EL and CM provided feedback regarding organization as well.

REFERENCES

- Bever, T. G., and Chiarello, R. J. (1974). Cerebral dominance in musicians and nonmusicians. *Science* 185, 537–539.
- Hickok, G., and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138. doi: 10.1016/S1364-6613(00)01463-7
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Miranda, R. A., and Ullman, M. T. (2007). Double dissociation between rules and memory in music: an event-related potential study. *Neuroimage* 38, 331–345. doi: 10.1016/j.neuroimage.2007.07.034
- Norman-Haignere, S., Kanwisher, N. G., and McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88, 1281–1296. doi: 10.1016/j.neuron.2015.11.035
- Peretz, I., Vuvan, D., Lagrois, M. É., and Armony, J. L. (2015). Neural overlap in processing music and speech. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140090. doi: 10.1098/rstb.2014.0090
- Rogalsky, C., Rong, F., Saberi, K., and Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J. Neurosci.* 31, 3843–3852. doi: 10.1523/JNEUROSCI.4515-10.2011
- 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
- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., and Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Hum. Brain Mapp.* 32, 771–783. doi: 10.1002/hbm.21060
- Tallal, P., and Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends Neurosci.* 29, 382–390. doi: 10.1016/j.tins.2006.06.003
- Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *J. Psycholinguist. Res.* 30, 37–69. doi: 10.1023/A:1005204207369
- 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
- Wong, P., Perrachione, T. K., and Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Hum. Brain Mapp.* 28, 995–1006. doi: 10.1002/hbm.20330

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 Jantzen, Large and Magne. 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 relationship between the neural computations for speech and music perception is context-dependent: an activation likelihood estimate study

Arianna N. LaCroix, Alvaro F. Diaz and Corianne Rogalsky*

Communication Neuroimaging and Neuroscience Laboratory, Department of Speech and Hearing Science, Arizona State University, Tempe, AZ, USA

OPEN ACCESS

Edited by:

McNeel Gordon Jantzen,
Western Washington University, USA

Reviewed by:

Lutz Jäncke,
University of Zurich, Switzerland

Yi Du,
McGill University, Canada

*Correspondence:

Corianne Rogalsky,
Department of Speech and Hearing
Science, Arizona State University, PO
Box 570102, Tempe,
AZ 85287-0102, USA
corianne.rogalsky@asu.edu

Specialty section:

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

Received: 08 April 2015

Accepted: 22 July 2015

Published: 11 August 2015

Citation:

LaCroix AN, Diaz AF and Rogalsky C
(2015) The relationship between the
neural computations for speech and
music perception is
context-dependent: an activation
likelihood estimate study.
Front. Psychol. 6:1138.
doi: 10.3389/fpsyg.2015.01138

The relationship between the neurobiology of speech and music has been investigated for more than a century. There remains no widespread agreement regarding how (or to what extent) music perception utilizes the neural circuitry that is engaged in speech processing, particularly at the cortical level. Prominent models such as Patel's Shared Syntactic Integration Resource Hypothesis (SSIRH) and Koelsch's neurocognitive model of music perception suggest a high degree of overlap, particularly in the frontal lobe, but also perhaps more distinct representations in the temporal lobe with hemispheric asymmetries. The present meta-analysis study used activation likelihood estimate analyses to identify the brain regions consistently activated for music as compared to speech across the functional neuroimaging (fMRI and PET) literature. Eighty music and 91 speech neuroimaging studies of healthy adult control subjects were analyzed. Peak activations reported in the music and speech studies were divided into four paradigm categories: passive listening, discrimination tasks, error/anomaly detection tasks and memory-related tasks. We then compared activation likelihood estimates within each category for music vs. speech, and each music condition with passive listening. We found that listening to music and to speech preferentially activate distinct temporo-parietal bilateral cortical networks. We also found music and speech to have shared resources in the left pars opercularis but speech-specific resources in the left pars triangularis. The extent to which music recruited speech-activated frontal resources was modulated by task. While there are certainly limitations to meta-analysis techniques particularly regarding sensitivity, this work suggests that the extent of shared resources between speech and music may be task-dependent and highlights the need to consider how task effects may be affecting conclusions regarding the neurobiology of speech and music.

Keywords: music perception, speech perception, fMRI, meta-analysis, Broca's area

Introduction

The relationship between the neurobiology of speech and music has been investigated and debated for nearly a century. (Henschen, 1924; Luria et al., 1965; Frances et al., 1973; Peretz, 2006; Besson et al., 2011). Early evidence from case studies of brain-damaged individuals suggested a dissociation of aphasia and amusia (Yamadori et al., 1977; Basso and Capitani, 1985;

Peretz et al., 1994, 1997; Steinke et al., 1997; Patel et al., 1998b; Tzortzis et al., 2000; Peretz and Hyde, 2003). However, more recent patient work examining specific aspects of speech and music processing indicate at least some overlap in deficits across the two domains. For example, patients with Broca's aphasia have both linguistic and harmonic structure deficits, and patients with amusia exhibit pitch deficits in both speech and music (Patel, 2003, 2005, 2013). Electrophysiological (e.g., ERP) studies also suggest shared resources between speech and music; for example, syntactic and harmonic violations elicit indistinguishable ERP responses such as the P600 response, which is hypothesized to originate from anterior temporal or inferior frontal regions (Patel et al., 1998a; Maillard et al., 2011; Sammler et al., 2011). Music perception also interacts with morphosyntactic representations of speech: the early right anterior negativity (ERAN) ERP component sensitive to chord irregularities interacts with the left anterior negativity's (LAN's) response to morphosyntactic violations or irregularities (Koelsch et al., 2005; Steinbeis and Koelsch, 2008b; Koelsch, 2011).

Several studies of trained musicians and individuals with absolute pitch also suggest an overlap between speech and music as there are carry-over effects of musical training onto speech processing performance (e.g., Oechslin et al., 2010; Elmer et al., 2012; for a review see Besson et al., 2011).

There is a rich literature of electrophysiological and behavioral work regarding the relationship between music and language (for reviews see Besson et al., 2011; Koelsch, 2011; Patel, 2012, 2013; Tillmann, 2012; Slevc and Okada, 2015). This work has provided numerous pieces of evidence of overlap between the neural resources of speech and music, including in the brainstem, auditory cortex and frontal cortical regions (Koelsch, 2011). This high degree of interaction between speech and music coincides with Koelsch et al.'s view that speech and music, and therefore the brain networks supporting them, cannot be separated because of their numerous shared properties, i.e., there is a "music-speech continuum" (Koelsch and Friederici, 2003; Koelsch and Siebel, 2005; Koelsch, 2011). However, evidence from brain-damaged patients suggests that music and speech abilities may dissociate, although there are also reports to the contrary (see above). Patel's (2003, 2008, 2012) Shared Syntactic Integration Resource Hypothesis (SSIRH) is in many ways a remedy to the shared-vs.-distinct debate in the realm of structural/syntactic processing. Stemming in part from the patient and electrophysiological findings, Patel proposes that language and music utilize overlapping cognitive resources but also have unique neural representations. Patel proposes that the shared resources reside in the inferior frontal lobe (i.e., Broca's area) and that distinct processes for speech and music reside in the temporal lobes (Patel, 2003).

The emergence of functional neuroimaging techniques such as fMRI have continued to fuel the debate over the contributions of shared vs. distinct neural resources for speech and music. fMRI lacks the high temporal resolution of electrophysiological methods and can introduce high levels of ambient noise potentially contaminating recorded responses to auditory stimuli. However, the greater spatial resolution of fMRI may provide additional information regarding the neural correlates

of speech and music, and MRI scanner noise can be minimized using sparse sampling scanning protocols and reduced-noise continuous scanning techniques (Pelle et al., 2010). Hundreds of fMRI papers have investigated musical processes, and thousands have investigated the neural substrates of speech. Conversely, to our knowledge and as Slevc and Okada (2015) noted, only a few studies have directly compared activations to hierarchical speech and music (i.e., sentences and melodies) using fMRI (Abrams et al., 2011; Fedorenko et al., 2011; Rogalsky et al., 2011). Findings from these studies conflict with the ERP literature (e.g., Koelsch, 2005; Koelsch et al., 2005) in that the fMRI studies identify distinct neuroanatomy and/or activation response patterns for music and speech processing, although there are notable differences across these studies, particularly relating to the involvement of Broca's area in speech and music.

The differences found across neuroimaging studies regarding the overlap of the neural correlates of speech and music likely arise from the tasks used in each of these studies. For example, Rogalsky et al. used passive listening and found no activation of Broca's area to either speech or music compared to rest. Conversely, Fedorenko et al. used a reading/memory probe task for sentences and an emotional ranking for music and found Broca's area to be preferentially activated by speech but also activated by music compared to rest. There is also evidence that the P600, the ERP component that is sensitive to both speech and music violations, is only present when subjects are actively attending to the stimulus (Besson and Faita, 1995; Brattico et al., 2006; Koelsch, 2011). The inclusion of a task may affect not only the brain regions involved, but also reliability of results: an fMRI study of visual tasks reported that tasks with high attentional loads also had the highest reliability measures compared to passive conditions (Specht et al., 2003). This finding in the visual domain suggests the possibility that greater (within and between) subject variability in passive listening conditions may lead to null effects in group-averaged results.

Given the scarcity of within-subject neuroimaging studies of speech and music, it is particularly critical to examine across-study, between-subjects findings to build a better picture regarding the neurobiology of speech and music. A major barrier in interpreting between-subject neuroimaging results is the variety of paradigms and tasks used to investigate speech and music neural resources. Most scientists studying the neurobiology of speech and/or music would likely agree that they are interested in understanding the neural computations employed in naturalistic situations that are driven by the input of speech or music, and the differences between the two. However, explicit tasks such as discrimination or error detection are often used to drive brain responses in part by increasing the subject's attention to the stimuli and/or particular aspects of the stimuli. This may be problematic: the influence of task demands on the functional neuroanatomy recruited by speech is well documented (e.g., Baker et al., 1981; Noesselt et al., 2003; Scheich et al., 2007; Geiser et al., 2008; Rogalsky and Hickok, 2009) and both speech and music processing engage domain-general cognitive, memory, and motor networks in likely distinct, but overlapping ways (Besson et al., 2011). Task effects are known to alter inter and intra hemisphere activations to speech

(Noesselt et al., 2003; Tervaniemi and Hugdahl, 2003; Scheich et al., 2007; Geiser et al., 2008; Rogalsky and Hickok, 2009). For example, there is evidence that right hemisphere fronto-temporal-parietal networks are significantly activated during an explicit task (rhythm judgment) with speech stimuli but not during passive listening to the same stimuli (Geiser et al., 2008). The neurobiology of speech perception, and auditory processing more generally, also can vary based on the type of explicit task even when the same stimuli are used across tasks (Platel et al., 1997; Ni et al., 2000; Von Kriegstein et al., 2003; Geiser et al., 2008; Rogalsky and Hickok, 2009). This phenomenon is also well documented in the visual domain (Corbetta et al., 1990; Chawla et al., 1999; Cant and Goodale, 2007). For example, in the speech domain, syllable discrimination and single-word comprehension performance (as measured by a word-picture matching task) doubly dissociate in stroke patients with aphasia (Baker et al., 1981). Syllable discrimination implicates left-lateralized dorsal frontal-parietal networks, while speech comprehension and passive listening tasks engage mostly mid and posterior temporal regions (Dronkers et al., 2004; Schwartz et al., 2012; Rogalsky et al., 2015). Similarly, contextual effects have been reported regarding pitch: when pitch is needed for linguistic processing, such as in a tonal language, there is a left hemisphere auditory cortex bias, while pitch processing in a melody discrimination task yields a right hemisphere bias (Zatorre and Gandour, 2008). Another example of the importance of context in pitch processing is in vowel perception: vowels and tones have similar acoustic features and when presented in isolation (i.e., just a vowel, not in a consonant-vowel (CV) pair as would typically be perceived in everyday life) no significant differences have been found in temporal lobe activations (Jäncke et al., 2002). However, there is greater superior temporal activation for CVs than tones suggesting that the context of the vowel modulates the temporal networks activated (Jäncke et al., 2002).

One way to reduce the influence of a particular paradigm or task is to use meta-analysis techniques to identify areas of activation that consistently activate to a particular stimulus (e.g., speech, music) across a range of tasks and paradigms. Besson and Schön (2001) noted that meta-analyses of neuroimaging data would provide critical insight into the relationship between the neurobiology of language and music. They also suggested that meta-analyses of music-related neuroimaging data were not feasible due to the sparse number of relevant studies. Now, almost 15 years later, there is a large enough corpus of neuroimaging work to conduct quantitative meta-analyses of music processing with sufficient power. In fact, such meta-analyses have begun to emerge, for specific aspects of musical processing, in relation to specific cognitive functions [e.g., Slevc and Okada's (2015) cognitive control meta-analysis in relation to pitch and harmonic ambiguity], in addition to extensive qualitative reviews (e.g., Tervaniemi, 2001; Jäncke, 2008; Besson et al., 2011; Grahn, 2012; Slevc, 2012; Tillmann, 2012).

The present meta-analysis addresses the following outstanding questions: (1) has functional neuroimaging identified significant distinctions between the functional neuroanatomy of speech and music and (2) how do specific types of tasks affect how music recruits speech-processing

networks? We then discuss the implications of our findings for future investigations of the neural computations of language and music.

Materials and Methods

An exhaustive literature search was conducted via Google Scholar to locate published fMRI and PET studies reporting activations to musical stimuli. The following search terms were used to locate papers about music: "fMRI music," "fMRI and music," "fMRI pitch," and "fMRI rhythm." To the best of our knowledge, all relevant journal research articles have been collected for the purposes of this meta-analysis.

All journal articles that became part of the meta-analysis reported peak coordinates for relevant contrasts. Peak coordinates reported in the papers identified by the searches were divided into four categories that encompassed the vast majority of paradigms used in the articles: music passive listening, music discrimination, music error detection, and music memory¹. Passive listening studies included papers in which participants listened to instrumental melodies or tone sequences with no explicit task as well as studies that asked participants to press a button when the stimulus concluded. Music discrimination studies included those that asked participants to compare two musical stimuli (e.g., related/unrelated, same/different). Music error detection studies included studies that instructed participants to identify a dissonant melody, unexpected note or deviant instrument. The music memory category included papers that asked participants to complete an n-back task, familiarity judgment, or rehearsal (covert or overt) of a melodic stimulus.

Only coordinates from healthy adult, non-musician, control subjects were included. In studies that included a patient group and a control group, only the control group's coordinates were included. Studies were excluded from the final activation likelihood estimate (ALE) if the data did not meet the requirements for being included in ALE calculations, including for the following reasons: coordinates not reported, only approximate anatomical location reported, stereotaxic space not reported, inappropriate contrasts (e.g., speech > music only), activations corresponding to participant's emotional reactions to music, studies of professional/trained musicians, and studies of children.

In addition to collecting the music-related coordinates via an exhaustive search, we also gathered a representative sample of fMRI and PET studies that reported coordinates for passive listening to intelligible speech compared to some type of non-speech control (e.g., tones, noise, rest, visual stimuli).

¹The music categories included studies with stimuli of the following types: instrumental unfamiliar and familiar melodies, tone sequences and individual tones. In comparison, the speech categories described below included studies with stimuli such as individual phonemes, vowels, syllables, words, pseudowords, sentences, and pseudoword sentences. For the purposes of the present study, we have generated two distinct groups of stimuli to compare. However, music and speech are often conceptualized as being two ends of continuum with substantial gray area between the two extremes (Koelsch, 2011). For example, naturally spoken sentences contain rhythmic and pitch-related prosodic features and a familiar melody likely automatically elicits a mental representation of the song's lyrics.

Coordinates corresponding to the following tasks were also extracted: speech discrimination, speech detection, and speech memory. The purpose of these speech conditions is to act as comparison groups for the music groups. Coordinates for this purpose were extracted from six sources: five well-cited review papers, Price (2010), Zheng et al. (2010), Turkeltaub and Coslett (2010), Rogalsky et al. (2011), and Adank (2012) and the brain imaging meta-analysis database Neurosynth.org. The Price (2010), Zheng et al. (2010), Turkeltaub and Coslett (2010), Rogalsky et al. (2011), and Adank (2012) papers yielded a total of 42 studies that fit the aforementioned criteria. An additional 49 relevant papers were found using the Neurosynth.org database with the search criteria “speech perception,” “speech processing,” “speech,” and “auditory working memory.” These methods resulted in 91 studies in which control subjects passively listened to speech or completed an auditory verbal memory, speech discrimination, or speech detection task. The passive listening speech condition included studies in which participants listened to speech stimuli with no explicit task as well as studies that asked participants to press a button when the stimulus concluded. Papers were included in the speech discrimination category if they asked participants to compare two speech stimuli (e.g., a same/different task). The speech detection category contained papers that asked participants to detect semantic, intelligibility, or grammatical properties or detect phonological, semantic, or syntactic errors. Studies included in the speech memory category were papers that instructed participants to complete an n-back task or rehearsal (covert or overt) of a speech (auditory verbal) stimulus.

Analyses were conducted using the meta-analysis software GingerALE to calculate ALEs for each condition based on the coordinates collected (Eickhoff et al., 2009, 2012; Turkeltaub et al., 2012). All results are reported in Talairach space. Coordinates originally reported in MNI space were transformed to Talairach space using GingerALE’s stereotaxic coordinate converter. Once all coordinates were in Talairach space, each condition was analyzed individually using the following GingerALE parameters: less conservative (larger) mask size, Turkeltaub nonadditive ALE method (Turkeltaub et al., 2012), subject-based FWHM (Eickhoff et al., 2009), corrected threshold of $p < 0.05$ using false discovery rate (FDR), and a minimum cluster volume of 200 mm^3 . We obtained subtraction contrasts between two given conditions by directly comparing activations between two conditions. To correct for multiple comparisons, each contrast’s threshold was set to $p < 0.05$, whole-brain corrected following the FDR algorithm with p value permutations set at 10,000, and a minimum cluster size of 200 mm^3 (Eickhoff et al., 2009). ALE statistical maps were rendered onto the Colin Talairach template brain using the software MRIcron (Rorden and Brett, 2000).

Results

Search Results

The literature search yielded 80 music studies (76 fMRI studies, 4 PET studies) and 91 relevant speech papers (88 fMRI, 3 PET studies) meeting the inclusion criteria described above. **Table 1**

TABLE 1 | Activations included in the present meta-analysis.

Condition	Number of studies	Number of subjects	Number of coordinates
Music passive listening	41	540	526
Music discrimination	12	211	168
Music error detection	25	355	489
Music memory	14	190	207
Speech passive listening	31	454	337
Speech discrimination	31	405	318
Speech detection	17	317	248
Speech memory	19	259	324

indicates the number of studies, subjects, and coordinates in each of the four music conditions, as well as for each of the four speech conditions.

Passive Listening To Music vs. Passive Listening To Speech

The music passive listening ALE identified large swaths of voxels bilaterally, spanning the length of the superior temporal gyri (STG), as well as additional smaller clusters, including in the bilateral inferior frontal gyrus (pars opercularis), bilateral postcentral gyrus, bilateral insula, left inferior parietal lobule, left medial frontal gyrus, right precentral gyrus, and right middle frontal gyrus (**Figure 1A, Table 2**). The speech passive listening ALE also identified bilateral superior temporal regions as well as bilateral precentral and inferior frontal (pars opercularis) regions. Notably, the speech ALE identified bilateral anterior STG, bilateral superior temporal sulcus (i.e., both banks, the middle and superior temporal gyri) and left inferior frontal gyrus (pars triangularis) regions not identified by the music ALE (**Figure 1A, Table 2**). ALEs used a threshold of $p < 0.05$, FDR corrected.

Pairwise contrasts of passive listening to music vs. passive listening to speech were calculated to identify any brain regions that were significantly activated more by speech or music, respectively. Results were as follows ($p < 0.05$, FDR corrected): the speech > music contrast identified significant regions on both banks of the bilateral superior temporal sulcus extending the length of the left temporal lobe and mid/anterior right temporal lobe, left inferior frontal lobe (pars triangularis), left precentral gyrus, and left postcentral gyrus regions. Music > speech identified bilateral insula and bilateral superior temporal/parietal operculum clusters as well as a right inferior frontal gyrus region (**Figure 1B, Table 2**). These results coincide with previous reports of listening to speech activating a lateral temporal network particularly in the superior temporal sulcus and extending into the anterior temporal lobe, while listening to music activated a more dorsal medial temporal-parietal network (Jäncke et al., 2002; Rogalsky et al., 2011). These results also coincide with Fedorenko et al.’s (2011) finding that Broca’s area, the pars triangularis in particular, is preferentially responsive to language stimuli.

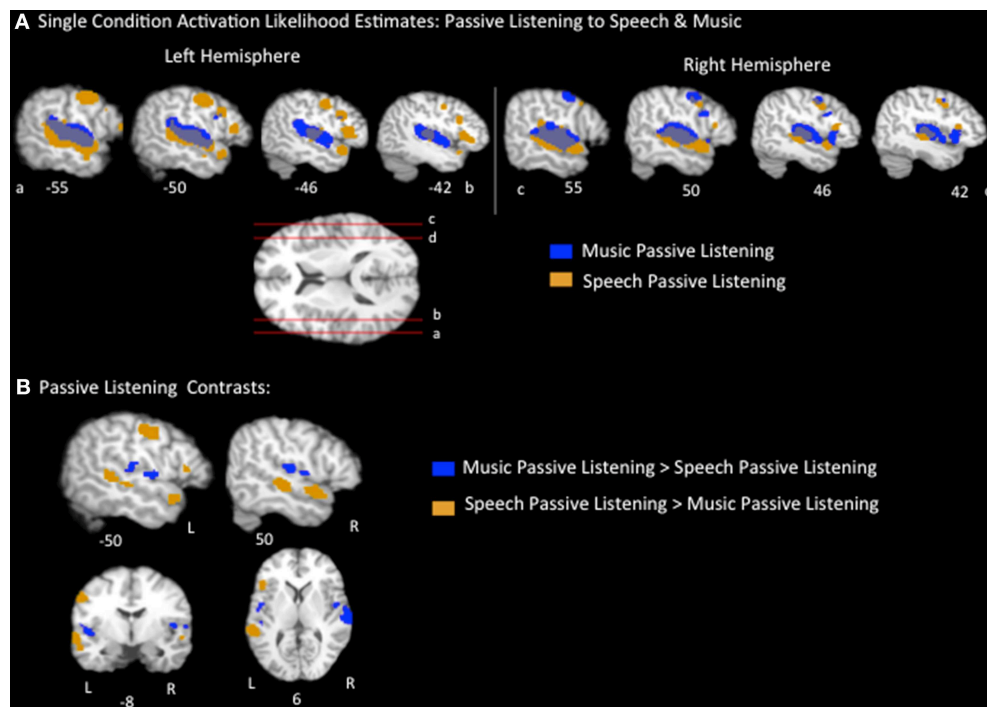


FIGURE 1 | (A) Representative sagittal slices of the ALE for passive listening to speech, $p < 0.05$, corrected, overlaid on top of the passive music listening ALE. **(B)** Speech vs. music passive listening contrasts results, $p < 0.05$ corrected.

Music Tasks vs. Speech Tasks

The passive listening ALE results identify distinct and overlapping regions of speech and music processing. We now turn to the question of how do these distinctions change as a function of the type of task employed? First, ALEs were computed for each music task condition, $p < 0.05$ FDR corrected (**Figure 1**, **Table 2**). The music task conditions' ALEs all significantly identified bilateral STG and bilateral precentral gyrus, and inferior parietal regions, overlapping with the passive listening music ALE (**Figure 2**). The tasks also activated additional inferior frontal and inferior parietal regions not identified by the passive listening music ALE; these differences are discussed in a subsequent section.

To compare the brain regions activated by each music task to those activated by speech in similar tasks, pairwise contrasts of the ALEs for each music task vs. its corresponding speech task group were calculated (**Figure 3**, **Table 2**). Music discrimination > speech discrimination identified regions including bilateral inferior frontal gyri (pars opercularis), bilateral pre and postcentral gyri, bilateral medial frontal gyri, left inferior parietal lobule, and left cerebellum, whereas speech discrimination > music discrimination identified bilateral regions in the anterior superior temporal sulci (including both superior and middle temporal gyri). Music detection > speech detection identified a bilateral group of clusters spanning the superior temporal gyri, bilateral precentral gyri, bilateral insula and bilateral inferior parietal regions, as well as clusters in the right middle frontal gyrus. Speech detection > music detection

identified bilateral superior temporal sulci regions as well as left inferior frontal regions (pars triangularis and pars opercularis). Music memory > speech memory identified a left posterior superior temporal/inferior parietal region and bilateral medial frontal regions; speech memory > music memory identified left inferior frontal gyrus (pars opercularis and pars triangularis) and bilateral superior and middle temporal gyri.

In sum, the task pairwise contrasts in many ways mirror the passive listening contrast: music tasks activated more dorsal/medial superior temporal and inferior parietal regions, while speech tasks activated superior temporal sulcus regions, particularly in the anterior temporal lobe. In addition, notable differences were found in Broca's area and its right hemisphere homolog: in discrimination tasks music significantly activated Broca's area (specifically the pars opercularis) more than speech. However, in detection and memory tasks speech activated Broca's area (pars opercularis and pars triangularis) more than music. The right inferior frontal gyrus responded equally to speech and music in both detection and memory tasks, but responded more to music than speech in discrimination tasks. Also notably, in the memory tasks, music activated a lateral superior temporal/inferior parietal cluster (in the vicinity of Hickok and Poeppel's "area Spt") more than speech while an inferior frontal cluster including the pars opercularis was activated more for speech than music. Both area Spt and the pars opercularis previously have been implicated in a variety of auditory working memory tasks (including speech and pitch working memory) in both lesion patients and control subjects (Koelsch and Siebel,

TABLE 2 | Locations, peaks and cluster size for significant voxel clusters for each condition's ALE and for each contrast of interest.

Condition	Anatomical locations	Peak coordinates	Voxels
Music passive listening	Left inferior frontal gyrus (pars opercularis)*	−46, 10, 26	32
	Left medial frontal gyrus*, left subcallosal gyrus	−2, 26, −14	65
	Left medial frontal gyrus*	−2, 2, 62	48
	Left postcentral gyrus*, left inferior parietal lobule	−34, −36, 54	27
	Left superior temporal gyrus*, left transverse temporal gyrus, left middle temporal gyrus, left insula	−52, −20, 6	2073
	Right inferior frontal gyrus*	48, 10, 28	43
	Right precentral gyrus*, right postcentral gyrus, right middle frontal gyrus	52, −2, 44	173
	Right superior temporal gyrus*, right transverse temporal gyrus, right middle temporal gyrus, right insula	58, −20, 6	2154
	Right insula*, right inferior frontal gyrus, right precentral gyrus	42, 14, 0	206
	Right lingual gyrus*, right culmen	16, −54, −2	27
Music discrimination	Left medial frontal gyrus*, left middle frontal gyrus	−8, −4, 58	224
	Left precentral gyrus*, left postcentral gyrus, left inferior parietal lobule	−48, −12, 48	259
	Left precentral gyrus*, left inferior frontal gyrus (pars opercularis)	−50, 2, 26	67
	Left superior temporal gyrus*, left transverse temporal gyrus, left precentral gyrus	−54, −16, 8	239
	Left superior temporal gyrus*, left middle temporal gyrus	−58, −34, 8	92
	Left insula*, left inferior frontal gyrus (pars triangularis)	−34, 22, 2	48
	Left cerebellum*	−28, −62, −24	127
	Right inferior frontal gyrus*, right middle frontal gyrus	52, 12, 28	58
	Right precentral gyrus*, right middle frontal gyrus	46, −6, 44	170
	Right superior temporal gyrus*, right middle temporal gyrus	62, −24, 8	310
Music error detection	Right superior temporal gyrus*, right precentral gyrus, right insula	50, 6, −2	91
	Left medial frontal gyrus*	−4, −4, 58	49
	Left superior temporal gyrus*, Left transverse temporal gyrus, Left postcentral gyrus, left insula	−50, −18, 8	1448
	Left inferior parietal lobule*, left supramarginal gyrus, left angular gyrus	−40, −48, 40	41
	Left lentiform nucleus*, left putamen	−22, 6, 10	263
	Right middle frontal gyrus*	36, 42, 18	43
	Right middle frontal gyrus*, right precentral gyrus	32, −4, 56	35
	Right superior frontal gyrus*, right medial frontal gyrus, left superior frontal gyrus, left medial frontal gyrus	2, 10, 52	95
	Right superior temporal gyrus*, right transverse temporal gyrus, right insula, right precentral gyrus, right middle temporal gyrus, right claustrum	50, −18, 6	1228
	Right parahippocampal gyrus*	22, −14, −12	36
	Right inferior parietal lobule*, right supramarginal gyrus	36, −44, 40	103
	Right insula*, right inferior frontal gyrus	32, 22, 12	329
	Right lentiform nucleus*, right putamen, right caudate	18, 6, 12	144
	Right thalamus*	12, −16, 8	33
	Right cerebellum*	26, −50, −26	28
Music memory	Left inferior frontal gyrus (pars opercularis)*, left precentral gyrus, left middle frontal gyrus	−50, 4, 26	206
	Left inferior frontal gyrus (pars triangularis*, pars orbitalis), left insula	−34, 24, −2	57
	Left inferior frontal gyrus (pars triangularis)*	−44, 26, 10	25
	Left medial frontal gyrus*	−4, 52, 12	31
	Left middle frontal gyrus*	−32, 4, 54	29
	Left precentral gyrus*	−44, −10, 42	33
	Left superior frontal gyrus*, left medial frontal gyrus, right superior frontal gyrus, right medial frontal gyrus	−0, 12, 50	373
	Left middle temporal gyrus*	−50, −20, −10	72
	Left middle temporal gyrus*, left superior temporal gyrus	−46, 4, −18	35

(Continued)

TABLE 2 | Continued

Condition	Anatomical locations	Peak coordinates	Voxels
	Left inferior parietal lobule*, left superior temporal gyrus, left middle temporal gyrus, left supramarginal gyrus	−48, −44, 22	224
	Left thalamus*	−14, −14, 14	37
	Right inferior frontal gyrus*, right insula, right claustrum	32, 26, 8	90
	Right middle frontal gyrus*	38, 44, 14	27
	Right superior temporal gyrus*, right middle temporal gyrus	54, −38, 10	35
	Right parahippocampal gyrus*, right hippocampus	30, −10, −20	35
	Right cerebellum*	30, −56, −18	47
Speech passive listening	Left inferior frontal gyrus (pars triangularis, pars opercularis)*, left insula, left precentral gyrus	−44, 20, 8	296
	Left inferior frontal gyrus (pars triangularis*, pars opercularis), left middle frontal gyrus, left precentral gyrus	−48, 10, 28	162
	Left precentral gyrus*, left postcentral gyrus	−52, −10, 40	294
	Left medial frontal gyrus*, left superior frontal gyrus, left medial frontal gyrus, left cingulate gyrus	−8, 8, 50	164
	Left superior temporal gyrus*, Left middle temporal gyrus, left postcentral gyrus, left transverse temporal gyrus	−58, −14, −2	2101
	Left superior temporal gyrus*	−46, 12, −14	107
	Left fusiform gyrus*, left inferior occipital gyrus, left middle occipital gyrus	−38, −78, −12	35
	Right inferior frontal gyrus*, right insula, right precentral gyrus	44, 18, 10	81
	Right middle frontal gyrus*, right precentral gyrus, right inferior frontal gyrus	46, 2, 38	118
	Right superior temporal gyrus*, right middle temporal gyrus, Right insula, right precentral gyrus, right transverse temporal gyrus	−52, 20, 0	1800
Speech discrimination	Left inferior frontal gyrus (pars orbitalis*, pars triangularis), left insula, left middle frontal gyrus	−38, 26, −4	115
	Left inferior frontal gyrus (pars triangularis*, pars opercularis), left precentral gyrus	−44, 20, 10	44
	Left middle frontal gyrus*, left inferior frontal gyrus (pars triangularis, pars opercularis)	−46, 16, 30	187
	Left middle frontal gyrus*, left precentral gyrus	−46, −0, 42	26
	Left superior temporal gyrus*, left postcentral gyrus, left transverse temporal gyrus, left middle temporal gyrus	−58, −20, 4	1737
	Left thalamus*, left caudate	−14, −16, 10	147
	Left cerebellum	−38, −60, −16	36
	Right inferior frontal gyrus*, right precentral gyrus, right insula	46, 20, 4	38
	Right superior temporal gyrus*, right middle temporal gyrus, right transverse temporal gyrus, right insula	58, −14, 0	1223
	Right precuneus*, right cuneus	4, −78, 38	34
Speech detection	Left inferior frontal gyrus (pars opercularis)*, left middle frontal gyrus, left insula	−48, 10, 22	361
	Left inferior frontal gyrus (pars triangularis)*	−48, 28, 12	101
	Left inferior frontal gyrus (pars triangularis*, pars orbitalis), left insula	−34, 24, 2	61
	Left postcentral gyrus*, left precentral gyrus	−50, −12, 46	92
	Left medial frontal gyrus*, left superior frontal gyrus	−6, −6, 60	54
	Left superior temporal gyrus*, left middle temporal gyrus, left transverse temporal gyrus	−60, −22, −2	1010
	Left superior temporal gyrus*, left supramarginal gyrus, left inferior parietal lobule	−60, −42, 20	66
	Left superior temporal gyrus*, left middle temporal gyrus	−50, 12, −14	34
	Left superior temporal gyrus*	−42, 18, −24	28
	Left transverse temporal gyrus*, left superior temporal gyrus	−36, −30, 12	38
	Left precuneus*, left superior parietal lobule, left inferior parietal lobule	−30, −62, 40	66
	Right inferior frontal gyrus*, right insula	34, 24, 6	62
	Right inferior frontal gyrus*	40, 24, −4	31
	Right inferior frontal gyrus*	52, 8, 22	29
	Right superior temporal gyrus*, right transverse temporal gyrus, right middle temporal gyrus	58, −14, 4	788
	Right middle frontal gyrus*	48, 14, 32	36

(Continued)

TABLE 2 | Continued

Condition	Anatomical locations	Peak coordinates	Voxels
Speech memory	Left middle frontal gyrus*, left inferior frontal gyrus (pars triangularis, pars opercularis), left precentral gyrus	−50, 22, 22	476
	Left superior frontal gyrus*, left medial frontal gyrus, right medial frontal gyrus, right superior frontal gyrus	−2, 4, 56	73
	Left precentral gyrus*, left postcentral gyrus	−50, −10, 44	127
	Left insula*, left inferior frontal gyrus (pars triangularis), left claustrum	−30, 18, 4	39
	Left superior temporal gyrus*, left middle temporal gyrus, left insula	−62, −24, 6	937
	Left superior temporal gyrus*, left middle temporal gyrus	−50, 10, −10	62
	Left superior parietal lobule*, left precuneus, left inferior parietal lobule	−30, −62, 48	109
	Left inferior parietal lobule*	−40, −46, 50	93
	Left caudate*, left thalamus	−16, −2, 16	36
	Left cerebellum*, left fusiform gyrus	−40, −44, −20	67
	Right superior temporal gyrus*, right middle temporal gyrus, right transverse temporal gyrus	58, −14, 0	773
	Right superior temporal gyrus*, right middle temporal gyrus	48, 8, −14	58
	Right cerebellum	24, −64, −16	50
Music passive > speech passive	Left insula*, left superior temporal gyrus	−44, −6, 2	148
	Left superior temporal gyrus*, left insula, left middle temporal gyrus	−42, −40, 14	146
	Left subcallosal gyrus*, left medial frontal gyrus, left anterior cingulate	−4, 22, −14	53
	Right inferior frontal gyrus*, right insula	44, 18, −2	49
	Right superior temporal gyrus*, right postcentral gyrus, right transverse temporal gyrus, right precentral gyrus, right insula	66, −20, 10	457
Music passive < speech passive	Left inferior frontal gyrus (pars triangularis)*, left precentral gyrus	−42, 30, 2	177
	Left precentral gyrus*, left postcentral gyrus	−56, −10, 40	191
	Left middle temporal gyrus*, left inferior temporal gyrus, left superior temporal gyrus	−56, −12, −12	856
	Left middle temporal gyrus*, left superior temporal gyrus	−50, 6, −18	91
	Left cingulate gyrus*, left medial frontal gyrus, left superior frontal gyrus	−10, 4, 46	70
	Right middle temporal gyrus*, right superior temporal gyrus, right insula	56, −22, −8	277
	Right middle temporal gyrus*, right superior temporal gyrus	52, 2, −12	167
Music discrimination > speech discrimination	Left inferior frontal gyrus (pars opercularis)*, left precentral gyrus	−52, 4, 24	56
	Left postcentral gyrus*, left inferior parietal lobule, left precentral gyrus	−48, −18, 44	253
	Left medial frontal gyrus*, left superior frontal gyrus, right medial frontal gyrus, right superior frontal gyrus	−8, −6, 54	224
	Left superior temporal gyrus*, left transverse temporal gyrus, left precentral gyrus	−52, −10, 8	122
	Left cerebellum	−28, −64, −28	114
	Right inferior frontal gyrus*, right middle frontal gyrus	50, 8, 26	53
	Right precentral gyrus*, right middle frontal gyrus	36, −6, 42	170
	Right precentral gyrus*, right insula, right superior temporal gyrus	48, 4, 8	91
	Right superior temporal gyrus*, right transverse temporal gyrus	66, −26, 10	93
Music discrimination < speech discrimination	Left middle temporal gyrus*, left superior temporal gyrus	−62, −18, −8	456
Music detection > speech detection	Right middle temporal gyrus*, right superior temporal gyrus	66, −8, −4	38
	Left insula*, left superior temporal gyrus, left precentral gyrus	−40, −16, 8	126
	Left insula*, left superior temporal gyrus	−42, 4, −6	76
	Left superior temporal gyrus*, left transverse temporal gyrus	−48, −34, 16	131
	Right insula*, right transverse temporal gyrus, right superior temporal gyrus	44, −10, −4	507
	Right middle frontal gyrus*, right insula	38, 16, 24	78
	Right middle frontal gyrus*, right precentral gyrus	32, −4, 54	35
Music detection < speech detection	Left inferior frontal gyrus (par opercularis)*	−56, 16, 20	240

(Continued)

TABLE 2 | Continued

Condition	Anatomical locations	Peak coordinates	Voxels
	Left inferior frontal gyrus (pars triangularis)*	−52, 28, 12	101
	Left middle temporal gyrus*, left superior temporal gyrus, left transverse temporal gyrus	−60, −32, −2	561
	Left superior temporal gyrus*	−44, 18, −24	28
	Right middle temporal gyrus*, right superior temporal gyrus	62, −12, −4	361
Music memory > speech memory	Left cingulate gyrus*, left superior frontal gyrus, left medial frontal gyrus, right cingulate gyrus	−6, 20, 32	161
	Left superior temporal gyrus*, left supramarginal gyrus, left inferior parietal lobule	−46, −48, 14	45
Music memory < speech memory	Left inferior frontal gyrus (pars triangularis*, pars opercularis)	−54, 24, 20	80
	Left superior temporal gyrus*, left middle temporal gyrus	−60, −16, 6	606
	Right superior temporal gyrus*, right middle temporal gyrus, right transverse temporal gyrus	52, −26, 2	506
Music passive listening > music discrimination	Left insula*, left superior temporal gyrus	−42, −12, −4	116
	Left superior temporal gyrus*, left insula	−42, −42, 12	261
	Right superior temporal gyrus*, right insula	52, −12, 4	157
Music passive listening < music discrimination	Left medial frontal gyrus*, right medial frontal gyrus	−8, −6, 54	165
	Left precentral gyrus*, left superior temporal gyrus	−52, 2, 8	80
	Left postcentral gyrus*, left inferior parietal lobule, left precentral gyrus	−46, −18, 46	228
	Left cerebellum*	−24, −62, −24	90
	Right precentral gyrus*, right middle frontal gyrus	44, −6, 42	105
	Right precentral gyrus*, right insula, right superior temporal gyrus	50, 6, 6	30
Music passive > music error detection	Left middle temporal gyrus*	−58, −32, 0	82
	Left superior temporal gyrus*	−58, −10, 4	81
	Right precentral gyrus*, right middle frontal gyrus	50, 2, 48	64
	Right postcentral gyrus*, right superior temporal gyrus	62, −24, 16	44
	Right superior temporal gyrus*, right transverse temporal gyrus, right middle temporal gyrus, right precentral gyrus	60, −16, 0	336
Music passive < music error detection	Left medial frontal gyrus*	−4, −8, 56	30
	Left superior frontal gyrus*, left medial frontal gyrus, right superior frontal gyrus	−0, 8, 48	93
	Left postcentral gyrus*, left transverse temporal gyrus, left precentral gyrus	−52, −22, 16	79
	Left inferior parietal lobule*, left supramarginal gyrus	−40, −48, 38	37
	Left superior temporal gyrus*, left precentral gyrus	−52, 2, 4	92
	Left insula*, left superior temporal gyrus, left transverse temporal gyrus	−40, −28, 14	67
	Left lentiform nucleus*, left caudate	−18, 10, 8	211
	Right inferior parietal lobule*, right supramarginal gyrus	36, −50, 42	101
	Right insula*, right inferior frontal gyrus, right middle frontal gyrus	38, 18, 16	227
	Right insula*, right superior temporal gyrus	40, −20, 16	139
	Right insula*, right superior temporal gyrus	42, −8, 0	42
	Right caudate*, right lentiform nucleus	14, 6, 14	143
	Right thalamus*	14, −18, 6	32
	Right cerebellum*	28, −54, −26	28
Music passive listening > music memory	Left superior temporal gyrus*, left middle temporal gyrus, left insula	−54, −22, 6	943
	Right superior temporal gyrus*, right insula, right postcentral gyrus, right precentral gyrus right transverse temporal gyrus, right middle temporal gyrus	52, −20, 4	1350
	Right insula*, right inferior frontal gyrus, right superior temporal gyrus	46, 10, 2	32
Music passive listening < music memory	Left inferior frontal gyrus (pars opercularis)*, left precentral gyrus	−44, 4, 30	79
	Left inferior frontal gyrus (pars orbitalis)*, left insula	−32, 24, −6	53
	Left middle frontal gyrus*, left inferior frontal gyrus (pars triangularis)	−42, 18, 28	29
	Left middle frontal gyrus*	−32, 6, 54	29

(Continued)

TABLE 2 | Continued

Condition	Anatomical locations	Peak coordinates	Voxels
	Left superior frontal gyrus*, left medial frontal gyrus, right medial frontal gyrus, right superior frontal gyrus	−0, 8, 48	329
	Left superior temporal gyrus*, left middle temporal gyrus	−44, −20, −10	69
	Left inferior parietal lobule*, left supramarginal gyrus, left superior temporal gyrus	−54, −44, 28	89
	Left thalamus*	−10, −16, 14	37
	Right inferior frontal gyrus*, right insula, right claustrum	32, 26, 4	83
	Right parahippocampal gyrus*, right hippocampus	32, −12, −24	35
Speech passive listening > music discrimination	Left middle temporal gyrus*, left superior temporal gyrus	−56, −20, −8	298
	Right middle temporal gyrus*, right superior temporal gyrus	56, −18, −4	308
Speech passive listening < music discrimination	Left precentral gyrus*, left superior temporal gyrus	−52, 2, 8	105
	Left postcentral gyrus*, left precentral gyrus, left inferior parietal lobule	−50, −14, 52	199
	Left cerebellum	−28, −64, −28	127
	Right inferior frontal gyrus*, right middle frontal gyrus	50, 10, 30	50
	Right medial frontal gyrus*, right superior frontal gyrus, left medial frontal gyrus, left superior frontal gyrus	2, −6, 62	166
	Right precentral gyrus*, right middle frontal gyrus	38, −8, 42	67
	Right superior temporal gyrus*	64, −26, 8	76
	Right superior temporal gyrus*, right precentral gyrus, right superior temporal gyrus	50, 6, 4	47
Speech passive listening > music detection	Left inferior frontal gyrus (pars triangularis*, pars opercularis),	−50, 22, 10	107
	Left middle frontal gyrus*, left precentral gyrus, left postcentral gyrus,	−54, 2, 40	138
	Left middle temporal gyrus*, left superior temporal gyrus, left inferior temporal gyrus	−60, −8, −10	1052
	Left superior temporal gyrus*	−48, 16, −16	29
	Right middle temporal gyrus*, right superior temporal gyrus	60, −18, −8	651
Speech passive listening < music detection	Left precentral gyrus*, left superior temporal gyrus	−52, 2, 6	54
	Left insula*, left superior temporal gyrus	−50, −20, 16	430
	Left insula*, left superior temporal gyrus	−40, 6, −4	31
	Left inferior parietal lobule*, left supramarginal gyrus	−42, −50, 38	40
	Left lentiform nucleus*, left claustrum, left insula	−20, 6, 6	203
	Right middle frontal gyrus*, right inferior frontal gyrus, right insula	42, 16, 32	220
	Right middle frontal gyrus*, right precentral gyrus	30, −6, 56	35
	Right superior frontal gyrus*, right middle frontal gyrus	32, 44, 16	40
	Right superior frontal gyrus*	4, 12, 54	36
	Right insula*, right transverse temporal gyrus, right superior temporal gyrus, right precentral gyrus	44, −12, 12	519
	Right inferior parietal lobule*, right supramarginal gyrus	40, −48, 38	103
	Right thalamus*, right caudate	8, −2, 10	142
	Right thalamus*	10, −14, 6	33
Speech passive listening > music memory [−10pt]	Left middle temporal gyrus*, left middle temporal gyrus, left transverse temporal gyrus	−58, −38, −4	1256
	Right superior temporal gyrus*, right transverse temporal gyrus, right middle temporal gyrus, right postcentral gyrus	58, −2, 2	1056
Speech passive listening < music memory [−10pt]	Left inferior frontal gyrus (pars orbitalis*, pars triangularis)	−32, 24, −6	31
	Left medial frontal gyrus*, left superior frontal gyrus, left cingulate, right superior frontal gyrus, right medial frontal gyrus, right cingulate	−0, 22, 46	336
	Left precentral gyrus*, left inferior frontal gyrus (pars opercularis)	−52, 2, 30	34
	Left precentral gyrus*	−38, −10, 38	32
	Left supramarginal gyrus*, left inferior parietal lobule, left superior temporal gyrus	−46, −46, 26	113
	Left inferior parietal lobule*, left postcentral gyrus	−48, −36, 48	71
	Left middle temporal gyrus*, left superior temporal gyrus	−46, −24, −10	44
	Right insula*, right inferior frontal gyrus, right claustrum	34, 22, 4	48

The x, y, z coordinates are in Talairach space and refer to the peak voxel activated in each contrast. All contrasts are thresholded at $p = 0.05$. Asterisks indicate anatomical location of peak voxel.

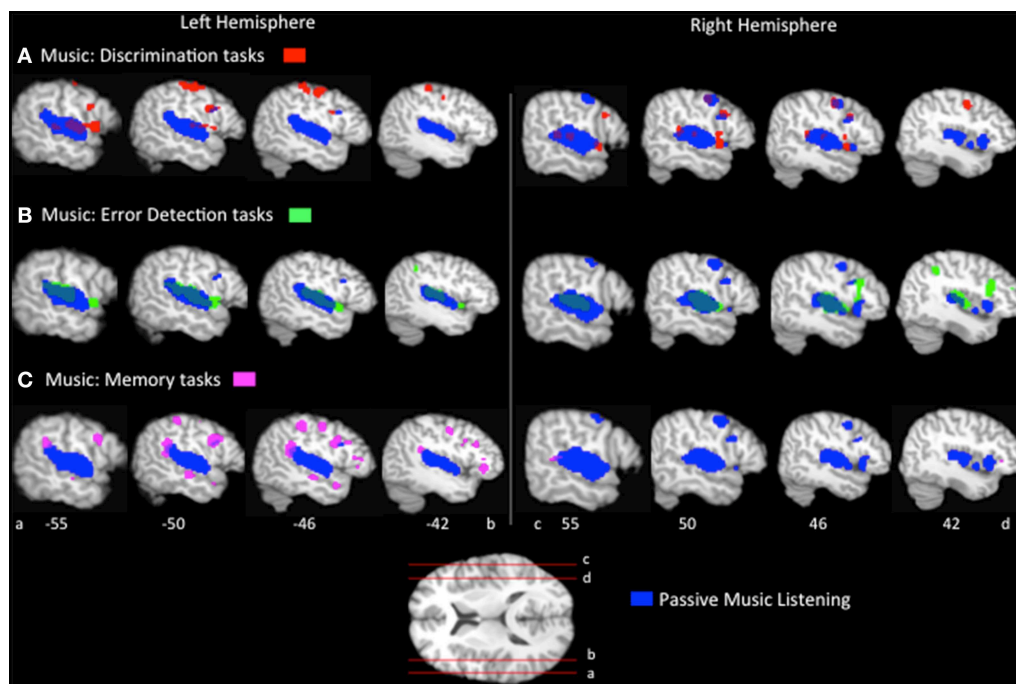


FIGURE 2 | Representative sagittal slices of the ALEs for the (A) music discrimination, (B) music error detection and (C) music memory task conditions, $p < 0.05$, corrected, overlaid on top of the passive music listening ALE for comparison.

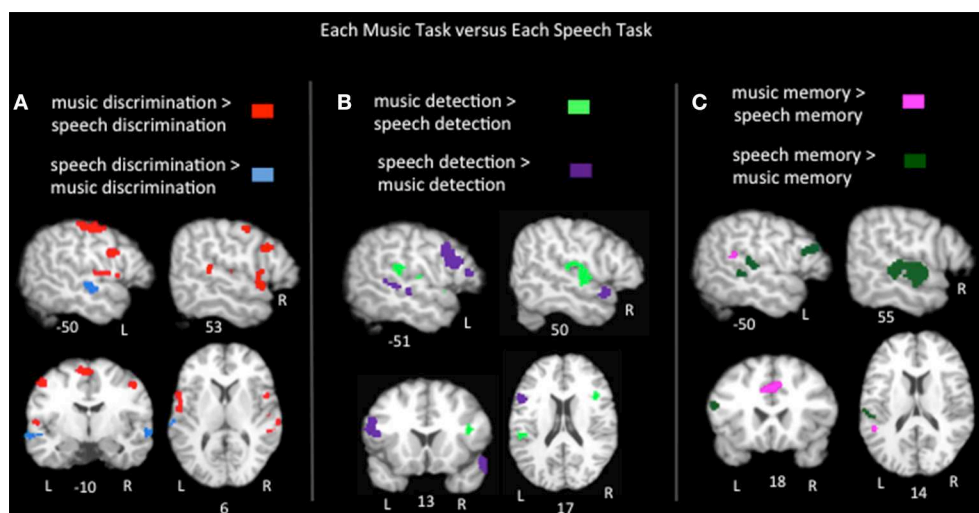


FIGURE 3 | Representative slices of the contrast results for the comparison of (A) music discrimination, (B) music error detection, and (C) music memory task conditions, compared to the corresponding speech task, $p < 0.05$, corrected.

2005; Koelsch et al., 2009; Buchsbaum et al., 2011) and are considered to be part of an auditory sensory-motor integration network (Hickok et al., 2003; Hickok and Poeppel, 2004, 2007).

Music Tasks vs. Passive Listening To Speech

Findings from various music paradigms and tasks are often reported as engaging language networks because of location; a music paradigm activating Broca's area or superior temporal

regions is frequently described as recruiting classic language areas. However, it is not clear if these music paradigms are in fact engaging the language networks engaged in the natural, everyday process of listening to speech. Thus, pairwise contrasts of the ALEs for listening to speech vs. the music tasks were calculated (Figure 4; Table 2). Music discrimination > speech passive listening identified regions in bilateral precentral gyri, bilateral medial frontal gyri, left postcentral gyrus, left inferior parietal

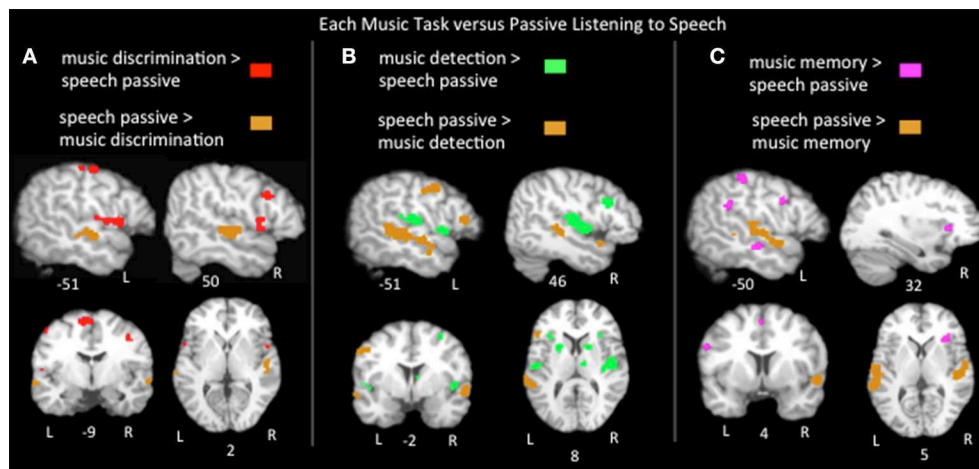


FIGURE 4 | Representative slices of the contrast results for the comparison of (A) music discrimination, (B) music error detection, (C) music memory task conditions, compared to passive listening to speech, $p < 0.05$, corrected.

lobule, left cerebellum, right inferior and middle frontal gyri, and right superior temporal gyrus. Music error detection > speech identified bilateral precentral gyri, bilateral superior temporal gyri, bilateral insula, bilateral basal ganglia, left postcentral gyrus, left cerebellum, bilateral inferior parietal lobe, right middle frontal gyrus, right inferior frontal gyrus and the right thalamus. Music memory > speech identified portions of bilateral inferior frontal gyri, bilateral medial frontal gyri, left inferior parietal lobe, left pre and postcentral gyri, and right insula. Compared to all three music tasks, speech significantly activated bilateral superior temporal sulcus regions and only activated Broca's area (specifically the pars triangularis) more than music detection. The recruitment of Broca's area and adjacent regions for music was task dependent: compared to listening to speech, music detection and discrimination activated additional bilateral inferior precentral gyrus regions immediately adjacent to Broca's area and music memory activated the left inferior frontal gyrus more than speech (in all three subregions: pars opercularis, pars triangularis, and pars orbitalis). In the right hemisphere homolog of Broca's area, all three music tasks activated this region more than listening to speech as well as adjacent regions in the right middle frontal gyrus. All together these results suggest that the recruitment of neural resources used in speech for music processing depends on the experimental paradigm. The finding of music memory tasks eliciting widespread activation in Broca's area compared to listening to speech is likely due to the inferior frontal gyrus, and the pars opercularis in particular being consistently implicated in articulatory rehearsal and working memory (Hickok et al., 2003; Buchsbaum et al., 2011, 2005), resources that are likely recruited by the music memory tasks.

Music Tasks vs. Passive Listening To Music

Lastly we compared the music task ALEs to the music passive listening ALE using pairwise contrasts to better characterize task-specific activations to music. Results ($p < 0.05$, FDR corrected)

include: (1) music discrimination > music listening identified bilateral inferior precentral gyri, bilateral medial frontal regions, left postcentral gyrus, left inferior parietal lobule, left cerebellum, right middle frontal gyrus and right insula (2) music error detection > music listening identified bilateral medial frontal, bilateral insula, bilateral inferior parietal areas, bilateral superior temporal gyri, bilateral basal ganglia, left pre and post central gyri, right inferior and middle frontal gyri and right cerebellum; (3) music memory > passive listening identified bilateral inferior frontal gyri (pars opercularis, triangularis and orbitalis in the left hemisphere, only the latter two in the right hemisphere), bilateral medial frontal gyri, bilateral insula, bilateral cerebellum, left middle frontal gyrus, left inferior parietal lobe, left superior and middle temporal gyri, right basal ganglia, right hippocampus and right parahippocampal gyrus (Figure 5, Table 2). The medial frontal and inferior parietal activations identified in the tasks compared to listening likely reflect increased vigilance and attention due to the presence of a task, as activation in these regions is known to increase as a function of effort and performance on tasks across a variety of stimuli types and domains (Petersen and Posner, 2012; Vaden et al., 2013). To summarize the findings in Broca's area and its right hemisphere homolog, music memory tasks activated Broca's area more than just listening to music, while music discrimination and detection tasks activated right inferior frontal gyrus regions more than listening to music. Also note that all three music tasks compared to listening to music implicate regions on the anterior bank of the inferior portion of the precentral gyrus immediately adjacent to Broca's area. Significant clusters more active for music passive listening than for each of the three task conditions are found in the bilateral superior temporal gyri (Table 2).

Discussion

The present meta-analysis examined data from 80 functional neuroimaging studies of music and 91 studies of speech

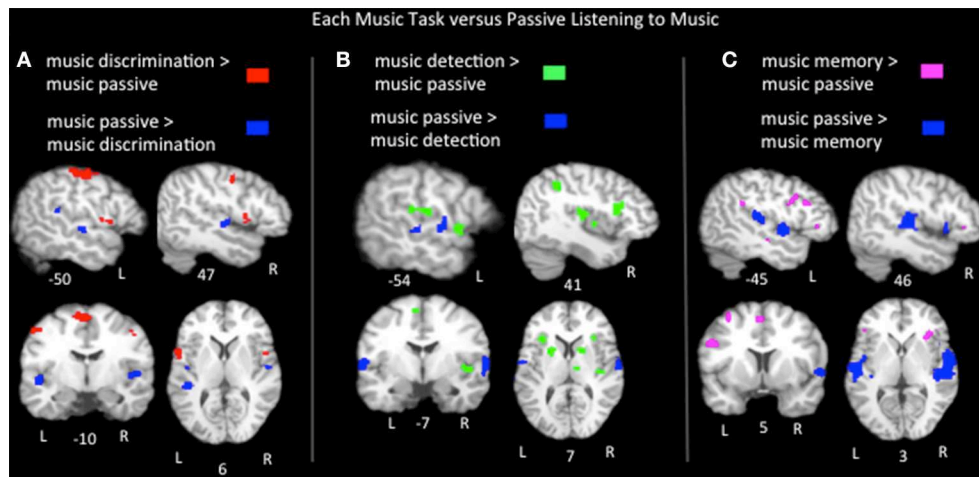


FIGURE 5 | Representative slices of the contrast results for the comparison of (A) music discrimination, (B) music error detection, (C) music memory task conditions, compared to passive listening to music, $p < 0.05$, corrected.

to characterize the relationship between the brain networks activated by listening to speech vs. listening to music. We also compared the brain regions implicated in three frequently used music paradigms (error detection, discrimination, and memory) to the regions implicated in similar speech paradigms to determine how task effects may change how the neurobiology of music processing is related to that of speech. We replicated across a large collection of studies' previous within-subject findings that speech activates a predominately lateral temporal network, while music preferentially activates a more dorsal medial temporal network extending into the inferior parietal lobe. In Broca's area, we found overlapping resources for passive listening to speech and music in the pars opercularis, but speech "specific" resources in pars triangularis; the right hemisphere homolog of Broca's area was equally responsive to listening to speech and music. The use of a paradigm containing an explicit task (error detection, discrimination or memory) altered the relationship between the brain networks engaged in music and speech. For example, speech discrimination tasks do not activate the pars triangularis (i.e., the region identified as "speech specific" by the passive listening contrast) more than music discrimination tasks, and both speech detection and memory tasks activate the pars opercularis (i.e., the region responding equally to music and speech passive listening) more than the corresponding music tasks, while music discrimination activates pars opercularis more than speech discrimination. These findings suggest that inferior frontal contributions to music processing, and their overlap with speech resources, may be modulated by task. The following sections discuss these findings in relation to neuroanatomical models of speech and music.

Hemispheric Differences for Speech and Music

The lateralization of speech and music processing has been investigated for decades. While functional neuroimaging studies report bilateral activation for both speech and music (Jäncke et al., 2002; Abrams et al., 2011; Fedorenko et al., 2011; Rogalsky

et al., 2011), evidence from amusia, aphasia and other patient populations have traditionally identified the right hemisphere as critical for music and the left for basic language processes in most individuals (Gazzaniga, 1983; Peretz et al., 2003; Damasio et al., 2004; Hyde et al., 2006). Further evidence for hemispheric differences comes from asymmetries in early auditory cortex: left hemisphere auditory cortex has better temporal resolution and is more sensitive to rapid temporal changes critical for speech processing, while the right hemisphere auditory cortex has higher spectral resolution and is more modulated by spectral changes, which optimize musical processing (Zatorre et al., 2002; Poeppel, 2003; Schönwiesner et al., 2005; Hyde et al., 2008). Thus, left auditory cortex has been found to be more responsive to phonemes than chords, while right auditory cortex is more responsive to chords than phonemes (Tervaniemi et al., 1999, 2000). This hemispheric specialization coincides with evidence from both auditory and visual domains, suggesting that the left hemisphere tends to be tuned to local features, while the right hemisphere is tuned to more global features (Sergent, 1982; Ivry and Robertson, 1998; Sanders and Poeppel, 2007).

Hemispheric differences in the present study for speech and music vary by location. We did not find any qualitative hemispheric differences between speech and music in the temporal lobe. Speech bilaterally activated lateral superior and middle temporal regions, while music bilaterally activated more dorsal medial superior temporal regions extending into the inferior parietal lobe. However, these bilateral findings should not be interpreted as evidence against hemispheric asymmetries for speech vs. music. The hemispheric differences widely reported in auditory cortex almost always are a matter of degree, e.g., phonemes and tones both activate bilateral superior temporal regions, but a direct comparison indicates a left hemisphere preference for the speech and a right hemisphere preference for the tones (Jäncke et al., 2002; Zatorre et al., 2002). These differences would not be reflected in our ALE results because both conditions reliably activate the same regions although to

different degrees and the ALE method does not assign weight to coordinates (i.e., all the significant coordinates reported for contrasts of interest in the studies used) based on their beta or statistical values.

The frontal lobe results, however, did include some laterality differences of interest: passive listening to speech activated portions of the left inferior frontal gyrus (i.e., Broca's area), namely in the pars triangularis, significantly more than listening to music. A right inferior frontal gyrus cluster, extending into the insula, was activated significantly more for listening to music than speech. These findings in Broca's area coincide with Koelsch's neurocognitive model of music perception, in that right frontal regions are more responsive to musical stimuli and that the pars opercularis, but not the pars triangularis, is engaged in structure building of auditory stimuli (Koelsch, 2011). It is also noteworthy that the inclusion of a task altered hemispheric differences in the frontal lobes: the music discrimination tasks activated the left pars opercularis more than speech discrimination, while speech detection and memory tasks activated all of Broca's area (pars opercularis and pars triangularis) more than music detection and memory tasks; music detection and discrimination tasks, but not music memory tasks, activated the right inferior frontal gyrus more than corresponding speech tasks. These task-modulated asymmetries in Broca's area for music are particularly important when interpreting the rich electrophysiological literature of speech and music interactions. For example, both the early right anterior negativity (ERAN) and early left anterior negativity (ELAN) are modulated by speech and music, and are believed to have sources in both Broca's area and its right hemisphere homolog (Friederici et al., 2000; Maess et al., 2001; Koelsch and Friederici, 2003). Thus, the lateralization patterns found in the present study emphasize the need to consider that similar ERP effects for speech and music may arise from different underlying lateralization patterns that may be task-dependent.

Speech vs. Music in the Anterior Temporal Lobe

Superior and middle posterior temporal regions on the banks of the superior temporal sulcus were preferentially activated in each speech condition compared to each corresponding music condition in the present meta-analysis. This is not surprising, as these posterior STS regions are widely implicated in lexical semantic processing (Price, 2010) and STS regions have been found to be more responsive to syllables than tones (Jäncke et al., 2002). Perhaps more interestingly, the bilateral anterior temporal lobe (ATL) also was activated more for each speech condition than by each corresponding music condition. The role of the ATL in speech processing is debated (e.g., Scott et al., 2000 cf. Hickok and Poeppel, 2004, 2007), but the ATL is reliably sensitive to syntactic structure in speech compared to several control conditions including word lists, scrambled sentences, spectrally rotated speech, environmental sounds sequences, and melodies (Mazoyer et al., 1993; Humphries et al., 2001, 2005, 2006; Xu et al., 2005; Spitsyna et al., 2006; Rogalsky and Hickok, 2009; Friederici et al., 2010; Rogalsky et al., 2011). One hypothesis is that the ATL is implicated in combinatorial semantic processing (Wong and Gallate, 2012; Wilson et al.,

2014), although pseudoword sentences (i.e., sentences lacking meaningful content words) also activate the ATL (Humphries et al., 2006; Rogalsky et al., 2011). Several of the speech activation coordinates included in the present meta-analysis were from studies that used sentences and phrases as stimuli (with and without semantic content). It is likely that these coordinates are driving the ATL findings. Our finding that music did not activate the ATL supports the idea that the ATL is not responsive to hierarchical structure *per se* but rather needs linguistic and/or semantic information for it to be recruited.

Speech vs. Music in Broca's Area

There is no consensus regarding the role of Broca's area in receptive speech processes (e.g., Fedorenko and Kanwisher, 2011; Hickok and Rogalsky, 2011; Rogalsky and Hickok, 2011). Results from the present meta-analysis indicate that listening to speech activated both the pars opercularis and pars triangularis portions of Broca's area, while listening to music only activated the pars opercularis. The pars triangularis has been proposed to be involved in semantic integration (Hagoort, 2005) as well as in cognitive control processes such as conflict resolution (Novick et al., 2005; Rogalsky and Hickok, 2011). It is likely that the speech stimuli contain more semantic content than the music stimuli, and thus semantic integration processes may account for the speech-only response in pars triangularis. However, there was no significant difference in activations in the pars triangularis for the music discrimination and music detection tasks vs. passive listening to speech, and the music memory tasks activated portions of the pars triangularis more than listening to speech. These music task-related activations in the pars triangularis may reflect the use of semantic resources for categorization or verbalization strategies to complete the music tasks, but may also reflect increased cognitive control processes to support reanalysis of the stimuli to complete the tasks. The activation of the left pars opercularis for both speech and music replicates numerous individual studies implicating the pars opercularis in both speech and musical syntactic processing (e.g., Koelsch and Siebel, 2005; Rogalsky and Hickok, 2011) as well as in a variety of auditory working memory paradigms (e.g., Koelsch and Siebel, 2005; Buchsbaum et al., 2011).

Implications for Neuroanatomical Models of Speech and Music

It is particularly important to consider task-related effects when evaluating neuroanatomical models of the interactions between speech and music. It has been proposed that inferior frontal cortex (including Broca's area) is the substrate for shared speech-music executive function resources, such as working memory and/or cognitive control (Patel, 2003; Slevc, 2012; Slevc and Okada, 2015) as well as auditory processes such as structure analysis, repair, working memory and motor encoding (Koelsch and Siebel, 2005; Koelsch, 2011). Of particular importance here is Slevc and Okada's (2015) proposal that cognitive control may be one of the shared cognitive resources for linguistic and musical processing when reanalysis and conflict resolution is necessary. Different tasks likely recruit cognitive control resources to different degrees, and thus may explain task-related differences

for the frontal lobe's response to speech and music. There is ample evidence to support Slevc and Okada's hypothesis: classic cognitive control paradigms such as the Stroop task (Stroop, 1935; MacLeod, 1991) elicit overlapping activations in Broca's area when processing noncanonical sentence structures (January et al., 2009). Unexpected harmonic and melodic information in music interfere with Stroop task performance (Masataka and Perlovsky, 2013). The neural responses to syntactic and sentence-level semantic ambiguities in language also interact with responses to unexpected harmonics in music (Koelsch et al., 2005; Steinbeis and Koelsch, 2008b; Slevc et al., 2009; Perruchet and Poulin-Charronnat, 2013). The present results suggest that this interaction between language and music possibly via cognitive control mechanisms, localized to Broca's area, may be task driven and not inherent to the stimuli themselves. In addition, many language/music interaction studies use a reading language task with simultaneous auditory music stimuli; it is possible that a word-by-word presentation reading paradigm engages additional reanalysis mechanisms that may dissociate from resources used in auditory speech processing (Tillmann, 2012).

Slevc and Okada suggest that future studies should use tasks designed to drive activation of specific processes, presumably including reanalysis. However, the present findings suggest it is possible that these task-induced environments may actually drive overlap of neural resources for speech and music not because they are taxing shared sensory computations but rather because they are introducing additional processes that are not elicited during typical, naturalistic music listening. For example, consider the present findings in the left pars triangularis: this region is not activated during listening to music, but is activated while listening to speech. However, by presumably increasing the need for reanalysis mechanisms via discrimination or memory tasks, music does recruit this region.

There may be inferior frontal shared mechanisms that are stimulus driven while others are task driven: Broca's area is a diverse region in terms of its cytoarchitecture, connectivity and response properties (Amunts et al., 1999; Anwander et al., 2007; Rogalsky and Hickok, 2011; Rogalsky et al., in press). It is possible that some networks are task driven and some are stimulus driven. The hypotheses of Koelsch et al. are largely grounded in behavioral and electrophysiology studies that indicate an interaction between melodic and syntactic information (e.g., Koelsch et al., 2005; Fedorenko et al., 2009; Hoch et al., 2011). It is not known if these interactions are stimulus driven; a variety of tasks have been used in this literature, including discrimination, anomaly/error detection, (Koelsch et al., 2005; Carrus et al., 2013), grammatical acceptability (Patel et al., 1998a; Patel, 2008), final-word lexical decision (Hoch et al., 2011), and memory/comprehension tasks (Fedorenko et al., 2009, 2011). In addition, there is substantial variability across individual subjects, both functionally and anatomically, within Broca's area (Amunts et al., 1999; Schönwiesner et al., 2007; Rogalsky et al., in press). Thus, future within-subject studies are needed to better understand the role of cognitive control and other domain-general resources in musical processing independent of task.

Different tasks, regardless of the nature of the stimuli, may require different attentional resources (Shallice, 2003). Thus, it is possible that the inferior frontal differences between the music tasks and passive listening to music and speech are due to basic attentional differences, not the particular task *per se*. However, we find classic domain-general attention systems in the anterior cingulate and medial frontal cortex to be significantly activated across all conditions: music tasks, speech tasks, passive listening to music and passive listening to speech. These findings support Slevc and Okada's (2015) claim that domain-general attention mechanisms facilitated by anterior cingulate and medial frontal cortex are consistently engaged for music as they are for speech. Each of our music task conditions do activate these regions significantly more than the passive listening, suggesting that the midline domain-general attention mechanisms engaged by music can be further activated by explicit tasks.

Limitations and Future Directions

One issue in interpreting our results may be the proximity of distinct networks for speech and music (Peretz, 2006; Koelsch, 2011). Overlap in fMRI findings, particularly in a meta-analysis, does not necessarily mean that speech and music share resources in those locations. It is certainly possible that the spatial resolution of fMRI is not sufficient to visualize separation occurring at a smaller scale (Peretz and Zatorre, 2005; Patel, 2012). However, our findings of spatially distinct regions for music and speech clearly suggest the recruitment of distinct brain networks for speech and music.

Another potential issue related to the limitations of fMRI is that of sensitivity. Continuous fMRI scanning protocols (i.e., stimuli are presented simultaneously with the noise of scanning) and sparse temporal sampling fMRI protocols (i.e., stimuli are presented during silent periods between volume acquisitions) are both included in the present meta-analyses. It has been suggested that the loud scanner noise may reduce sensitivity to detecting hemodynamic response to stimuli, particularly complex auditory stimuli such as speech and music (Peelle et al., 2010; Elmer et al., 2012). Thus, it is possible that effects only detected by a sparse or continuous paradigm are not represented in our ALE results. However, a comparison of continuous vs. sparse fMRI sequences found no significant differences in speech activations in the frontal lobe between the pulse sequences (Peelle et al., 2010).

Priming paradigms measuring neurophysiological responses (ERP, fMRI, etc.) are one way to possibly circumvent task-related confounds in understanding the neurobiology of music in relation to that of speech. Tillmann (2012) suggests that priming paradigms may provide more insight into an individual's implicit musical knowledge than is demonstrated by performance on an explicit, overt task (e.g., Schellenberg et al., 2005; Tillmann et al., 2007). In fact, there are ERP studies that indicate that musical chords can prime processing of target words if the prime and target are semantically (i.e., emotionally) similar (Koelsch et al., 2004; Steinbeis and Koelsch, 2008a). However, most ERP priming studies investigating music or music/speech interactions have included an explicit task (e.g., Schellenberg et al., 2005; Tillmann et al., 2007; Steinbeis and Koelsch, 2008a). It is not known how

the presence of an explicit task may affect priming mechanisms via top-down mechanisms. Priming is not explored in the present meta-analysis; to our knowledge there is only one fMRI priming study of music and speech, which focused on semantic (i.e., emotion) relatedness (Steinbeis and Koelsch, 2008a).

The present meta-analysis examines networks primarily in the cerebrum. Even though almost all of the studies included in our analyses focused on cortical structures, we still identified some subcortical task-related activations: music detection compared to music passive listening activated the basal ganglia and music memory tasks activated the thalamus, hippocampus and basal ganglia compared to music passive listening. No significant differences between passive listening to speech and music were found in subcortical structures. These findings (and null results) in subcortical regions should be interpreted cautiously: given the relatively small size of these structures, activations in these areas are particularly vulnerable to spatial smoothing filters and group averaging (Raichle et al., 1991; White et al., 2001). There is also strong evidence that music and speech share subcortical resources in the brainstem (Patel, 2011), which are not addressed by the present study. For example, periodicity is a critical aspect of both speech and music and known to modulate networks between the cochlea and inferior colliculus of

the brainstem (Cariani and Delgutte, 1996; Patel, 2011). Further research is needed to better understand where speech and music processing networks diverge downstream from these shared early components.

Conclusion

Listening to music and listening to speech engage distinct temporo-parietal cortical networks but share some inferior and medial frontal resources (at least at the resolution of fMRI). However, the recruitment of inferior frontal speech-processing regions for music is modulated by task. The present findings highlight the need to consider how task effects may be interacting with conclusions regarding the neurobiology of speech and music.

Acknowledgments

This work was supported by a GRAMMY Foundation Scientific Research Grant (PI Rogalsky) and Arizona State University. We thank Nicole Blumenstein and Dr. Nancy Moore for their help in the preparation of this manuscript.

References

- Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., and Menon, V. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cereb. Cortex* 21, 1507–1518. doi: 10.1093/cercor/bhq198
- Adank, P. (2012). Design choices in imaging speech comprehension: an activation likelihood estimation (ALE) meta-analysis. *Neuroimage* 63, 1601–1613. doi: 10.1016/j.neuroimage.2012.07.027
- Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B. M., and Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* 412, 319–341.
- Anwander, A., Tittgemeyer, M., von Cramon, D. Y., Friederici, A. D., and Knösche, T. R. (2007). Connectivity-based parcellation of Broca's area. *Cereb. Cortex* 17, 816–825. doi: 10.1093/cercor/bhk034
- Baker, E., Blumstein, S. E., and Goodglass, H. (1981). Interaction between phonological and semantic factors in auditory comprehension. *Neuropsychology* 19, 1–15. doi: 10.1016/0028-3932(81)90039-7
- Basso, A., and Capitani, E. (1985). Spared musical abilities in a conductor with global aphasia and ideomotor apraxia. *J. Neurol. Neurosurg. Psychiatry* 48, 407–412. doi: 10.1136/jnnp.48.5.407
- Besson, M., Chobert, J., and Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *F. Psychol.* 2:94. doi: 10.3389/fpsyg.2011.00094
- Besson, M., and Faita, F. (1995). An event-related potential (ERP) study of musical expectancy: comparison of musicians with nonmusicians. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 1278–1296. doi: 10.1037/0096-1523.21.6.1278
- Besson, M., and Schön, D. (2001). Comparison between language and music. *Ann. N.Y. Acad. Sci.* 930, 232–258. doi: 10.1111/j.1749-6632.2001.tb05736.x
- Brattico, E., Tervaniemi, M., Näätänen, R., and Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. *Brain Res.* 1117, 162–174. doi: 10.1016/j.brainres.2006.08.023
- Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., et al. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory - an aggregate analysis of lesion and fMRI data. *Brain Lang.* 119, 119–128. doi: 10.1016/j.bandl.2010.12.001
- Buchsbaum, B. R., Olsen, R. K., Koch, P., and Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* 48, 687–697. doi: 10.1016/j.neuron.2005.09.029
- Cant, J. S., and Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cereb. Cortex* 17, 713–731. doi: 10.1093/cercor/bhk022
- Cariani, P. A., and Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* 76, 1698–1716.
- Carrus, E., Pearce, M. T., and Bhattacharya, J. (2013). Melodic pitch expectation interacts with neural responses to syntactic but not semantic violations. *Cortex* 49, 2186–2200. doi: 10.1016/j.cortex.2012.08.024
- Chawla, D., Rees, G., and Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nat. Neurosci.* 2, 671–676. doi: 10.1038/10230
- Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., and Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science* 248, 1556–1559. doi: 10.1126/science.2360050
- Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., and Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition* 92, 179–229. doi: 10.1016/j.cognition.2002.07.001
- Dronkers, N. F., Wilkins, D. P., Van Valin, R. D. Jr., Redfern, B. B., and Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition* 92, 145–177. doi: 10.1016/j.cognition.2003.11.002
- Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., and Fox, P. T. (2012). Activation likelihood estimation revisited. *Neuroimage* 59, 2349–2361. doi: 10.1016/j.neuroimage.2011.09.017
- Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., and Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Hum. Brain Mapp.* 30, 2907–2926. doi: 10.1002/hbm.20718
- Elmer, S., Meyer, S., and Jäncke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cereb. Cortex* 22, 650–658. doi: 10.1093/cercor/bhr142

- Fedorenko, E., Behr, M. K., and Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16428–16433. doi: 10.1073/pnas.1112937108
- Fedorenko, E., and Kanwisher, N. (2011). Some regions within Broca's area do respond more strongly to sentences than to linguistically degraded stimuli: a comment on Rogalsky and Hickok (2010). *J. Cogn. Neurosci.* 23, 2632–2635. doi: 10.1162/jocn_a_00043
- Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., and Gibson, E. (2009). Structural integration in language and music: evidence for a shared system. *Mem. Cogn.* 37, 1–9. doi: 10.3758/MC.37.1.1
- Frances, R., Lhermitte, F., and Verdy, M. F. (1973). Le deficit musical des aphasiques. *Appl. Psychol.* 22, 117–135. doi: 10.1111/j.1464-0597.1973.tb00391.x
- Friederici, A. D., Kotz, S. A., Scott, S. K., and Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Hum. Brain Mapp.* 31, 448–457. doi: 10.1002/hbm.20878
- Friederici, A. D., Wang, Y., Herrmann, C. S., Maess, B., and Oertel, U. (2000). Localization of early syntactic processes in frontal and temporal cortical areas: a magnetoencephalographic study. *Hum. Brain Mapp.* 11, 1–11. doi: 10.1002/1097-0193(200009)11:1<1::AID-HBM10>3.0.CO;2-B
- Gazzaniga, M. S. (1983). Right hemisphere language following brain bisection: a 20-year perspective. *Am. Psychol.* 38, 525–537. doi: 10.1037/0003-066X.38.5.525
- Geiser, E., Zaehle, T., Jancke, L., and Meyer, M. (2008). The neural correlate of speech rhythm as evidenced by metrical speech processing. *J. Cogn. Neurosci.* 20, 541–552. doi: 10.1162/jocn.2008.20029
- Grahn, J. A. (2012). “Advances in neuroimaging techniques: Implications for the shared syntactic integration resource hypothesis,” in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohrmeier, J. Hawkins, and I. Cross (Oxford: Oxford University Press), 235–241.
- Hagoort, P. (2005). On Broca, brain and binding: a new framework. *Trends Cogn. Sci.* 9, 416–423. doi: 10.1016/j.tics.2005.07.004
- Henschen, S. E. (1924). On the function of the right hemisphere of the brain in relation to the left in speech, music and calculation. *Brain* 44, 110–123.
- Hickok, G., Buchsbaum, B., Humphries, C., and Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J. Cogn. Neurosci.* 15, 673–682. doi: 10.1162/08989290323207393
- Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi: 10.1016/j.cognition.2003.10.011
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Hickok, G., and Rogalsky, C. (2011). What does Broca's area activation to sentences reflect? *J. Cogn. Neurosci.* 23, 2629–2631. doi: 10.1162/jocn_a_00044
- Hoch, L., Poulin-Charronnat, B., and Tillmann, B. (2011). The influence of task-irrelevant music on language processing: syntactic and semantic structures. *Front. Psychol.* 2:112. doi: 10.3389/fpsyg.2011.00112
- Humphries, C., Binder, J. R., Medler, D. A., and Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *J. Cogn. Neurosci.* 18, 665–679. doi: 10.1162/jocn.2006.18.4.665
- Humphries, C., Love, T., Swinney, D., and Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Hum. Brain Mapp.* 26, 128–138. doi: 10.1002/hbm.20148
- Humphries, C., Willard, K., Buchsbaum, B., and Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *Neuroreport* 12, 1749–1752. doi: 10.1097/00001756-200106130-00046
- Hyde, K. L., Peretz, I., and Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia* 46, 632–639. doi: 10.1016/j.neuropsychologia.2007.09.004
- Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., and Peretz, I. (2006). Morphometry of the amusic brain: a two-site study. *Brain* 129, 2562–2570. doi: 10.1093/brain/awl204
- Ivry, R. B., and Robertson, L. C. (1998). *The Two Sides of Perception*. Cambridge, MA: MIT Press.
- Jäncke, L. (2008). Music, memory and emotion. *J. Biol.* 7, 21. doi: 10.1186/jbiol82
- Jäncke, L., Wüstenberg, T., Scheich, H., and Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage* 15, 733–746. doi: 10.1006/nimg.2001.1027
- January, D., Trueswell, J. C., and Thompson-Schill, S. L. (2009). Co-localization of stroop and syntactic ambiguity resolution in Broca's area: implications for the neural basis of sentence processing. *J. Cogn. Neurosci.* 21, 2434–2444. doi: 10.1162/jocn.2008.21179
- Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Curr. Opin. Neurobiol.* 15, 207–212. doi: 10.1016/j.conb.2005.03.005
- Koelsch, S. (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., and Friederici, A. D. (2003). Toward the neural basis of processing structure in music. *Ann. N.Y. Acad. Sci.* 999, 15–28. doi: 10.1196/annals.1284.002
- 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., Kasper, E., Sammler, D., Schulze, K., Gunter, T., and Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nat. Neurosci.* 7, 302–307. doi: 10.1038/nn1197
- Koelsch, S., 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
- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., and Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an fMRI study. *Hum. Brain Mapp.* 30, 859–873. doi: 10.1002/hbm.20550
- Luria, A. R., Tsvetkova, L., and Futer, D. S. (1965). Aphasia in a composer. *J. Neurol. Sci.* 1, 288–292. doi: 10.1016/0022-510X(65)90113-9
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychol. Bull.* 109, 163–203. doi: 10.1037/0033-2909.109.2.163
- Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nat. Neurosci.* 4, 540–545. doi: 10.1038/87502
- Maillard, L., Barbeau, E. J., Baumann, C., Koessler, L., Bénar, C., Chauvel, P., et al. (2011). From perception to recognition memory: time course and lateralization of neural substrates of word and abstract picture processing. *J. Cogn. Neurosci.* 23, 782–800. doi: 10.1162/jocn.2010.21434
- Masataka, N., and Perlovsky, L. (2013). Cognitive interference can be mitigated by consonant music and facilitated by dissonant music. *Sci. Rep.* 3, 1–6. doi: 10.1038/srep02028
- Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., et al. (1993). The cortical representation of speech. *J. Cogn. Neurosci.* 5, 467–479. doi: 10.1162/jocn.1993.5.4.467
- Ni, W., Constable, R. T., Mencl, W. E., Pugh, K. R., Fulbright, R. K., Shaywitz, S. E., et al. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *J. Cogn. Neurosci.* 12, 120–133. doi: 10.1162/08989290051137648
- 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
- Novick, J. M., Trueswell, J. C., and Thompson-Schill, S. L. (2005). Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. *Cogn. Affect. Behav. Neurosci.* 5, 263–281. doi: 10.3758/CABN.5.3.263
- Oechslin, M. S., Meyer, M., and Jäncke, L. (2010). Absolute pitch – functional evidence of speech-relevant auditory acuity. *Cereb. Cortex* 20, 447–455. doi: 10.1093/cercor/bhp113
- Patel, A. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681. doi: 10.1038/nn1082
- Patel, A. D. (2005). The relationship of music to the melody of speech and to syntactic processing disorders in aphasia. *Ann. N.Y. Acad. Sci.* 1060, 59–70. doi: 10.1196/annals.1360.005
- Patel, A. D. (2008). *Music, Language, and the Brain*. New York, NY: Oxford Univ. Press.
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychol.* 2:142. doi: 10.3389/fpsyg.2011.00142
- Patel, A. D. (2012). “Language, music, and the brain: a resource-sharing framework,” in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohrmeier, J. Hawkins, and I. Cross (New York, NY: Oxford University Press), 204–223.

- Patel, A. D. (2013). "Sharing and nonsharing of brain resources for language and music," in *Language, Music, and the Brain*, ed M. Arbib (Cambridge, MA: MIT Press), 329–355.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. (1998a). Processing syntactic relations in language and music: an event-related potential study. *J. Cogn. Neurosci.* 10, 717–733. doi: 10.1162/089892998563121
- Patel, A. D., Peretz, I., Tramo, M., and Labreque, R. (1998b). Processing prosodic and musical patterns: a neuropsychological investigation. *Brain Lang.* 61, 123–144. doi: 10.1006/brln.1997.1862
- Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., and Davis, M. H. (2010). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *Neuroimage* 52, 1410–1419. doi: 10.1016/j.neuroimage.2010.05.015
- 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., Belleville, S., and Fontaine, S. (1997). Dissociations between music and language functions after cerebral resection: a new case of amusia without aphasia. *Can. J. Exp. Psychol.* 51, 354–368. doi: 10.1037/1196-1961.51.4.354
- Peretz, I., Champod, A. S., and Hyde, K. (2003). Varieties of musical disorders. *Ann. N.Y. Acad. Sci.* 999, 58–75. doi: 10.1196/annals.1284.006
- Peretz, I. and Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends. Cog. Sci.* 7, 362–367. doi: 10.1016/S1364-6613(03)00150-5
- Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., et al. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain* 117, 1283–1302. doi: 10.1093/brain/117.6.1283
- Peretz, I., and Zatorre, R. J. (2005). Brain organization for music processing. *Annu. Rev. Psychol.* 56, 89–114. doi: 10.1146/annurev.psych.56.091103.070225
- Perruchet, P., and Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychon. Bull. Rev.* 20, 310–317. doi: 10.3758/s13423-012-0344-5
- Petersen, S. E., and Posner, M. I. (2012). The attention system of the human brain: 20 years later. *Annu. Rev. Neurosci.* 35, 73–89. doi: 10.1146/annurev-neuro-062111-150525
- Platel, H., Price, C., Baron, J. C., Wise, R., Lambert, J., Frackowiak, R. S., et al. (1997). The structural components of music perception: a functional anatomical study. *Brain* 120, 229–243. doi: 10.1093/brain/120.2.229
- Poeppl, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time. *Speech Commun.* 41, 245–255. doi: 10.1016/S0167-6393(02)00107-3
- Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. N.Y. Acad. Sci.* 1191, 62–88. doi: 10.1111/j.1749-6632.2010.05444.x
- Raichle, M. E., Mintun, M. A., Shertz, L. D., Fusselman, M. J., and Miezen, F. (1991). The influence of anatomical variability on functional brain mapping with PET: a study of intrasubject versus intersubject averaging. *J. Cereb. Blood Flow Metab.* 11, S364.
- Rogalsky, C., Almeida, D., Sprouse, J., and Hickok, G. (in press). Sentence processing selectivity in Broca's area: evident for structure but not syntactic movement. *Lang. Cogn. Neurosci.*
- Rogalsky, C., and Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cereb. Cortex* 19, 786–796. doi: 10.1093/cercor/bhn126
- Rogalsky, C., and Hickok, G. (2011). The role of Broca's area in sentence comprehension. *J. Cogn. Neurosci.* 23, 1664–1680. doi: 10.1162/jocn.2010.21530
- Rogalsky, C., Poppa, N., Chen, K. H., Anderson, S. W., Damasio, H., Love, T., et al. (2015). Speech repetition as a window on the neurobiology of auditory-motor integration for speech: a voxel-based lesion symptom mapping study. *Neuropsychologia* 71, 18–27. doi: 10.1016/j.neuropsychologia.2015.03.012
- Rogalsky, C., Rong, F., Saberi, K. and Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J. Neurosci.* 31, 3843–3852. doi: 10.1523/JNEUROSCI/4515-10.2011
- Rorden, C., and Brett, M. (2000). Stereotaxic display of brain lesions. *Behav. Neurol.* 12, 191–200. doi: 10.1155/2000/421719
- Sammler, D., Koelsch, S., and Friederici, A. D. (2011). Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? *Cortex* 47, 659–673. doi: 10.1016/j.cortex.2010.04.007
- Sanders, L. D., and Poeppel, D. (2007). Local and global auditory processing: behavioral and ERP evidence. *Neuropsychologia* 45, 1172–1186. doi: 10.1016/j.neuropsychologia.2006.10.010
- Scheich, H., Brechmann, A., Brosch, M., Budinger, E., and Ohl, F. W. (2007). The cognitive auditory cortex: task-specificity of stimulus representations. *Hear. Res.* 229, 213–224. doi: 10.1016/j.heares.2007.01.025
- Schellenberg, E. G., Bigand, E., Poulin-Charronnat, B., Garnier, C., and Stevens, C. (2005). Children's implicit knowledge of harmony in Western music. *Dev. Sci.* 8, 551–566. doi: 10.1111/j.1467-7687.2005.00447.x
- Schönwiesner, M., Novitski, N., Pakarinen, S., Carlson, S., Tervaniemi, M., and Näätänen, R. (2007). Heschl's gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of acoustic changes. *J. Neurophysiol.* 97, 2075–2082. doi: 10.1152/jn.01083.2006
- Schönwiesner, M., Rübsem, R., and von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur. J. Neurosci.* 22, 1521–1528. doi: 10.1111/j.1460-9568.2005.04315.x
- Schwartz, M. F., Faseyitan, O., Kim, J., and Coslett, H. B. (2012). The dorsal stream contribution to phonological retrieval in object naming. *Brain* 135(Pt 12), 3799–3814. doi: 10.1093/brain/aww300
- Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406. doi: 10.1093/brain/123.12.2400
- Sergent, J. (1982). About face: left-hemisphere involvement in processing physiognomies. *J. Exp. Psychol. Hum. Percept. Perform.* 8, 1–14.
- Shallice, T. (2003). Functional imaging and neuropsychology findings: how can they be linked? *Neuroimage* 20, S146–S154. doi: 10.1016/j.neuroimage.2003.09.023
- Slevc, R. L. (2012). Language and music: sound, structure and meaning. *WIREs Cogn. Sci.* 3, 483–492. doi: 10.1002/wcs.1186
- Slevc, L. R., Rosenberg, J. C., and Patel, A. D. (2009). Making psycholinguistics musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychon. Bull. Rev.* 16, 374–381. doi: 10.3758/16.2.374
- Slevc, L. R., and Okada, B. M. (2015). Processing structure in language and music: a case for shared reliance on cognitive control. *Psychon. Bull. Rev.* 22, 637–652. doi: 10.3758/s13423-014-0712-4
- Specht, K., Willmes, K., Shah, N. J., and Jäncke, L. (2003). Assessment of reliability in functional imaging studies. *J. Magn. Reson. Imag.* 17, 463–471. doi: 10.1002/jmri.10277
- Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., and Wise, R. J. (2006). Converging language streams in the human temporal lobe. *J. Neurosci.* 26, 7328–7336. doi: 10.1523/JNEUROSCI.0559-06.2006
- Steinbeis, N., and Koelsch, S. (2008a). Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. *PLoS ONE* 3:e2226. doi: 10.1371/journal.pone.0002226
- Steinbeis, N., and Koelsch, S. (2008b). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cereb. Cortex* 18, 1169–1178. doi: 10.1093/cercor/bhm149
- Steinke, W. R., Cuddy, L. L., and Holden, R. R. (1997). Dissociation of musical tonality and pitch memory from nonmusical cognitive abilities. *Can. J. Exp. Psychol.* 51:316. doi: 10.1037/1196-1961.51.4.316
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651
- Tervaniemi, M. (2001). Musical sound processing in the human brain. Evidence from electric and magnetic recordings. *Ann. N.Y. Acad. Sci.* 930, 259–272. doi: 10.1111/j.1749-6632.2001.tb05737.x
- Tervaniemi, M., and Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Res. Brain Res. Rev.* 43, 231–246. doi: 10.1016/j.brainresrev.2003.08.004
- Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R., and Näätänen, R. (1999). Functional specialization of the human auditory cortex in processing

- phonetic and musical sounds: a magnetoencephalographic (MEG) study. *Neuroimage* 9, 330–336. doi: 10.1006/nimg.1999.0405
- Tervaniemi, M., Mendvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., Van Zuijen, T. L., et al. (2000). Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Hum. Brain Mapp.* 10, 74–79. doi: 10.1002/(SICI)1097-0193(200006)10:2<74::AID-HBM30>3.0.CO;2-2
- Tillmann, B. (2012). Music and language perception: expectations, structural integration, and cognitive sequencing. *Top. Cogn. Sci.* 4, 568–584. doi: 10.1111/j.1756-8765.2012.01209.x
- Tillmann, B., Peretz, I., Bigand, E., and Gosselin, N. (2007). Harmonic priming in an amusic patient: the power of implicit tasks. *Cogn. Neuropsychol.* 24, 603–622. doi: 10.1080/02643290701609527
- Turkeltaub, P. E., and Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain Lang.* 114, 1–15. doi: 10.1016/j.bandl.2010.03.008
- Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., and Fox, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Hum. Brain Mapp.* 33, 1–13. doi: 10.1002/hbm.21186
- Tzortzis, C., Goldblum, M. C., Dang, M., Forette, F., and Boller, F. (2000). Absence of amusia and preserved naming of musical instruments in an aphasic composer. *Cortex* 36, 227–242. doi: 10.1016/S0010-9452(08)70526-4
- Vaden, K. I. Jr., Kuchinsky, S. E., Cute, S. L., Ahlstrom, J. B., Dubno, J. R., and Eckert, M. A. (2013). The cingulo-opercular network provides word-recognition benefit. *J. Neurosci.* 33, 18979–18986. doi: 10.1523/JNEUROSCI.1417-13.2013
- Von Kriegstein, K., Eiger, E., Kleinschmidt, A., and Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cogn. Brain Res.* 17, 48–55. doi: 10.1016/S0926-6410(03)00079-X
- White, T., O'Leary, D., Magnotta, V., Arndt, S., Flaum, M., and Andreasen, N. C. (2001). Anatomic and functional variability: the effects of filter size in group fMRI data analysis. *Neuroimage* 13, 577–588. doi: 10.1006/nimg.2000.0716
- Wilson, S. M., DeMarco, A. T., Henry, M. L., Gesierich, B., Babiak, M., Mandelli, M. L., et al. (2014). What role does the anterior temporal lobe play in sentence-level processing? Neural correlates of syntactic processing in semantic variant primary progressive aphasia. *J. Cogn. Neurosci.* 26, 970–985. doi: 10.1162/jocn_a_00550
- Wong, C., and Gallate, J. (2012). The function of the anterior temporal lobe: a review of the empirical evidence. *Brain Res.* 1449, 94–116. doi: 10.1016/j.brainres.2012.02.017
- Xu, J., Kemeny, S., Park, G., Frattali, C., and Braun, A. (2005). Language in context: emergent features of word, sentence, and narrative comprehension. *Neuroimage* 25, 1002–1015. doi: 10.1016/j.neuroimage.2004.12.013
- Yamadori, A., Osumi, Y., Masuhara, S., and Okubo, M. (1977). Preservation of singing in Broca's aphasia. *J. Neurol. Neurosurg. Psychiatry* 40, 221–224. doi: 10.1136/jnnp.40.3.221
- 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
- Zatorre, R. J., and Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1087–1104. doi: 10.1098/rstb.2007.2161
- Zheng, Z. Z., Munhall, K. G., and Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *J. Cogn. Neurosci.* 22, 1770–1781. doi: 10.1162/jocn.2009.21324

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 © 2015 LaCroix, Diaz and Rogalsky. 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.



Studying Musical and Linguistic Prediction in Comparable Ways: The Melodic Cloze Probability Method

Allison R. Fogel^{1*}, Jason C. Rosenberg², Frank M. Lehman³, Gina R. Kuperberg^{1,4,5} and Aniruddh D. Patel¹

¹ Department of Psychology, Tufts University, Medford, MA, USA, ² Department of Arts and Humanities, Yale-NUS College, Singapore, Singapore, ³ Department of Music, Tufts University, Medford, MA, USA, ⁴ MGH/HST Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA, USA, ⁵ Department of Psychiatry, Massachusetts General Hospital, Charlestown, MA, USA

OPEN ACCESS

Edited by:

McNeel Gordon Jantzen,
Western Washington University, USA

Reviewed by:

Elizabeth Hellmuth Margulis,
University of Arkansas, USA
E. Glenn Schellenberg,
University of Toronto, Canada

*Correspondence:

Allison R. Fogel
allison.fogel@tufts.edu

Specialty section:

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

Received: 01 June 2015

Accepted: 26 October 2015

Published: 12 November 2015

Citation:

Fogel AR, Rosenberg JC,
Lehman FM, Kuperberg GR
and Patel AD (2015) Studying Musical
and Linguistic Prediction
in Comparable Ways: The Melodic
Cloze Probability Method.
Front. Psychol. 6:1718.
doi: 10.3389/fpsyg.2015.01718

Prediction or expectancy is thought to play an important role in both music and language processing. However, prediction is currently studied independently in the two domains, limiting research on relations between predictive mechanisms in music and language. One limitation is a difference in how expectancy is quantified. In language, expectancy is typically measured using the cloze probability task, in which listeners are asked to complete a sentence fragment with the first word that comes to mind. In contrast, previous production-based studies of melodic expectancy have asked participants to sing continuations following only one to two notes. We have developed a melodic cloze probability task in which listeners are presented with the beginning of a novel tonal melody (5–9 notes) and are asked to sing the note they expect to come next. Half of the melodies had an underlying harmonic structure designed to constrain expectations for the next note, based on an implied authentic cadence (AC) within the melody. Each such ‘authentic cadence’ melody was matched to a ‘non-cadential’ (NC) melody matched in terms of length, rhythm and melodic contour, but differing in implied harmonic structure. Participants showed much greater consistency in the notes sung following AC vs. NC melodies on average. However, significant variation in degree of consistency was observed within both AC and NC melodies. Analysis of individual melodies suggests that pitch prediction in tonal melodies depends on the interplay of local factors just prior to the target note (e.g., local pitch interval patterns) and larger-scale structural relationships (e.g., melodic patterns and implied harmonic structure). We illustrate how the melodic cloze method can be used to test a computational model of melodic expectation. Future uses for the method include exploring the interplay of different factors shaping melodic expectation, and designing experiments that compare the cognitive mechanisms of prediction in music and language.

Keywords: music, language, prediction, music cognition, melodic expectation, cloze probability

INTRODUCTION

Recent years have seen growing interest in cognitive and neural relations between music and language. Although there are clear differences between the two—for example, language can convey specific semantic concepts and propositions in a way that instrumental music cannot (Slevc and Patel, 2011)—they share several features. For example, both language and music involve the

generation and comprehension of complex, hierarchically structured sequences made from discrete elements combined in principled ways (Patel, 2003; Koelsch et al., 2013), and both rely heavily on implicit learning during development (Tillmann et al., 2000).

While neuropsychology has provided clear cases of selective deficits in linguistic or musical processing following brain damage (e.g., Peretz, 1993), several neuroimaging studies of healthy individuals suggest overlap in the brain mechanisms involved in processing linguistic and musical structure. One early demonstration of this overlap came from event-related potential (ERP) research, which revealed that a component known as the P600 is observed in response to syntactically challenging or anomalous events in both domains (Patel et al., 1998). Later research using MEG and fMRI provided further suggestions of neural overlap in structural processing, e.g., by implicating Broca's region in the processing of tonal-harmonic structure (e.g., Maess et al., 2001; Tillmann et al., 2003; LaCroix et al., 2015; Musso et al., 2015; though see Fedorenko et al., 2012). To resolve the apparent contradiction between evidence from neuropsychology and neuroimaging, Patel (2003) proposed the shared syntactic integration resource hypothesis (SSIRH). The SSIRH posits a distinction between domain-specific representations in long-term memory (e.g., stored knowledge of words and their syntactic features, and of chords and their harmonic features), which can be separately damaged, and shared neural resources which act upon these representations as part of structural processing. This "dual-system" model proposes that syntactic integration of incoming elements in language and music involves the interaction (via long-distance neural connections) of shared "resource networks" and domain-specific "representation networks" (see Patel, 2013 for a detailed discussion, including relations between the SSIRH and Hagoort's (2005) "memory, unification, and control" model of language processing).

The SSIRH predicted that simultaneous demands on linguistic and musical structural integration should produce interference. This prediction has been supported by behavioral and neural research (for a review, see Kunert and Slevc, 2015). For example, behavioral studies by Fedorenko et al. (2009) and Slevc et al. (2009) have shown that it is particularly difficult for participants to process complex syntactic structures in both language and music simultaneously (see also Hoch et al., 2011; Carrus et al., 2013; though cf. Perruchet and Poulin-Charronnat, 2013). Additionally, Koelsch et al. (2005) conducted an ERP study that observed an interaction between structural processing in language and music, as reflected by effects of music processing on the left anterior negativity (LAN, associated with processing syntax in language) and effects of language processing on the early right anterior negativity (ERAN, associated with processing musical syntax).

In addition to structural integration, it has been suggested that prediction may be another process that operates similarly in language and music (Koelsch, 2012a,b; Patel, 2012). Prediction is increasingly thought to be a fundamental aspect of human cognition (Clark, 2013), and is a growing topic of research in psycholinguistics (Van Petten and Luka, 2012; see Kuperberg

and Jaeger, in press for a recent review). It has become clear that we regularly use context to predict upcoming words when comprehending language (Tanenhaus et al., 1995; Altmann and Kamide, 1999; Wicha et al., 2004; DeLong et al., 2005). This has been demonstrated using ERPs, a brain measure with millisecond-level temporal resolution that allows one to study cognitive processing during language comprehension. Recent evidence from ERP research has suggested that prediction in language processing occurs at multiple distinguishable levels (e.g., syntactic, semantic, phonological) (Pickering and Garrod, 2007; Kuperberg and Jaeger, in press).

Strong lexical predictions for a specific word occur when multiple types of information within a linguistic context constrain strongly for the semantic features, the syntactic properties, and the phonological form of a specific word. For example, the sentence "*The piano is out of ____*" leads to a strong expectation for the word "tune", so one can refer to this as a high lexical constraint sentence. It is well established that unexpected words following these contexts evoke a larger N400 ERP component (occurring 300–500 ms after the presentation of the final word) than expected words (Kutas and Hillyard, 1980; Kutas and Hillyard, 1984; Kutas and Federmeier, 2011). Such unexpected words do not necessarily need to be anomalous to produce an N400: predictions can also be violated with words that are perfectly coherent and non-anomalous. For example, if the final word delivered in the above sentence is "place" (i.e., "*The piano is out of place*") this word still violates a lexical prediction for the highly expected word "tune." As in the previous example, the N400 elicited by "place" would be larger than that elicited by "tune," as it is less expected. Moreover, in recent ERP research, violations of specific lexical predictions with other plausible words have also been observed to elicit a late anteriorly distributed positive component. This late frontal positivity has been observed at various time points after the N400, often peaking around 500–900 ms after the presentation of a critical item (Federmeier et al., 2007; Van Petten and Luka, 2012). Importantly, unlike the N400, the late frontal positivity is not produced by words that follow non-constraining contexts, when comprehenders have no strong prediction for a particular word (e.g., "place" following the context, "After a while, the boy saw the...").

Predictions in language are not always at the level of specific lexical items: they can also be generated at the level of semantic-syntactic statistical contingencies that determine the structure of an event ('who does what to whom') (Kuperberg, 2013). For example, at a certain point in a sentence we might expect a certain syntactic category of word, like a noun-phrase, with certain coarse conceptual features, such as animacy. For example, in the sentence "*Mary went outside to talk to the ____*" there is no strong indication of which word will come next, but it is clear that it must be an animate noun-phrase (Mary would likely not talk to an inanimate object like a truck). Violations of these semantic-syntactic structural predictions have been observed to elicit a different neural response from the anterior positivity discussed above, namely the P600 (a late posterior positivity, peaking from around 600 ms after onset of the violating word; see Kuperberg, 2007 for a review). This provides evidence that distinct neural

signatures may be associated with violations of strong predictions at different representational levels (e.g., a late anterior positivity evoked by violations of strong lexical predictions, Federmeier et al., 2007; a late posterior positivity evoked by violations of strong semantic-syntactic predictions, Kuperberg, 2007; see Kuperberg, 2013 for discussion). The functional significance of these late positivities (both frontal and posterior) evoked by strong prediction violations remains unclear. One possibility, however, is that they reflect the neural consequences of suppressing the predicted (but not presented) information and adapting one's internal representation of context in order to generate more accurate predictions in the future (e.g., see Kuperberg, 2013; Kuperberg and Jaeger, in press, for discussion).

Turning to music, expectation has long been a major theme of music cognition research. Meyer (1956) first suggested a strong connection between the thwarting of musical expectations and the arousal of emotion in listeners. In recent years, theories of musical expectation have been brought into a modern cognitive science framework (e.g., Margulis, 2005; Huron, 2006; Huron and Margulis, 2010; Pearce et al., 2010), and expectation has been studied empirically with both behavioral and neural methods (e.g., Steinbeis et al., 2006). It is increasingly recognized that multiple sub-processes are involved in musical expectation (see Huron, 2006, for one theoretical treatment). Empirical research has shown that predictions are generated for multiple aspects of music, such as harmony, rhythm, timbre, and meter (Rohrmeier and Koelsch, 2012). Such expectations are thought to be automatically generated by enculturated listeners (Koelsch et al., 2000; Koelsch, 2012a).

Here, we focus on melodic prediction, and specifically on expectations for upcoming notes in monophonic (single-voice) melodies based on implicit knowledge of the melodic and harmonic structures of Western tonal music (Tillmann et al., 2000). For those interested in relations between predictive mechanisms in music and language, melodic expectancy provides an interesting analog to linguistic expectancy in sentence processing. Like sentences, monophonic melodies consist of a single series of events created by combining perceptually discrete elements in principled ways to create hierarchically structured sequences (Jackendoff and Lerdahl, 2006). Sentences and melodies have regularities at multiple levels, including local relations between neighboring elements and larger-scale patterns, e.g., due to underlying linguistic-grammatical or tonal structure.

In order to study relations between the cognitive mechanisms of prediction in sentences and melodies, it is necessary to measure prediction in these two types of sequences in comparable ways. In sentence processing, lexical expectancy has typically been measured using the cloze probability task, in which participants are asked to complete a sentence fragment with the first word that comes to mind (Taylor, 1953). For a given context, the percentage of participants providing a given continuation is taken as the “cloze probability” of that response. The cloze probability of an item is therefore a straightforward measure of how expected or probable it is. In addition to measuring the cloze probability of a particular word in relation to its context, it is also possible to use the cloze task to measure the ‘lexical constraint’ of a particular context by calculating the proportion

of participants who produce a given word (see Federmeier et al., 2007). For example, a sentence such as “*The day was breezy so the boy went outside to fly a . . .*” would likely elicit the highly expected continuation “kite” from most participants, and thus be a ‘strongly lexically constraining’ context. In contrast, a sentence such as “*Carol always wished that she'd had a . . .*” would elicit a more varied set of responses, and thus be a ‘weakly lexically constraining’ context.

While expectancy in music has been measured in various ways over the years, to date there has been nothing comparable to the standard cloze probability method in language, i.e., a production-based task in which a person is presented with the beginning of a short coherent sequence and then asked to produce the event she thinks comes next.¹ Most behavioral studies of expectancy in music have used perceptual paradigms, such as harmonic priming paradigms or ratings of how well a tone continues an initial melodic fragment. Harmonic priming paradigms consist of a prime context followed by a target event, in which the degree of tonal relatedness between the two is manipulated. Typically, harmonically related targets are processed faster and more accurately than unrelated targets (Tillmann et al., 2014). These studies have shown that chords that are more harmonically related to the preceding context are easier to process, while there is a cost of processing chords that are less related or unrelated to the context (Tillmann et al., 2003). Another genre of priming studies has shown that timbre identification is improved when a pitch is close in frequency to the preceding pitch and harmonically congruent with the preceding context (Margulis and Levine, 2006). In studies using explicit ratings of expectancy, listeners are asked to rate how well a target note continues a melodic opening, e.g., on a scale of 1 (very bad continuation) to 7 (very good continuation) (e.g., Schellenberg, 1996). More recently, a betting paradigm has been used in which participants place bets on a set of possible continuations for a musical passage, and bets can be distributed across multiple possible outcomes (Huron, 2006). The betting paradigm has the advantage of providing a measure of the *strength* of an expectation for a specific item. However, like the “continuation rating” task, this task requires *post hoc* judgments, and is therefore not an online measure of participants’ real-time expectations. ERPs and measures of neural oscillatory activity can provide online measures of expectation in musical sequences (e.g., Pearce et al., 2010; Fujioka et al., 2012), but such studies have focused on perception, not production.

A handful of studies have used production tasks to measure musical expectancy, but they differ in important ways from the standard linguistic cloze probability task. Some studies have used extremely short contexts, in which participants are asked to sing a continuation after hearing only a single two-note interval, or even a single note (Carlsen, 1981; Unyk and Carlsen, 1987; Povel, 1996; Thompson et al., 1997; Schellenberg et al., 2002). Lake (1987) presented two-note intervals after establishing a tonal context

¹Waters et al. (1998) used what they refer to as a “musical ‘cloze’ task,” but theirs was a multiple-choice task where participants selected one of several pre-composed sections of musical notation.

consisting of major chords and a musical scale. However, no prior singing-based study of melodic expectation has used coherent melodies as the context (some studies using piano performance have used very long contexts, in which pianists have been asked to improvise extended continuations for entire piano passages, Schmuckler, 1989, 1990). Also, in all of these studies (and unlike in the linguistic cloze probability task), participants were asked to produce continuations of whatever length they chose in response to brief stimuli. The closest analog to a musical cloze task comes from a study of implicit memory for melody, in which listeners first heard a set of novel tonal melodies and then heard melodic stems of several notes and were asked to “sing the note that they thought would come next musically” (Warker and Halpern, 2005). However, the structure of the melodic stems was not manipulated, and the focus of the study was on implicit memory, not on expectation.

In order to advance the comparative study of prediction in language and music, it is necessary to develop comparable methods for studying prediction in the two domains. To this end, we have developed a melodic cloze probability task. In this task, participants are played short melodic openings drawn from novel coherent tonal melodies, and are asked to sing a single-note continuation. In an attempt to manipulate the predictive constraint of the melodies, the underlying harmonic structure of each opening (henceforth, ‘melodic stem’) was designed to either lead to a strong expectancy for a particular note, or not (see Materials and Methods for details). For each melodic stem, the cloze probability of a given note is calculated as the percentage of participants producing that note. The predictive constraint of a melodic stem is determined by examining the degree of agreement between participants’ responses. For example, if all participants sing the same note after a particular stem, the stem has 100% constraint. On the other hand, if the most commonly sung note is produced by 40% of the participants, then the stem has 40% constraint.

The melodic cloze probability method allows the cloze probabilities of notes to be quantitatively measured, and thus provides a novel way to study how different structural factors (e.g., local melodic interval patterns vs. larger-scale harmonic structure) interact in shaping melodic expectation. As demonstrated below, the method can also be used to test quantitative models of melodic expectation, such as Narmour’s (1990) “Implication-Realization” model, using naturalistic musical materials. In the future, the method can facilitate the design of studies comparing predictive mechanisms in language and music, e.g., by systematically manipulating constraint and cloze probabilities across linguistic and musical stimuli in behavioral or ERP studies of expectancy (cf. Tillmann and Bigand, 2015).

MATERIALS AND METHODS

Participants

Fifty participants (29 female, 21 male, age range 18–25 years, mean age 20.3 years) took part in the experiment and were included in the data analysis (eight further participants were

excluded due to difficulties with singing on pitch; see “Data Analysis”). All participants were self-identified musicians with no hearing impairment who had a minimum of 5 years of musical experience within the past 10 years (playing an instrument, singing, or musical training); 22 (44%) reported “voice” as one of their instruments. Participants had received a mean of 9.0 years of formal musical training on Western musical instruments ($SD = 4.8$) and reported no significant exposure to non-Western music. Participants were compensated for their participation and provided informed consent in accordance with the procedures of the Institutional Review Board of Tufts University.

Materials

The stimuli consisted of 45 pairs of short novel tonal melodies created by the second author (JCR), a professional composer. Stimuli were truncated in the middle, creating “melodic stems.” The melodies ranged across all 12 major keys and employed variety of meters (3/4, 4/4, and 6/8 time signatures). Each stem was 5–9 notes long ($M = 8.38$ notes, $SD = 0.83$), and was played at a tempo of 120 beats per minute (bpm). Note durations varied from eighth notes (250 ms) to half notes (1000 ms). Stems contained no rests, articulation indications, dynamic variability, or non-diatonic pitches. All stimuli were created using Finale software with sampled grand piano sounds. Across all melodies, the highest and lowest pitch were A5 (880.0 Hz) and D3 (146.8 Hz), respectively, and the mean pitch was near E4 (329.6 Hz). On average, stems had a pitch range of 11.4 semitones (distance between the highest and lowest pitch in the stem, $SD = 3.2$ st). Male participants heard the melodic stems transposed down one octave. The average stem duration was 5.02 s ($SD = 1.23$).

Each stimulus pair consisted of two stems in the same musical key: one was an “authentic cadence” version, which was designed to create a strong expectation for a particular note, and the other was a “non-cadence” (NC) version, which was designed to *not* generate a strong expectation for a particular note. AC stems ended preceding a strong beat within the meter on the 2nd, 5th, or 7th scale degree and with an implied AC that would typically be expected to resolve to a tonic function. NC stems ended with an implied IV, iv, or ii harmony, with the last presented note never on the 2nd or 7th scale degree and rarely on the 5th. The two stems in each pair were identical in length, rhythm, and melodic contour; they differed only in the pitch of some of their notes, which influenced their underlying harmonic structure (see **Figure 1** for an example). On average, the two stems of an AC-NC melodic pair differed in 48.3% of their notes ($SD = 28.5\%$). When notes of an AC-NC pair differed, they remained close in overall pitch height, on average 1.90 semitones apart ($SD = 0.38$).

The extent to which the two groups of stems projected a sense of key was compared using the Krumhansl-Schmuckler key-finding algorithm (Krumhansl, 1990). This model is based on “key-profiles” of each potential key, which represent the stability of each pitch in the key, i.e., how well it fits in a tonal context (Krumhansl and Kessler, 1982). The pitch distribution of a given melody, weighted by duration, is compared to the key-profile of each key, and a correlation value is calculated. When



FIGURE 1 | (A) Authentic cadence (AC) and **(B)** Non-cadence (NC) versions of one melodic pair (see text for explanation). The figure shows the AC and NC stems in Western music notation. Shown beneath each stem is a possible interpretation of the underlying implied harmonic progression (e.g., D, G, D, A chords in the AC stem), and harmonic functions (I, or tonic chord; IV, or subdominant chord; V, or dominant chord; vi or submediant chord). The stems of a pair are identical in length, key, rhythm, and melodic contour, and each consists of a single stream of notes with no accompaniment. In this pair the stems differ only in the identity of the final two notes, which are slightly lower in the second stem. Crucially, this small physical change alters the underlying harmonic progression.

correlations with the profiles of each potential key were calculated for each stem, the mean correlation with the correct key for AC stems [$r(22) = 0.70$] did not differ significantly from the mean correlation with the correct key of NC stems [$r(22) = 0.73$], $t(44) = 1.24$, $p = 0.22$ (averaging and statistics were performed on Fisher transformed correlation coefficients). The two groups of stems therefore did not differ in the degree to which they projected a sense of key.

Procedure

Stimuli were played to participants over Logitech Z200 computer speakers at a comfortable listening volume within a sound attenuated room. The experiment was presented using PsychoPy (v1.79.01) on a MacBook Pro laptop, and sung responses were recorded as .wav files using the computer's built-in microphone.

Each participant was instructed that s/he would hear the beginnings of some unfamiliar melodies and would need to “sing the note you think comes next.” Participants were asked to *continue* the melody—not necessarily complete it—on the syllable “la.” Each trial began when the participant pressed a button to hear a melodic stem. Immediately after the end of the last note of each stem, the word “Sing” appeared on the screen and participants were given 5 s to sing the continuation, after which they rated their confidence in their response on a 7-point Likert scale (1 = *low*, 7 = *high*).

Each participant was presented with 24 AC and 24 NC melodic stems (only one version from each AC-NC pair) in one of eight randomized presentation orders. (Three pairs were removed from analysis due to differences in the melodic contours of the two stems, hence data from 45 pairs was analyzed.) At the beginning of the experiment, each participant completed a pitch-matching task in which they heard and were asked to sing back a series of individual tones (F4, A4, B3, G#4, A#3, D4, C#4, and E^b4 [corresponding to 349.2, 440.0, 246.9, 415.3, 233.1, 293.7, 277.2, 311.1 Hz, respectively]; one octave lower for male participants). This

was used to evaluate participants' singing accuracy. Before the experimental trials began, participants were familiarized with the experimental procedure with a block of practice items, which ranged from simple scales and familiar melodies to unfamiliar melodies.

Data Analysis

We extracted the mean fundamental frequency of the sung note using Praat (Boersma, 2002). The pitch of the sung note was determined by rounding the measured mean fundamental frequency to the closest semitone in the Western chromatic scale (e.g., A4 = 440 Hz), with the deviation from the frequency of this chromatic scale tone recorded (in cents, i.e., in hundredths of a semitone). The sung response was also represented in terms of its scale degree within the key of the stem in question. Responses were generalized across octaves for the purpose of this study. Participants' responses to the pitch-matching portion of the experiment were also analyzed; if any participant's pitch-matching responses did not round to the same note that was presented, or if their responses to at least 25% of the experimental trials were more than 40 cents away from the nearest semitone, the participant's responses were excluded from further analysis (eight participants were omitted for these reasons). Additionally, reaction times were measured using a sound onset measurement script in Praat (a sound's onset was detected when the sound reached a level −25 dB below its maximum intensity for a minimum of 50 ms) to determine how quickly the continuation was sung after the offset of the last note of the stem.

RESULTS

Participants found the task intuitive and uncomplicated, suggesting that the melodic cloze probability task provides a naturalistic way to measure melodic expectations. On average, participants sang a continuation note with a reaction time of 899 ms ($SD = 604$ ms), and their sung notes were an average of

1896 ms long ($SD = 808$ ms). Given that the melodies had a tempo of 120 BPM, this corresponds to an average time interval of 1.80 beats after the offset of the stem, and a sung note duration of 3.79 beats.

Constraint

The primary dependent variable in our study was the predictive constraint of a melodic stem, as measured by the percentage of participants that sang the most common note after the stem. **Figure 2** illustrates how this was computed, based on the AC-NC melodic pair in **Figure 1**. **Figures 2A,B** show the distributions of sung notes after the AC and NC stems in **Figure 1**, respectively. **Figure 2A** shows that 92% of participants that heard the AC stem produced the most commonly sung note (the tonic, D), while **Figure 2B** shows that no more than 24% of participants that heard the NC stem produced any one note (in this case, there was a tie between C# and A, but in most cases, one pitch class was most common). Thus the constraint of this melodic pair was 92% (or 0.92) for the AC melody and 24% (or 0.24) for the NC melody. For this pair, the AC melody was indeed far more constraining than the NC melody, as predicted.

For each AC and NC stem, we computed the constraint as described above. After AC stems, the average constraint was 69% (i.e., on average, 69% of participants sang the same note after hearing an AC stem), while after NC stems, the average constraint was 42% (i.e., on average, only 42% of participants sang the same note after hearing an NC stem). Thus on average, melodic stems in the AC condition did prove to be more constraining than NC stems (AC $M = 0.692$, $SD = 0.171$; NC $M = 0.415$, $SD = 0.153$), [$t(44) = 7.79$, $p < 0.001$]. This pattern of higher constraint for the AC vs. NC stem was observed in 38 of the 45 item pairs (**Figure 3**).

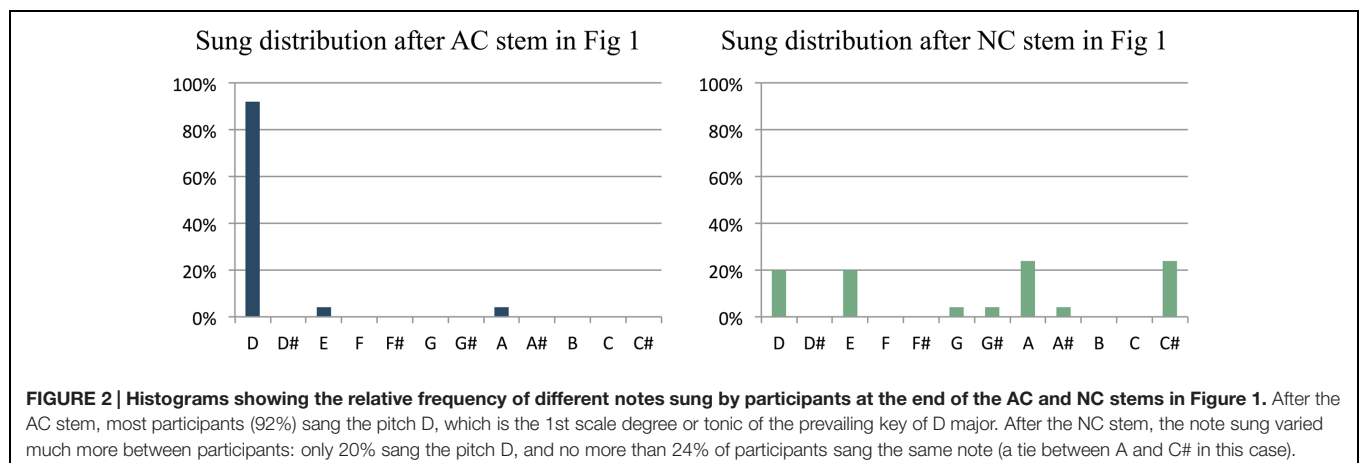
On average, participants responded significantly more quickly after AC stems (mean $RT = 767$ ms, $SD = 265$ ms) than after NC stems (mean $RT = 1033$ ms, $SD = 302$ ms), $t(49) = 9.78$, $p < 0.001$. Additionally, on average participants were significantly more confident in their responses to AC stems ($M = 5.14$, $SD = 0.95$) than to NC stems ($M = 4.36$, $SD = 1.04$), $t(49) = 9.60$, $p < 0.001$.

Scale Degree

When responses were represented in terms of their scale degree in the key of the stem in question, and compiled across all items in each condition, the distributions for AC and NC items were strikingly different. For six of the seven diatonic scale degrees, the frequency of response differed significantly between AC and NC items based on t -tests of each scale degree with a Bonferroni correction applied (see **Figure 4** for p -values). For AC items, responses were heavily weighted around the first note of the scale, or tonic (known as 'do' in solfege). For NC items, responses were more widely distributed; however, they were mainly restricted to in-key diatonic scale degrees.

Variability

While AC stems were on average significantly more constraining than their matched NC stems, there was considerable variability across AC-NC pairs in the degree of difference in constraint between members of a pair (see **Figure 3**). Thirty-eight out of 45 pairs demonstrated the expected pattern, with the AC stem proving more constraining than the NC stem. For instance, the stem pair in **Figure 5A** has a highly constraining AC stem, with 92% of participants singing the same note, the melody's tonic pitch, C (in **Figures 5** and **6**, the most commonly sung note is shown as a red note head after the end of each stem). Why might this be? This stem is short, contains only one rhythmic value, and has very clear harmonic implications, beginning with an unambiguously arpeggiated tonic triad (C-E-G) and concluding with a similarly outlined complete dominant triad (G-B-D). This stem also ends on the leading tone of B, i.e., the seventh scale degree of the diatonic major scale, which customarily resolves to the tonic scale degree, particularly near the end of a phrase. Further structural factors that may contribute to the high degree of agreement on the final pitch are (1) the melody's consistent downward contour, which seems to close in on middle C, and (2) the fact that the tonic note is heard very close to the end of the phrase, which may make it more likely to be replicated. Turning to the NC stem in **Figure 5B**, it is similar in many respects to the AC stem, yet very different in constraint, with the most commonly sung note (F) being produced by just 24% of participants who heard this stem. What might



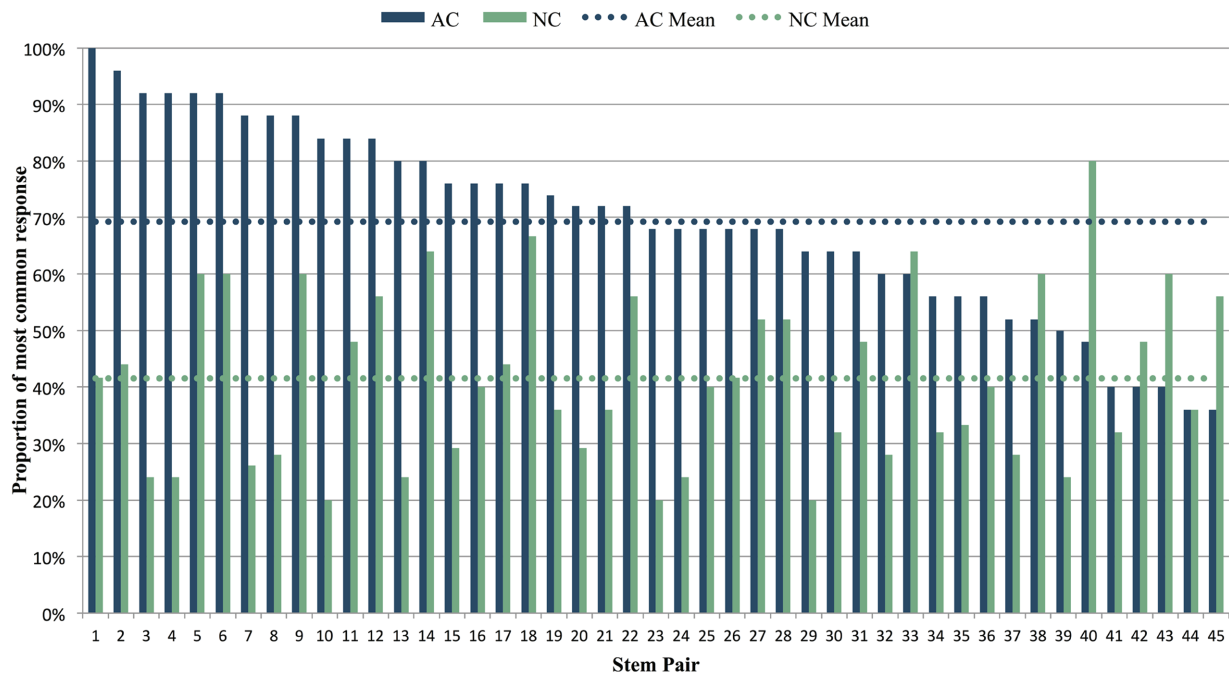


FIGURE 3 | Constraint of AC and NC stems, as calculated by the percentage of participants providing the most common response for each stem. Stem pairs are ranked in order of decreasing constraint for AC stems. The melodic pair shown in **Figure 1** corresponds to stem pair 3 in this graph. Dotted horizontal lines show the mean constraint across all AC and NC stems.

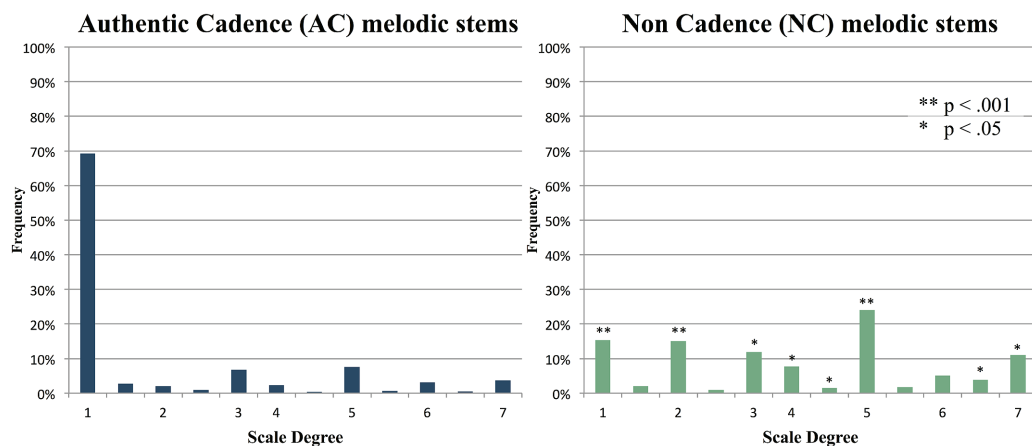


FIGURE 4 | Average of all response distributions to AC and NC stems, shown as scale degrees. Numbers represent diatonic (major) scale degrees (e.g., 1 = tonic, 7 = leading tone, etc.), with asterisks indicating scale degrees with significantly different frequencies between the two conditions.

account for this? The NC stem does not have any resolution-demanding dominant pitches at its conclusion, and as a result lacks a clear sense of harmonic direction. Instead, the melody follows a downward pattern of melodic thirds (E-G, C-E, A-C) whose continuation is ambiguous. The most commonly chosen completion of F could be explained as the next logical pitch in the chain of descending thirds, after A-C. Thus, when faced with a stem where harmonic direction is underdetermined, subjects may have recruited an alternative strategy of melodic pattern continuation.

Another example of an AC stem that proved to be highly constraining is shown in **Figure 5B** (same melodic pair as in **Figure 1**). As with the melody in **Figure 5A**, the AC stem begins on the tonic note and returns to it as the most expected continuation, with an overall melodic range that emphasizes the octave generated above the first scale degree. The melody's interior arpeggiates two chords, first the tonic (D-F#-A) in measure 1, then the subdominant (G-B-D) in measure 2. The subdominant chord frequently serves a syntactic role of “predominant,” a harmonic function that signals the initiation of



FIGURE 5 | Examples of two melodic stem pairs (A,B) with an AC stem that was much more constraining than the non-cadence (NC) stem. Stems are shown in black and white, and for each stem the most frequently sung note is shown as a red note head at the end of the stem. The pitch class name of this note and the proportion of listeners who sang the note (i.e., the measured melodic constraint of the stem) are printed next to the red note. These two pairs correspond to stem pairs 4 and 3 in **Figure 3** [the stems in panel (B) are the same as in **Figure 1**].

a cadence. This is indeed how measure three is structured, with a heavily implied dominant harmony via scale degrees 2 and 7, and a melodic contour that insures D as a plausible completion due to an implied F#-E-D melodic descent and a unresolved leading tone of C#. The less constraining NC stem in **Figure 5B**, by contrast, ends on the sixth scale degree (the submediant). Unlike the leading tone, this note lacks a strong tendency to resolve in a particular way. It may plausibly serve as part of a stepwise motion to or away from the dominant, or as part of an arpeggiation of a predominant harmony; in either case, it negates the cadential function of the third measure and points to no obvious melodic completion.

Contrasting with these stems, where subjects' responses to stems adhered to the AC/NC designations, there were several items where the constraint of the NC stem unexpectedly *exceeded* that of the AC stem. For example, after the NC stem in **Figure 6A**, 80% of participants sang the same note (F#, the 5th scale degree). In this particular melody, we believe this reflects the tendency for a large melodic interval to be followed by stepwise motion in the opposite direction. This "gap-fill" pattern (Meyer, 1956; Narmour, 1990) likely strongly influenced the continuation most participants chose, which involved singing a note (F#) one step down from the last note of the stem (G#), following a large leap of a sixth to an already contextually unstable note (scale degree six). Additionally, this stem has a strongly implied compound melody, wherein most of the topmost notes form a rising, stepwise pattern of B-C#-D#-E, which leads to an F# if this pattern is continued. Meanwhile, the unexpectedly low constraint of the AC stem in **Figure 6A** was perhaps due to the lack of a strong tendency note (like the leading tone) as its last pitch, and the obscuring of the underlying harmonic implications by the relative rhythmic complexity of the melody. That is, the unpredictable and syncopated rhythm may have reduced the strength of the expectancy for the tonic scale degree (Schmuckler and Boltz, 1994). Similarly, in the stem pair in **Figure 6B**, the most common continuation for the NC stem was a gap-filling motion to fill

the exceptionally wide upward leap of an octave from Bb4–Bb5. Landing on Ab, which 56% of subjects agreed on, helps close that gap with a downward step and continues the melody on the more stable pitch of scale degree 5. This note also has the advantage of mirroring the first note of the melody, thus promoting melodic symmetry. The AC stem of this melodic pair presented no such clearly determined ending. If subjects opted to fill in the large upward octave gap to Ab with a downward step, they would land on the unstable fourth scale degree (Gb). On the other hand, if they were to resolve the melody with a cadence on the tonic note (Db), they would land far from the final note of the stem, going against a general tendency in melodic expectation for pitches that are proximate in frequency to the previous note (see section on modeling below).

Based on the above observations, it is clear that underlying harmonic structure, which was manipulated in the AC vs. NC stems, does not alone determine melodic expectation. Melodic factors that likely contributed to increased constraint in our melodies include (but are not limited to) rhythmic simplicity, gap-fill pattern, compound-line implication, leading-tone resolution, and pattern completion. In this way, stems in which linear, contrapuntal, rhythmic and harmonic parameters were closely coordinated produced reliable agreement on melodic completions, while examples with a conflict or ambiguity between those factors were prone to considerably less consensus.

Musical Experience

Prior research suggests that musical training enhances sensitivity to underlying harmonic structure (Koelsch et al., 2002). Since implicit harmony was used to guide the listeners' expectation for a tonic note after AC stems, we sought to determine if participants with greater degrees of musical training were more likely to sing the tonic after AC stems. Thus across AC stems, we correlated each participant's total years of formal musical training with their frequency of responding with the tonic. (Thus for example, if a participant sang the tonic after half of the AC stems they



FIGURE 6 | Examples of two melodic stem pairs (A,B) with an AC stem that was less constraining than the NC stem. Stems are shown in black and white, and for each stem the most frequently sung note is shown as a red note head at the end of the stem. The pitch class name of this note and the proportion of listeners who sang the note (i.e., the measured melodic constraint of the stem) are printed next to the red note. These two pairs correspond to stem pairs 40 and 45 in Figure 3.

heard, their frequency of responding with the tonic to an AC stem would be 0.5.) When all AC items were included in the analysis, there was no significant correlation with years of formal musical training, $r(48) = 0.035$, $p = 0.812$. However, when we divided AC stems according to the scale degree of their final note, an interesting pattern emerged. On average, after AC stems that ended on the 7th scale degree, participants sang the tonic 81% of the time, and in these melodies, there was a significant correlation between participants' years of formal training and their frequency of responding with the tonic, $r(48) = 0.45$, $p = 0.001$ (see Figure 7). This relationship with musical training was also observed with AC stems that ended on the 5th scale degree, where participants sang the tonic 55% of the time on average, $r(48) = 0.33$, $p = 0.02$. (The relationship was not seen for AC stems that ended on the 2nd scale degree, where participants sang the tonic 57% of the time on average.)

Model Comparison

One potential use of the melodic cloze probability task is to test models of melodic expectation. While different forms of musical expectancy (e.g., melodic, rhythmic, harmonic) have been the subject of many important theoretical and empirical investigations (e.g., Schmuckler, 1989; Narmour, 1990; Schellenberg, 1996; Krumhansl et al., 1999; Large and Jones, 1999; Huron, 2006), melodic expectancy in particular has been a focus for quantitative modeling (e.g., Schellenberg, 1996; Krumhansl et al., 1999; Pearce et al., 2010). While comparison of behavioral and modeling data is not the primary focus of this paper, we present one such comparison to illustrate how melodic cloze data can be used for this purpose. We focus on the simplified version of the implication-realization (I-R) model of melodic expectancy (Narmour, 1990) developed by Schellenberg (1997).

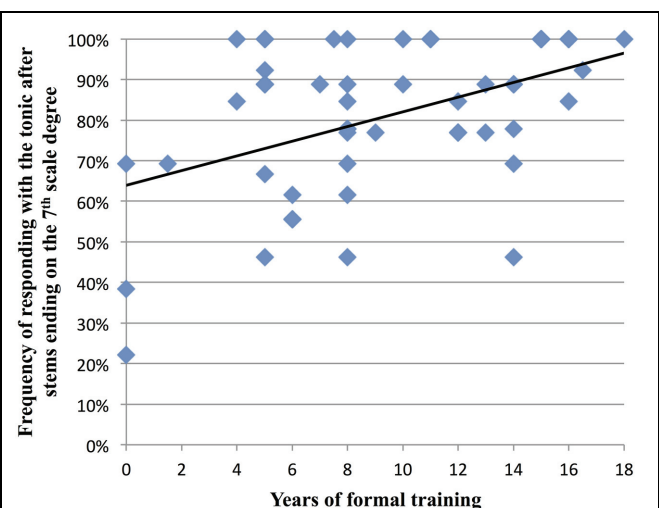


FIGURE 7 | Relationship between participants' years of formal musical training and how often they sang the tonic after melodic stems that ended on the 7th scale degree. On average, participants sang the tonic 81% of the time after these stems (data for all 50 participants are shown: due to some data points lying directly on top of each other, fewer than 50 data points are visible on the graph).

This model computes the probability of each possible continuation of a melody based on two factors. The first of these factors is "pitch proximity," which states that listeners expect the next tone of a melody to be proximate in pitch to the last tone heard. (Another way of stating this is that listeners generally expect melodies to move by small steps.) The second factor is "pitch reversal," which states that after a leap, listeners expect the

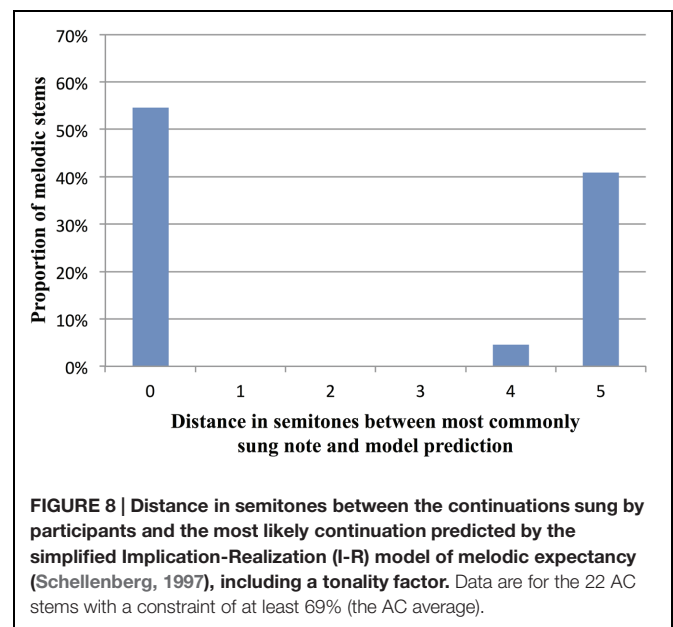
next tone to reverse direction (e.g., after an upward leap, they expect a downward pitch interval), and also expect the upcoming tone to land in a pitch region proximate to the penultimate tone (the first tone of the leap). A third factor relating to tonal stability was also included, based on values from the probe-tone profiles of Krumhansl and Kessler (1982). This factor reflects expectation for notes that fit well into the existing key context, with higher values for structurally more important/stable notes in key. Based on the equations for the simplified I-R model as codified in Schellenberg (1997) these three factors (proximity, reversal, and tonality) were weighted evenly by equalizing their maximum values, and were used to compute expectancies for all notes within two octaves of the final note of each melodic stem, using the MIDI toolbox (Eerola and Toivainen, 2004).

In order to compare the model's predictions to cases where humans had strong expectations, we focused on high-constraint stems where most participants sang the same continuation (stems with constraint >69%, the mean of all AC stems). In the 22 stems satisfying this criterion, the simplified I-R model (including the tonality factor) correctly predicted the note most often sung by participants in 12 stems, i.e., 54.5% of the time. In the remaining 10 of these high-constraint stems (i.e., 45.5% of the time), the model's predictions were an average of 4.9 semitones away from participants' sung note ($SD = 0.32$ st) (see **Figure 8** for the distribution of distances between human data and model predictions). For the 10 stems where the model's predictions differed from the mostly commonly sung note, we checked if the note predicted by the model was the *second*-most-commonly produced note by participants. This was true in only one stem. Overall, the model's performance suggests that our data cannot be accounted for solely by local factors of proximity and reversal, combined with tonality. This suggests that larger-scale factors need to be taken into account, as further discussed below.

DISCUSSION

We introduce the melodic cloze probability task, in which participants hear the opening of a short, novel tonal melody and sing the note they expect to come next. This task, which is modeled on the well-known cloze probability task in psycholinguistics, has not previously been used to study expectancy in the field of music cognition. Participants found the melodic cloze task easy to do, demonstrating that expectancy can be measured in a comparable way across linguistic and musical domains.

Prior work using singing to study melodic expectancy has focused on responses to two-note intervals (see introduction for references). Of these studies, the closest task to ours is Lake (1987), who had participants sing extended continuations in response to a two-note interval preceded by a tonal context. Unlike the current study, the tonal context was not the opening of a novel coherent melody, but a sequence of notes consisting of a major chord, a scale, and another major chord, which served to establish a strong sense of key before the two-note interval. One might ask how our results compare to those of Lake, since one



can conceive of our stimuli as also consisting of a key-inducing context followed by a final two-tone interval (i.e., the final two tones of the melodic stem).

While the last two notes of our stems clearly contribute to our results, our findings cannot be attributed to only hearing this final interval in a generic tonal context. A number of our stems are identical in the scale degrees of their final two notes, yet they elicit very different patterns of results from participants (see **Figure 9** for an example). This different pattern of responding to the same final interval reflects differences in the *structure* of the preceding notes. Thus our paradigm and results are not simply a replication of Lake (1987), and show the relevance of using melodically coherent materials as contexts for production-based studies of melodic expectation. Similarly, we note that our results are not simply a replication of the well-known probe-tone results of Krumhansl and Kessler (1982), since the pattern of responding was not just a reflection of the tonal hierarchy, and depended on the structure of the heard melody (e.g., **Figures 1** and **2**).

In addition to being the first study to obtain cloze probabilities for musical notes, to our knowledge the current study is the also the first to manipulate the predictive *constraint* of musical sequences as part of research on melodic expectation. By using pairs of monophonic melodic openings (or 'stems') matched in length, rhythm, and melodic contour, but differing in implied harmonic structure, we show that underlying harmonic progressions can strongly guide melodic expectations. Specifically, there was significantly more consistency in participants' responses to melodic stems ending on an implied authentic cadence (AC condition) than in their responses to stems ending non-cadentially (NC condition), as reflected by a higher percentage of participants singing the most common continuation for items in the AC condition. In other words, AC stems were more highly constraining than NC stems on average.

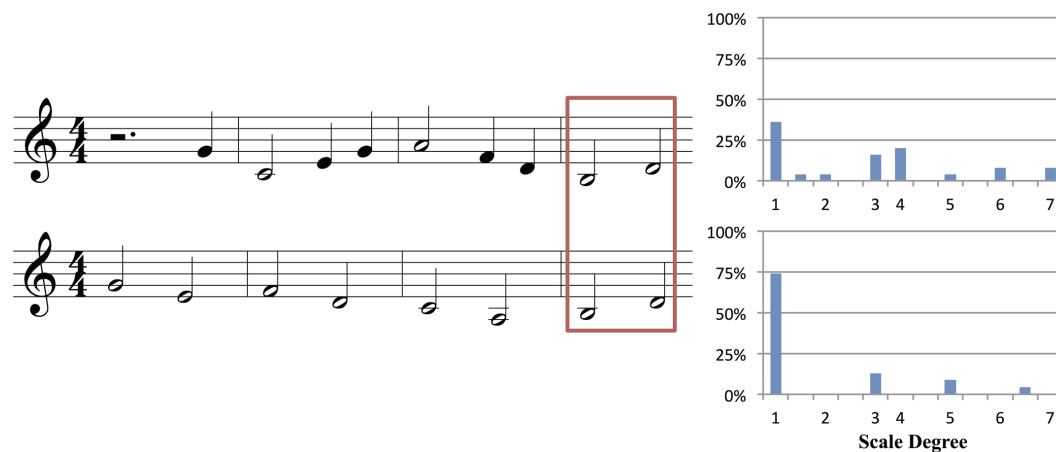


FIGURE 9 | Example of stems that have the same final two notes but elicit different patterns of responses from participants. Stems have been transposed from their original keys to C major in order to facilitate comparison. Both stems end with scale degrees 7 and 2. The distribution of sung responses (expressed as scale degrees) is shown to the right of each stem.

However, our data also clearly indicate that expectations based on larger-scale implied harmony interact with expectations based on melodic structure. That is, despite the fact that the harmonic differences between the AC and NC melodies in each pair were similar, we observed considerable variability in the constraint of melodies. In some pairs, the AC stem was considerably more constraining than the NC stem, but in other pairs the difference in constraint was mild, and in seven pairs the NC stem was actually equal to or more constraining than the AC stem (Figure 3). Analysis of two such ‘reversed constraint’ pairs (Figure 6) suggested that factors related to rhythmic simplicity, gap-fill pattern, compound line implication, and pattern completion may have been involved in overwhelming harmonic expectations. Further investigation of the factors driving the observed large variation in constraint among melodies is clearly warranted. From our results it is clear that expectancies related to melodic patterns (e.g., gap-fill) may sometimes trump those related to tonality.

Indeed, the variability in constraint observed in our data (Figure 3) suggests that the melodic cloze task is well suited for use in future studies aimed at exploring the relative contributions of melodic and harmonic patterns in shaping melodic expectation. Such studies can help test and improve quantitative models of melodic expectation (e.g., Schellenberg, 1996, 1997; Krumhansl et al., 1999; Eerola and Toiviainen, 2004; Margulis, 2005; Pearce, 2005; Pearce and Wiggins, 2006). In the current study, we compared human melodic expectations to predictions based on Schellenberg’s (1997) simplified version of Narmour’s (1990) Implication-Realization (I-R) model of melodic expectation, with an added tonality factor. For the 22 AC melodies with a high degree of measured constraint (i.e., where >69% of participants sang the same note), the model correctly predicted the sung pitch in 54.5% of these melodies. In the remaining 45.5% of these melodies, the model predicted a pitch that was on average 4.9 semitones from the pitch actually sung by participants. This discrepancy

between human expectations and model predictions likely stems from the fact that the simplified I-R model focuses on just the last interval of a melody, and does not take larger-scale structural patterns into account (such as harmonic progressions and recurring motivic patterns). Successful models of melodic expectation will almost certainly need to operate at multiple timescales, reflecting the human tendency to integrate both local and global information in processing melodic sequence structure (Dowling, 2010). In the future, it will be interesting to use the melodic cloze method to test models which are sensitive to patterns at multiple timescales, including Margulis’ (2005) model of melodic expectation, and Pearce’s (2005) IDyOM model (cf. Pearce and Wiggins, 2006). Such models can be tested and improved by comparing their predictions with observed cloze probabilities from human participants.

The musical cloze probability task has further uses in the field of music cognition. For example, this paradigm can be used to investigate how different factors influence melodic expectancy. While we manipulated only the harmonic structure of melodies in the present experiment, the influence of any other factor (e.g., melodic contour, rhythm, dynamics, etc.) on musical expectations could be explored in subsequent studies by composing melodies in pairs and manipulating the one factor while keeping other factors constant. Additionally, the task could be varied to have participants sing multiple-note continuations, as has been done in previous studies (Carlsen, 1981; Lake, 1987; Unyk and Carlsen, 1987; Thompson et al., 1997; Schellenberg et al., 2002). This would allow responses to be examined on longer timescales than just the first sung note. In addition, it would reduce the possibility that participants are responding by *completing* the melodic sequences with the sung note, instead of *continuing* them (as instructed). This is an important issue, as the note sung after the stem may differ depending on whether listeners treat it as a continuation or a completion (Aarden, 2003, cf. Huron, 2006).

Of course, the melodic cloze paradigm does have its limitations. By focusing on what pitch a person sings, it cannot give independent measures of all the different types of expectations which may be at play at a given point in a melody, such as timbral expectations (if listening to complex textures) or rhythmic expectations. To study these sorts of expectations, modifications of the paradigm presented here would be necessary. For example, if studying rhythmic expectations, at the end of each stem one could ask participants to press a bar for as long as they think the next note will last.

The melodic cloze task can also be used to examine musical expectations in different populations. We observed a significant correlation between formal musical training and a tendency to sing the tonic after AC stems that ended on the 7th or 5th scale degrees. It has been suggested that having more musical experience leads to greater sensitivity to harmonic cues, which is consistent with our finding and with neural research on harmonic processing (Koelsch et al., 2002). Future studies could use the melodic cloze method to investigate how different kinds of musical experience might impact expectancy formation. For example, expectations may differ between musicians who have been educated in music theory vs. those who have experience

singing or improvising without reading music. Additionally, the melodic cloze paradigm could be used in studies with children, to investigate how melodic expectations develop (cf. Corrigan and Trainor, 2014).

Obtaining melodic cloze probabilities is crucial for future research comparing predictive processing in music and language, as it allows for the comparison of the effects of violating predictions of comparable strength in the two domains (cf. Tillmann and Bigand, 2015). Previous studies comparing expectancy violations in music and language have typically chosen violations that are intuitively thought to be comparable in the two domains. By using a cloze paradigm to quantify cloze probabilities for possible continuations in both domains, it is possible to compare effects of violations of the same degree, using normed stimuli (cf. Featherstone et al., 2012). For example, this will allow comparison of brain responses to plausible violations of expectations, instead of to frank structural violations (which rarely occur in naturalistic sequences). Also, studies that probe interactions between simultaneously presented music and language expectancy violations can be more precisely calibrated, in order to further elucidate cognitive and neural relations between language and music processing.

REFERENCES

- Aarden, B. (2003). *Dynamic Melodic Expectancy*. Ph.D. thesis, Dissertation, School of Music, The Ohio State University, Columbus, OH.
- Altmann, G. T., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/S0010-0277(99)00059-1
- Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot Int.* 5, 341–345.
- Carlsen, J. C. (1981). Some factors which influence melodic expectancy. *Psychomusicol. J. Res. Music Cogn.* 1, 12–29. doi: 10.1037/h0094276
- Carrus, E., Pearce, M. T., and Bhattacharya, J. (2013). Melodic pitch expectation interacts with neural responses to syntactic but not semantic violations. *Cortex* 49, 2186–2200. doi: 10.1016/j.cortex.2012.08.024
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477
- Corrigan, K. A., and Trainor, L. J. (2014). Enculturation to musical pitch structure in young children: evidence from behavioral and electrophysiological methods. *Dev. Sci.* 17, 142–158. doi: 10.1111/desc.12100
- DeLong, K., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121. doi: 10.1038/nn1504
- Dowling, W. J. (2010). “Music perception,” in *Oxford Handbook of Auditory Science: Hearing*, ed. C. Plack (Oxford: Oxford University Press), 231–248.
- Eerola, T., and Toivianen, P. (2004). *MIDI Toolbox: MATLAB Tools for Music Research*. Jyväskylä: University of Jyväskylä.
- Featherstone, C. R., Waterman, M. G., and Morrison, C. M. (2012). Norming the odd: creation, norming, and validation of a stimulus set for the study of incongruities across music and language. *Behav. Res. Methods* 44, 81–94. doi: 10.3758/s13428-011-0137-1
- Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., and Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Res.* 1146, 75–84. doi: 10.1016/j.brainres.2006.06.101
- Fedorenko, E., McDermott, J. H., Norman-Haignere, S., and Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *J. Neurophysiol.* 108, 3289–3300. doi: 10.1152/jn.00209.2012
- Fedorenko, E., Patel, A. D., Casasanto, D., Winawer, J., and Gibson, E. (2009). Structural integration in language and music: evidence for a shared system. *Mem. Cogn.* 37, 1–9. doi: 10.3758/MC.37.1.1
- Fujioka, T., Ween, J. E., Jamali, S., Stuss, D. T., and Ross, B. (2012). Changes in neuromagnetic beta-band oscillation after music-supported stroke rehabilitation. *Ann. N. Y. Acad. Sci.* 1252, 294–304. doi: 10.1111/j.1749-6632.2011.06436.x
- Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9, 416–423. doi: 10.1016/j.tics.2005.07.004
- Hoch, L., Poulin-Charronnat, B., and Tillmann, B. (2011). The influence of task-irrelevant music on language processing: syntactic and semantic structures. *Front. Psychol.* 2:112. doi: 10.3389/fpsyg.2011.00112
- Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.
- Huron, D., and Margulis, E. H. (2010). “Music, expectancy, and thrills,” in *Handbook of Music and Emotion: Theory, Research, Applications*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 575–604.
- Jackendoff, R., and Lerdahl, F. (2006). The capacity for music: what is it, and what’s special about it? *Cognition* 100, 33–72. doi: 10.1016/j.cognition.2005.11.005
- Koelsch, S. (2012a). *Brain and Music*. Hoboken, NJ: John Wiley & Sons.
- Koelsch, S. (2012b). “Response to target article “Language, music, and the brain: a resource-sharing framework,”” in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohrmeier, J. A. Hawkins, and I. Cross (Oxford: Oxford University Press), 224–234.
- Koelsch, S., Gunter, T., Friederici, A. D., and Schroger, E. (2000). Brain indices of music processing: “Nonmusicians” are musical. *J. Cogn. Neurosci.* 12, 520–541. doi: 10.1162/089892900562183
- 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., Rohrmeier, M., Torrecuso, R., and Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15443–15448. doi: 10.1073/pnas.1300272110
- Koelsch, S., Schmidt, B. H., and Kansok, J. (2002). Effects of musical expertise on the early right anterior negativity: an event-related brain potential study. *Psychophysiology* 39, 657–663. doi: 10.1111/1469-8986.3950657
- Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. New York, NY: Oxford University Press.

- Krumhansl, C. L., and Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychol. Rev.* 89, 334–368. doi: 10.1037/0033-295X.89.4.334
- Krumhansl, C. L., Louhivuori, J., Toivainen, P., Järvinen, T., and Eerola, T. (1999). Melodic expectation in Finnish spiritual folk hymns: convergence of statistical, behavioral, and computational approaches. *Music Percept.* 17, 151–195. doi: 10.2307/40285890
- Kunert, R., and Slevc, L. R. (2015). A commentary on: “Neural overlap in processing music and speech” (Peretz et al., 2015). *Front. Hum. Neurosci.* 9:330. doi: 10.3389/fnhum.2015.00330
- Kuperberg, G. R., and Jaeger, T. F. (in press). What do we mean by prediction in language comprehension. *Lang. Cogn. Neurosci.*
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Res.* 1146, 23–49. doi: 10.1016/j.brainres.2006.12.063
- Kuperberg, G. R. (2013). “The proactive comprehender: what event-related potentials tell us about the dynamics of reading comprehension,” in *Unraveling the Behavioral, Neurobiological, and Genetic Components of Reading Comprehension*, eds B. Miller, L. Cutting, and P. McCardle (Baltimore: Paul Brookes Publishing), 176–192.
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657
- Kutas, M., and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163. doi: 10.1038/307161a0
- LaCroix, A. N., Diaz, A. F., and Rogalsky, C. (2015). The relationship between the neural computations for speech and music perception is context-dependent: an activation likelihood estimate study. *Front. Psychol.* 6:1138. doi: 10.3389/fpsyg.2015.01138
- Lake, W. E. (1987). *Melodic Perception and Cognition: The Influence of Tonality*. Ann Arbor, MI: University of Michigan
- Large, E. W., and Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106, 119–159. doi: 10.1037/0033-295X.106.1.119
- Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001). Musical syntax is processed in Broca’s area: an MEG study. *Nat. Neurosci.* 4, 540–545.
- Margulis, E. H. (2005). A model of melodic expectation. *Music Percept.* 22, 663–714. doi: 10.1525/mp.2005.22.4.663
- Margulis, E. H., and Levine, W. H. (2006). Timbre priming effects and expectation in melody. *J. New Music Res.* 35, 175–182. doi: 10.1080/09298210600835042
- Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago, IL: University of Chicago Press.
- Musso, M., Weiller, C., Horn, A., Glauche, V., Umarova, R., Hennig, J., et al. (2015). A single dual-stream framework for syntactic computations in music and language. *Neuroimage* 117, 267–283
- Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago: University of Chicago Press.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681. doi: 10.1038/nn1082
- Patel, A. D. (2012). “Advancing the comparative study of linguistic and musical syntactic processing,” in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohrmeier, J. Hawkins, and I. Cross (Oxford: Oxford University Press), 248–253.
- Patel, A. D. (2013). “Sharing and nonsharing of brain resources for language and music,” in *Language, Music, and the Brain: A Mysterious Relationship*, ed. M. Arbib (Cambridge, MA: MIT Press), 329–355.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. (1998). Processing syntactic relations in language and music: an event-related potential study. *J. Cogn. Neurosci.* 10, 717–733. doi: 10.1162/089892998563121
- Pearce, M. T. (2005). *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. Ph.D. thesis, Department of Computing, City University, London.
- 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. doi: 10.1525/mp.2006.23.5.377
- Peretz, I. (1993). Auditory atonia for melodies. *Cogn. Neuropsychol.* 10, 21–56. doi: 10.1080/02643299308253455
- Perruchet, P., and Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychon. Bull. Rev.* 20, 310–317. doi: 10.3758/s13423-012-0344-5
- Pickering, M. J., and Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends Cogn. Sci.* 11, 105–110. doi: 10.1016/j.tics.2006.12.002
- Povel, D. J. (1996). Exploring the elementary harmonic forces in the tonal system. *Psychol. Res.* 58, 274–283. doi: 10.1007/BF00447073
- Rohrmeier, M. A., and Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *Int. J. Psychophysiol.* 83, 164–175. doi: 10.1016/j.ijpsycho.2011.12.010
- Schellenberg, E. G. (1996). Expectancy in melody: tests of the implication-realization model. *Cognition* 58, 75–125. doi: 10.1016/0010-0277(95)00665-6
- Schellenberg, E. G. (1997). Simplifying the implication-realization model of melodic expectancy. *Music Percept.* 14, 295–318. doi: 10.2307/40285723
- Schellenberg, E. G., Adachi, M., Purdy, K. T., and McKinnon, M. C. (2002). Expectancy in melody: tests of children and adults. *J. Exp. Psychol. Gen.* 131, 511–537. doi: 10.1037/0096-3445.131.4.511
- Schmuckler, M. A. (1989). Expectation in music: investigation of melodic and harmonic processes. *Music Percept.* 7, 109–149. doi: 10.2307/40285454
- Schmuckler, M. A. (1990). The performance of global expectations. *Psychomusicol. J. Res. Music Cogn.* 9, 122–147. doi: 10.1037/h0094151
- Schmuckler, M. A., and Boltz, M. G. (1994). Harmonic and rhythmic influences on musical expectancy. *Percept. Psychophys.* 56, 313–325. doi: 10.3758/BF03209765
- Slevc, L. R., and Patel, A. D. (2011). Meaning in music and language: three key differences. Comment on “Towards a neural basis of processing musical semantics” by Stefan Koelsch. *Phys. Life Rev.* 8, 110–111.
- Slevc, L. R., Rosenberg, J. C., and Patel, A. D. (2009). Making psycholinguistics musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychon. Bull. Rev.* 16, 374–381. doi: 10.3758/16.2.374
- 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
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634. doi: 10.1126/science.7777863
- Taylor, W. (1953). ‘Cloze’ procedure: a new tool for measuring readability. *J. Q.* 30, 415–433.
- Thompson, W. F., Cuddy, L. L., and Plaus, C. (1997). Expectancies generated by melodic intervals: evaluation of principles of melodic implication in a melody-completion task. *Percept. Psychophys.* 59, 1069–1076. doi: 10.3758/BF03205521
- 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., and Bigand, E. (2015). Response: a commentary on: “Neural overlap in processing music and speech.” *Front. Hum. Neurosci.* 9:491. doi: 10.3389/fnhum.2015.00491
- Tillmann, B., Janata, P., Birk, J., and Bharucha, J. J. (2003). The costs and benefits of tonal centers for chord processing. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 470–482. doi: 10.1037/0096-1523.29.2.470
- Tillmann, B., Poulin-Charronnat, B., and Bigand, E. (2014). The role of expectation in music: from the score to emotions and the brain. *Wiley Interdiscip. Rev. Cogn. Sci.* 5, 105–113. doi: 10.1002/wcs.1262

- Unyk, A. M., and Carlsen, J. C. (1987). The influence of expectancy on melodic perception. *Psychomusicol. J. Res. Music Cogn.* 7, 3–23. doi: 10.1037/h0094189
- Van Petten, C., and Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *Int. J. Psychophysiol.* 83, 176–190. doi: 10.1016/j.ijpsycho.2011.09.015
- Warker, J. A., and Halpern, A. R. (2005). Musical stem completion: humming that note. *Am. J. Psychol.* 118, 567–585.
- Waters, A. J., Townsend, E., and Underwood, G. (1998). Expertise in musical sight reading: a study of pianists. *Br. J. Psychol.* 89, 123–149. doi: 10.1111/j.2044-8295.1998.tb02676.x
- Wicha, N. Y., Moreno, E. M., and Kutas, M. (2004). Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *J. Cogn. Neurosci.* 16, 1272–1288. doi: 10.1162/0898929041920487
- 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 © 2015 Fogel, Rosenberg, Lehman, Kuperberg and Patel. 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.



Rhythmic Effects of Syntax Processing in Music and Language

Harim Jung, Samuel Sontag, YeBin S. Park and Psyche Loui*

Music, Imaging, and Neural Dynamics Lab, Psychology and Neuroscience and Behavior, Wesleyan University, Middletown, CT, USA

OPEN ACCESS

Edited by:

Edward W. Large,
University of Connecticut, USA

Reviewed by:

Theodor Rueber,
Bonn University Hospital, Germany
Reyna L. Gordon,
Vanderbilt University Medical Center,
USA

*Correspondence:

Psyche Loui
ploui@wesleyan.edu

Specialty section:

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

Received: 31 January 2015

Accepted: 03 November 2015

Published: 23 November 2015

Citation:

Jung H, Sontag S, Park YS and Loui P
(2015) Rhythmic Effects of Syntax
Processing in Music and Language.
Front. Psychol. 6:1762.
doi: 10.3389/fpsyg.2015.01762

Music and language are human cognitive and neural functions that share many structural similarities. Past theories posit a sharing of neural resources between syntax processing in music and language (Patel, 2003), and a dynamic attention network that governs general temporal processing (Large and Jones, 1999). Both make predictions about music and language processing over time. Experiment 1 of this study investigates the relationship between rhythmic expectancy and musical and linguistic syntax in a reading time paradigm. Stimuli (adapted from Slevc et al., 2009) were sentences broken down into segments; each sentence segment was paired with a musical chord and presented at a fixed inter-onset interval. Linguistic syntax violations appeared in a garden-path design. During the critical region of the garden-path sentence, i.e., the particular segment in which the syntactic unexpectedness was processed, expectancy violations for language, music, and rhythm were each independently manipulated: musical expectation was manipulated by presenting out-of-key chords and rhythmic expectancy was manipulated by perturbing the fixed inter-onset interval such that the sentence segments and musical chords appeared either early or late. Reading times were recorded for each sentence segment and compared for linguistic, musical, and rhythmic expectancy. Results showed main effects of rhythmic expectancy and linguistic syntax expectancy on reading time. There was also an effect of rhythm on the interaction between musical and linguistic syntax: effects of violations in musical and linguistic syntax showed significant interaction only during rhythmically expected trials. To test the effects of our experimental design on rhythmic and linguistic expectancies, independently of musical syntax, Experiment 2 used the same experimental paradigm, but the musical factor was eliminated—linguistic stimuli were simply presented silently, and rhythmic expectancy was manipulated at the critical region. Experiment 2 replicated effects of rhythm and language, without an interaction. Together, results suggest that the interaction of music and language syntax processing depends on rhythmic expectancy, and support a merging of theories of music and language syntax processing with dynamic models of attentional entrainment.

Keywords: syntax, music, harmony, language, rhythm, expectancy

INTRODUCTION

Music and language are both universal human cognitive functions, but the degree to which they share cognitive resources is a long-standing debate in cognition. Theorists have argued for a shared evolutionary origin (Mithen, 2006), as well as extensive structural similarities between music and language (Lerdahl and Jackendoff, 1983; Botha, 2009), while others have argued for significant

differences between music and language processing and domain specificity of the two domains (Peretz and Coltheart, 2003). Although syntax usually refers to the rules that govern how words and phrases are arranged in language, syntactic structure also exists in other domains, such as music. Musical syntax can be understood as the rules that define how pitches are organized to form melody and harmony. Western tonal harmony, like language, is organized in hierarchical structures that are built upon discrete and combined elements (Lerdahl and Jackendoff, 1983). Syntax in Western music can be realized in the structured organization of the 12 chromatic tones into diatonic scale degrees within tonal centers, which form chords within harmonic progressions. Both musical and linguistic structures unfold syntactically over time.

One theory that has influenced research in the structures of music and language is the Shared Syntactic Integration Resource Hypothesis (SSIRH), which postulates an “overlap in the neural areas and operations which provide the resources for syntactic integration” (Patel, 2003). The hypothesis reconciles contrasting findings between neuropsychology and neuroimaging studies on syntax processing, by suggesting that the same syntactic processing mechanisms act on both linguistic and musical syntax representations. The SSIRH predicts that the syntactic processing resources are limited, and thus studies with tasks combining musical and linguistic syntactic integration will show patterns of neural interference (Patel, 2003). While topics of ongoing debate concern the nature of the resources that are shared (Slevc and Okada, 2015) and the extent to which such resources are syntax-specific (Perruchet and Poulin-Charronnat, 2013), convergent studies do provide evidence for some shared processing of music and language, with evidence ranging from behavioral manipulations of syntactic expectancy violations in music and language (e.g., Fedorenko et al., 2009; Slevc et al., 2009; Hoch et al., 2011) to cognitive neuroscience methods such as ERP and EEG studies that track the neural processing of syntax and its violations (e.g., Koelsch et al., 2005; Steinbeis and Koelsch, 2008; Fitzroy and Sanders, 2012).

One piece of evidence in support of the shared processing of musical and linguistic syntax comes from a reading time study in which musical and linguistic syntax were manipulated simultaneously (Slevc et al., 2009). Reading time data for a self-paced reading paradigm showed interactive effects when linguistic and musical syntax were simultaneously violated, suggesting the use of the same neural resources for linguistic and musical syntax processing. In this self-paced reading paradigm, linguistic syntax was violated using garden path sentences, whereas musical syntax was violated using harmonically unexpected musical chord progressions.

As both musical and linguistic syntax unfold over time, the timing of both musical and linguistic events may affect such sharing of their processing resources. Rhythm, defined as the pattern of time intervals in a stimulus sequence, is usually perceived as the time between event onsets (Grahn, 2012a). As a pattern of durations that engenders expectancies, rhythm may represent its own form of syntax and thus be processed similarly to both musical and linguistic syntax in the brain (Fitch, 2013). It has also been suggested that rhythm is an implicitly

processed feature of environmental events that affects attention and entrainment to events in various other domains such as music and language (Large and Jones, 1999). Specifically, the Dynamic Attending Theory (DAT) posits a mechanism by which internal neural oscillations, or attending rhythms, synchronize to external rhythms (Large and Jones, 1999). In this entrainment model, rhythmic processing is seen as a fluid process in which attention is involuntarily entrained, in a periodic manner, to a dynamically oscillating array of external rhythms, with attention peaking with stimuli that respect the regularity of a given oscillator (Large and Jones, 1999; Grahn, 2012a). This process of rhythmic entrainment has been suggested to occur via neural resonance, where neurons form a circuit that is periodically aligned with the stimuli, allowing for hierarchical organization of stimuli with multiple neural circuits resonating at different levels, or subdivisions, of the rhythm (Large and Snyder, 2009; Grahn, 2012a; Henry et al., 2015). One piece of evidence in support of the DAT comes from Jones et al. (2002), in which a comparative pitch judgment task was presented with interleaving tones that were separated temporally by regular inter-onset intervals (IOIs) that set up a rhythmic expectancy. Pitch judgments were found to be more accurate when the tone to be judged was separated rhythmically from the interleaving tones by a predictable IOI, compared to an early or late tone that was separated by a shorter or longer IOI, respectively. The temporal expectancy effects from this experiment provide support for rhythmic entrainment of attention within a stimulus sequence.

Both SSIRH and DAT make predictions about how our cognitive system processes events as they unfold within a stimulus sequence, but predictions from SSIRH pertain to expectations for linguistic and musical structure, whereas those from DAT pertain to expectations for temporal structure. The two theories should converge in cases where expectations for music, language, and rhythm unfold simultaneously.

Aims and Overall Predictions

The current study aims to examine the simultaneous cognitive processing of musical, linguistic, and rhythmic expectancies. We extend the reading time paradigm of Slevc et al. (2009), by borrowing from the rhythmic expectancy manipulations of Jones et al. (2002), to investigate how the introduction of rhythmic expectancy affects musical and linguistic syntax processing. Rhythmic expectancy was manipulated through rhythmically early, on-time, or late conditions relative to a fixed, expected onset time. As previous ERP data that have shown effects of temporal regularity in linguistic syntax processing (Schmidt-Kassow and Kotz, 2008), it is expected that rhythmic expectancy does affect syntax processing. The current behavioral study more specifically assesses how rhythmic expectancy may differentially modulate the processing of musical and linguistic syntax.

EXPERIMENT 1

Methods

Participants read sentences that were broken down into segments, each of which was paired with a chord from a harmonic chord progression. Linguistic syntax expectancy was

manipulated using syntactic garden-path sentences, musical expectancy was manipulated using chords that were either in key or out of key, and rhythmic expectancy was manipulated by presenting critical region segments early, on time, or late.

Participants

Fifty six undergraduate students from Wesleyan University participated in this study in return for course credit. A recording error resulted in the loss of data for 8 out of the 56 total students, and so 48 participants' data were used in the final analysis. Of the remaining participants, all reported normal hearing. Twenty eight participants (58.3%) reported having prior music training, averaging 6.8 years ($SD = 3.4$). Twenty five (52%) participants identified as female, and 23 as male. Thirty eight (79.1%) reported that their first language was English, three were native speakers of English and one other language, and seven had a language other than English as their first language. Other than English, participants' first languages included Chinese (Mandarin), Arabic, Thai, Japanese, Spanish, French, German, Vietnamese, and Bengali. Sixteen participants (33.3%) spoke more than one language. All participants had normal or corrected-to-normal vision and reported being free

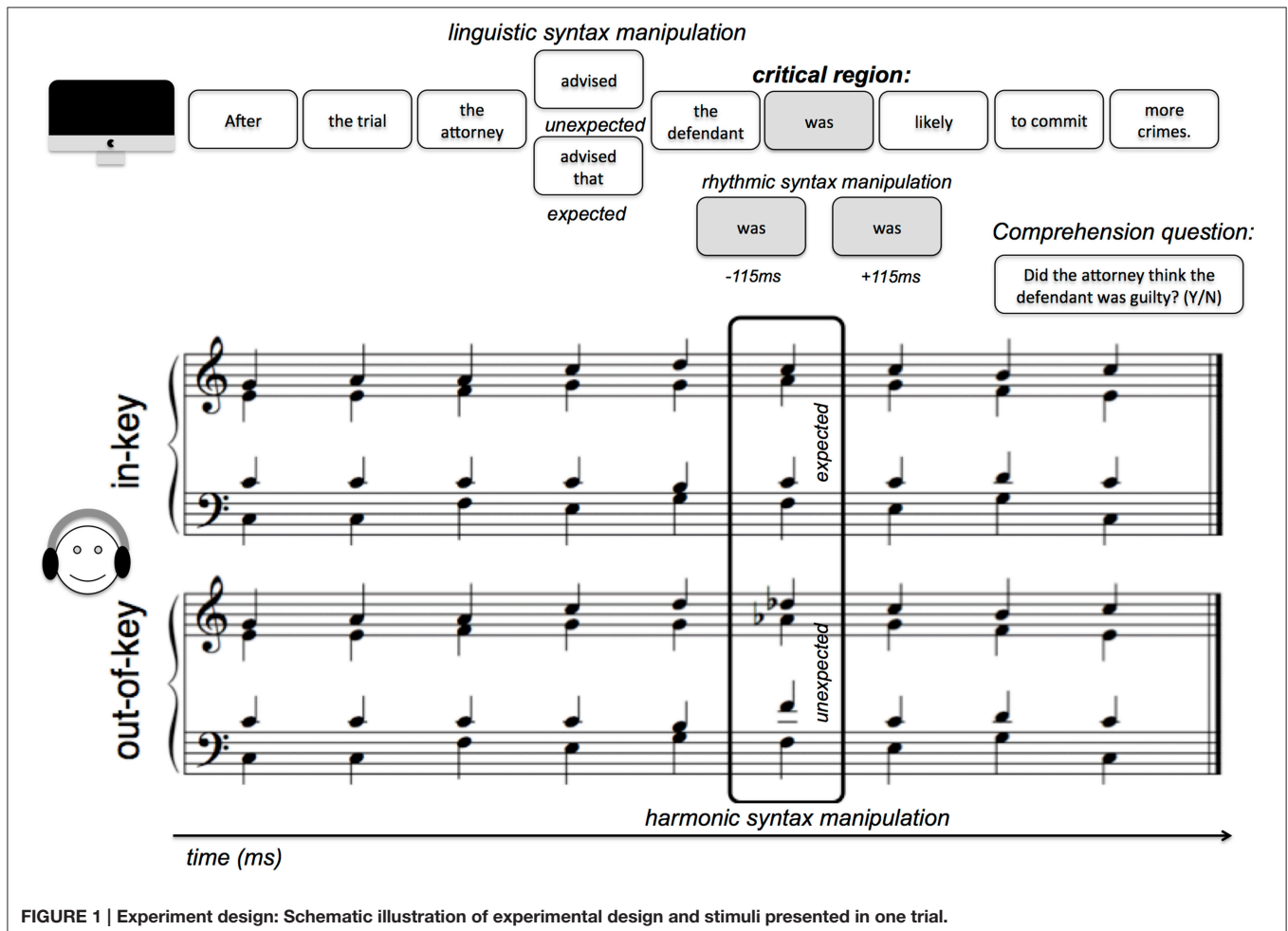
of psychiatric or neurological disorders. Informed consent was obtained from all subjects as approved by the Ethics Board of Psychology at Wesleyan University.

Materials

All experiments were conducted in Judd Hall of Wesleyan University. An Apple iMac and Sennheiser HD280 pro headphones were used for the experiments, with MaxMSP software (Zicarelli, 1998) for all stimulus presentation and response collection.

Stimuli

The current study used 48 sentences from Slevc et al. (2009). These sentences were divided into segments of one or several words, and presented sequentially on the iMac screen using MaxMSP. Twelve of the sentences were syntactic garden paths, which were manipulated to be either syntactically expected or unexpected at the critical region (by introducing a garden path effect—see **Figure 2**). Reading time (RT) comparisons between different conditions were controlled for length of segment because the critical regions are always the same number of words (as shown in **Figure 1**) in the different conditions. Sentence segments with the paired harmonic progression were presented



at a critical region, either on-time (at the regular inter-onset interval of 1200 ms) or “jittered” to be either early or late. The early jitter was 115 ms earlier than the on-time presentation, and the late jitter was 115 ms later than the on-time presentation. Thus, the IOIs were either $1200 - 115 = 1085$ ms (early), 1200 ms (on-time), or $1200 + 115 = 1315$ ms (late; **Figure 2**). 115 ms was selected as the temporal jitter based on pilot testing and the IOIs used in Experiment 2 of Jones et al. (2002) in their manipulation of temporal expectancy. Accompanying chord progressions were played in MIDI using a grand piano timbre. These 48 different progressions were also from Slevc et al. (2009) and followed the rules of Western tonal harmony, and were all in the key of C major. Out-of-key chords violated harmonic expectancy given the context, but were not dissonant chords by themselves (**Figure 1**). A yes-or-no comprehension question was presented at the end of each trial (sentence). Participants’ task was to press the spacebar on the keyboard as soon as they had read each sentence segment, and to answer “yes” or “no” to the comprehension questions. Ninety six unique comprehension questions, two for each sentence, were written so each sentence would have one comprehension question written to have a correct answer “yes,” and another to have a correct answer “no.” The comprehension questions are now given in the Supplementary Materials accompanying this manuscript.

Twelve unique experimental modules were created in order to counterbalance the experimental design. Each module contained all 48 sentences, with violation and filler conditions rotated through the sentences in order to control for systematic effects of content, length, and sentence order. Each module contained: 4 rhythmic violation trials (2 early and 2 late), 3 musical syntax violation trials, 1 linguistic syntax violation trial, 5 musical syntax plus rhythmic violation trials, 1 linguistic plus musical syntax violation trial, 2 linguistic syntax plus rhythmic violation trial, 2 trials with all 3 violations, and 30 sentences with no violations. Therefore, in a given module only 37.5% of trials

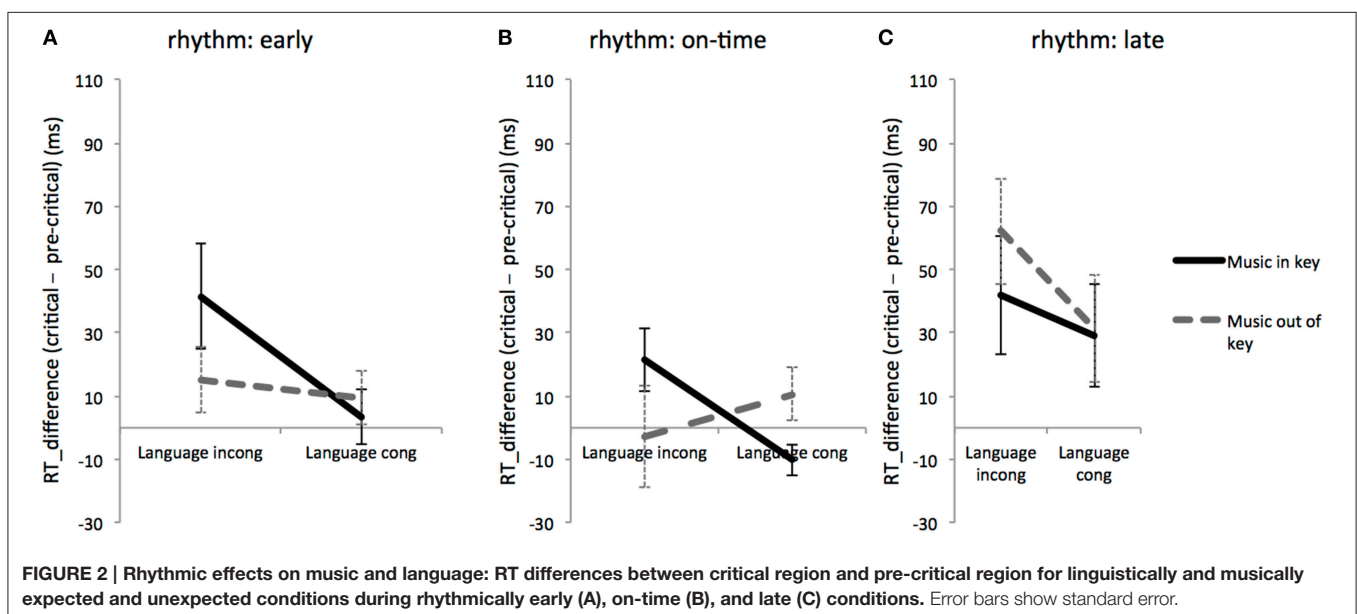
contained any violation. Half of the sentences in a given module were assigned a “yes” question, the other half were assigned a “no.” The order of the trials was randomized for each subject.

Procedure

Before beginning the experiment, the participants gave informed consent and completed a short background survey. The participants were then instructed to pay close attention to the sentences being read, rather than the chord progressions that were heard over the headphones. Then, the participants ran through a set of practice trials. After the practice trials, in the actual experiment the experimenter selected one of the 12 possible experimental modules at random. Participants were instructed to press the spacebar on the keyboard as soon as they had read the sentence segment, and then wait for the next segment to be presented. Pressing the spacebar caused the current sentence segment to disappear and an indicator button labeled “I read it” to light up. The following segment appeared at a fixed IOI regardless of when the current segment disappeared. After the end of each sentence, a yes-or-no comprehension question was displayed, at which point participants answered the question by pressing Y or N on the keyboard. Answering the comprehension question cued a new trial. The experiment lasted ~20 min. Examples of different types of trials are shown in a video demo in the Supplementary Materials accompanying this manuscript.

Data Analysis

RT and response data were saved as text files from MaxMSP, and imported into Microsoft Excel and SPSS for statistical analysis. RTs were log-transformed to normal distribution for statistical testing. Only RTs pre-critical, critical, and post-critical regions for each trial were used for analysis. Filler trials were, therefore, excluded from analysis (21 trials per subject). Of the remaining trials, trials with RTs that were two or more standard deviations



from the mean of log-transformed critical region RTs were excluded as outliers, resulting in a range of 102.76–816.74 ms. These criteria led to the exclusion of 92 (7.20%) of observations from critical regions in Experiment 1.

No significant differences were observed in log-transformed RTs between native English speakers ($n = 41$) and non-native English speakers [non-native $n = 7$, $t_{(46)} = 0.42$, n.s.]. Similarly, no significant differences were observed between participants who reported musical training ($n = 29$) and those who reported no musical training [$n = 19$, $t_{(46)} = 1.53$, n.s.]. To check for interactions between linguistic syntax and native English speaker experience, an ANOVA was run on the dependent variable of log-transformed RT with the fixed factor of linguistic syntax (congruent vs. incongruent) and the random factor of native English speaker status (native vs. non-native English speaker). No significant interaction between native English speaker status and linguistic syntax was observed [$F_{(1, 92)} = 0.53$, $MSE = 0.01$, $p = 0.47$]. Similarly, to check for interactions between musical syntax and musical training, an ANOVA with the fixed factor of musical syntax (congruent vs. incongruent) and the random factor of musical training (musically trained vs. no musical training) showed no interaction between musical syntax and musical training [$F_{(1, 92)} = 0.091$, $MSE = 0.008$, $p = 0.764$]. As we observed no main effects or interactions that were explainable by native English speaking experience or musical training, results were pooled between native and non-native English speakers, and between musically trained and untrained subjects.

Results

On comprehension questions, participants performed significantly above chance in all conditions [overall $M = 78.95\%$, $s = 12.24$, two-tailed t -test against chance level of 50% correct: $t_{(47)} = 16.38$, $p < 0.0001$].

A Three-way ANOVA on the dependent variable of log-transformed RT during the critical region (log_RT_CR) was

run with fixed factors of language (two levels: congruent and incongruent), music (two levels: congruent vs. incongruent), and rhythm (three levels: early, on-time, and late), with subject number as a random factor. Results showed a significant three-way interaction among the factors of linguistic, musical and rhythmic expectancies [$F_{(2, 52)} = 5.02$, $MSE = 0.008$, $p = 0.01$], as well as a significant main effect of language [$F_{(1, 54)} = 12.5$, $MSE = 0.006$, $p = 0.001$] and a significant main effect of rhythm [$F_{(2, 99)} = 13.2$, $MSE = 0.01$, $p < 0.001$] and a marginally significant effect of music [$F_{(1, 53)} = 3.7$, $MSE = 0.01$, $p = 0.059$]. Means and SDs of RTs are given in **Table 1** for each condition, and in **Table 2** for each cell.

To investigate any possible interactive effects between music and language syntax at different rhythmic conditions, an RT difference was computed between RTs for critical region and for pre-critical region. Two-way ANOVAs with fixed factors of language and music were used to test for interactions between music and language at each of the three rhythm conditions (early, on-time, and late). Results showed that for the rhythmically on-time condition, there was an interaction between language and music [$F_{(1, 170)} = 4.9$, $MSE = 4776.9$, $p = 0.027$]. In contrast, the interaction between language and music was not significant at the rhythmically early condition [$F_{(1, 170)} = 0.27$, $MSE = 12882.0$, $p = 0.603$] or the rhythmically late condition [$F_{(1, 170)} = 2.34$, $MSE = 5155.2$, $p = 0.127$] (see **Figure 2**). These results suggest that the interaction between linguistic and musical syntax varies by rhythmic expectancy.

Further investigation of the degree to which factors interacted at the critical region required comparing RTs across the pre-critical, critical, and post-critical time regions. For this comparison, difference scores of linguistically congruent from linguistically incongruent RTs were calculated, and these difference scores were compared for musically in-key and out-of-key trials across time regions for each rhythmic condition (see **Figure 3**). We found a significant effect of time region:

TABLE 1 | Mean critical region RTs (ms) under different conditions of linguistic syntax, musical syntax, and rhythmic expectancies.

	Lang			Music			Rhythm	
	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>
Congruent	311.8	63.15	In-key	315.91	65.85	Early	327.1	80.62
Incongruent	339.12	84.81	Out-of-key	322.67	69.78	On-Time	301.12	67.92
						Late	351.6	71.24

TABLE 2 | Mean critical region RTs (ms) under different combinations of conditions of linguistic syntax, musical syntax, and rhythmic expectancies.

Language	Early				On time				Late			
	Music				Music				Music			
	In key		Out key		In key		Out key		In key		Out key	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Congruent	326.22	100.59	313.81	88.72	294.53	65.86	307.84	106.92	369	101.44	334.62	89.4
Incongruent	361.72	126.53	316.09	76.31	331.16	102.5	310.03	116.35	365.37	203.76	388.32	139.02

RT was longer in the critical region in the rhythmically early condition only [$F_{(2, 92)} = 4.67, p = 0.012$]. In the rhythmically late condition only, musical syntax violations produced larger difference scores at the critical region; however this difference was not significant. In the rhythmically early condition and on-time conditions, musically in-key trials yielded larger difference scores than musically out-of-key trials at the critical regions, although these differences were not significant (see **Figure 3**).

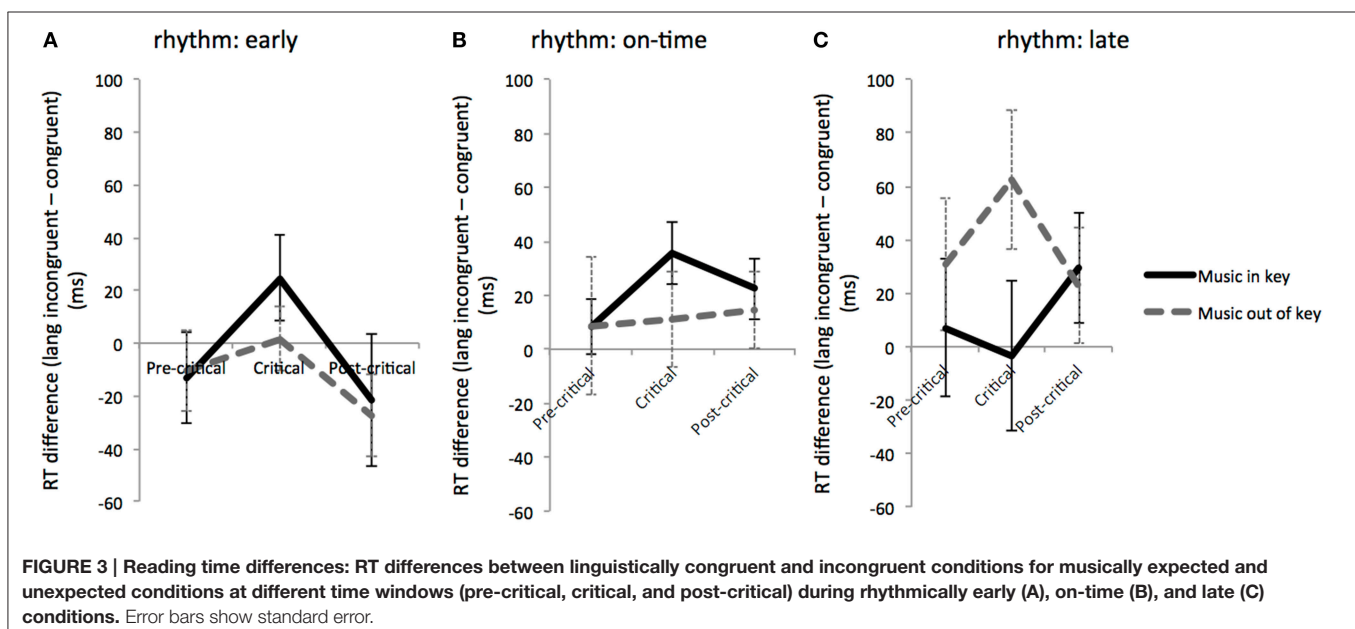
Discussion

Experiment 1 tested to see how rhythmic expectancy affected the processing of musical and linguistic syntax. Results from log-transformed RTs during the critical region (**Table 2**) and RT differences between critical and pre-critical regions (**Figure 2**) showed significant main effects of language and rhythm, a significant three-way interaction of language, music, and rhythm, and a significant two-way interaction between linguistic and musical syntax in the on-time condition only. These findings extend the results of past research (Slevc et al., 2009) to show that the sharing of cognitive resources for music and language appear specific to rhythmically expected events.

In contrast to critical region RTs, however, RT differences between linguistically incongruent and congruent trials (**Figure 3**) showed slower RTs within the critical region only during rhythmically early trials. The interaction patterns between musical and linguistic syntax over different time regions were inconclusive. This differs from the original findings of Slevc et al. (2009), who observed a synergistic interaction between musical syntax and time region on the reaction time difference between linguistically congruent minus incongruent trials, suggestive of a language and music interaction specifically during the critical region, when rhythm was not a factor. The less robust effect of critical region in this experiment may arise from spillover effects of linguistic incongruence that last beyond the critical region.

While neither SSIRH nor DAT makes specific predictions about this possible spillover effect, the main findings of a three-way interaction among language, music, and rhythm is generally consistent with both theoretical accounts and does suggest that any synergy or sharing of neural resources between music and language depends on rhythmic expectancy. Violations in rhythmic expectancy may disrupt the shared resources that are generally recruited for syntax processing, such as cognitive control (Slevc and Okada, 2015). As music and language both unfold over time, it stands to reason that our expectations for rhythm—defined here as the pattern of time intervals within a stimulus sequence (Grahn, 2012a)—would govern any sharing of neural resources between music and language, as is consistent with the DAT (Large and Jones, 1999), as well as prior behavioral data on rhythmic entrainment (Jones et al., 2002) and studies on the neural underpinnings of rhythmic entrainment (Henry et al., 2015) and their effects on linguistic syntax processing (Schmidt-Kassow and Kotz, 2008).

The three-way interaction between language, music, and rhythm is accompanied by significant main effects of language and rhythm, and marginally significant main effect of musical expectancy. The main effect of rhythm is similar to Jones et al. (2002) and others, in which perturbed temporal expectations resulted in longer RTs. Incongruent garden-path sentences elicit longer RTs during the critical region compared to their counterparts. This is consistent with Slevc et al. (2009) and Perruchet and Poulin-Charronnat, (2013) as well as with previous uses of the self-paced reading time paradigm (Ferreira and Henderson, 1990). The main effect of musical expectancy was only marginally significant. While it is worth noting that Slevc et al. (2009) also did not report a significant main effect of musical expectancy, this weak effect may also be due to task instructions to pay close attention to the sentence segments rather than to the chord progressions heard over headphones. To determine whether music generally taxed cognitive or attentional resources



away from subjects' monitoring of the sentence segments, it was necessary to compare comprehension accuracy with and without musical stimuli. This was a motivation for Experiment 2, in which the experiment was re-run without musical stimuli.

While previous studies that used a self-paced reading paradigm (Ferreira and Henderson, 1990; Trueswell et al., 1993; Slevc et al., 2009; Perruchet and Poulin-Charronnat, 2013) required subjects to activate the next sentence segment as part of the task, in order to implement a factor of rhythmic expectancy our design featured a fixed inter-onset interval of sentence segments, and subjects were asked instead to press a button to indicate that they had read each segment. To our knowledge this type of implementation is new for psycholinguistic studies. One of the goals of Experiment 2 is to check for the validity of this type of implementation by testing for an effect of linguistic congruency with fixed IOI presentations of sentence segments, even in the absence of musical stimuli.

EXPERIMENT 2

Our modification of the standard self-paced reading paradigm resulted in fixed IOIs with the task of indicating that subjects had read the displayed sentence segment. This was a different task from the standard self-paced reading paradigm in which subjects' task was to advance the following sentence segment, and our task had yet to be confirmed as effective in detecting effects of linguistic syntax, even without the presence of musical stimuli. Furthermore, it was possible that the three-way and two-way interactions from Experiment 1 resulted from the complexity of our experimental design, and that the processing of multiple violations could affect attending and development of expectancy to task-irrelevant stimuli, as well as syntax processing *per se*. Experiment 2 thus follows up on Experiment 1 by investigating effects of rhythmic violations on comprehension and the processing of linguistic syntax stimuli, removing the variable of musical stimuli. A significant effect of linguistic syntax as well as rhythmic expectancy could validate the current manipulation of the self-paced reading paradigm, and a significant interaction between language and rhythm would suggest that the two domains tap into the same specific neural resources whereas no interaction might suggest more parallel processing.

Methods

In experiment 2, participants again read sentences broken down into segments. Linguistic syntax expectancy was manipulated using syntactic garden-path sentences, and rhythmic expectancy was manipulated by presenting critical region segments early, on-time, or late.

Participants

A new group of 35 undergraduate students from Wesleyan University participated in Experiment 2 in return for course credit. From these participants, all reported normal hearing, normal or corrected-to-normal vision, and no psychiatric or neurological disorders. Twenty-five participants (71.4%) reported having prior music training, averaging 5.9 years

($SD = 3.0$). Twenty (57.1%) participants identified as female, and 15 (42.3%) as male. Twenty-eight (80%) reported that their first language was English, and seven had a language other than English as their first language. Other than English, participants' first languages included Spanish, Chinese, and Thai. Twenty-four participants (68.6%) spoke more than one language. Informed consent was obtained from all subjects as approved by the Ethics Board of Psychology at Wesleyan University.

Materials

The second experiment was conducted in the Music, Imaging, and Neural Dynamics (MIND) Lab Suite in Judd Hall at Wesleyan University. An Apple iMac was used for the experiment, with MaxMSP software for all stimulus presentation and response collection.

Stimuli

The same experimental patch on MaxMSP and 12 experimental modules with the 48 sentences borrowed from Slevc et al. (2009) were used from the first experiment. However, to investigate how rhythmic violations would affect reading and interact with violations in linguistic syntax, independent of violations in musical syntax, the experimental patch was muted, so that chords were not heard with each sentence segment. The IOIs of sentence segments remained unaltered, and the same "yes" or "no" comprehension questions were also asked at the end of each trial, with randomized order of the trials for each subject.

Procedure

Similar to Experiment 1, participants were instructed to read sentences carefully, and hit the spacebar as soon as they had read a sentence segment. After running through a practice set, the participants began the actual experiment. The experimenter selected one of the twelve possible experimental modules at random. At the end of each trial, participants answered the "yes" or "no" comprehension question, queuing the next trial.

Data Analysis

RTs and comprehension question responses were saved as text files from MaxMSP, and imported into Microsoft Excel, and SPSS for statistical analysis. Only RTs at the pre-critical, critical, and post-critical regions for each trial were used for analysis. Filler trials were, again, excluded from analysis (21 trials per subject). The same parameters and methods of outlier exclusion were used from the previous experiment, resulting in a RT range of 123.63–1121.40 ms. These criteria led to the exclusion of 19 (1.97%) of observations in Experiment 2. RTs were also log-transformed to normal distribution for statistical tests.

Results between musically trained and non-musically trained subjects were pooled because music was not a factor in this experiment. No significant differences were observed in log-transformed RTs between native English speakers and non-native English speakers [$t_{(34)} = 0.96$, n.s.]. Similarly, an ANOVA with the fixed factor of linguistic syntax and the random factor of native English experience showed no significant interaction [$F_{(1, 523)} = 1.059$, $MSE = 0.018$, $p = 0.30$]. As we observed no differences that were explainable by native English speaking

experience, results were pooled between native and non-native English speakers.

Results

Participants performed significantly above chance ($M = 86.93\%$, $s = 6.21$) on comprehension questions in all conditions. To compare comprehension accuracy with and without musical stimulus presentation, a One-way ANOVA on average comprehension accuracy as the dependent variable was run with the factor of experiment, comparing average comprehension accuracy for subjects between Experiment 1 and 2. Results showed a significant main effect of experiment on comprehension accuracy, with subjects from Experiment 2 performing better on average on comprehension questions than those from Experiment 1 [$F_{(1, 81)} = 12.51$, $MSE = 0.01$, $p = 0.001$]. This suggests that the added variable of musical expectancy further taxed participants' attention from the task-relevant comprehension questions in Experiment 1.

A Two-way ANOVA on the dependent variable of log-transformed RT during the critical region was run with the factors of language and rhythm. Results showed a significant main effect of language [$F_{(1, 34)} = 7.69$, $MSE = 0.001$, $p = 0.009$], a significant effect of rhythm [$F_{(2, 68)} = 9.69$, $MSE = 0.001$, $p < 0.001$], and no significant two-way interaction [$F_{(2, 68)} = 1.07$, $MSE = 0.001$, $p = 0.83$]. Mean and SD RTs are shown for each condition in **Table 3** and for each cell in **Table 4**.

Discussion

Results from Experiment 2 showed main effects of language and rhythm, validating the use of this novel task. There was also a higher comprehension accuracy compared to Experiment 1, but no interactions between the two factors of linguistic syntax and rhythmic expectancy (see **Table 4**).

Experiment 2 further investigates the effects of rhythmic expectancy on linguistic syntax processing. When the factor of music was removed, main effects of language and rhythm were still observed. RTs were longer for syntactically unexpected

sentences, replicating results from Experiment 1 as well as previous experiments that used the self-paced reading time paradigm (Ferreira and Henderson, 1990; Trueswell et al., 1993). Notably, this finding of longer RTs during syntactically unexpected critical regions within the garden path sentences provides a validation of the current adaptation of the self-paced reading time paradigm: while previous studies that used the self-paced reading time paradigm (Ferreira and Henderson, 1990; Trueswell et al., 1993; Slevc et al., 2009; Perruchet and Poulin-Charronnat, 2013) required subjects to advance the sentence segments manually, in the current study we adapted the paradigm with fixed IOIs to enable simultaneous investigations of rhythmic and linguistic syntax expectancy.

Effects of rhythmic expectancy were also observed, as participants were slower to respond to critical regions presented earlier or later than the expected IOI. This replicates results from Experiment 1 and suggests that temporal entrainment was possible even with a visual-only reading task, and thus is not limited to the auditory modality. This effect of rhythm on visual processing is consistent with prior work on rhythmic effects of visual detection (Landau and Fries, 2012) and visual discrimination (Grahn, 2012b).

Although main effects of language and rhythm were observed, there was no significant interaction. An explanation for this lack of interaction could be that removing the factor of music resulted in the implemented violations no longer being sufficiently attention-demanding to lead to an interaction between the remaining factors, resulting in parallel processing of language and rhythm. In this view, the data suggests that rhythm affects a general, rather than a syntax-specific, pool of attentional resources. When the factor of music was removed, attentional resources were less demanded from the available pool, reducing the interactive effects of language and rhythm on each other and resulting in no interaction and higher comprehension accuracy. Alternately, it could be that the rhythm only affected peripheral visual processing, without also affecting syntax processing at a central level. While the present experiment cannot tease apart these possible explanations, considering the extant literature on relationships between rhythm and grammar (Schmidt-Kassow and Kotz, 2009; Gordon et al., 2015b) it is clear that rhythm can affect central cognitive processes such as syntactical or grammatical computations.

Finally, another finding from Experiment 2 is that comprehension accuracy was higher compared to Experiment 1, suggesting that eliminating the factor of music restored some attentional resources to the task of comprehension. When the primary task was to read sentence segments for comprehension, musical stimuli in the background could have functioned as a distractor in a seeming dual-task condition of comprehending the entire sentence while responding to each segment (by pressing the spacebar).

Taken together, Experiment 2 helps to validate the paradigm used in Experiment 1. By simplifying the experiment to remove the factor of music, some attentional resources may have been restored, resulting in higher comprehension accuracy overall, as well as main effects of language and rhythm with no interaction between the two.

TABLE 3 | Mean critical region RTs (ms) under different conditions of linguistic and rhythmic expectancies.

	Lang			Rhythm	
	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>
Congruent	387.34	57.2	Early	415.26	64.21
Incongruent	414.13	87.64	On-Time	381.39	62.18
			Late	399.11	75.97

TABLE 4 | Mean critical region RTs (ms) under different combinations of conditions of linguistic syntax and rhythmic expectancies.

Language	Early		On time		Late	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Congruent	407.41	68.06	377.65	59.45	398.16	82.17
Incongruent	434.84	116.69	397.2	109.48	412.56	125.85

GENERAL DISCUSSION

The goal of the current study is to examine how rhythmic expectancy affects the processing of musical and linguistic syntax. Experiment 1 shows main effects of language, music, and rhythm, and specificity of the interaction between musical and linguistic syntax in the rhythmically expected condition only. These data patterns confirm that rhythm affects the sharing of cognitive resources for music and language, and is largely consistent with SSIRH (Patel, 2003) and DAT (Large and Jones, 1999). However, some of the follow-up analyses are inconclusive as to the exact nature of these interactions over time. In particular, only in rhythmically early trials did we find that the critical region significantly affected the difference in RT between incongruent and congruent language trials, with no significant interactions with musical expectancy unlike in Slevc et al. (2009). The reason for this specific effect of critical region in rhythmically early trials is unclear. It might arise from some spillover effects of linguistic incongruence that last beyond the critical region in rhythmically on-time and late trials. Alternately, it might be a consequence of the complexity of our task in this experiment design. Although the significant main effects suggest that our manipulations were effective, this inconclusive data pattern may nevertheless result from low power due to relatively few trials per cell in the experiment design of Experiment 1.

As it is possible that results were due to the complexity of our design, Experiment 2 simplifies the design by eliminating the factor of music altogether. Results of Experiment 2 show superior comprehension accuracy compared to Experiment 1, and main effects of language and rhythm without an interaction between the two factors. The main effects help to validate our adaptation of the original self-paced reading time paradigm (Ferreira and Henderson, 1990; Trueswell et al., 1993) for research in rhythmic expectancy. The null interaction, when accompanied by significant main effects, suggests that given the task conditions and attentional allocation in Experiment 2, rhythm and language were processed in parallel and did not affect each other.

The superior comprehension accuracy in Experiment 2 may be explained by an increase in general attentional resources that are now available to subjects in Experiment 2 due to the removal of music as a factor. While it was not specifically tested whether these general attentional mechanisms may be the same or different from the temporal attention that is taxed by temporal perturbations of rhythmic expectancy, other literature on voluntary (endogenous) vs. involuntary (exogenous) attention might shed light on this distinction (Haft et al., 2008; Prinzmetal et al., 2009). Voluntary or endogenous attention, such as that tested in dual-task situations when the task is to attend to one task while ignoring another, is similar to the general design of the present studies where subjects are instructed to pay attention to sentence segments while ignoring music that appears simultaneously. Involuntary or exogenous attention, in contrast, is driven by stimulus features such as rhythmic properties as tapped by our rhythmic expectancy manipulations. Previous research has shown that voluntary attention tends to affect accuracy whereas involuntary attention affects reaction time

(Prinzmetal et al., 2005). This fits with our current findings where comprehension accuracy is affected by the removal of music as a factor (by comparing Experiments 1 and 2), whereas reading time is affected by rhythmic perturbations of the presentation of sentence segments.

In both experiments, effects of rhythm were observed in response to visually-presented sentence segments. While the rhythmic aspect of language might generally manifest itself more readily in the auditory than the visual modality, this effect observed from the visual manipulations suggests that rhythmic expectation for language is not limited to auditory processing, but may instead pervade the cognitive system in a modality-general manner, affecting even the visual modality. As visual detection and discrimination are both modulated by rhythm (Grahn, 2012b; Landau and Fries, 2012) and musical expectation can cross-modally affect visual processing (Escoffier and Tillmann, 2008), the current study provides support for the view that rhythmic, musical, and linguistic expectations are most likely not tied to the auditory modality, but instead affect the cognitive system more centrally.

Results appear to be independent of musical training and native English speaker experience. The link between linguistic and musical grammar processing could have been expected to vary by musical and linguistic expertise: children who perform well on phonemic or phonological tasks also outperform their counterparts in rhythmic discrimination as well as pitch awareness (Loui et al., 2011; Gordon et al., 2015b). At a neural level, brain areas and connections that subserve language are different in their structure and function among professional musicians (Sluming et al., 2002; Halwani et al., 2011), and some highly trained populations, such as jazz drummers, process rhythmic patterns in the supramarginal gyrus, a region of the brain that is thought to be involved in linguistic syntax (Herdener et al., 2014). Despite these effects of training and expertise, the current study found no effects of musical training or linguistic background, converging with the original study (Slevc et al., 2009) as well as prior reports of the language-like statistical learning of musical structure (Loui et al., 2010; Rohrmeier et al., 2011). It is possible that only some types of task performance, such as those that tap more sensory or perceptual resources, might be affected by music training via selective enhancement of auditory skills (Kraus and Chandrasekaran, 2010).

In sum, the current study demonstrates that rhythmic expectancy plays an important role in the shared processing of musical and linguistic structure. The subject of shared processing of musical and language structure has been central to music cognition, as is the question of how rhythm affects attentional entrainment. While providing support for an overlap in processing resources for musical and linguistic syntax, the current results also suggest that perturbations in rhythmicity of stimuli presentation tax these attentional resources. By offering a window into how perturbations of rhythmic and temporal expectancy affect musical and linguistic processing, results may be translatable toward better understanding and possibly designing interventions for populations with speech and language difficulties, such as children with atypical language

development (Przybylski et al., 2013; Gordon et al., 2015a). Toward that goal, the specific neural underpinnings of these shared processing resources still remain to be addressed in future studies.

ACKNOWLEDGMENTS

Supported by startup funds from Wesleyan University, a grant from the Grammy Foundation and the Imagination Institute to

PL., and the Ronald E. McNair Scholars Program to HJ. We thank all our participants and L. Robert Slevc for helpful comments at an early stage of this project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01762>

REFERENCES

- Botha, R. (2009). On musilanguage/"Hmmmm" as an evolutionary precursor to language. *Lang. Commun.* 29, 61–76. doi: 10.1016/j.langcom.2008.01.001
- Escoffier, N., and Tillmann, B. (2008). The tonal function of a task-irrelevant chord modulates speed of visual processing. *Cognition* 107, 1070–1083. doi: 10.1016/j.cognition.2007.10.007
- Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., and Gibson, E. (2009). Structural integration in language and music: evidence for a shared system. *Mem. Cognit.* 37, 1–9. doi: 10.3758/MC.37.1.1
- Ferreira, F., and Henderson, J. M. (1990). Use of verb information in syntactic parsing: evidence from eye movements and word-by-word self-paced reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 555–568. doi: 10.1037/0278-7393.16.4.555
- Fitch, W. T. (2013). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Front. Syst. Neurosci.* 7:68. doi: 10.3389/fnsys.2013.00068
- Fitzroy, A. B., and Sanders, L. D. (2012). Musical expertise modulates early processing of syntactic violations in language. *Front. Psychol.* 3:603. doi: 10.3389/fpsyg.2012.00603
- Gordon, R. L., Jacobs, M. S., Schuele, C. M., and McAuley, J. D. (2015a). Perspectives on the rhythm-grammar link and its implications for typical and atypical language development. *Ann. N.Y. Acad. Sci.* 1337, 16–25. doi: 10.1111/nyas.12683
- Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., and McAuley, J. D. (2015b). Musical rhythm discrimination explains individual differences in grammar skills in children. *Dev. Sci.* 18, 635–644. doi: 10.1111/desc.12230
- Grahn, J. A. (2012a). Neural mechanisms of rhythm perception: current findings and future perspectives. *Top. Cogn. Sci.* 4, 585–606. doi: 10.1111/j.1756-8765.2012.01213.x
- Grahn, J. A. (2012b). See what i hear? Beat perception in auditory and visual rhythms. *Exp. Brain Res.* 220, 56–61. doi: 10.1007/s00221-012-3114-8
- Haft, E. R., Sarampalis, A., and Loui, P. (2008). "Auditory attention and filters," in *Auditory Perception of Sound Sources*, eds W. A. Yost, A. N. Popper, and R. R. Fay (New York, NY: Springer), 115–142. doi: 10.1007/978-0-387-71305-2_5
- Halwani, G. F., Loui, P., Rieber, T., and Schlaug, G. (2011). Effects of practice and experience on the arcuate fasciculus: comparing singers, instrumentalists, and non-musicians. *Front. Psychol.* 2:156. doi: 10.3389/fpsyg.2011.00156
- Henry, M. J., Herrmann, B., and Obleser, J. (2015). Selective attention to temporal features on nested time scales. *Cereb. Cortex* 25, 450–459. doi: 10.1093/cercor/bht240
- Herdener, M., Humbel, T., Esposito, F., Habermeyer, B., Cattapan-Ludewig, K., and Seifritz, E. (2014). Jazz drummers recruit language-specific areas for the processing of rhythmic structure. *Cereb. Cortex* 24, 836–843. doi: 10.1093/cercor/bhs367
- Hoch, L., Poulin-Charronnat, B., and Tillmann, B. (2011). The influence of task-irrelevant music on language processing: syntactic and semantic structures. *Front. Psychol.* 2:112. doi: 10.3389/fpsyg.2011.00112
- Jones, M. R., Moynihan, H., MacKenzie, N., and Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychol. Sci.* 13, 313–319. doi: 10.1111/1467-9280.00458
- 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
- Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605. doi: 10.1038/nrn2882
- Landau, A. N., and Fries, P. (2012). Attention samples stimuli rhythmically. *Curr. Biol.* 22, 1000–1004. doi: 10.1016/j.cub.2012.03.054
- Large, E. W., and Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106:119. doi: 10.1037/0033-295X.106.1.119
- Large, E. W., and Snyder, J. S. (2009). Pulse and meter as neural resonance. *Ann. N. Y. Acad. Sci.* 1169, 46–57. doi: 10.1111/j.1749-6632.2009.04550.x
- Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- Loui, P., Kroog, K., Zuk, J., Winner, E., and Schlaug, G. (2011). Relating pitch awareness to phonemic awareness in children: implications for tone-deafness and dyslexia. *Front. Psychol.* 2:111. doi: 10.3389/fpsyg.2011.00111
- Loui, P., Wessel, D. L., and Hudson Kam, C. L. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Percept.* 27, 377–388. doi: 10.1525/mp.2010.27.5.377
- Mithen, S. J. (2006). *The Singing Neanderthals: The origins of Music: Language, Mind and Body*. Cambridge, MA: Harvard University Press.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681. doi: 10.1038/nm1082
- Peretz, I., and Coltheart, M. (2003). Modularity of music processing. *Nat. Neurosci.* 6, 688–691. doi: 10.1038/nm1083
- Perruchet, P., and Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychon. Bull. Rev.* 20, 310–317. doi: 10.3758/s13423-012-0344-5
- Prinzmetal, W., McCool, C., and Park, S. (2005). Attention: reaction time and accuracy reveal different mechanisms. *J. Exp. Psychol. Gen.* 134, 73–92. doi: 10.1037/0096-3445.134.1.73
- Prinzmetal, W., Zvinyatskovskiy, A., Gutierrez, P., and Dilem, L. (2009). Voluntary and involuntary attention have different consequences: the effect of perceptual difficulty. *Q. J. Exp. Psychol.* 62, 352–369. doi: 10.1080/17470210801954892
- Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Léculier, L., et al. (2013). Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology* 27, 121–131. doi: 10.1037/a0031277
- Rohrmeier, M., Rebuschat, P., and Cross, I. (2011). Incidental and online learning of melodic structure. *Conscious. Cogn.* 20, 214–222. doi: 10.1016/j.concog.2010.07.004
- Schmidt-Kassow, M., and Kotz, S. A. (2008). Entrainment of syntactic processing? ERP-responses to predictable time intervals during syntactic reanalysis. *Brain Res.* 1226, 144–155. doi: 10.1016/j.brainres.2008.06.017
- Schmidt-Kassow, M., and Kotz, S. A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *J. Cogn. Neurosci.* 21, 1693–1708. doi: 10.1162/jocn.2008.21.153
- Slevc, L. R., and Okada, B. M. (2015). Processing structure in language and music: a case for shared reliance on cognitive control. *Psychon. Bull. Rev.* 22, 637–652. doi: 10.3758/s13423-014-0712-4
- Slevc, L. R., Rosenberg, J. C., and Patel, A. D. (2009). Making psycholinguistics musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychon. Bull. Rev.* 16, 374–381. doi: 10.3758/16.2.374
- Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., and Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in broca's area in male symphony orchestra musicians. *Neuroimage* 17, 1613–1622. doi: 10.1006/nimg.2002.1288

- Steinbeis, N., and Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cereb. Cortex* 18, 1169–1178. doi: 10.1093/cercor/bhm149
- Trueswell, J. C., Tanenhaus, M. K., and Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 528–553. doi: 10.1037/0278-7393.19.3.528
- Zicarelli, D. D. (1998). “An extensible real-time signal processing environment for MAX,” *Paper Presented at the Proceedings of the International Computer Music Conference*, (Ann Arbor, MI: University of Michigan).

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 © 2015 Jung, Sontag, Park and Loui. 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.



Pronunciation difficulty, temporal regularity, and the speech-to-song illusion

Elizabeth H. Margulis*, Rhimmon Simchy-Gross and Justin L. Black

Music Cognition Lab, University of Arkansas, Fayetteville, AR, USA

Edited by:

McNeel Gordon Jantzen, Western Washington University, USA

Reviewed by:

Erin E. Hannon, University of Nevada, USA

Cyrille Magne, Middle Tennessee State University, USA

*Correspondence:

Elizabeth H. Margulis, Music Cognition Lab, University of Arkansas, MUSC 201, Fayetteville, AR 72701, USA
e-mail: ehm@uark.edu

The speech-to-song illusion (Deutsch et al., 2011) tracks the perceptual transformation from speech to song across repetitions of a brief spoken utterance. Because it involves no change in the stimulus itself, but a dramatic change in its perceived affiliation to speech or to music, it presents a unique opportunity to comparatively investigate the processing of language and music. In this study, native English-speaking participants were presented with brief spoken utterances that were subsequently repeated ten times. The utterances were drawn either from languages that are relatively difficult for a native English speaker to pronounce, or languages that are relatively easy for a native English speaker to pronounce. Moreover, the repetition could occur at regular or irregular temporal intervals. Participants rated the utterances before and after the repetitions on a 5-point Likert-like scale ranging from “sounds exactly like speech” to “sounds exactly like singing.” The difference in ratings before and after was taken as a measure of the strength of the speech-to-song illusion in each case. The speech-to-song illusion occurred regardless of whether the repetitions were spaced at regular temporal intervals or not; however, it occurred more readily if the utterance was spoken in a language difficult for a native English speaker to pronounce. Speech circuitry seemed more liable to capture native and easy-to-pronounce languages, and more reluctant to relinquish them to perceived song across repetitions.

Keywords: speech-to-song illusion, repetition, music and language, music perception, meter

INTRODUCTION

Music and speech offer excellent comparative cases to illuminate the mechanisms subserving human communication (cf. Patel, 2008). They share many acoustic features, but differ in salient ways too – music tends to feature slower pitch changes, more stable fundamental frequencies, and rhythmic structure that gives rise to the perception of an isochronous beat. Music and speech may share not only developmental origins (McMullen and Safran, 2004), but also common evolutionary origins (Wallin et al., 2001), yet they often seem quite phenomenologically distinct. It can seem that music is heard as music, and speech is heard as speech, and that is that. Several years ago, however, Deutsch et al. (2008) reported a striking illusion where repeating a segment of speech could engender a perceived transformation from speech to song. In this illusion, participants first hear an ordinary spoken utterance. Then they hear a segment from this utterance repeated 10 times in succession. Finally, they rehear the original utterance, but on this hearing, the segment that had been repeated sounds as if it were being sung rather than spoken. Rhythmic and pitch content comes strikingly to the fore, and this change in perceptual orientation results in a change in the category to which listeners attribute the stimulus.

Since the discovery of this illusion, various studies have sought to examine what qualities must be in place for this perceptual transformation to occur. Deutsch et al. (2011) showed that no illusory change to song occurred if the repetitions were inexact – if they were slightly differently transposed in pitch on each repetition, or if the syllables were jumbled into different orderings

on each repetition. Tierney et al. (2013) were able to collect a set of spoken utterances that tended to transform to song after repetition, and a set of spoken utterances that did not tend to transform. The utterances that did transform were distinguished from the others by slightly more stable fundamental frequency contours within syllables, and by more regular spacing of inter-accent intervals. When speech was perceived as song, regions associated with pitch processing such as the superior temporal gyrus and regions associated with auditory-motor integration such as the precentral gyrus were differentially activated. These results suggest that not only does a shift from speech to song reflect increased attention to pitch, but it might also entail more imagined motor involvement. When we hear a song, we tend to sing along in our heads in a way that is quite different from how we listen to speech.

Falk et al. (2014) showed that when the utterance's pitch contour was made up of stable tonal targets, people perceived the transformation to song earlier and more frequently. Rhythmic aspects of the utterance did not play as big of a role. They also manipulated the regularity of the pause between utterances, but found it had no effect on the speech to song transformation. These findings are consistent with Mantell and Pfordresher (2013), which used a vocal imitation task to show that people could replicate the absolute pitch of song more accurately than the absolute pitch of speech, but there was no difference in accuracy between song and speech on replication of timing. People with and without formal musical training experienced the illusion the same way.

Given the increased auditory-motor integration for song perception revealed in Tierney et al. (2013), we wondered whether

part of what distinguishes attending to music from attending to speech is a participatory stance, where the listener begins to sing through a tune in her head while it is playing after she has heard it a few times – a hypothesis explored in Margulis (2013). To address this hypothesis, languages of varying pronunciation difficulty were used. It should be easy to imaginatively reproduce native language speech after a few repetitions, but progressively harder as the language gets more difficult to pronounce relative to the native language. For example, since Catalan might be judged by English speakers to be easier to pronounce than Hindi, a few repetitions of a Catalan sentence might allow English speakers more accurate auditory imagery of the phrase than a few repetitions of a Hindi sentence, resulting in a stronger tendency for the Catalan sentence to transform to music. This hypothesis suggests, then, that the differences between pre and post repetition speech-to-song ratings should be greatest in the native language (English), and progressively smaller as the languages get more difficult to pronounce.

We also wondered whether the higher-level temporal regularity produced by spacing the repetitions at identical intervals was necessary for the illusion to occur. Falk et al. (2014) found that temporal regularity was not necessary, but we used a different method for making the repetitions temporally irregular, a method that made the difference between the regular and irregular versions more salient. We sought to confirm that higher-level temporal regularity was not required for repetition to transform speech into song.

Our study used recordings from an archive of native speakers telling the same story in different languages as stimuli. Half of the recordings were from languages hypothesized to be easier for English speakers to pronounce, and half were from languages hypothesized to be harder for English speakers to pronounce. The English language recording of the story was also included for comparison. Half of the participants heard these recordings in a temporally regular condition, where each repetition followed after an identical temporal interval, and half of the participants heard them in a temporally irregular condition, where the repetition occurred at unpredictable intervals. They rated a phrase from each utterance on a 5-point scale from speech to song both before and after the repetitions. The difference in ratings was taken as an index of the transformation from speech to song. At the end of the session, participants responded to various questions about the languages in the study, including how difficult each might be to pronounce, so that the results could be interpreted in terms of participants' actual ratings of pronunciation difficulty, in addition to the hypothesized categories.

MATERIALS AND METHODS

PARTICIPANTS

The 24 participants (8 male, 16 female) ranged in age from 18 to 22 with a mean age of 19.6 years ($SD = 1.2$). In exchange for participating, they received extra credit in a general music appreciation course aimed at non-majors called Music Lecture. Only one participant reported being enrolled as a music major. Only six of the participants reported formal training in music; all of it at a young age and all of it short lived. Thus, unlike Deutsch et al. (2011), which used musically trained listeners as participants,

this study focuses predominantly on people without formal musical training. Since results did not change when the one music major participant was excluded, we retained all participants in the reported analyses. All participants were native English speakers, and all reported normal hearing.

One participant reported fluency in each of the following languages: Vietnamese, Japanese, Chinese, and Swedish. 12 participants reported some experience with Spanish. Of these, three reported they were fluent, three reported their level of Spanish ability to be advanced, three reported it to be at the beginner level, and the rest reported an intermediate ability. One participant had studied beginning Japanese, and one participant reported proficiency in Vietnamese. The Chinese speaker, the Japanese speaker and one Spanish speaker reported using the language in childhood. None of the participants reported receiving training in any of the languages used in the experiment.

All participants signed an informed consent form before starting the experiment. The protocol was approved by the University of Arkansas Institutional Review Board.

MATERIALS

Seven excerpts from non-tonal languages were selected from the examples used in the Handbook of the International Phonetic Association (1999), available at http://web.uvic.ca/ling/resources/ipa/handbook_downloads.htm. Each excerpt consisted of a person speaking the following utterance “The north wind and the sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take his cloak off, should be considered stronger than the other” in one of seven languages: English, Catalan, Portuguese, French, Croatian, Hindi, or Irish. Aside from English, three languages (Catalan, Portuguese, and French) were hypothesized to be easier for English speakers to pronounce, and three languages (Croatian, Hindi, and Irish) were hypothesized to be harder for English speakers to pronounce. All the languages except for Catalan were spoken by a female.

The mean utterance length was 12.1 s ($SD = 2.7$). For each language, a segment was extracted from the utterance using Audacity 2.0.3. The segment extraction was made at about the three-quarter mark of each utterance. The mean segment length was 2.7 s ($SD = 0.3$).

Two stimuli were created for each language: a temporally regular and a temporally irregular version. The regular versions consisted of the full utterance followed by 10 segment repetitions, each separated by 1000 ms. The irregular versions were more complex, consisting of the full utterance followed by 10 segments, each separated by time intervals that were random percentages (between 1% and 50%) shorter and longer than 1000 ms. For each randomly selected percentage, one interonset interval was created by shortening the 1000 ms span the appropriate amount, and another was created by lengthening it. To increase the salience of the temporal shifts, 400 ms was subtracted from each of the sub-1000 ms values, and added to each of the over-1000 ms values. For example, the randomly selected percentage 15% generated the interonset intervals 450 ms (850–400 ms) and 1550 ms (1150 ms + 400 ms).

A total of 10 interonset intervals were generated from 5 randomly selected percentages. The order of the 10 time interval lengths was randomized. In a few cases, this randomization resulted in two similar time intervals placed back to back (e.g., 510 ms followed by 520 ms); when this happened, one of the time intervals was moved to a different position in the sequence. The object was to create a series of time intervals that made the extraction of meter as unlikely as possible. One advantage of the procedure is that the total duration of all the repetitions was the same in the regular and the irregular condition, eliminating an explanation based on exposure length rather than temporal regularity.

PROCEDURE

Participants were seated at a computer terminal in a Whisper-Room 4' by 4' Enhanced, Double Wall Isolation Booth and outfitted with Sennheiser HD 600 headphones. Instructions were presented on screen and stimuli were presented over the headphones. Participants made all responses using the keyboard and mouse.

Participants were randomly assigned to one of two groups. Group one heard the repetitions in temporally regular form; group two heard the repetitions in temporally irregular form. All other procedures for the two groups were the same.

First, participants answered a series of demographic questions. Next, they performed the task for each of the 7 languages, with the language order randomized. For each language, they were told to listen carefully to an utterance. After the utterance was complete, they were told they would hear a segment from the utterance and be asked to rate it on a scale from 1 to 5, with 1 signifying “sounds exactly like speech” and five signifying “sounds exactly like singing.” They were played the segment, and asked to rate it. Next, they were told they would rehear the utterance, followed by 10 repetitions of the segment, followed by a restatement of the entire utterance. It was explained that they should then rate how the segment sounded within that utterance on the same 1 to 5 scale. Thus, participants rated the segment twice—once before and once after the repetitions. Finally, participants answered a series of questions about each of the languages in the experiment, with the language order randomized. For each language, they were replayed the utterance and asked to enter the name of the language. Next, they rated the familiarity of the language on a scale from 1 to 5. Finally, they were asked how easy they thought it would be to pronounce the words in the language accurately, on a scale from 1 to 5.

RESULTS

A linear mixed model was used with the difference in speech–song ratings pre and post repetition as the dependent variable, language difficulty (native, easy, and hard) and temporal structure (regular vs. irregular) as fixed factors and language (English, Catalan, Portuguese, French, Croatian, Hindi, Irish) as a repeated variable. As shown in **Figure 1**, there was a main effect of the hypothesized pronunciation difficulty of the language on the change in speech–song ratings, $F(2,140) = 6.45$, $p = 0.002$; however, there was no main effect of temporal regularity $F(1,27) = 0.03$, $p = 0.87$.

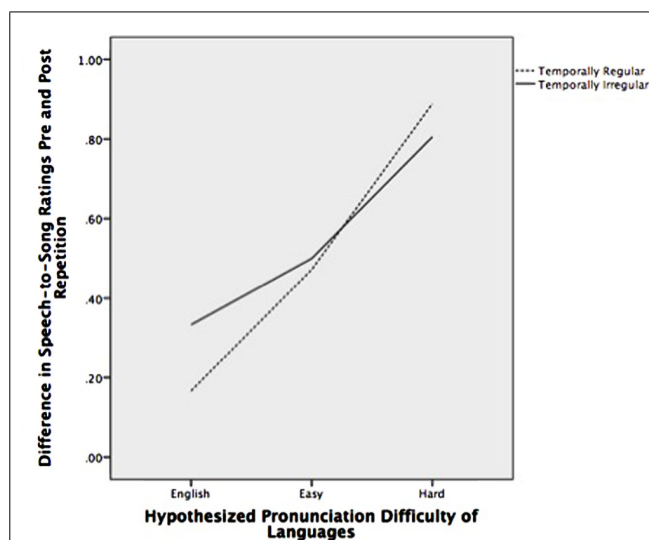


FIGURE 1 | Difference in speech-to-song ratings before and after repetition for temporally regular and irregular stimuli by hypothesized pronunciation difficulty of languages.

Table 1 shows the speech–song ratings for each language category before and after the repetitions. The pre and post repetition ratings were different for every category except English, signifying that a transformation from speech to song occurred in every foreign language, but not the native one. Rating changes from pre to post repetition increased from the native to easy to hard categories, signifying an intensification of the speech-to-song illusion for languages hypothesized to be difficult for native English speakers to pronounce.

Table 2 shows the mean rating change for each language in the temporally regular and irregular condition. Patterns were broadly similar between the two groups, with easier to pronounce languages engendering less dramatic transformation from speech to song and harder to pronounce languages engendering more dramatic transformation. **Table 3** summarizes this effect for each of the three language difficulty categories.

As shown in **Table 4**, participants rated how difficult they thought each language would be to pronounce accurately. Participants' ratings generally correlated with the hypothesized difficulty ratings, but their judgments tended to group into four categories rather than three – Native (English); Easy (Catalan, Portuguese); Medium (French, Croatian); and Hard (Hindi, Irish). In the hypothesized categories, French was grouped with Easy and Croatian with Hard. The data were reanalyzed using these four categories rather than the original three as predictors. Rating change varied significantly according to the difficulty of each language as rated by the participants; $F(3,138) = 5.30$, $p = 0.002$, as shown in **Figure 2**. **Table 5** lists the means for these categories.

Figure 2 shows the speech–song ratings before and after the repetition for each of the three hypothesized language difficulty categories. **Figure 3** shows the same trend for speech–song ratings within each of the four participant-rated language difficulty categories. For each breakdown of the categories, harder-to-pronounce languages were rated as more songlike to begin with;

Table 1 | Mean speech-to-song ratings for each language difficulty category before and after repetition.

Language difficulty	Repetition	Mean	SE	Rate change (post–pre)	Pairwise comparison, pre to post, using sidak adjustment for multiple tests
Native	Pre	1.08	0.18	0.25	$p = 0.17$
	Post	1.33	0.18		
Easy	Pre	1.38	0.11	0.48	$p < 0.001$
	Post	1.86	0.11		
Hard	Pre	1.56	0.11	0.84	$p < 0.001$
	Post	2.40	0.11		

Table 2 | Mean changes in speech-to-song rating from pre to post repetition for each language in each condition.

Temporal structure	Language	Mean rating change
Regular	English	0.17
	Catalan	0.42
	Portuguese	0.25
	French	0.75
	Croatian	0.92
	Hindi	0.83
	Irish	0.92
Irregular	English	0.33
	Catalan	0.33
	Portuguese	0.50
	French	0.67
	Croatian	0.67
	Hindi	0.75
	Irish	1.00

Table 3 | Mean changes in speech-to-song rating from pre to post repetition for each category in each condition.

Temporal structure	Language difficulty	Mean rating change
Regular	English	0.17
	Easy	0.47
	Hard	0.89
Irregular	English	0.33
	Easy	0.50
	Hard	0.81

however, harder-to-pronounce languages also experienced a larger speech-to-song transformation than easier-to-pronounce languages. The native language experienced the least transformation, the easy and medium more, and the hard the most.

Participants also rated the familiarity of each language (also shown in Table 4). These ratings were marginally predictive of speech–song rating changes post repetition, $F(1,155) = 3.80$, $p = 0.05$.

Ninety-six percent (all but one) of the participants correctly identified the English language. Eighty-three percent correctly identified the French language. Every other language was correctly identified by one participant (4% of respondents), except Portuguese, which was correctly identified by 4 (16%). A large percentage of participants misidentified Catalan as Spanish, potentially accounting for the high familiarity ratings for Catalan despite the low success with identifying its name. This pattern underscores the distinction between perceived pronunciation difficulty and mere familiarity; the sound of the French language was quite familiar to participants, and most were able to identify it correctly; however, they still rated the language as moderately difficult to pronounce. Responses did not differ by gender.

In order to ascertain whether there was something inherently more music-like about stimuli in some categories, we used two measures from Tierney et al. (2013), one to assess the degree of fundamental frequency stability across syllables in each utterance, and the other to assess the degree of temporal regularity among syllable stresses. Following the procedure outlined in that paper, we first used Praat to assess the fundamental frequency stability across each syllable in each utterance, by calculating the average fundamental frequency change in semitones per second. In Tierney et al. (2013), utterances that were more likely to transform to song had more within-syllable fundamental frequency stability (less change). Table 6 lists the mean fundamental frequency change for each syllable in each of the seven languages. These means do not vary significantly between pronunciation difficulty categories,

Table 4 | Participants’ difficulty and familiarity ratings for each language compared with their hypothesized categories.

Language	Participants’ difficulty rating	Participants’ familiarity rating	Hypothesized difficulty categories	Participants’ difficulty categories
English	1.38	4.58	Native	Native
Catalan	2.75	3.38	Easy	Easy
Portuguese	3.08	2.79	Easy	Easy
French	3.50	3.17	Easy	Medium
Croatian	3.50	2.83	Hard	Medium
Hindi	4.13	2.29	Hard	Hard
Irish	4.04	2.46	Hard	Hard

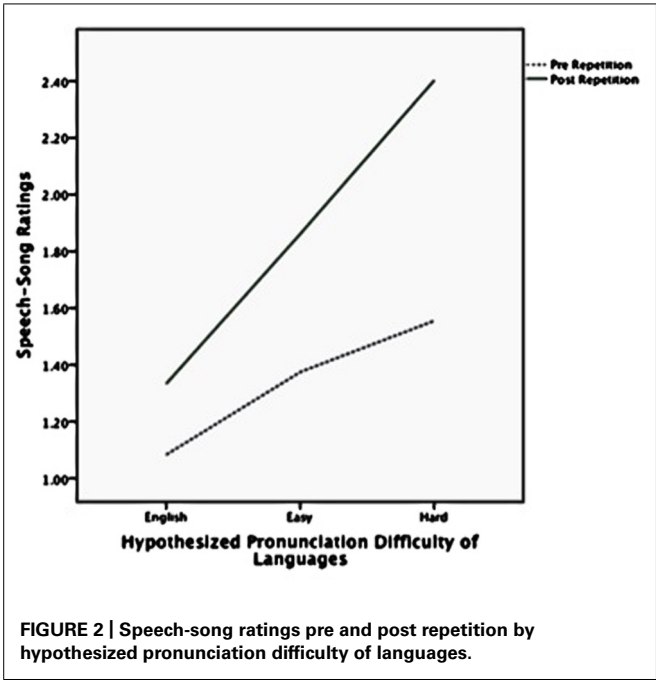


FIGURE 2 | Speech-song ratings pre and post repetition by hypothesized pronunciation difficulty of languages.

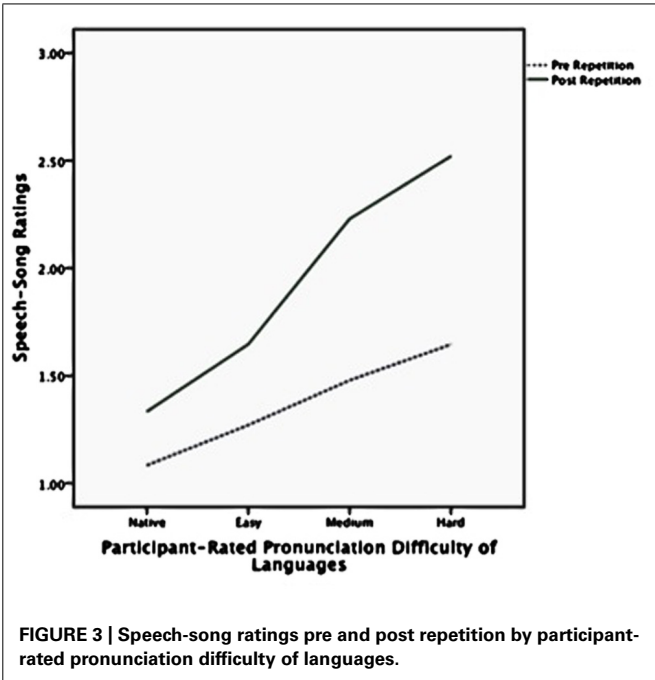


FIGURE 3 | Speech-song ratings pre and post repetition by participant-rated pronunciation difficulty of languages.

except between the Native and Easy participant-rated categories. Stimuli in the category rated by participants as Easy exhibited more fundamental frequency variability per syllable than stimuli in the Native category. If acoustic characteristics were driving the effect, we would expect to see the stimuli with less intrasyllable frequency variability (the Native stimuli) transform to song more easily; however, the opposite effect occurred. This reinforces the notion that the pronunciation difficulty, rather than some more basic acoustic characteristic, influenced the degree to which particular utterances were susceptible to the speech-to-song illusion.

Next, following another procedure in Tierney et al. (2013), we identified the timepoints of the onsets of stressed syllables in each utterance. To assess the temporal regularity of the speech segment, we measured the SD of the duration between successive onsets of stressed syllables. The results for each language are shown in Table 7.

If the results were driven by these acoustic characterizations rather than by pronunciation difficulty, we would have expected to see fundamental frequency change correlate negatively with the size of the speech–song rating change across repetitions; languages with large intrasyllable fundamental frequency changes (lower frequency stability) should transform to song less easily, as shown by smaller speech–song rating changes. Instead, no consistent pattern emerged ($p > 0.05$). We would also have expected to see the standard deviations of stressed syllable onsets vary negatively with speech–song rating change across repetitions; more temporally irregular utterances should transform to song less easily. Again, however, no consistent pattern emerged ($p > 0.05$).

DISCUSSION

Contrary to our initial hypothesis, utterances spoken in languages more difficult to pronounce relative to the listener’s native tongue were actually *more* susceptible to the speech-to-song illusion. Since

Table 5 | Mean speech-to-song ratings for each of the participants’ language difficulty categories.

Participants’ difficulty categories	Repetition	Mean	Rate change (post–pre)	Pairwise comparison, pre to post, using sidak adjustment for multiple tests
Native	Pre	1.08	0.25	$p = 0.17$
	Post	1.33		
Easy	Pre	1.27	0.37	$p = 0.04$
	Post	1.64		
Medium	Pre	1.48	0.75	$p < 0.001$
	Post	2.23		
Hard	Pre	1.65	0.87	$p < 0.001$
	Post	2.52		

Table 6 | Intrasyllable fundamental frequency change for each language.

Language	Mean intrasyllable fundamental frequency change	SD
English	16.61	13.38
Catalan	31.70	13.24
Portuguese	36.34	19.24
French	19.04	10.53
Croatian	32.61	17.24
Hindi	20.52	5.78
Irish	18.09	6.43

Table 7 | Variability of duration between stressed syllable onsets for each language.

Language	SD of durations between stressed syllable onsets
English	0.07
Catalan	0.13
Portuguese	0.18
French	0.07
Croatian	0.16
Hindi	0.11
Irish	0.13

it should have been easier to imaginatively simulate the pronunciation of the syllables in easier to pronounce languages, it seems on first pass that this kind of virtual participation must not be essential to musical attending. Yet there is another way of understanding this result.

The pre-repetition ratings from this experiment show that harder to pronounce languages *started out* sounding more musical to listeners, even before any repetitions had contributed the illusory transformation. When the data are reanalyzed using the same methods except substituting initial ratings rather than rating differences as the dependent variable, there is a main effect of hypothesized language difficulty, $F(2,40) = 5.05, p = 0.008$. If the speech-to-song illusion had been independent of the pronunciation difficulty, the solid lines on **Figures 2** and **3** would have moved up in parallel to the dotted lines, signifying that languages in each of the categories transformed to song after repetition to roughly the same degree. But instead the slope of the solid lines is steeper; the languages that were more difficult to pronounce, and more songlike to start with, became *even more* songlike after repetition than did the easier to pronounce languages. This suggests that when speech circuitry captures acoustic input, it is more resistant to releasing it to other perceptual mechanisms. Speech circuitry seems more likely to capture acoustic input when it is easy to pronounce than when it is hard to pronounce.

To imagine what this release might entail, consider the semantic satiation effect (Severance and Washburn, 1907). It is normally very difficult to perceive a word independently of its semantic

correlate. It takes many repetitions before the meaning starts to disintegrate and the sounds can be heard on their own terms. Across the course of these repetitions, it is almost possible to feel the release as the lexicon’s grip on the word recedes. The harder to pronounce languages may not have elicited as strong a grip by language regions in the first place, allowing repetition to effect a starker shift to song.

Our results supported those in Falk et al. (2014) showing that the illusion occurred whether the repetitions were spaced regularly or irregularly. Temporal regularity does not seem to be a necessary factor in the speech-to-song illusion. The illusion seems to be driven by repetition itself rather than by the emergence of larger-scale temporal regularity.

Additionally, the transformation to song does not seem to be driven by the acoustic characteristic of fundamental frequency stability within syllables, or the acoustic characteristic of regularity between stressed syllable onsets. This strengthens the case that pronunciation difficulty—and perhaps associatedly, the degree to which an utterance is captured by speech circuitry—can influence any particular utterance’s susceptibility to the speech-to-song illusion.

Because the speech-to-song illusion exposes a border between the perception of language and the perception of music, it is especially useful for illuminating how different aspects of acoustic input get emphasized in different contexts. Listeners may start with more acute perception of the prosody and songlike aspects of foreign languages, especially if they are very difficult to pronounce relevant to their native tongue. The more closely acoustic input conforms to the sounds of their native language, the tighter a grip the language circuitry may have on that input, and the less accessible language-irrelevant (or less language relevant) aspects of the sound may be.

To return to the initial hypothesis, although the harder-to-pronounce languages may have been difficult to imaginatively *speak* along with, they might have actually been easier to imaginatively *sing* along with. If language circuitry was less dominant in the processing of utterances in these languages, it may have been easier to disregard formant transitions and tune into the prosodic contour and timing of the pitch changes, features that are already more traditionally musical. Future work might examine people’s capacity for vocal imitation in languages relatively easier or harder to pronounce, similar to Mantell and Pfordresher (2013), to investigate this hypothesis.

REFERENCES

Deutsch, D., Henthorn, T., and Lapidis, R. (2011). Illusory transformation from speech to song. *J. Acoust. Soc. Am.* 129, 2245–2252. doi: 10.1121/1.3562174

Deutsch, D., Lapidis, R., and Henthorn, T. (2008). The speech-to-song illusion. *J. Acoust. Soc. Am.* 124:2471. doi: 10.1121/1.4808987

Falk, S., Rathcke, T., and Dalla Bella, S. (2014). When speech sounds like music. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 1491–1506. doi: 10.1037/a0036858

International Phonetic Association. (1999). *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

Mantell, J. T., and Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition* 127, 177–202. doi: 10.1016/j.cognition.2012.12.008

Margulis, E. H. (2013). *On Repeat: How Music Plays the Mind*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199990825.001.0001

McMullen, E., and Saffran, J. R. (2004). Music and language: a developmental comparison. *Music Percept.* 21, 289–311. doi: 10.1525/mp.2004.21.3.289

- Patel, A. D. (2008). *Music, Language, and the Brain*. New York, NY: Oxford University Press.
- Severance, E., and Washburn, M. F. (1907). Minor studies from the psychological laboratory of Vassar College: the loss of associative power in words after long fixation. *Am. J. Psychol.* 18, 182–186. doi: 10.2307/1412411
- Tierney, A., Dick, F., Deutsch, D., and Sereno, M. (2013). Speech versus song: multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cereb. Cortex* 23, 249–254. doi: 10.1093/cercor/bhs003
- Wallin, N. L., Merker, B., and Brown, S. (2001). *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.

Received: 08 October 2014; accepted: 09 January 2015; published online: 29 January 2015.

Citation: Margulis EH, Simchy-Gross R and Black JL (2015) Pronunciation difficulty, temporal regularity, and the speech-to-song illusion. *Front. Psychol.* 6:48. doi: 10.3389/fpsyg.2015.00048

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

Copyright © 2015 Margulis, Simchy-Gross and Black. 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.



Sex Differences in Music: A Female Advantage at Recognizing Familiar Melodies

Scott A. Miles^{1,2*}, Robbin A. Miranda^{1,3*} and Michael T. Ullman^{1*}

¹ Brain and Language Laboratory, Department of Neuroscience, Georgetown University, Washington, DC, USA,

² Interdisciplinary Program in Neuroscience, Georgetown University, Washington, DC, USA, ³ Infinimetrics Corporation, Vienna, VA, USA

OPEN ACCESS

Edited by:

McNeel Gordon Jantzen,
Western Washington University, USA

Reviewed by:

Cyrille Magne,
Middle Tennessee State University,
USA

Aaron J. Newman,
Dalhousie University, Canada

*Correspondence:

Michael T. Ullman
michael@georgetown.edu;
Scott A. Miles
sam337@georgetown.edu;
Robbin A. Miranda
raw25@georgetown.edu

[†] These authors have contributed
equally to the work.

Specialty section:

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

Received: 30 July 2015

Accepted: 12 February 2016

Published: 01 March 2016

Citation:

Miles SA, Miranda RA and Ullman MT
(2016) Sex Differences in Music:
A Female Advantage at Recognizing
Familiar Melodies.
Front. Psychol. 7:278.
doi: 10.3389/fpsyg.2016.00278

Although sex differences have been observed in various cognitive domains, there has been little work examining sex differences in the cognition of music. We tested the prediction that women would be better than men at recognizing familiar melodies, since memories of specific melodies are likely to be learned (at least in part) by declarative memory, which shows female advantages. Participants were 24 men and 24 women, with half musicians and half non-musicians in each group. The two groups were matched on age, education, and various measures of musical training. Participants were presented with well-known and novel melodies, and were asked to indicate their recognition of familiar melodies as rapidly as possible. The women were significantly faster than the men in responding, with a large effect size. The female advantage held across musicians and non-musicians, and across melodies with and without commonly associated lyrics, as evidenced by an absence of interactions between sex and these factors. Additionally, the results did not seem to be explained by sex differences in response biases, or in basic motor processes as tested in a control task. Though caution is warranted given that this is the first study to examine sex differences in familiar melody recognition, the results are consistent with the hypothesis motivating our prediction, namely that declarative memory underlies knowledge about music (particularly about familiar melodies), and that the female advantage at declarative memory may thus lead to female advantages in music cognition (particularly at familiar melody recognition). Additionally, the findings argue against the view that female advantages at tasks involving verbal (or verbalizable) material are due solely to a sex difference specific to the verbal domain. Further, the results may help explain previously reported cognitive commonalities between music and language: since declarative memory also underlies language, such commonalities may be partly due to a common dependence on this memory system. More generally, because declarative memory is well studied at many levels, evidence that music cognition depends on this system may lead to a powerful research program generating a wide range of novel predictions for the neurocognition of music, potentially advancing the field.

Keywords: music, music cognition, melody, declarative memory, recognition, sex differences, musical training, language

INTRODUCTION

Sex differences have been observed in various cognitive domains. For example, it has been suggested that boys and men have advantages at aspects of visuospatial cognition, while girls and women are better at aspects of verbal cognition (Kimura, 1999; Halpern, 2013). Sex differences in a variety of other domains have also been examined, though inconsistent findings and variability in the magnitude of the effects have led to questions about the existence of sex differences in cognition (Hyde, 2005).

There has been little examination, however, of sex differences in the cognition of music. This seems somewhat surprising, given the surge of research on music cognition in recent decades (Levitin and Tirovolas, 2009; Tirovolas and Levitin, 2011), as well as the apparent sex differences found in verbal cognition. Recent evidence suggests that the processing of language and music may be subserved by at least partially overlapping neural substrates (Patel, 2003; Brown et al., 2006). It is possible that some of the sex differences observed in language are driven by sex differences in these common substrates, suggesting they may extend to music cognition as well.

A relatively small number of neurocognitive studies have examined behavioral sex differences in aspects of music cognition. These studies have focused mainly on the low-level perception of single auditory events, such as those involved in spontaneous and click-evoked otoacoustic emissions (Snihur and Hampson, 2011), transient evoked otoacoustic emissions (Cassidy and Ditty, 2001), and pitch memory (Gaab et al., 2003). Music is, however, a complex phenomenon, consisting of several such events unfolding and interacting in time. It is possible that this focus on the low-level perception of single auditory events has left undetected behavioral sex differences in higher-level aspects of music cognition.

A useful distinction can be made between two higher level aspects of music cognition: knowledge of the general patterns of a musical system, often referred to as knowledge of musical syntax (Koelsch and Friederici, 2003; Koelsch et al., 2013; Sammler et al., 2013; Matsunaga et al., 2014) or *schematic* knowledge (Bharucha, 1994; Tillmann and Bigand, 2001; Huron, 2006); and knowledge of the idiosyncratic representations in music, such as of specific melodies, sometimes referred to as *veridical* knowledge (Bharucha, 1994; Huron, 2006). It has been proposed that much of the aesthetic value of music comes from the adherence to and violation of expectations generated by each of these two types of knowledge (Bharucha, 1994). It has also been proposed that the two types of knowledge can be dissociated, and may depend on different memory systems in the brain (Huron, 2006; Miranda and Ullman, 2007). This proposal is supported by an event-related potential (ERP) study demonstrating a double dissociation between the processing of violations of musical syntax and violations of familiar melodies, which involve idiosyncratic representations (Miranda and Ullman, 2007). Given these dissociations, it is possible that sex differences may be found in either syntactic (schematic) or idiosyncratic (veridical) aspects of music cognition, but not in both.

We are aware of two studies that have examined behavioral sex differences in higher-level aspects of music cognition (Koelsch et al., 2003a,b). Both of these focused on musical syntax, probing responses to violations of syntactic expectations. Though sex differences in electrophysiological brain responses (as measured by ERPs) were observed in both studies, neither found sex differences in performance. Of course, such null effects could be due to many factors. The possibility remains, however, that there are indeed performance advantages for one sex over the other in tasks of higher-level music cognition, but that these involve knowledge of idiosyncratic aspects of music rather than knowledge of musical syntax.

Indeed, as we shall see, some previous evidence suggests that knowledge regarding specific aspects of melodies is stored, at least in part, in declarative memory, a general-purpose memory system that is critical for learning idiosyncratic information in general, including in language. Crucially, declarative memory also shows sex differences, in particular a female advantage, including in the recognition of previously learned idiosyncratic verbal material such as vocabulary items. Thus it is possible that this female advantage might extend to aspects of music cognition that depend on this memory system. Specifically, a female advantage may be expected in the recognition of familiar melodies, which involve idiosyncratic representations. We tested this prediction in the present study by examining the performance of men and women in a familiar melody recognition task.

In the remainder of the Introduction, we first briefly summarize the nature of declarative memory and evidence suggesting sex differences in this system. We then lay out the evidence suggesting that in music cognition, the storage and retrieval of knowledge about specific melodies depends, at least in part, on declarative memory. Finally, we summarize the present study.

Declarative Memory: Overview and Sex Differences

Declarative memory is quite well understood (for reviews, see Ullman, 2004, 2016; Henke, 2010; Squire and Wixted, 2011; Eichenbaum, 2012; Cabeza and Moscovitch, 2013). As its name suggests, this memory system underlies the learning, storage, and retrieval of explicit knowledge, which is available to conscious awareness – although increasing evidence indicates that it also subserves implicit knowledge (Henke, 2010; Ullman, 2016). The system is rooted in the hippocampus and other medial temporal lobe structures. These structures are critical for the learning and consolidation of new knowledge. The subsequent storage of much of this knowledge, however, eventually relies largely on neocortical regions, especially in the temporal lobes. Declarative memory may be specialized for learning arbitrary bits of information and binding them together (Henke, 2010; Squire and Wixted, 2011). Indeed, the system may be necessary for learning such idiosyncratic information. This may help explain evidence that damage to the declarative memory system can severely impair or even

prevent the learning of knowledge about words and other idiosyncratic information (Squire and Wixted, 2011; Ullman, 2016).

Increasing evidence suggests a female advantage at declarative memory, including in idiosyncratic aspects of language (for a discussion and review of the literature, see Ullman et al., 2008). Studies have shown female advantages for a wide variety of episodic memory tasks (which crucially depend on declarative memory), including those testing verbal material, landmarks, objects, object locations, novel faces, and complex abstract patterns (Ullman et al., 2008). A female advantage has also been reported for word learning (Kaushanskaya et al., 2011) and for the retrieval of well-established (previously learned) knowledge, including in tests of vocabulary, lexical retrieval, and verbal fluency (Ullman et al., 2008). These behavioral female advantages are consistent with anatomical sex differences (Ullman et al., 2008). For example, the hippocampus seems to develop at a faster rate, with respect to the rest of the brain, in girls than in boys between the ages of one and sixteen (Pfluger et al., 1999). The behavioral and anatomical sex differences may be at least partly mediated by estrogen, which is found in higher levels in girls and (pre-menopausal) women than in boys and men (Wilson et al., 1998), and affects declarative memory and hippocampal structure and function, through both organizational effects *in utero* and activation effects later on (Phillips and Sherwin, 1992).

Given the dependence of idiosyncratic (and other) aspects of language on declarative memory (Ullman, 2004, 2016), many if not most of the previously reported sex differences in language may in fact be explained by broader, domain-independent sex differences in the declarative memory system (Ullman, 2004, 2016; Ullman et al., 2008). Accordingly, the female advantage at the storage and retrieval of idiosyncratic representations may extend beyond previously studied verbal and non-verbal domains and functions to music cognition – in particular to the storage and retrieval of knowledge about specific melodies.

Melodies, Declarative Memory, and Expected Sex Differences

As we have seen, the cognition of music, like that of language, requires the memorization of specific, idiosyncratic representations, including of familiar melodies. Melodies contain specific sequences of notes that must be veridically learned, even though the sequences are also schematically constrained by the syntax of a musical system – much like words involve particular sequences of phonemes that are also constrained by the rules of phonotactics. Given that declarative memory seems to underlie the learning and storage of knowledge about words, and more generally may be necessary for learning arbitrary bits of information and binding them together, it may be expected that this system is also critical for learning idiosyncratic representations in music, including knowledge about specific melodies.

Some evidence already suggests that this may be the case. In an electrophysiological study, an ERP component characterized as an N400 was observed in response to expectation violations resulting from altered notes within melodies that were well

known (and thus likely to be familiar to participants), but not to violations of notes within novel melodies (Miranda and Ullman, 2007). N400s, which originate in part in the medial temporal lobe (McCarthy et al., 1995; Meyer et al., 2005), and are found in response to a variety of lexical stimuli, as well as to idiosyncratic non-verbal stimuli such as objects and faces (Kutas and Federmeier, 2011), have been linked to declarative memory (Ullman, 2001, 2016). The findings of the music ERP study (Miranda and Ullman, 2007) thus suggest that, like knowledge of these various types of non-musical idiosyncratic information, knowledge about familiar melodies may also be stored in and retrieved from declarative memory.

Given the female advantages observed in other tasks involving declarative memory, including in both the learning of new knowledge and the retrieval of previously learned information, such advantages might also extend to knowledge of idiosyncratic representations in music, including of familiar melodies. We thus predicted a female advantage at recognizing familiar melodies.

The Present Study

To test this prediction we examined the recognition of well-known melodies in adults. We focused on the recognition of already-known melodies, rather than the learning of new melodies, because previous evidence suggests that consolidation – even over the course of months or longer – can significantly affect outcomes (Marshall and Born, 2007; Morgan-Short et al., 2012).

Healthy men and women were presented with both well-known and novel melodies. Participants were asked to indicate as quickly and accurately as possible during the presentation of each melody whether they were familiar with it. Response time (RT) as well as accuracy measures were obtained. RTs typically provide greater variability than accuracy, and minimize the likelihood of ceiling effects. In addition, some previous evidence suggests that the time element may be important in revealing the hypothesized female advantages (Walenski et al., 2008).

We examined both musicians and non-musicians. This allowed us to test how broadly the findings may hold across musical training. Testing across musicians and non-musicians is also important because previous studies examining neural sex differences have found interactions between sex and musical training (Evers et al., 1999; Hutchinson et al., 2003). Musicians might be expected to show stronger representations of familiar melodies simply due to greater exposure (Besson and Faïta, 1995). It is also plausible that members of either sex might have had greater previous exposure to the well-known melodies than members of the other sex. To attempt to address these issues, after each of their timed recognition responses, participants were asked to report a familiarity rating for the melody. By covarying out these ratings in our analyses, we were able to test whether any group differences in performance held even when familiarity was held constant.

All of the stimuli were presented instrumentally. However, since many of the melodies in the study are commonly associated with lyrics, any observed female advantages could in principle be due to advantages in the verbal domain, rather than in familiar

melody recognition itself. We therefore separated the melodies into those that are or are not associated with lyrics, to be able to test whether any sex differences might hold across both.

Finally, it is possible that any observed sex differences in the recognition of melodies might be due to sex differences in basic motor processes, rather than differences in music cognition. To help rule out this possibility, we also gave participants a control task, in which they were asked to respond to single tones as quickly as possible. If the sex differences were due to lower-level motor processes, any differences in the experimental task might also be reflected in the results of the control task.

Overall, given the hypothesis that the female advantage in declarative memory should extend to knowledge about familiar melodies, we predicted that women would show faster and perhaps more accurate recognition of well-known melodies than men. Moreover, we expected this advantage to hold broadly, over both musicians and non-musicians, and across melodies with and without lyrics, and that the advantage would not be fully explained by sex differences in familiarity or in basic motor processes.

MATERIALS AND METHODS

Participants

Participants were right-handed native speakers of American English. They had no known developmental, neurological, or psychiatric disorders. Since familiarity with the well-known melodies used in this study is largely culture-dependent, we selected only participants who had not lived outside of the United States for more than 6 months before the age of 18. Research methods were approved by the Institutional Review Board at Georgetown University. All participants gave written informed consent and received monetary compensation for their participation.

Two groups of participants were tested: 24 men and 24 women. Half of the participants within each group were musicians and half were non-musicians. The musicians had at least 4 years of formal musical training, which was defined as private instrument or voice lessons, or participation in a musical ensemble. The non-musicians had 1 year or less of formal musical training. In our initial analysis of RTs to well-known melodies (described below), we found that two of the participants were outliers (one female musician and one female non-musician), each having a mean RT greater than two standard deviations from the mean RT for their respective participant subgroup. The data from these two participants were excluded and replaced with data from two newly tested participants: one female musician and one female non-musician.

The final two groups of participants therefore also consisted of 24 participants each. **Table 1** shows information for each of the four 12-member subgroups regarding age, years of education, handedness (Oldfield, 1971), years of formal musical training, and (for the musicians only) age when formal musical training began, number of years since last formal musical training, number of instruments played (including voice), and number of participants who still regularly played an instrument or

sang at the time of testing. Results from 2×2 analyses of variance (ANOVAs), with the factors Sex (male/female) and Musical Training (musician/non-musician), confirmed that the four subgroups did not differ significantly in age [Sex: $F(1,44) = 0.20$, $p = 0.656$, Musical Training: $F(1,44) = 0.20$, $p = 0.656$, Sex by Musical Training: $F(1,44) = 0.04$, $p = 0.848$], years of education [Sex: $F(1,44) = 0.34$, $p = 0.561$, Musical Training: $F(1,44) = 0.18$, $p = 0.677$, Sex by Musical Training: $F(1,44) = 0.03$, $p = 0.868$], or handedness [Sex: $F(1,41) = 0.03$, $p = 0.870$, Musical Training: $F(1,41) = 0.46$, $p = 0.500$, Sex by Musical Training: $F(1,41) = 3.27$, $p = 0.078$; note that values were missing from three participants; see **Table 1**]. Importantly, the male and female musicians did not differ significantly in the number of years of formal musical training [$t(22) = 0.46$, $p = 0.653$]; the same was true for male and female non-musicians [$t(22) = 0.67$, $p = 0.511$]. Furthermore, there were no significant differences between male and female musicians regarding the age when musical training began [$t(22) = 1.47$, $p = 0.156$], the number of years since last formal musical training [$t(22) = 1.11$, $p = 0.278$], the number of instruments (including voice) played by each participant [$t(22) = 0.99$, $p = 0.335$], or the number of participants who regularly played a musical instrument or sang at the time of the experiment [$t(22) = 0.80$, $p = 0.430$].

Stimuli

The musical stimuli consisted of 260 melodies ranging from 4.1 to 15.8 s in length (mean = 8.2 s, $SE = 0.17$). The stimuli were created in MIDI format using Finale Version 3.5.1 (Coda Music) and then converted to WAV files with a “grand piano” sound font using MidiSyn Version 1.9 (Future Algorithms). All melodies were in the key of C-major or C-minor. Half of the melodies (130) were segments from well-known tunes (see Appendix, in Supplementary Material), including traditional, folk, children’s, patriotic, holiday, classical, and pop music, as well as themes from movies, television, and Broadway musicals. The other half (130) were novel melodies composed by one of the authors (RM). The novel melodies served only as foils for the familiar melody recognition task, and are not reported or analyzed here. Each novel melody was composed to correspond to one of the well-known melodies. More specifically, the tempo and implied harmony (possible accompanying chords that are not present, but strongly suggested by the sequence of notes in the melody) of each novel melody were identical to those of its corresponding well-known melody; moreover, pitch range was closely matched. Distinctive rhythms were slightly altered in some of the novel melodies in order to minimize false recognition of these melodies based on rhythm. False recognition of novel melodies based on rhythm was not of great concern, in any case, since pitch structure has been found to be a better cue for the recognition of melodies than rhythmic structure (Hébert and Peretz, 1997).

Experimental Task

For the purpose of counterbalancing, the 260 melodies were presented over the course of three runs, with each run containing a similar number (43 or 44) of well-known and novel melodies. Any given well-known melody and its matched novel melody were always presented in separate runs. The order of the three

TABLE 1 | Participant information on age, education, and musical training.

	Age in years	Years of education	Handedness	Years of formal musical training	Age when musical training began	Years since last musical training	Number of instruments played (including voice)	Number of participants currently engaged in instrument or vocal activities
Male musicians	20.2 (2.0)	14.4 (1.5)	93.6 (8.1)	7.8 (2.4)	8.8 (2.7)	3.2 (3.2)	2.3 (0.8)	6
Female musicians	20.3 (1.8)	14.7 (1.8)	88.3 (12.7)	8.3 (2.6)	7.3 (2.0)	4.6 (3.1)	1.9 (1.2)	4
Male non-musicians	20.3 (3.0)	14.1 (2.1)	90.0 (12.6)	0.3 (0.4)	NA	NA	NA	NA
Female non-musicians	20.8 (2.0)	14.5 (1.5)	96.4 (9.2)	0.2 (0.4)	NA	NA	NA	NA

All values are presented as means (with standard deviations), except for the number of participants currently engaged in instrument or vocal activities. Handedness values were computed from the Edinburgh Handedness Inventory (Oldfield, 1971), with a value of 100 indicating maximally right handed; the means and standard deviations for male musicians, male non-musicians, and female non-musicians are based on data from 11 rather than 12 participants, since Edinburgh Handedness Inventory data were missing from one participant in each of these subgroups. NA, not applicable. See main text for details.

runs was counterbalanced across participants, such that for every six participants in each of the subgroups, the runs were presented in all possible orders. The presentation order of well-known and novel melodies was randomized within each run for each participant. Completion time for each run was approximately 15 min.

Melodies were presented on a laptop computer running Microsoft Windows, using Meds 2002 Revision B-1 (UCLA, Los Angeles). Participants were instructed to listen to each melody and to press the space bar as soon as the melody sounded familiar. If the melody was not recognized as familiar, the participant was instructed to wait until the end of the melody and then press the space bar to advance (only the keystrokes that occurred prior to the end of the melody were analyzed as responses). The full melody was presented regardless of when the space bar was pressed.

Immediately after the melody was completed and the space bar was pressed (whichever came last), the participant was prompted to rate the familiarity of the melody from 0 to 100, with 0 being most familiar (we selected this rating scale due to software constraints). Prior to testing, each participant received written instructions specifying that a rating of “0” should indicate “very familiar” melodies that the participant would be able to hum along with, whereas a rating of “100” should indicate melodies that were not familiar at all to the participant. The rating scale was shown on the screen as a horizontal scroll bar with “0” on the left and “100” on the right, with the words “Familiar” and “Unfamiliar” positioned under the left and right sides of the bar, respectively. The participant used a mouse to move a marker on the scroll bar to select the rating of his or her choice. As expected, the participants were indeed broadly familiar with the well-known melodies (mean rating of 17.9, $SD = 9.0$).

All participants were instructed to press the space bar with the left hand and to operate the mouse with the right hand, keeping the left hand just over the space bar at all times in order to minimize RTs. Before starting the experiment, each participant was given a practice run that included eight melodies, four of which were well known and four of which were novel.

Control Task

After five participants had been tested on the experimental task, a control task was added to determine whether possible RT differences between participant groups could be attributed to group-wide differences in basic motor functions. The remaining participants (9 male musicians, 11 male non-musicians, 11 female musicians, and 12 female non-musicians) were given this task after completing all three runs of the experimental task. The control task included 20 tones of different pitches, each 500 ms long, presented at staggered intervals (between 0.3 and 2.1 s) after the participant's previous response. Each participant was instructed to press the space bar with the left hand as soon as s/he heard a tone. Analysis of these RTs for each participant group revealed that three participants (one male musician, one female musician, and one female non-musician) were outliers, each having a mean RT greater than two standard deviations from the mean RT of their corresponding participant subgroup. Data from these participants were excluded

from analyses of this task, and the data from the remaining eight male musicians, 11 male non-musicians, 10 female musicians, and 11 female non-musicians were subjected to full analysis.

RESULTS

Response Times to Well-Known Melodies

Means for the recognition RTs to well-known melodies – that is, the latencies of responses registered during the presentation of these melodies – are shown for each of the four subgroups in the first column of **Table 2**. Prior to analysis, these were natural log transformed. Next, we eliminated very slow trials, which might result from diminished attention to the task. Specifically, for each participant, we eliminated trials with RTs that were greater than two standard deviations (SDs) above that participant's mean. This resulted in the exclusion of a total of 2.69% of responses as outliers (135 out 5,012 correct responses to well-known melodies). To maintain an overall Type I Error probability of 0.05, we applied the Bonferroni correction: since six AN(C)OVAs were performed on the data from the experimental task, the significance level was set at $0.05/6 = 0.0083$.

These transformed and filtered RTs were then entered into a 2×2 ANOVA, with Sex (male/female) and Musical Training (musician/non-musician) as between-group factors. The ANOVA yielded a significant (i.e., following Bonferroni correction) main effect of Sex [$F(1, 44) = 11.09, p = 0.002, \eta_p^2 = 0.201$], with a large effect size (Cohen, 1988), indicating that women were significantly faster than men at responding to well-known melodies; see **Figure 1**. There was no significant main effect of Musical Training [$F(1,44) = 6.27, p = 0.016, \eta_p^2 = 0.125$] (though there was a tendency for musicians to respond faster than non-musicians), nor was there any interaction between Sex and Musical Training [$F(1,44) = 0.001, p = 0.981, \eta_p^2 < 0.001$], suggesting that the female advantage held similarly for musicians and non-musicians.

Familiarity as a Possible Confound

There was a significant correlation between participants' mean recognition RTs and their mean familiarity ratings for well-known melodies [$r(46) = 0.62, p < 0.001$]. Accordingly, it is possible that women were faster at responding to well-known melodies simply because they were more familiar with the melodies, as compared to men. If this were the case, then

including familiarity ratings as a covariate in the analysis would be expected to eliminate the finding of sex differences in RTs.

To examine this issue, a 2 (Sex) \times 2 (Musical Training) analysis of covariance (ANCOVA) was performed on recognition RTs, with the covariate constituting each participant's mean familiarity rating over all of the well-known melodies. The pattern of significance was identical to that described above. The analysis yielded a main effect of Sex [$F(1,44) = 9.79, p = 0.003, \eta_p^2 = 0.185$], with a large effect size, but there was no significant effect of Musical Training [$F(1,44) = 4.04, p = 0.051, \eta_p^2 = 0.086$], nor an interaction between Sex and Musical Training [$F(1,44) = 1.22, p = 0.276, \eta_p^2 = 0.028$]. These results suggest that the effects of Sex on recognition RTs could not be explained by group differences in familiarity. [Note that an ANOVA on mean familiarity ratings over all well-known melodies revealed no effects of Sex ($F(1,44) = 1.57, p = 0.217, \eta_p^2 = 0.034$) or Musical Training ($F(1,44) = 1.96, p = 0.168, \eta_p^2 = 0.043$), nor an interaction between them ($F(1,44) = 2.23, p = 0.142, \eta_p^2 = 0.048$)].

Response Bias as a Possible Confound

It is possible that the observed sex difference in RTs could be explained by differential response biases between the men and women. In particular, if the women had a greater tendency to respond with a recognition key press to all stimuli (novel as well as well-known melodies), this might account for their RT advantage in recognizing familiar melodies.

To address this concern, we performed a 2 (Sex) \times 2 (Musical Training) ANOVA on bias scores [$c = -0.5 * z(\text{Hit rate}) + z(\text{False Alarm rate})$]. This analysis revealed no main effects and no interaction [Sex: $F(1,44) = 2.19, p = 0.146, \eta_p^2 = 0.048$; Musical Training: $F(1,47) = 2.95, p = 0.093, \eta_p^2 = 0.063$; Sex by Musical Training: $F(1,44) = 0.12, p = 0.726, \eta_p^2 = 0.003$], suggesting that there were no differences between the groups in their response biases. This in turn suggests that the advantage for women over men at RTs in recognizing familiar melodies could not be explained by group differences in response biases.

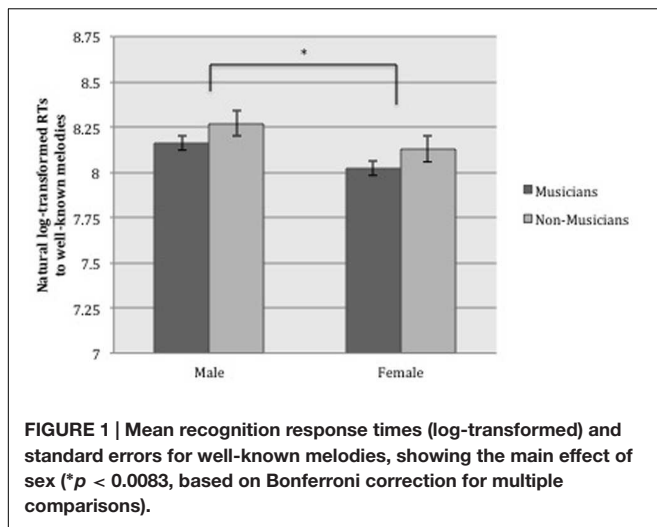
Verbal Ability as a Possible Confound

As mentioned above, although the musical stimuli were presented without lyrics, many of the well-known melodies used in the study are often associated with lyrics. Thus, it might be argued that the female participants' speed advantage at recognizing familiar melodies may have been specifically due to faster RTs for

TABLE 2 | Performance at the melody recognition and control tasks for each subgroup of participants.

	Recognition RTs to well-known melodies (ms)	RTs to tones in the control task (ms)	Recognition accuracy for well-known melodies
Male musicians	3607 (430)	404 (35)	81.2% (9.9%)
Male non-musicians	4037 (729)	352 (13)	70.5% (20.0%)
Female musicians	3169 (451)	390 (24)	88.7% (8.5%)
Female non-musicians	3483 (305)	386 (35)	81.5% (9.8%)

Means (and standard deviations), computed over participants' untransformed data (i.e., without natural log or arcsine transformations). ms, milliseconds.



those melodies associated with lyrics, which women recognized more quickly because of their verbal associations. On this view, the sex differences observed here might be explained by a female advantage at processing verbal information, rather than an advantage at recognizing purely musical aspects of familiar melodies. If this were the case, we might expect to see an interaction between the factors of Sex and “Lyricness” (i.e., whether or not melodies are associated with lyrics). On the other hand, no such interaction would be expected if the sex difference held similarly across melodies that are associated with lyrics and those that are not.

To examine this issue, we first assessed each well-known melody’s association with lyrics by testing six native speakers of American English (four women, two men), ages 19–36 (mean = 23.8 years), with 1–14 years of musical training (mean = 9.2 years), none of whom had lived outside the United States for more than 6 months before age 18. None of these six participants were included in the larger experiment. The participants listened to all of the 130 well-known melodies. After each melody, they were presented with two questions, to which they responded “Yes” or “No.” The questions were presented, one after another, on a computer screen: “(1) Are you familiar with this melody?” and “(2) Do you associate this melody with any lyrics?” For the second question, participants were instructed to answer “Yes” to any melody for which they thought they knew either the actual lyrics or any other (informal) lyrics (e.g., any lyric that they had ever heard or sung with that particular melody). To determine the strength of the association between each melody and its possible lyrics, a “lyric familiarity” score was calculated as the percentage of participants who associated lyrics with the melody, only out of those participants who were familiar with the melody itself (since unfamiliarity with a melody inevitably resulted in unfamiliarity with that melody’s lyrics). Of the 130 melodies, 105 received a lyric familiarity score of 50% or higher (mean = 86.0%) and were considered “lyrics” melodies, while the remaining 25 melodies received a score below 50% (mean = 8.6%) and were considered “no-lyrics” melodies.

We then performed an ANOVA with the between-group factors Sex and Musical Training, and the within-group factor Lyricness (lyrics/no-lyrics melodies). This yielded a main effect of Sex [$F(1,42) = 8.87$, $p = 0.005$, $\eta_p^2 = 0.168$] as well as of Musical Training [$F(1,42) = 7.78$, $p = 0.008$, $\eta_p^2 = 0.150$], both with large effect sizes, but no interaction between Sex and Musical Training [$F(1,42) = 0.05$, $p = 0.833$, $\eta_p^2 = 0.001$]. Importantly, there was no significant main effect of Lyricness [$F(1,42) = 5.450$, $p = 0.024$, $\eta_p^2 = 0.110$], nor any significant interactions between Sex and Lyricness [$F(1,42) = 0.930$, $p = 0.340$, $\eta_p^2 = 0.021$], between Musical Training and Lyricness [$F(1,42) = 0.390$, $p = 0.536$, $\eta_p^2 = 0.009$], nor among Sex, Musical Training, and Lyricness [$F(1,42) = 0.532$, $p = 0.470$, $\eta_p^2 = 0.012$]. This analysis suggests that the RT advantage for women at the recognition of familiar melodies held similarly for melodies that were associated with lyrics and those that were not.

Basic Motor Processes as Possible Confounds

To test for the possibility that sex differences in basic motor processes could account for the women’s RT advantage over men, we administered a control task (see Materials and Methods for details, and **Table 2** for mean RTs by subgroup). Prior to analyses, the RTs were natural log transformed. Next, negative RTs (1.9% of all responses) resulting from premature responses were excluded from analysis. There were no very slow RTs (RTs greater than two SDs above each participant’s mean), so none were eliminated.

The 2 (Sex) \times 2 (Musical Training) ANOVA on these RTs yielded no main effects of Sex [$F(1,36) = 0.094$, $p = 0.760$, $\eta_p^2 = 0.003$] or of Musical Training [$F(1,36) = 0.778$, $p = 0.383$, $\eta_p^2 = 0.021$], and no interaction between them [$F(1,36) = 0.736$, $p = 0.396$, $\eta_p^2 = 0.020$]. This suggests that the group differences in recognition RTs to well-known melodies are not likely to be explained by group differences in basic motor processes (at least those measured by this task).

Accuracy

To examine whether the findings of a female advantage might extend beyond RTs, we also examined accuracy. Each participant’s percentage of correct recognition responses to all well-known melodies constituted the dependent variable in this analysis; see **Table 2**. These percentages were arcsine-transformed prior to analyses. A 2 (Sex) \times 2 (Musical Training) ANOVA revealed no significant main effects, that is, neither of Sex [$F(1,47) = 6.755$, $p = 0.013$, $\eta_p^2 = 0.132$], nor of Musical Training [$F(1,47) = 6.189$, $p = 0.017$, $\eta_p^2 = 0.123$], nor an interaction between them [$F(1,47) = 0.008$, $p = 0.928$, $\eta_p^2 < 0.001$].

DISCUSSION

This study examined the prediction that women would have an advantage at recognizing familiar melodies, as compared to men. Indeed, women were significantly faster than men at recognizing familiar melodies, based on a Bonferroni corrected significance level. This sex difference yielded a large effect size (defined as $\eta_p^2 \geq 0.138$; Cohen, 1988). The result held across musicians

and non-musicians, as reflected by the absence of an interaction between sex and musicianship.

Unlike in the case of recognition RTs, we did not find a significant female advantage in our measure of accuracy, after correcting for multiple comparisons. However, as discussed above, accuracy is a less sensitive indicator of performance than RT. Perhaps for this reason, a female advantage was found for RTs but not accuracy in a recent study of lexical retrieval (Walenski et al., 2008). Indeed, it is possible that women are more accurate than men in their familiarity recognition responses, but our sample sizes (two groups of 24 participants each) were not large enough to demonstrate this effect. The finding of a significant female advantage in accuracy prior to correction for multiple comparisons is consistent with this view – especially since Bonferroni correction is quite conservative.

The female RT advantage was not explained by a number of potentially confounding factors. First, there were no significant group differences in various demographic variables that might have otherwise accounted for the observed advantages. The four subgroups (male musicians, male non-musicians, female musicians, and female non-musicians) did not differ in age, years of education, or handedness. Additionally, the male and female musicians did not differ in years of formal musical training, and likewise for the male and female non-musicians. The male and female musicians also did not differ regarding the age when their musical training began, the years since their last musical training, the number of instruments played (including voice), or the number of participants in each subgroup who were currently engaged in instrumental or vocal activities. Second, the advantages were not explained by group differences in familiarity with the well-known melodies. It might be suggested that the women were faster at recognizing well-known melodies because they were simply more familiar with the melodies than the men. However, the female advantage was observed even when familiarity ratings were covaried out. Third, since there were no group differences or interactions on bias scores, group differences in bias are also not likely to explain the observed female advantage. Fourth, the advantages could not be fully accounted for by associations between the melodies and lyrics. It might be argued that a female advantage in the verbal domain could explain the sex difference observed here, rather than an advantage in the recognition of familiar melodies *per se*. In particular, since quite a few of the melodies in the study are associated with lyrics, it might have been the case that the female advantage held only or mainly for these items. However, there were no significant interactions between lyricness and sex, suggesting the speed advantage for women held across melodies that are and are not commonly associated with lyrics. This in turn suggests that the findings cannot be explained by a female advantage purely in the verbal domain. Fifth, it is not likely that group differences in basic motor processes accounted for the female advantage in melody recognition, since there were no significant differences between the groups in performance during a simple tone detection control task. This suggests that at least the basic motor processes examined in this

task did not differ between the groups, and thus were not likely to have explained the observed differences in melody recognition.

We suggest instead that the sex differences in recognition RTs are at least partly explained by the previously reported female advantage at declarative memory. As discussed in Section “Introduction,” this advantage has been found not only for learning new material, but also for the retrieval of previously learned material, as was tested in the present study. Together with independent electrophysiological evidence suggesting that the processing of familiar melodies depends at least in part on declarative memory (see Introduction, and Miranda and Ullman, 2007), the data from the present study suggest that the female advantage at declarative memory may indeed extend to music cognition, in particular to the retrieval of stored knowledge about melodies. However, given that this is the first study to examine sex differences in familiar melody recognition, some caution in interpreting the findings is warranted; see Section “Limitations and Future Studies” below for further discussion.

The claim that knowledge about familiar melodies depends on declarative memory does not presuppose that this is the only memory or other cognitive system involved in the learning, storage, or retrieval of such knowledge. For example, attention and working memory systems may be expected to play roles, at least in part because of their interactions with the declarative memory system for learning and retrieval (Ullman, 2004, 2016).

A role for declarative memory in stored knowledge of melodies also would not preclude additional roles for this system in music cognition. One interesting possibility is that declarative memory might, to some extent, play redundant roles with procedural memory in certain aspects of music cognition – for example, in learning and processing syntactic (schematic) knowledge, that is, knowledge about the regularities of musical systems. Increasing evidence suggests that such redundancy between declarative and procedural memory exists for language and other domains (Ullman, 2004, 2016). For example, individuals or groups with declarative memory advantages, or with deficits of procedural memory, appear to rely more on declarative memory, relative to procedural memory, for various grammatical functions (Ullman and Pullman, 2015; Ullman, 2016). Of particular interest here, girls and women seem to rely more on declarative memory than boys and men for aspects of grammar, likely due in part to the female advantages at declarative memory (Ullman et al., 2008; Ullman, 2016). It is plausible that such a sex difference might be found analogously for syntactic aspects of music cognition. Intriguingly, two studies have reported more bilateral negativities in girls and women than boys and men in response to syntactic anomalies within musical stimuli (Koelsch et al., 2003a,b). Although these negativities had primarily anterior distributions, their bilaterality suggests the possibility that they may be related to N400s, consistent with a greater dependence of musical syntactic processing on declarative memory in females than males. Indeed, such redundancy is consistent with the lack of sex differences in performance reported in these studies, since the errors

may be processed equally well in the two systems (Ullman, 2004, 2016). However, this interpretation of these studies should be treated with caution, and future research is needed.

Although the goal of the present study was to test sex differences in melody recognition, and the observed female advantage was indeed the most robust effect, an advantage of musicians over non-musicians was also found. Musicians showed a significant (i.e., following Bonferroni correction) RT advantage in the analysis with lyricness as a factor, as well as RT and accuracy advantages that were significant or borderline significant prior to Bonferroni correction, in other analyses. The cause of this apparent effect is not entirely clear. One possibility is that musicians simply have greater familiarity with the melodies. Another possibility is that the training involved in learning to perform music results in improvements in declarative memory. Indeed, some evidence hints at declarative memory improvements from other types of training (Draganski et al., 2006; Woollett and Maguire, 2011). Alternatively (or in addition), perhaps individuals with better declarative memory (and maybe other advantages as well) are more likely to become musicians, or to stick with musical training. Finally, the fact that a significant musician advantage only emerged in the analysis with lyricness may be attributed to a reduction of the error term in this analysis due the inclusion of this factor. Future studies examining the apparent musician advantage at familiar melody recognition seem warranted.

Implications

The present study has implications for various disciplines and endeavors. In the domain of music cognition, together with the ERP results of Miranda and Ullman (2007), it provides evidence suggesting that knowledge of melodies depends at least in part on declarative memory. This, in turn, has further implications. First of all, it suggests that, like language, music cognition may depend on general-purpose brain systems. We emphasize, however, that portions of these systems could become subspecialized for aspects of music cognition, both evolutionarily and developmentally, as has been suggested for language (Ullman, 2004, 2016).

Importantly, because declarative memory has been well studied at multiple levels (including its behavioral, computational, neuroanatomical, physiological, cellular, molecular, genetic, and pharmacological correlates), this vast independent knowledge about the memory system could also pertain to music cognition (Ullman, 2004, 2016). Thus, as with language, linking music cognition to declarative memory could generate a wide range of novel predictions that there might be no independent reason to make based on the more circumscribed study of music cognition alone (Ullman, 2016). For example, the anatomical, developmental and genetic correlates of declarative memory might also be expected to underlie music, in particular ways. An understanding of the dependence of music cognition on declarative memory may therefore provide important insights regarding the evolution and development of music cognition. Overall, linking music to

declarative memory could prove to be a powerful approach that may lead to substantial advances in the understanding of the neurocognition of music. These advances could include efforts to understand how knowledge about specific melodies contributes to the development, within the brains of listeners, of musical expectations. Such an understanding is crucial to the effort to understand how music is able to evoke powerful emotions and pleasure in listeners.

Linking aspects of music cognition to declarative memory could also help clarify commonalities between the cognition of music and language. Unlike proposals that have suggested that music cognition has ‘piggybacked’ on language circuitry (e.g., Pinker, 1997), here we suggest that the language/music neurocognitive commonality lies at least in part with declarative memory (also see Miranda and Ullman, 2007). On this view, this general-purpose system may underlie the cognition of both language and music, rather than music cognition depending directly on language circuitry. Of course, such a common dependence on declarative memory does not preclude any additional ‘piggybacking’ of music cognition on language (or vice versa) – either in portions of declarative memory that have evolutionarily or developmentally become specialized for language, or in any additional circuitry that might be specific to language (Ullman, 2004, 2016). Moreover, a joint language/music dependence on declarative memory does not preclude any additional joint dependence on other brain systems, including working memory and procedural memory (Miranda and Ullman, 2007).

From the perspective of memory systems, the findings presented here and in Miranda and Ullman (2007) underscore the view that declarative memory seems to underlie a wide range of types of knowledge, functions, domains, and modalities, and is not limited to episodic (event) and semantic (fact) knowledge as has traditionally been suggested (for discussion, see Ullman and Pullman, 2015; Ullman, 2016).

From a language perspective, the findings of the present study underscore the plausibility that highly specialized areas of knowledge, which are moreover found across human cultures, may depend importantly on general-purpose brain systems. This underscores the plausibility of the reliance of language on declarative memory and other general-purpose cognitive systems (Ullman, 2004, 2016).

The findings also have important implications for the study of sex differences. They reveal, for the first time, that women seem to have an advantage at recognizing familiar melodies, as compared to men. The findings also show for the first time that there are behavioral sex differences in higher-level aspects of music cognition. Importantly, the observed female superiority does *not* seem to be due to an exclusively verbal advantage, since the female advantage did not interact with lyricness. This not only strengthens the evidence of an overall female advantage at tasks involving declarative memory, and evidence of its extension to the domain of music, but also crucially throws doubt on the claim that the female advantage at many verbal tasks is specific to the verbal domain. Rather, many if not most of the previously observed female

advantages at verbal tasks may instead be partly if not largely due to female advantages in declarative memory (Ullman et al., 2008). This controversial issue seems to warrant further research.

The findings of the present study also have educational and clinical implications. Pedagogical techniques that have been shown to improve learning and retention in declarative memory, such as spaced presentation and the testing (retrieval practice) effect (Cepeda et al., 2006; Roediger and Butler, 2011; Ullman and Lovelett, under review) may also be expected to enhance music learning, in particular the learning of specific melodies, just as they seem to enhance language, in particular the learning of words (Ozemir et al., in preparation; Ullman and Lovelett, under review). Also, understanding the neural substrates of the learning of knowledge about specific melodies could help guide music therapy, an approach that has been shown to be effective in helping patients with conditions involving deficits of both language and memory, such as aphasia and Alzheimer's disease (Norton et al., 2009; Ueda et al., 2013).

Limitations and Future Studies

This study has various limitations. Perhaps most importantly, it does not *directly* tie the observed sex differences in melody recognition to female advantages in declarative memory. Thus, some other factor or factors could at least partially account for the findings. For example, it is possible that females generally make quicker decisions than males regarding information on which confidence is not high, or that sex differences in other aspects of music cognition involved in melody recognition could lead to the observed findings.

However, the sex differences found here were *predicted* on the basis of independent findings of sex differences in declarative memory, and moreover, analyses suggested they were not due to a wide range of potentially confounding factors or alternative explanations. Additionally, previous evidence has linked knowledge of familiar melodies to declarative memory (Miranda and Ullman, 2007). Together, this suggests that the study provides initial support for the view that the female advantage at declarative memory extends to music cognition, and can at least partly explain the observed sex differences in melody recognition.

Importantly, the findings constitute a useful foundation for future studies to more directly examine the issue. For example, further studies might examine whether participants' ability at melody recognition correlates with their ability at various declarative memory tasks. One could also examine the neural underpinnings of the observed sex differences, for example with fMRI or ERPs. Further research should also probe how broadly the apparent female advantage might hold, for example across different musical systems (e.g., in the Javanese or North Indian classical musical systems), age groups, and so on. One might also examine whether the female advantage would also hold in the actual identification of melodies (as in the game show "Name that Tune"), or whether it might be limited to binary familiarity judgments. Given the importance of sex hormones

on cognition, including declarative memory (Hausmann et al., 2000; Hausmann, 2005; Ullman et al., 2008), the influence of estrogen and other sex hormones, and their variability throughout the menstrual cycle, also warrant investigation. For example, further studies may examine whether the findings obtained here might be due in part to elevated levels of estrogen during particular points along the menstrual cycle. The possibility of cultural influences (Hoffman et al., 2011) on the observed sex differences should also be investigated. Although the control task examined very simple aspects of auditory processing (i.e., the participants heard various tones and responded with a simple key press to any tone), the task did not directly examine pitch processing (since the same response was made to any pitch), nor other aspects of auditory processing such as rhythm. Future studies could control for such aspects of auditory processing, for example with different responses for different pitches or rhythms. It would additionally be highly informative to examine the *learning* of new specific melodies, and whether and how this depends on declarative memory. Finally, future studies might extend the investigation of music cognition to procedural memory, to examine whether and how the learning or use of musical syntax, or other aspects of music, might depend on this system.

CONCLUSION

This study revealed, for the first time, a female advantage at recognizing familiar melodies, as compared to males. This pattern, which showed a large effect size, held across musicians and non-musicians, and over melodies with and without commonly associated lyrics. We predicted the female advantage based on independent evidence suggesting both a female advantage at declarative memory and a dependence of knowledge of familiar melodies on this system. Although some caution is warranted because this is the first study to examine sex differences in melody recognition, the findings lend support to the hypothesis that knowledge pertaining to specific melodies indeed depends on declarative memory, which in turn leads to a female advantage at familiar melody recognition, thanks to a more general female advantage at declarative memory. The finding that the female advantage held across melodies that are and are not associated with lyrics argues against the view that the commonly observed female advantage at tasks involving verbal (or verbalizable) material is best explained by a sex difference specific to the verbal domain. Additionally, because declarative memory also underlies language, it seems likely that the cognitive commonalities between music and language may be explained, at least in part, by a common dependence on declarative memory. More generally, because declarative memory is well studied at many levels, evidence that aspects of music cognition rely on this system could lead to a powerful research program capable of generating a wide range of novel predictions for the neurocognition of music.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

We thank Matthew Walenski, Matthew Moffa, Jocelyn Curchack, Marco Piñeyro, Harriet Bowden, João Veríssimo, Natasha Janfaza, Rochelle Tractenberg, and Benson Stevens for help on

various aspects of this study. Support was provided to RAM from an NSF Graduate Research Fellowship, to SM from an NSF Graduate Research Fellowship, and to MU from NIH R01 MH58189 and R01 HD049347.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00278>

REFERENCES

- Besson, M., and Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: comparison of musicians with nonmusicians. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 1278–1296. doi: 10.1037/0096-1523.21.6.1278
- Bharucha, J. J. (1994). "Tonality and expectation," in *Musical Perceptions*, eds R. Aiello and J. Sloboda (Oxford: Oxford University Press), 213–239.
- Brown, S., Martinez, M. J., and Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *Eur. J. Neurosci.* 23, 2791–2803. doi: 10.1111/j.1460-9568.2006.04785.x
- Cabeza, R., and Moscovitch, M. (2013). Memory systems, processing modes, and components: functional neuroimaging evidence. *Perspect. Psychol. Sci.* 8, 49–55. doi: 10.1177/1745691612469033
- Cassidy, J. W., and Ditty, K. M. (2001). Gender differences among newborns on a transient otoacoustic emissions test for hearing. *J. Music Ther.* 38, 28–35. doi: 10.1093/jmt/38.1.28
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., and Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychol. Bull.* 132, 354–380. doi: 10.1037/0033-2909.132.3.354
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Draganski, B., Gaser, C., Kempermann, G., Kuhn, H. G., Winkler, J., Buchel, C., et al. (2006). Temporal and spatial dynamics of brain structure changes during extensive learning. *J. Neurosci.* 26, 6314–6317. doi: 10.1523/JNEUROSCI.4628-05.2006
- Eichenbaum, H. (2012). *The Cognitive Neuroscience of Memory: An Introduction*, 2 Edn. Oxford: Oxford University Press.
- Evers, S., Dannert, J., Rödding, D., Rötter, G., and Ringelstein, E. B. (1999). The cerebral haemodynamics of music perception. *Brain* 122, 75–85. doi: 10.1093/brain/122.1.75
- Gaab, N., Keenan, J. P., and Schlaug, G. (2003). The effects of gender on the neural substrates of pitch memory. *J. Cogn. Neurosci.* 15, 810–820. doi: 10.1162/089892903322370735
- Halpern, D. F. (2013). *Sex Differences in Cognitive Abilities*, 4th Edn. Abingdon: Psychology Press.
- Hausmann, M. (2005). Hemispheric asymmetry in spatial attention across the menstrual cycle. *Neuropsychologia* 43, 1559–1567. doi: 10.1016/j.neuropsychologia.2005.01.017
- Hausmann, M., Slabekorn, D., Van Goozen, S. H. M., Cohen-Kettenis, P. T., and Güntürkün, O. (2000). Sex hormones affect spatial abilities during the menstrual cycle. *Behav. Neurosci.* 114, 1245–1250. doi: 10.1037/0735-7044.114.6.1245
- Hébert, S., and Peretz, I. (1997). Recognition of music in long-term memory: are melodic and temporal patterns equal partners? *Mem. Cognit.* 25, 518–533. doi: 10.3758/BF03201127
- Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nat. Rev. Neurosci.* 11, 523–532. doi: 10.1038/nrn2850
- Hoffman, M., Gneezy, U., and List, J. A. (2011). Nurture affects gender differences in spatial abilities. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14786–14788. doi: 10.1073/pnas.1015182108
- Huron, D. B. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT press.
- Hutchinson, S., Lee, L. H. L., Gaab, N., and Schlaug, G. (2003). Cerebellar volume of musicians. *Cereb. Cortex* 13, 943–949. doi: 10.1093/cercor/13.9.943
- Hyde, J. S. (2005). The gender similarities hypothesis. *Am. Psychol.* 60, 581–592. doi: 10.1037/0003-066X.60.6.581
- Kaushanskaya, M., Marian, V., and Yoo, J. (2011). Gender differences in adult word learning. *Acta Psychol.* 137, 24–35. doi: 10.1016/j.actpsy.2011.02.002
- Kimura, D. (1999). *Sex and Cognition*. Cambridge, MA: The MIT Press.
- Koelsch, S., and Friederici, A. D. (2003). Toward the neural basis of processing structure in music. Comparative results of different neurophysiological investigation methods. *Ann. N. Y. Acad. Sci.* 999, 15–28. doi: 10.1196/annals.1284.002
- Koelsch, S., Grossmann, T., Gunter, T. C., Hahne, A., Schröger, E., and Friederici, A. D. (2003a). Children processing music: electric brain responses reveal musical competence and gender differences. *J. Cogn. Neurosci.* 15, 683–693. doi: 10.1162/jocn.2003.15.5.683
- Koelsch, S., Maess, B., Grossmann, T., and Friederici, A. D. (2003b). Electric brain responses reveal gender differences in music processing. *Neuroreport* 14, 709–713. doi: 10.1097/00001756-200304150-00010
- Koelsch, S., Rohrmeier, M., Torrecuso, R., and Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15443–15448. doi: 10.1073/pnas.1300272110
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Levitin, D. J., and Tirovolas, A. K. (2009). Current advances in the cognitive neuroscience of music. *Ann. N. Y. Acad. Sci.* 1156, 211–231. doi: 10.1111/j.1749-6632.2009.04417.x
- Marshall, L., and Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends Cogn. Sci.* 11, 442–450. doi: 10.1016/j.tics.2007.09.001
- Matsunaga, R., Yokosawa, K., and Abe, J. I. (2014). Functional modulations in brain activity for the first and second music: a comparison of high- and low-proficiency bimusicals. *Neuropsychologia* 54C, 1–10. doi: 10.1016/j.neuropsychologia.2013.12.014
- McCarthy, G., Nobre, A. C., Bentin, S., and Spencer, D. D. (1995). Language-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. *J. Neurosci.* 15, 1080–1089.
- Meyer, P., Mecklinger, A., Grunwald, T., Fell, J., Elger, C. E., and Friederici, A. D. (2005). Language processing within the human medial temporal lobe. *Hippocampus* 15, 451–459. doi: 10.1002/hipo.20070
- Miranda, R. A., and Ullman, M. T. (2007). Double dissociation between rules and memory in music: an event-related potential study. *Neuroimage* 38, 331–345. doi: 10.1016/j.neuroimage.2007.07.034
- Morgan-Short, K., Finger, I., Grey, S., and Ullman, M. T. (2012). Second language processing shows increased native-like neural responses after months of no exposure. *PLoS ONE* 7:e32974. doi: 10.1371/journal.pone.0032974
- Norton, A., Zipse, L., Marchina, S., and Schlaug, G. (2009). Melodic intonation therapy. *Ann. N. Y. Acad. Sci.* 1169, 431–436. doi: 10.1111/j.1749-6632.2009.04859.x
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

- Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681. doi: 10.1038/nn1082
- Pfluger, T., Weil, S., Weis, S., Vollmar, C., Heiss, D., Egger, J., et al. (1999). Normative volumetric data of the developing hippocampus in children based on magnetic resonance imaging. *Epilepsia* 40, 414–423. doi: 10.1111/j.1528-1157.1999.tb00735.x
- Phillips, S. M., Sherwin, B. B. (1992). Effects of estrogen on memory function in surgically menopausal women. *Psychoneuroendocrinology* 17, 485–495. doi: 10.1016/0306-4530(92)90007-T
- Pinker, S. (1997). *How the Mind Works*. New York, NY: W. W. Norton and Company.
- Roediger, H. L. III, and Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends Cogn. Sci.* 15, 20–27. doi: 10.1016/j.tics.2010.09.003
- Sammeler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H. J., et al. (2013). Co-localizing linguistic and musical syntax with intracranial EEG. *Neuroimage* 64, 134–146. doi: 10.1016/j.neuroimage.2012.09.035
- Snihur, A. W., and Hampson, E. (2011). Sex and ear differences in spontaneous and click-evoked otoacoustic emissions in young adults. *Brain Cogn.* 77, 40–47. doi: 10.1016/j.bandc.2011.06.004
- Squire, L. R., and Zola-Morgan, J. (1991). The cognitive neuroscience of human memory since H. M. *Ann. Rev. Neurosci.* 34, 259–288. doi: 10.1146/annurev-neuro-061010-113720
- Tillmann, B., and Bigand, E. (2001). Global context effect in normal and scrambled musical sequences. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 1185–1196. doi: 10.1037/0096-1523.27.5.1185
- Tirovolas, A. K., and Levitin, D. J. (2011). Music perception and cognition research from 1983 to 2010: a categorical and bibliometric analysis of empirical articles in Music Perception. *Music Percept.* 29, 23–36. doi: 10.1525/mp.2011.29.1.23
- Ueda, T., Suzukamo, Y., Sato, M., and Izumi, S. I. (2013). Effects of music therapy on behavioral and psychological symptoms of dementia: a systematic review and meta-analysis. *Ageing Res. Rev.* 12, 628–641. doi: 10.1016/j.arr.2013.02.003
- Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Biling. Lang. Cogn.* 4, 105–122. doi: 10.1017/S1366728901000220
- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92, 231–270. doi: 10.1016/j.cognition.2003.10.008
- Ullman, M. T. (2016). “The declarative/procedural model: a neurobiological model of language learning, knowledge and use,” in *The Neurobiology of Language*, eds G. Hickok and S. A. Small (Amsterdam: Elsevier), 953–968.
- Ullman, M. T., Miranda, R. A., and Travers, M. L. (2008). “Sex differences in the neurocognition of language,” in *Sex on the Brain: From Genes to Behavior*, eds J. B. Becker, K. J. Berkley, N. Geary, E. Hampson, J. Herman, and E. Young (New York, NY: Oxford University Press), 291–309.
- Ullman, M. T., and Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. *Neurosci. Biobehav. Rev.* 51, 205–222. doi: 10.1016/j.neubiorev.2015.01.008
- Walenski, M., Mostofsky, S. H., Gidley-Larson, J. C., and Ullman, M. T. (2008). Brief report: enhanced picture naming in autism. *J. Autism. Dev. Disord.* 38, 1395–1399. doi: 10.1007/s10803-007-0513-y
- Wilson, I. D., Foster, D. W., Kronenberg, H. M., Larsen, P. R. (eds). (1998). *Williams Textbook of Endocrinology*, 9th Edn (Philadelphia, PA: W. B. Saunders Company).
- Woollett, K., and Maguire, E. A. (2011). Acquiring “the Knowledge” of London’s layout drives structural brain changes. *Curr. Biol.* 21, 2109–2114. doi: 10.1016/j.cub.2011.11.018

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 Miles, Miranda and Ullman. 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.

Associations between musical abilities and precursors of reading in preschool aged children

Franziska Degé*, Claudia Kubicek and Gudrun Schwarzer

Department of Developmental Psychology, Justus Liebig University Giessen, Giessen, Germany

OPEN ACCESS

Edited by:

McNeel Gordon Jantzen,
Western Washington University, USA

Reviewed by:

Mireille Besson,
Centre National de la Recherche
Scientifique – Institut
de Neurosciences Cognitives
de la Méditerranée, France
Chantel Spring Prat,
University of Washington, USA

*Correspondence:

Franziska Degé,
Department of Developmental
Psychology, Justus Liebig University
Giessen, Otto-Behaghel-Straße 10F,
35394 Giessen, Germany
franziska.dege@psychol.uni-
giessen.de

Specialty section:

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

Received: 31 March 2015

Accepted: 03 August 2015

Published: 17 August 2015

Citation:

Degé F, Kubicek C and Schwarzer G
(2015) Associations between musical
abilities and precursors of reading
in preschool aged children.
Front. Psychol. 6:1220.
doi: 10.3389/fpsyg.2015.01220

The association between music and language, in particular, the overlap in their processing results in the possibility to use one domain for the enhancement of the other. Especially in the preschool years music may be a valuable tool to train language abilities (e.g., precursors of reading). Therefore, detailed knowledge about associations between musical abilities and precursors of reading can be of great use for designing future music intervention studies that target language-related abilities. Hence, the present study investigated the association between music perception as well as music production and precursors of reading. Thereby, not only phonological awareness, the mostly studied precursor of reading, was investigated, but also other precursors were examined. We assessed musical abilities (production and perception) and precursors of reading (phonological awareness, working memory, and rapid retrieval from long-term memory) in 55 preschoolers (27 boys). Fluid intelligence was measured and controlled in the analyses. Results showed that phonological awareness, working memory, and rapid retrieval from long-term memory were related to music perception as well as to music production. Our data suggest that *several* precursors of reading were associated with music perception as well as music production.

Keywords: musical abilities, precursors of reading, phonological awareness, working memory, preschoolers

Introduction

The non-musical benefits of music lessons have intrigued the public and fascinated researchers. However, the potential of music lessons to enhance cognitive abilities (e.g., IQ or memory) still remains under discussion. There are only few experiments that clearly demonstrated a causal relationship between music lessons and cognitive abilities (e.g., IQ; Schellenberg, 2004). The vast majority of studies only established an association between music lessons and cognitive abilities, but did not examine the effect of music lessons on cognitive abilities. However, regarding language-related abilities, several studies showed that music interventions could cause improvements. Particularly, intervention studies demonstrated that music training can enhance vocabulary (Moreno et al., 2011), reading (Moreno et al., 2009), and phonological awareness – a precursor of reading (Degé and Schwarzer, 2011). These results indicate that music interventions might be able to support the development of language-related abilities (e.g., reading and phonological awareness). Until now, the majority of studies have primarily focused on music perception abilities (i.e., rhythm, pitch, meter, and timbre) and their relationship to language-related abilities, mostly addressing only one precursor (i.e., phonological awareness). However, there are other precursors of reading that have been identified: working memory, and rapid retrieval from long-term memory

(Jansen et al., 2002). But up to now there has not been much research concerning musical abilities and precursors of reading other than phonological awareness. Another issue concerns the inclusion of musical abilities: although musical abilities comprise perception as well as production abilities, the association between musical production abilities and precursors of reading is still understudied. Therefore, our exploratory study is aimed at investigating the association between musical perception abilities as well as musical production abilities and *several* precursors of reading, such as phonological awareness, working memory, and rapid retrieval from long-term memory.

Explanations for the Associations between Music and Language

Explanations draw upon the functional overlap of brain structures that are involved in music and speech processing (Besson et al., 2011). It is assumed that domain-general abilities (abilities that are used for music as well as for speech processing) build the basis of the connection between music and language. Explanatory approaches only differ in the auditory features that they promote as the connecting domain-general abilities. Besson et al. (2011) assume that musicians have an enhanced sensitivity to auditory parameters (e.g., frequency and duration) that are important for music processing. Because these parameters are involved in music and speech processing, the higher sensitivity results in a more elaborated auditory perception of speech. This enhancement on lower levels of processing can also create an advantage for higher levels of speech processing (e.g., phonological processing).

While the former approach focuses on frequency and duration, another one postulates timing as the important acoustical feature. Tierney and Kraus (2014) put forward the precise auditory timing hypothesis that explains the connection between auditory motor entrainment and phonological skills. They assume that music training requires entrainment, and entrainment necessitates the precise perception of acoustic event timing. Hence, musical training over extended periods of time might result in higher timing precision in the automated representation of acoustic events in the auditory system. This higher precision also benefits speech sound perception, which is important for phonological skills. Taken together, it is highly likely that common auditory features trained by musical experience have a positive effect on speech processing. As different explanations focus on different auditory features, it might be possible that more mechanisms drive the effect of music training on language abilities (Tierney and Kraus, 2014).

Music Training and Vocabulary

There is correlational (i.e., quasi-experimental) and experimental evidence of associations between music training and vocabulary. Piro and Ortiz (2009) found in a study with second-grade students that *music training* was associated with improvements in vocabulary. Also, musically trained 10-year-old children outperformed their untrained counterparts in vocabulary tasks (Forgeard et al., 2008). However, these correlational studies allowed no inferences about causation; it therefore remained

unclear whether music training actually caused improvements in vocabulary. Nonetheless, there is one experimental study that used a computerized music-listening training as a potential intervention for vocabulary (Moreno et al., 2011). Preschoolers were pseudo-randomly assigned to music training or visual arts training. Before and after 4 weeks (20 days) of training, vocabulary was assessed. Only the music group showed increases from pre- to post-test on the vocabulary task, which unequivocally demonstrates that the music intervention indeed enhanced vocabulary.

Music Aptitude, Music Training, and Reading

Regarding *music aptitude*, Anvari et al. (2002) showed that music perception abilities (pitch and rhythm) were related to reading abilities in 4-year-old children. This association remained reliable when phonological awareness was held constant. In 5-year-old children only pitch perception was associated with reading, whereas rhythm was not related to reading. The relationship between pitch perception and reading was again independent of phonological awareness. This independence might indicate that, apart from phonological awareness, other precursors of reading might mediate the association between musical abilities and reading. In 7- to 8-year-old children, Douglas and Willatts (1994) demonstrated an association between reading and pitch perception as well as rhythm perception.

Correlational research regarding *music training* and reading skills demonstrated an association between music training and spelling in 8- to 9-year-old children (Hille et al., 2011). Additionally, an association between music training and reading was revealed; this association disappeared when socioeconomic status (SES) was controlled (Hille et al., 2011). However, a study with 6- to 9-year-old children revealed an association between length of music training and reading comprehension even when SES was controlled (Corrigall and Trainor, 2011). Even some experimental or longitudinal studies supported the assumption of a causal relation between music training and reading skills. Douglas and Willatts (1994) performed a small scale intervention study. They trained the music group and a discussion control group for 6 months, and measured reading skills before and after the training. The music group showed small improvements in reading, whereas the control group did not improve from pre- to post-test. In a recent study with a larger sample size, Moreno et al. (2009) trained 8-year-old children either in music or painting and tested reading skills (i.e., reading of inconsistent words; inconsistent with respect to phoneme grapheme mapping and pronunciation) before and after 6 months of training. The music group improved its reading skills, while the painting group showed no improvements.

Two meta-analyses were conducted on music training and reading skills. Butzlaff (2000) reported a strong association between music training and reading with respect to correlational studies. Though, for experimental studies he found no reliable results because some studies demonstrated a positive effect of music training on reading skills, while others did not. However, a more recently conducted meta-analysis (Standley, 2008) did reveal a modest but significant positive effect of music training on reading skills.

Music Aptitude, Music Training, and Phonological Awareness

Phonological awareness describes the insight into the phonological structure of language. It refers to the ability to analyze and manipulate language on two levels. On the word level, phonological awareness describes the ability to manipulate and analyze larger phonological units (e.g., rhyming and blending words). On the phoneme level, phonological ability refers to the ability to analyze and manipulate the individual sound units (phonemes) within a word. It has been shown that phonological awareness is an important precursor of later reading ability (Pratt and Brady, 1988; Bruck, 1992).

Phonological awareness is related to music aptitude as well as to music training. A few studies investigated the effect of *music training* on phonological awareness, and revealed that music training can indeed enhance phonological skills. In a quasi-experiment, Gromko (2005) investigated the effect of music training on phonological awareness. Children in the treatment kindergarten received music training for 4 months, while children in the control kindergarten received no treatment. Gromko (2005) revealed significantly greater gains in phonological awareness in the treatment kindergarten children than in the control kindergarten children. The pseudo-random assignment of the preschoolers to the treatment and the control group, however, precludes firm conclusions. Children were not assigned randomly on an individual basis, but the kindergartens were chosen to be the control or the treatment kindergarten. Therefore, children in the treatment group may have systematically differed (e.g., in SES) from children in the control group. Furthermore, the control group did not receive an alternative training. Degé and Schwarzer (2011) randomly assigned preschoolers to music production and perception training, phonological skills training, and sports training. Children in all three groups received training for 20 weeks. The phonological skills training as well as the music training enhanced phonological awareness, whereas the sports training did not. The advancement of phonological awareness in the music group and in the phonological skills group was mainly driven by improvements in phonological awareness on the word level. These results demonstrated unequivocally that music training could enhance phonological awareness.

Several studies revealed positive associations between *music aptitude* and phonological awareness. Huss et al. (2011) found a relationship between metrical perception and phonological awareness in a sample of 10-year-old dyslexic and non-dyslexic children. Also, Norton et al. (2005) found that audiation (the ability to hear, feel, and comprehend music for which the sound is not physically present) was correlated with phonological awareness. The test that was applied in this study comprises pitch perception and rhythm perception. Hence, phonological awareness was associated with a global music perception factor. Moreover, in a study with 4- and 5-year-old children, Lamb and Gregory (1993) investigated the association between phonological awareness and pitch as well as timbre perception. They found that pitch perception (but not timbre perception) was related to phonological awareness (Lamb and Gregory, 1993). The association between pitch perception and phonological

awareness was reliable even when age and fluid intelligence was controlled. By testing the same age group as Lamb and Gregory (1993), Anvari et al. (2002) investigated the association between musical abilities and phonological awareness. They assessed melody perception, chord perception, chord analysis, rhythm perception, and rhythm production. However, they ran factor analyses and found that one factor for the 4-year-olds and two factors for the 5-year-olds represented the musical abilities best. Thus, for their further analyses, they created one music macro-variable that contained all assessed musical abilities for the 4-year-old children and two music variables (i.e., pitch and rhythm) for the 5-year-old children. For the 4-year-old children, the music macro-variable was correlated with phonological awareness, and for the 5-year-olds both music variables (i.e., pitch and rhythm) were correlated with phonological awareness. In sum, these data suggest that pitch perception as well as rhythm perception is associated with phonological awareness.

Objectives

Music intervention studies showed that music training could enhance vocabulary, reading skills, and phonological awareness. An explanation could be that music interventions that mostly target music perception abilities enhanced music listening skills. This enhancement was then accompanied by improvements in speech perception and might have promoted some aspects of language processing (Corrigall et al., 2013). This language enhancing potential of music training might be particularly valuable. Therefore, analyzing the association between musical abilities and language-related abilities might be of great importance in order to understand relevant underlying mechanisms, and to design effective interventions for reading or phonological awareness. So far, studies investigating the association between musical abilities and language-related abilities focused on music perception abilities only. Thus, the first aim of our study was to examine the association between precursors of reading in preschoolers and music perception as well as music production. Furthermore, we applied a musical test battery that assessed music perception (melody perception, pitch perception, rhythm perception, meter perception, tone length perception) as well as music production (singing a song, rhythm production, meter production) on a detailed level. This detailed investigation represents a more complete picture of associations between musical abilities and precursors of reading. This will in turn help to identify reasonable tasks for music training programs that focus, for example, on the improvement of reading.

Anvari et al. (2002) revealed that there might be abilities other than phonological awareness involved in the association between musical abilities and reading skills. However, most studies have investigated only one precursor of reading (i.e., phonological awareness). Therefore, our second aim was to investigate the association between musical abilities and several precursors of reading (i.e., phonological awareness, working memory, rapid retrieval from long-term memory) in preschool children. This exploratory approach which comprises several correlational analyses will broaden our understanding of associations between musical abilities and precursors of reading.

Materials and Methods

The study was conducted in full accordance with the Ethical Guidelines of the German Association of Psychologists (DGPs). In accordance with the ethical guidelines mentioned above informed consent was obtained from the parents for each participant.

Participants

The sample comprised 55 preschoolers (27 boys, 28 girls; mean age = 75.13 months; SD = 4.02 months) from five different kindergartens in Giessen, Germany. Participants had a mean fluid intelligence score of $M = 113.25$ (SD = 11.06), see below. Hence, average fluid intelligence scores were higher than the published norms. The sample showed diversity with respect to parents' education: for 40% of the children neither parent had a university degree, for 27.3% of the children one parent had a university degree, and for 32.7% of the children both parents had a university degree.

Measures

Possible confounding variables such as age, gender, SES, and intelligence were assessed. As predictor variables musical abilities and as criterion variables precursors of reading were measured.

Parents completed a demographic questionnaire that asked for information about their education as one possible measure of SES. Mothers' and fathers' education was initially coded as a dichotomous variable (0 for "no university degree" and 1 for "a university degree"). For the statistical analyses, parents' education was collapsed into a single variable: 0, 1, or 2 parents with a university degree. Although parental income or parental profession could also be used as a measure of SES, we decided to ask for parents' education, because in former studies parents have been mostly willing to share this information. This questionnaire was also used to assess gender and age of the participants.

To measure intelligence, the culture fair test (CFT1; Weiß and Osterland, 1977), which measures fluid intelligence, was employed. The test consisted of five subtests (substitution, mazes, classification, similarities, and matrices) and was administered in groups that did not exceed six children. The duration of test administration was 60 min including instructions and breaks. Age norms were used to determine the intelligence score for each participant.

Precursors of reading were measured with the Bielefelder Screening (BISC; Jansen et al., 2002). The screening allows the assessment of different precursors of reading: phonological awareness, working memory, and rapid retrieval from long-term memory.

Phonological awareness was assessed with the following four subtests: rhymes, word segmentation, phoneme synthesis, and phoneme recognition. Two tests (rhymes and word segmentation) measured phonological awareness for large phonological units (words) and the other two subtests (phoneme synthesis and phoneme recognition) assessed phonological awareness for small phonological units (phonemes). Each subtest consisted of two to four practice items and 10 test items.

In the rhymes task, children were asked whether two words rhyme or do not rhyme (e.g., Do train and rain rhyme?). Children were asked to segment words by clapping their hands in the word segmentation task. The phoneme synthesis task requested the synthesis of the initial sound and the remaining word (e.g., m-ouse) into one word. The phoneme recognition task required recognition of a particular phoneme in a word (e.g., Is there a "u" in elephant?). A composite score of all of the subtest scores was calculated. For the statistical analyses the subtest scores as well as the composite score were used.

Working memory was assessed with recall of non-sense words (e.g., gor-ki-ra-si-mi). In this task the children had to listen to a non-sense word and recall it immediately after listening. The test consisted of two practice items and 10 test items. Seven of these test items were four syllables long, two were five syllables long, and one test item was six syllables long. The practice items were three and four syllables long, respectively. If any syllable was recalled incorrectly or omitted the non-sense word was marked as incorrect. For each correctly recalled word children received one point.

Rapid retrieval from long-term memory was assessed with a speeded naming task. This task consisted of two parts. In the first part, children were asked to name the appropriate color of black and white fruits as fast as possible. Reaction time was measured and the amount of correct answers was registered. This task was designed to assess rapid retrieval from long-term memory. In the second part, children were asked to name the appropriate color of wrongly colored fruits (e.g., yellow salad or blue lemon). As in the first part, reaction time and correctness were registered. This task should assess interference of rapid retrieval from long-term memory. The interference score was built by subtracting the "correct answer and time score" from part one from the "correct answer and time score" from part two. This difference score (small difference indicating little interference) was then transformed into the interference score (high score indicating little interference).

Musical abilities (music perception and production abilities) were measured with the music screening for children (Jungbluth and Hafen, 2005). We applied five subtests to measure music perception: melody perception, pitch perception, rhythm perception, tone length perception, and meter perception. Each subtest consisted of 10 items with increasing difficulty. All subtests required same-different discriminations, wherein the position or direction of changes should be indicated. In the melody perception subtest, children were asked to identify a change and the position of the change in two consecutive melodies. On the pitch perception task, children had to decide whether the second tone was higher, lower, or the same as the first. In the subtest rhythm perception, children had to decide whether two short rhythmic patterns were the same or different and they had to indicate the position of the difference. In the subtest tone length perception, two tones of the same pitch were played to the children and they had to indicate whether the second tone was longer, shorter, or of same duration as the first. In the subtest meter perception, two different meters were presented. Each consisted of five beats. The children

had to decide whether the second meter was faster, slower, or the same as the first. A music perception total score was built for each child by adding the scores reached in each subtest.

We used three subtests to assess music production abilities: singing a song, rhythm production, and meter execution. In the subtest singing a song, children learned and sang a 4-bar-song. Two independent raters analyzed the recorded performance. They rated melody contour, rhythm, starting tone, and intonation. The interrater reliability was $r = 0.94$. In the subtest rhythm production, 10 rhythms of increasing levels of difficulty were presented from a CD and reproduced by the children on a keyboard. This subtest was recorded and scored by two independent raters, as well. The interrater reliability was $r = 0.96$. On the meter execution subtest, children had to perform four different tasks, while they always listened to the same piece of music. Firstly, children had to walk in the meter of the musical piece. Secondly, children had to clap their hands in the meter of the musical piece. In task three and four, children had to clap their hands in the meter of the music and continue clapping in the correct meter when the music had stopped. They continued until they heard “stop” from the CD. All four tasks were recorded and were coded by two raters. The interrater reliability was $r = 0.80$. In addition to the subtest scores, a music production total score was built for each child by adding the scores reached in each subtest.

Procedure

Prior to testing, the informed consent of the parents was attained. Additionally, the demographic questionnaire was sent to the participants and they were asked to hand them back to a person working in the kindergarten. This way the experimenter could collect them. All test sessions took place in the kindergarten during their daily routine. All kindergartens provided a quiet room for the test sessions. The intelligence test was performed in groups of five to six children with two experimenters present. One experimenter instructed the children, while the second experimenter made sure that the children remained seated and concentrated on their own sheet of paper. The precursors of reading were assessed in individual sessions. Two experimenters applied the tests that measured music perception abilities in groups of five to six children. Again one experimenter instructed the group, while the other experimenter made sure that the children were focused on the tasks. The items of the music perception test were presented via speakers and the children indicated their responses on a sheet of paper. Items were coded with little cartoons or pictures to guide the children through the test. In case the children had to indicate the positions of differences (melody and rhythm) they could mark on the sheet of paper the notes or drums, respectively. Music production abilities were assessed in individual test sessions. The assessments were performed on consecutive days. All in all, children participated in four group sessions (two sessions intelligence test, two sessions music perception tasks) and two individual sessions (one session precursors of reading, one session music production tasks). At the end of the project, each child received a present and a certificate for participation.

Results

Preliminary Analyses

We correlated possible confounding variables (age, gender, SES, and IQ) with music perception as well as production abilities. Only a significant correlation between IQ and music production ($r = 0.321$, $p = 0.017$) and IQ and music perception ($r = 0.497$, $p = 0.000$) was found. Age, gender, and SES were not significantly correlated with musical abilities (Table 1). Also possible confounding variables and precursors of reading were correlated. Age was significantly correlated with interference of rapid retrieval ($r = 0.313$, $p = 0.020$). SES was significantly correlated with working memory ($r = 0.269$, $p = 0.047$). IQ was significantly correlated with working memory ($r = 0.325$, $p = 0.016$). Gender did not show any significant correlations with any precursor of reading, for details see Table 1. Because IQ was significantly correlated with musical abilities and precursors of reading, it was controlled in further statistical analyses.

Principal Analyses

Correlations between Musical Abilities and Precursors of Reading

Correlations (with IQ partialled out) between musical abilities (music perception and production total scores) and precursors of reading revealed significant associations between musical abilities and phonological awareness, working memory, and rapid retrieval from long-term memory (Table 2).

Phonological awareness was correlated with music perception ($r = 0.417$, $p = 0.002$) and music production ($r = 0.650$, $p = 0.000$). Also, working memory was correlated significantly with music perception ($r = 0.363$, $p = 0.007$) as well as music production abilities ($r = 0.280$, $p = 0.040$). We observed a significant correlation for interference of rapid retrieval and music perception ($r = 0.337$, $p = 0.013$). Higher scores in music perception (total) were associated with less interference

TABLE 1 | Correlations among possible confounding variables [age, gender, socioeconomic status (SES), and IQ], musical abilities (music perception and music production), and precursors of reading (phonological awareness, working memory, and rapid retrieval from long-term memory).

	Age	Gender	SES	IQ
Music perception	0.137	0.119	0.091	0.497**
Music production	0.206	0.245	0.074	0.321*
Phonological awareness	0.086	0.057	−0.046	0.243
Working memory	0.120	−0.051	0.269*	0.325*
Rapid retrieval	0.155	0.069	−0.051	0.041
Interference of rapid retrieval	0.313	0.136	0.070	0.143

* $p < 0.05$, ** $p < 0.001$.

TABLE 2 | Correlations between precursors of reading and music production (total) and music perception (total).

	Phonological awareness	Working memory	Rapid retrieval	Interference of rapid retrieval
Music perception	0.417 (0.002)	0.363 (0.007)	0.156 (0.260)	0.337 (0.013)
Music production	0.650 (0.000)	0.280 (0.040)	0.019 (0.889)	0.262 (0.056)

p-values in parentheses. Significant results in bold.

in rapid retrieval. Music production was not significantly correlated with interference of rapid retrieval ($r = 0.262$, $p = 0.056$). In further analyses the associations between musical abilities and precursors of reading were explored in more detail.

Correlations between Music Perception, Music Production, and Phonological Awareness

The more detailed (on subtest level) analyses showed that phonological awareness total score was significantly correlated (IQ controlled) with pitch perception ($r = 0.321$, $p = 0.018$), rhythm perception ($r = 0.335$, $p = 0.018$), and tone length perception ($r = 0.322$, $p = 0.018$). The subtests of phonological awareness on the word level were significantly correlated with rhythm perception (rhymes: $r = 0.320$, $p = 0.018$) and marginally significantly correlated with tone length perception (rhymes: $r = 0.264$, $p = 0.053$) and pitch perception (word segmentation: $r = 0.269$, $p = 0.050$). For the subtests regarding phonological awareness on the phoneme level only a correlation between phoneme recognition and tone length perception ($r = 0.347$, $p = 0.010$) was revealed. All the other correlations between phonological awareness and music perception were not significant (see **Table 3** for more details). The phonological awareness total score was significantly correlated with singing a song ($r = 0.529$, $p = 0.000$) and rhythm production ($r = 0.632$, $p = 0.000$). Both phonological awareness subtests on the word level were significantly correlated with singing a song (rhymes: $r = 0.347$, $p = 0.010$; word segmentation: $r = 0.344$, $p = 0.011$)

and rhythm production (rhymes: $r = 0.332$, $p = 0.014$; word segmentation: $r = 0.473$, $p = 0.000$). Additionally, rhymes were also significantly correlated with meter execution ($r = 0.443$, $p = 0.001$). Only one subtest operating on the phoneme level was correlated with singing a song. Phoneme synthesis was significantly correlated with singing a song ($r = 0.387$, $p = 0.004$), whereas the correlation between phoneme recognition and singing a song was not significant ($r = 0.257$, $p = 0.061$). However, phoneme recognition was significantly correlated with rhythm production ($r = 0.391$, $p = 0.003$). None of the other correlations between phonological awareness and music production reached significance (see **Table 3**).

Correlations between Music Perception, Music Production, and Working Memory

We calculated partial correlations with IQ controlled between working memory and the music perception subtests. Working memory was associated significantly with rhythm perception ($r = 0.435$, $p = 0.001$). For the other subtests (melody perception, pitch perception, tone length perception, and meter perception) no significant correlations were found (**Table 4**). Partial correlations between the music production subtests (singing a song, rhythm production, and meter execution) and working memory revealed a marginal significant relationship between rhythm production and working memory ($r = 0.265$, $p = 0.053$). Singing a song and meter execution were not significantly correlated with working memory (**Table 4**).

TABLE 3 | Associations between phonological awareness (total score, rhymes, word segmentation, phoneme synthesis, phoneme recognition) and the subtests of music perception as well as the subtests of music production.

	Music perception					Music production		
	Melody perception	Pitch perception	Rhythm perception	Tone length perception	Meter perception	Singing a song	Rhythm production	Meter execution
Phonological awareness total	0.093 (0.505)	0.321 (0.018)	0.335 (0.013)	0.322 (0.018)	0.246 (0.073)	0.529 (0.000)	0.632 (0.000)	0.205 (0.138)
Rhymes	0.013 (0.928)	0.189 (0.172)	0.320 (0.018)	0.264 (0.053)	0.125 (0.369)	0.347 (0.010)	0.332 (0.014)	0.443 (0.001)
Word segmentation	0.010 (0.945)	0.269 (0.050)	0.170 (0.219)	0.110 (0.427)	0.154 (0.265)	0.344 (0.011)	0.473 (0.000)	−0.014 (0.922)
Phoneme synthesis	0.138 (0.319)	0.104 (0.453)	0.108 (0.437)	0.139 (0.317)	0.214 (0.120)	0.387 (0.004)	0.227 (0.098)	−0.054 (0.700)
Phoneme recognition	0.164 (0.237)	0.114 (0.410)	0.231 (0.093)	0.347 (0.010)	0.166 (0.230)	0.257 (0.061)	0.391 (0.003)	0.175 (0.206)

p-values in parentheses. Significant results in bold.

TABLE 4 | Associations among working memory and the subtests of music perception as well as the subtests of music production.

	Music perception					Music production		
	Melody perception	Pitch perception	Rhythm perception	Tone length perception	Meter perception	Singing a song	Rhythm production	Meter execution
Working memory	0.147 (0.290)	0.102 (0.456)	0.435 (0.001)	0.225 (0.102)	0.214 (0.120)	0.208 (0.131)	0.265 (0.053)	0.154 (0.267)

p-values in parentheses. Significant results in bold.

Correlations between Music Perception, Music Production, and Interference of Rapid Retrieval from Long-Term Memory

Partial correlations (IQ controlled) between interference of rapid retrieval from long-term memory and music perception revealed a significant association with rhythm perception ($r = 0.344$, $p = 0.011$). No significant associations between interference of rapid retrieval from long-term memory and any other tested music perception ability were found (see **Table 5** for details). With respect to music production abilities, only rhythm production was significantly correlated with interference of rapid retrieval from long-term memory ($r = 0.295$, $p = 0.030$). Neither singing a song nor meter execution was significantly related with interference of rapid retrieval from long-term memory (see **Table 5**).

Although this study was exploratory in nature and therefore several correlations were calculated, it should be taken into account that this exploratory approach affects the alpha level. Because in the principal analyses 64 correlations were calculated, it might be reasonable to adjust the alpha level. On an adjusted alpha level ($p = 0.0008$) only four correlations remained significant: the correlation between music production and phonological awareness total score, the correlation between singing a song and phonological awareness total score, the correlation between rhythm production and phonological awareness total score, and the correlation between rhythm production and word segmentation.

Discussion

In the present study, we investigated associations between music perception as well as music production and precursors of reading. In particular, we assessed musical abilities on a detailed (subtest) level and examined their associations with several precursors of

reading (phonological awareness, working memory, and rapid retrieval from long-term memory).

The total scores of music perception as well as of music production were associated with phonological awareness. Furthermore, we found correlations between the music perception total score and working memory as well as between the music production total score and working memory. Finally, the music perception total score correlated significantly with interference of rapid retrieval from long-term memory.

All in all, our results indicated that music production as well as music perception was associated with several precursors of reading. Thus, our study complements already existing studies by showing associations with music production abilities. With respect to production abilities, rhythm production was associated with three of the precursors, singing a song was correlated with all phonological awareness subtests, and meter execution showed only one significant relationship with rhymes, one subtest of phonological awareness. Our assessment of several precursors of reading and their relationship to musical abilities showed that above and beyond phonological awareness, working memory, and interference of rapid retrieval from long-term memory were associated with musical abilities. Taken together, our data suggest that there are several links between musical abilities and precursors of reading.

In a next step we explored the revealed associations between musical abilities and phonological awareness, working memory, and interference of rapid retrieval from long-term memory in more detail (i.e., on the subtest level).

Phonological awareness on the word level (rhymes and word segmentation) was correlated with pitch perception (marginally), rhythm perception, and tone length perception (marginally). Regarding phonological awareness on the word level and music production abilities, we found associations between singing a song, rhythm production, and meter execution. Phonological awareness on the phoneme level (phoneme recognition) was

TABLE 5 | Associations among interference of rapid retrieval from long-term memory and the subtests of music perception as well as the subtests of music production.

	Music perception					Music production		
	Melody perception	Pitch perception	Rhythm perception	Tone length perception	Meter perception	Singing a song	Rhythm production	Meter execution
Interference of rapid retrieval from long-term memory	0.101 (0.465)	0.092 (0.510)	0.344 (0.011)	0.191 (0.167)	0.228 (0.098)	0.158 (0.253)	0.295 (0.030)	0.196 (0.157)

p-values in parentheses. Significant results in bold.

correlated with tone length perception. Furthermore, we revealed a significant correlation between phonological awareness on the phoneme level (phoneme recognition) and rhythm production and a significant correlation between phonological awareness on the phoneme level (phoneme synthesis) and singing a song. Thus, first of all, our study is in accordance with earlier findings that indicate a relationship between phonological awareness and musical abilities (Lamb and Gregory, 1993; Anvari et al., 2002). Furthermore, our results demonstrated that phonological awareness on the word level is involved in more associations with musical abilities than phonological awareness on the phoneme level. This fits to the results by Degé and Schwarzer (2011), who found an improvement especially for phonological awareness on the word level after a music training program. Our research indicates that more associations between phonological awareness on the word level and musical abilities are evident. Thus, it might have been easier to observe an effect of music training on phonological awareness on the word level in the Degé and Schwarzer study. Hence, musical abilities had more ways of interacting with phonological awareness on the word level. Therefore, a positive effect might have emerged earlier for phonological awareness on the word level than for phonological awareness on the phoneme level. Additionally, it is also possible that music training is not suitable to train phonological awareness on the phoneme level. However, we also found associations between phonological awareness on the phoneme level and musical abilities. Therefore, it should be possible to train phonological awareness on the phoneme level with a music training program. Because we found only few associations between musical abilities and phonological awareness on the phoneme level, it could be speculated that these associations might be weaker as compared to the word level, which in turn suggests that simply increasing training length might drive effects of music training on phonological awareness on the phoneme level. It remains to future research to test this specific hypothesis. Like Lamb and Gregory (1993), our results showed an association between phonological awareness and pitch perception that remained reliable after controlling for fluid intelligence. Hence, this association was not due to the influence of a third variable (i.e., fluid intelligence), but a direct association between pitch perception and phonological awareness. Contradictory to the results by Lamb and Gregory (1993), we found also reliable associations between phonological awareness and rhythm perception as well as rhythm production. This finding, though, is supported by the results of Anvari et al. (2002), who also found associations between pitch as well as rhythm and phonological awareness. Taken together, pitch and rhythm seem to be related to phonological awareness. Thus, both aspects of musical abilities might contribute to positive effects of music training on phonological awareness.

With respect to meter, we only found a significant relationship between meter production and phonological awareness. Meter perception was not significantly correlated with phonological awareness. Hence, our results are only partly in accordance with the study by Huss et al. (2011); in so far that both studies found a link between meter and phonological awareness. Huss et al. (2011) found correlations between phonological awareness and

metrical perception. Interestingly, we revealed an association between phonological awareness and meter production (i.e., meter execution). Possibly, the difference in the applied meter perception tasks was responsible for the slightly different results. In the study by Huss et al. (2011) beats per minutes remained stable and only the accents were changed between consecutive stimuli, whereas in the task we applied two sequences differed in beats per minutes and not in accents. In the light of the precise auditory timing hypothesis (Tierney and Kraus, 2014) it is surprising that meter did not show more associations with several aspects of phonological awareness. Because meter execution and meter perception heavily rely on a precise perception of auditory timing, someone might have expected that meter is strongly related to phonological awareness.

For working memory results regarding perception and production showed an association between working memory and the rhythm subtests. Working memory was correlated with rhythm perception and rhythm production (marginally significant). These findings are in line with the results of Anvari et al. (2002). They showed an association between musical abilities and working memory. Considering the applied tasks to assess rhythm perception (compare two drum sequences), rhythm production (reproducing a rhythm on a keyboard), and working memory (reproduce non-sense words), it seems reasonable to conclude that all of them were processed by the phonological loop (i.e., the subsystem responsible for auditory/verbal input of Baddeley's working memory model; Baddeley, 1986). Therefore, the phonological loop was possibly the common basis of these tasks and reflected in task outcome. This might indicate that in intervention studies music training might have trained phonological loop processes and thereby produced a benefit for language processing. Moreover, the rhythm perception and production tasks were to some extent quite similar. For both of them the children had to keep in mind a rhythm and either reproduce it or compare it to a second one. Thus, it comes as no surprise that rhythm perception and rhythm production tasks are related to the same precursor of reading; working memory. However, it might not be possible to generalize our findings to all kinds of working memory tasks. Because the task we applied uses non-sense words to assess working memory, it relies not only on working memory capacity but also on phonological processing and articulatory acuity. For example, if a child is not able to reproduce all letters of a syllable correctly this might be due to working memory capacity, but it is also possible that this child has difficulties with phonological processing. In the context of working memory as a precursor of reading it might be reasonable to use language material (non-sense words) in the working memory task. However, to claim a general link between rhythmical abilities and working memory future studies should investigate this association by using tests that do not rely on phonological abilities.

Interference of rapid retrieval from long-term memory was correlated with rhythm perception as well as with rhythm production. Hence, for interference of rapid retrieval from long-term memory associations with perception and production showed the same pattern of results. To the best of our knowledge our study is the first study that demonstrated an association

between musical abilities and interference of rapid retrieval from long-term memory. As mentioned above, the rhythm tasks placed similar cognitive demands on the children. Therefore, it is plausible that they showed comparable relationships. This result again points toward the importance of memory for language processing. Although the task assessed rapid retrieval from long-term memory, it might be speculated that again a subsystem of working memory (the central executive) could build the common ground of the music and the language task. The central executive is, among other things, responsible for providing a link between working memory and long-term memory (Baddeley, 2007). Furthermore, the nature of the task (speeded naming of color incongruent fruits) is comparable to other set shifting tasks that typically assess central executive. Hence, the central executive, in particular set shifting abilities, might be important for musical abilities and rapid retrieval from long-term memory. Indeed, there is correlational evidence of an association between set shifting and music lessons (Degé, et al., 2011).

With regard to the design of music interventions studies focusing on enhancing precursors of reading, our results point out that music training should target music listening skills as well as music production skills. A combination of perception and production tasks will probably provide a more successful training of precursors of reading. Our most important finding is that music production should be part of a music training program. Additionally, our results show that rhythm perception and production tasks may be a powerful tool to enhance memory-related precursors of reading. Moreover, the present results suggest that not only pitch perception in the sense of discriminating pitches of different frequencies, but also discriminating different length of pitches may be important to train in a music intervention. Lastly, body movements, as required in meter execution, may be helpful in training phonological awareness. However, they probably play only a minor role, because meter execution was only

related to the rhymes subtest of phonological awareness. It is important to note that all of the above mentioned suggestions were inspired by correlational data. Our study provides no evidence of a causal effect of music training on the mentioned precursors of reading, but sheds light on the path a music training might take to improve precursors of reading.

As already mentioned, our study aimed at analyzing associations between musical abilities and precursors of reading on a detailed level. Therefore, a detailed and exploratory approach was chosen. However, there is a tradeoff between the detailed picture we could show and the high amount of correlations tested. Due to several comparisons the level of significance should be adjusted. If the alpha level is adjusted, only associations between music production and phonological awareness remain reliable. Although in our analyses alpha inflation is a problem, we believe that this approach is seminal. It draws a more complete picture of association between musical abilities and precursors of reading than former research has done. Moreover, for future studies it is now possible to test specific relations and hypothesis.

Future studies should replicate our results with a larger sample size and should extend them by using different measures to assess musical abilities as well as precursors of reading to investigate the stability of the revealed associations. Finally, it remains to future research to employ an experimental design to allow inferences about causation.

Acknowledgments

This research was supported by a grant for educational research (TP9 6050085) from the Justus Liebig University Giessen. The authors would like to thank the participating kindergartens as well as all of the participants and their parents.

References

- Anvari, S. H., Trainor, L. J., Woodside, J., and Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *J. Exp. Child Psychol.* 83, 111–130. doi: 10.1016/S0022-0965(02)00124-8
- Baddeley, A. D. (1986). *Working Memory*. Oxford: Clarendon.
- Baddeley, A. (2007). *Working Memory, Thought, and Action*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198528012.001.0001
- Besson, M., Chobert, J., and Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *Front. Psychol.* 2:94. doi: 10.3389/fpsyg.2011.00094
- Bruck, M. (1992). Persistence of dyslexics' phonological awareness deficits. *Dev. Psychol.* 28, 874–886. doi: 10.1037/0012-1649.28.5.874
- Butzlaff, R. (2000). Can music be used to teach reading? *J. Aesthet. Educ.* 34, 167–178. doi: 10.2307/3333642
- Corrigall, K. A., Schellenberg, E. G., and Misura, N. M. (2013). Music training, cognition, and personality. *Front. Psychol.* 4:222. doi: 10.3389/fpsyg.2013.00222
- Corrigall, K. A., and Trainor, L. J. (2011). Associations between length of music training and reading skills in children. *Music Percept.* 29, 147–155. doi: 10.1525/mp.2011.29.2.147
- Degé, F., Kubicek, C., and Schwarzer, G. (2011). Music lessons and intelligence: a relation mediated by executive functions. *Music Percept.* 29, 195–201. doi: 10.1525/mp.2011.29.2.195
- Degé, F., and Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Front. Psychol.* 2:124. doi: 10.3389/fpsyg.2011.00124
- Douglas, S., and Willatts, P. (1994). The relationship between musical ability and literacy skills. *J. Res. Read.* 17, 99–107. doi: 10.1111/j.1467-9817.1994.tb00057.x
- Forgeard, M., Winner, E., Norton, A., and Schlaug, G. (2008). Practicing a musical instrument in childhood is associated with enhanced verbal ability and nonverbal reasoning. *PLoS ONE* 3:e3566. doi: 10.1371/journal.pone.0003566
- Gromko, J. E. (2005). The effect of music instruction on phonemic awareness in beginning readers. *J. Res. Music Educ.* 53, 199–209. doi: 10.1177/002242940505300302
- Hille, K., Gust, K., Bitz, U., and Kammer, T. (2011). Associations between music education, intelligence, and spelling ability in elementary school. *Adv. Cogn. Psychol.* 7, 1–6. doi: 10.2478/v10053-008-0082-4
- Huss, M., Verney, J. P., Fosker, T., Mead, N., and Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. *Cortex* 47, 674–689. doi: 10.1016/j.cortex.2010.07.010
- Jansen, H., Mannhaupt, G., Marx, H., and Skowronek, H. (2002). *Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten*. Göttingen: Hogrefe.
- Jungbluth, A., and Hafen, R. (2005). *Musik- Screening für Kinder*. Vechta: unpublished test material.

- Lamb, S. J., and Gregory, A. H. (1993). The relationship between music and reading in beginning readers. *Educ. Psychol.* 13, 19–27. doi: 10.1080/0144341930130103
 - Moreno, S., Bialyok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., and Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychol. Sci.* 22, 1425–1433. doi: 10.1177/0956797611416999
 - Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex* 19, 712–723. doi: 10.1093/cercor/bhn120
 - Norton, A., Winner, E., Cronin, K., Overy, K., Lee, D. J., and Schlaug, G. (2005). Are there pre-existing neural, cognitive, or motoric markers for musical ability? *Brain Cogn.* 59, 124–134. doi: 10.1016/j.bandc.2005.05.009
 - Piro, J., and Ortiz, C. (2009). The effect of piano lessons on the vocabulary and verbal sequencing skills of primary grade students. *Psychol. Music* 37, 325–347. doi: 10.1177/0305735608097248
 - Pratt, A. C., and Brady, S. (1988). Relation of phonological awareness to reading disability in children and adults. *J. Educ. Psychol.* 80, 319–323. doi: 10.1037/0022-0663.80.3.319
 - Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychol. Sci.* 15, 511–514. doi: 10.1111/j.0956-7976.2004.00711.x
 - Standley, J. M. (2008). Does music instruction help children to learn to read? Evidence of a meta-analysis. *Update Appl. Res. Music Educ.* 27, 17–32. doi: 10.1177/8755123308322270
 - Tierney, A., and Kraus, N. (2014). Auditory-motor entrainment and phonological skills: precise auditory timing hypothesis (PATH). *Front. Hum. Neurosci.* 8:949. doi: 10.3389/fnhum.2014.00949
 - Weiß, R., and Osterland, J. (1977). *Grundintelligenztest CFT 1*. Braunschweig: Westermann.
- 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 © 2015 Degé, Kubicek and Schwarzer. 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.



Does Music Training Enhance Literacy Skills? A Meta-Analysis

Reyna L. Gordon^{1,2*}, Hilda M. Fehd³ and Bruce D. McCandliss⁴

¹ Music Cognition Lab, Program for Music, Mind and Society, Department of Otolaryngology, Vanderbilt University Medical Center, Nashville, TN, USA, ² Vanderbilt Kennedy Center, Vanderbilt University Medical Center, Nashville, TN, USA, ³ Institute for Software Integrated Systems, School of Engineering, Vanderbilt University, Nashville, TN, USA, ⁴ Department of Psychology, Graduate School of Education, Stanford University, Stanford, CA, USA

OPEN ACCESS

Edited by:

McNeel Gordon Jantzen,
Western Washington University, USA

Reviewed by:

Virginia Penhune,
Concordia University, Canada
Franziska Degé,
Justus-Liebig-University, Germany

*Correspondence:

Reyna L. Gordon
reyna.gordon@vanderbilt.edu

Specialty section:

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

Received: 22 July 2015

Accepted: 05 November 2015

Published: 01 December 2015

Citation:

Gordon RL, Fehd HM and
McCandliss BD (2015) Does Music
Training Enhance Literacy Skills?
A Meta-Analysis.
Front. Psychol. 6:1777.
doi: 10.3389/fpsyg.2015.01777

Children's engagement in music practice is associated with enhancements in literacy-related language skills, as demonstrated by multiple reports of correlation across these two domains. Training studies have tested whether engaging in music training directly transfers benefit to children's literacy skill development. Results of such studies, however, are mixed. Interpretation of these mixed results is made more complex by the fact that a wide range of literacy-related outcome measures are used across these studies. Here, we address these challenges via a meta-analytic approach. A comprehensive literature review of peer-reviewed music training studies was built around key criteria needed to test the direct transfer hypothesis, including: (a) inclusion of music training vs. control groups; (b) inclusion of pre- vs. post-comparison measures, and (c) indication that reading instruction was held constant across groups. Thirteen studies were identified ($n = 901$). Two classes of outcome measures emerged with sufficient overlap to support meta-analysis: phonological awareness and reading fluency. Hours of training, age, and type of control intervention were examined as potential moderators. Results supported the hypothesis that music training leads to gains in phonological awareness skills. The effect isolated by contrasting gains in music training vs. gains in control was small relative to the large variance in these skills ($d = 0.2$). Interestingly, analyses revealed that transfer effects for rhyming skills tended to grow stronger with increased hours of training. In contrast, no significant aggregate transfer effect emerged for reading fluency measures, despite some studies reporting large training effects. The potential influence of other study design factors were considered, including intervention design, IQ, and SES. Results are discussed in the context of emerging findings that music training may enhance literacy development via changes in brain mechanisms that support both music and language cognition.

Keywords: music training, reading, literacy, phonological awareness, meta-analysis, brain development

INTRODUCTION

Acquiring fluency in reading requires children to transform symbolic information provided by print into mental representations based on their prior language experience. This literacy acquisition relies heavily on the process of phonological awareness. In particular, children's ability to focus their attention on sub-syllabic phonological units within words is a critical factor for mastering

the early challenge of alphabetic decoding. Phonological awareness has also been linked to neural mechanisms that help explain individual differences in early literacy (Schlaggar and McCandliss, 2007). Moreover, a growing number of studies have linked music skills and music training to differences in speech perception (Wong et al., 2007; François and Schön, 2011); basic auditory perception (Shahin et al., 2003; Hyde et al., 2009) and acquisition of second language or an artificial language (Slevc and Miyake, 2006; Brod and Opitz, 2012). Basic auditory processing appears to be a building block of phonological awareness (Walker et al., 2006), and music training is associated with both superior auditory perception (Seither-Preisler et al., 2014) and enhanced language skills (see Patel, 2008, for a review).

Understanding the potential connection between music training and literacy skills is informed by two areas of research literature. The first is a well-established body of research showing that some language-related skills, such as phonological awareness, are a fundamental pre-cursor of reading skills (see meta-analysis by Melby-Lervåg et al., 2012), and the second is an emerging literature investigating the potential role of music training as an activity that may induce plastic changes and perceptual enhancements within neural systems crucial for reading (e.g., Kraus et al., 2014b). Learning to play an instrument or to sing requires a complex series of neural transformations in order to process fine-grained acoustic variations in timing, frequency, spectral characteristics, and intensity into musically relevant auditory-motor actions to create rhythm, pitch, timbre, and dynamics. The OPERA hypothesis (Patel, 2011, 2014) provides a framework for highlighting the multiple perceptual demands musical training requires and the benefits such demands may bestow on neural systems that are important for literacy and language skills. Together, these two literatures provide constraints on understanding pathways through which musical training may enhance early literacy acquisition.

A rapidly accumulating body of evidence has shown associations between language and music skills in children. For instance, 7-to-9-year-old musicians outperformed their non-musician peers at detecting small prosodic (pitch) incongruities in sentences (Magne et al., 2006). Likewise, 9-year-olds musicians (vs. non-musicians) showed enhanced brain responses and behavioral performance on detection of deviants of the voice-onset-time, frequency, and duration of syllables (Chobert et al., 2011). Foreign language pronunciation skills and brain response to duration deviants (in music and speech) were better in 10- to 12-year-olds with musical training (Milovanov et al., 2009). Even without explicit music training, some of the variability in language skills can be accounted for by measuring individual differences in music aptitude. Measures of music aptitude have been found to account for over 40% of the variance in reading performance in typically developing 8- to 13-year-old children with little to no music training (Strait et al., 2011). Rhythm perception skills were robustly correlated with grammar production skills in 6-year-olds (Gordon et al., 2015b); a follow-up study of grammatical categories and musical rhythm revealed

that musical rhythm explains production of complex sentence structure in particular (Gordon et al., 2015a).

Reading is one language skill that has received recent attention in the neuroscience community regarding potential shared neural resources with music. Anvari et al. (2002) showed that pitch and rhythm skills in 4- and 5-year-olds correlated with phonological awareness and early reading skills, converging with prior findings of a correlation between pitch discrimination and both phonemic awareness and early reading abilities in a similar age group (Lamb and Gregory, 1993). Musical rhythm in particular has been linked to reading skills in prior work using a wide variety of methods for measuring rhythm in young children, across many native languages. American-English-speaking preschoolers who excelled at synchronizing to an acoustic beat ("Synchronizers") outperformed their "Non-synchronizer" peers at phonological awareness and rapid naming tasks (Woodruff Carr et al., 2014). A French study with large sample size ($n = 695$) showed that kindergarteners' ability to reproduce musical rhythms was significantly predictive of their second grade reading skills (Dellatolas et al., 2009). Interestingly, Banai and Ahissar (2013) found a stronger relationship between reading and auditory processing skills in Israeli children without musical training, while the musician children in the study showed better auditory processing but no advantage in reading skills.

The relation between rhythm and reading-related skills continues to be significant in later stages of language development. Tierney and Kraus (2013b) found that beat tapping variability (to an isochronous metronome at a 2 Hz rate) negatively correlated with reading skills in adolescents, such that those who tapped to the beat more consistently were more likely to have better performance on the reading measures. Correlational studies in adults have shown that musicians have greater sensitivity to speech rhythm (Marie et al., 2011), better reading-related skills (e.g., phoneme discrimination: Zuk et al., 2013a) and that individual differences in speech rhythm sensitivity is related to variability in musical aptitude when participants with a wide range across the continuum of musical abilities are studied (Magne et al., in revision). Over the course of aging, there is evidence that early musical training is associated with protection against age-related linguistic and cognitive declines (Parbery-Clark et al., 2011; Bidelman et al., 2014; Bidelman and Alain, 2015), even in adults with hearing loss (Parbery-Clark et al., 2013). However, as noted in Butera (2015), associations with musical training in these correlational studies cannot be interpreted in favor of causality in the absence of longitudinal data that rules out other genetic and environmental contributions to the observed findings of neural enhancements in individuals with musical training.

If enhanced language skills and musical skills are correlated, then would individuals with language disorders also have deficits in musical processing? Research on reading disabilities and language impairment suggests that this is often the case (e.g., Goswami, 2011; Gordon et al., 2015a). Seminal work by Overy (2003) revealed that a small group of children

with reading disability improved their phonological awareness and spelling skills faster during an 8-week period of music instruction than during the same amount of time with no music training. Sensitivity to musical rhythm predicted significant variance in phonological awareness concurrently and longitudinally in 10-year-olds with dyslexia (Huss et al., 2011; Goswami et al., 2013). Difficulties processing the prosodic aspect of speech (i.e., variations in timing and pitch that mark linguistic events) are thought to be reflected in both musical deficits and weaknesses in phonological awareness (Goswami et al., 2010; Power et al., 2013; Leong and Goswami, 2014) in individuals with reading disabilities. Given these connections, musical practice holds promise as a tool to contribute to reading skills, potentially via a pathway of enhancing children's sensitivity to prosodic aspects of speech.

Correlational evidence does not, of course, exclude potential effects of self-selection or environmental and genetic differences that could alternatively account for enhanced language skills in musicians (Schellenberg, 2015). Evidence from longitudinal studies that administer a controlled and specific amount of musical training is crucial for investigating a possible causal influence of music on non-musical skills. The potential that music training could enhance reading skills is especially pertinent now that there are ongoing debates in educational systems about the most effective strategies for impacting academic achievement in the core curriculum. However, it is important to note that much of this work has focused on training-related brain changes (rather than behavioral outcomes); the significance for academic achievement of these modifications in brain activity is difficult to ascertain in the absence of reporting of behavioral gains in language skills (as discussed in Evans et al., 2014; Schellenberg, 2015). As reviewed in the present study, a considerable collection of controlled training studies has provided positive evidence for the hypothesis that musical training transfers to literacy-related skills. Taken as a whole, however, the range of studies published to date present a rather mixed set of results, marked by a large range of potential outcome measures related to literacy skills. To assess and quantify the state of the evidence that may potentially support the hypothesis that musical training in children transfers into enhancements in literacy-related skills, we first set out to delineate the subset of peer-reviewed papers that directly address this issue via training and pre- post-assessment designs.

A meta-analytic approach is useful in assessing the efficacy of music training for language outcomes and identifying the attributes of music training paradigms that are relevant to specific reading outcomes. The present meta-analysis is thus aimed at synthesizing previous research on music training and reading-related outcomes. The following research questions were examined:

(1) Does music training improve reading-related outcomes when other reading instruction is controlled for? Are certain aspects of learning how to read (i.e., reading fluency and phonological awareness) particularly susceptible to transfer from music training?

- (2) Does the age of participants account for variability in the efficacy of the training?
- (3) Does the quantity of music training impact the efficacy of the training, and how many hours of training are needed to affect changes in reading-related outcomes?
- (4) Does the design of the control group condition moderate outcomes?

MATERIALS AND METHODS

Literature Search

Search Strategies

The goal of this meta-analysis is to evaluate the effectiveness of musical interventions on reading-related measures. To find all articles that met our criteria, we conducted a literature search using the PubMed, Web of Knowledge, and ProQuest article databases. ProQuest functioned as a meta-database, allowing us to search 12 databases simultaneously: ERIC, International Index to Music Periodicals Full Text, Linguistics and Language Behavior Abstracts, MLA International Bibliography, ProQuest Education Journals, ProQuest Psychology Journals, ProQuest Research Library, ProQuest Science Journals, ProQuest Social Science Journals, PsychARTICLES, PsycINFO, and RILM Abstracts of Music Literature. The search terms used in each of the three searches are listed in Supplementary Table 1. The initial search was conducted in November 2013, and it was repeated/updated in March 2014. In total, the search returned 4855 articles whose article titles were searched for relevance to the topic. Additionally, to pass this first screening phase, each article could not be a conference presentation, thesis or dissertation, or trade newspaper or magazine article, and had to be written in English. A preliminary search of these titles narrowed down the potentially relevant articles to 178. The abstracts of these remaining articles were then reviewed for inclusion criteria and relevance. The criteria in this second phase of screening required that articles not be a review or meta-analysis, that they have a music intervention with a control group, and that they investigated reading-related outcomes.

Inclusion and Exclusion Criteria

In our literature review, we defined inclusion and exclusion criteria based on meta-analysis guidelines for distinguishing features of studies (e.g., characteristics of the participants, key variables, research methods, and publication type; Lipsey and Wilson, 2001). Only articles that met the following criteria were included in the study:

1. *Had an intervention with a control group* (i.e., no within-group interventions, observational or correlational studies).
2. *Was a peer-reviewed publication* (i.e., no dissertations/thesis, conference proceedings, unpublished manuscripts, or secondary sources such as trade magazines or media coverage). This criteria was adapted to ensure a minimally acceptable level of quality and rigor. This approach coincides

with the National Reading Panel's standards for meta-analysis (Lonigan and Shanahan, 2009) and with previous meta-analysis on literacy education (e.g., Bus and van IJzendoorn, 1999).

3. *Reported phonological or reading-related outcomes.*
4. *Assessed outcomes pre- and post-intervention.*
5. *Provided sufficient data to extract effect sizes (means, SD, and N, pre- and post-intervention, for the same participants in each group).* For studies that met requirements 1 through 4 but did not report sufficient data in the published paper, corresponding authors were contacted via email and asked to provide the additional information. In two cases (Douglas and Willatts, 1994; Standley and Hughes, 1997), authors responded that the data was not available due to the long time period that has lapsed since the publication of their respective studies; thus these two studies were excluded.

Out of 178 studies that were reviewed at the abstract level (with full-text examination if necessary to determine inclusion based on above criteria), 17 articles met these criteria. The types of interventions used and contrasting control groups were found to vary substantially across the studies, with some showing confounds of uneven amounts of reading instruction across the groups or failed to provide more musical training to one of the groups. We thus added the following constraint to study design for inclusion:

1. The intervention group had to receive more music instruction than the control group.
2. Studies need to provide an indication of equivalent amounts of reading instruction across the intervention and control groups.

After applying this final design constraint, an additional 5 studies were excluded (Register, 2001; Register et al., 2007; Bolduc, 2009; Darrow, 2009; Bhide et al., 2013) and only 12 papers still qualified, as listed in **Table 1** (Register, 2004; Gromko, 2005; Myant et al., 2008; Moreno et al., 2009, 2011; Yazejian and Peisner-Feinberg, 2009; Degé and Schwarzer, 2011; Herrera et al., 2011; Bolduc and Lefebvre, 2012; Cogo-Moreira et al., 2013; Moritz et al., 2013; Thomson et al., 2013). Herrera et al. presented results from two independent samples (each with its own control group) that received the same intervention, and was thus coded as two separate studies in our analysis, giving a final study count of $k = 13$ for the meta-analysis.

Coding Procedures

Procedure and Outcome Variables

A custom data entry system was created for the study using the Research Electronic Data Capture (REDCap) tools (Harris et al., 2009) hosted at Vanderbilt University (REDCap is a secure, web-based application designed to support data capture for research studies, providing an intuitive interface for validated data entry and automated export procedures for seamless data downloads to common statistical packages). All study

characteristics and data were coded and entered into the custom forms.

The outcomes measures used within these 13 studies are somewhat variable; each can be classified into one of the two broad categories of Reading Fluency and Phonological Awareness. For studies that reported more than one measure in an outcome category, we selected the measure that most directly tapped into the category. For Reading Fluency, measures that emphasized fluent use of known words and letters were chosen over those that used non-words. Within Phonological awareness, two subcategories were identified: Rhyming and Other Phonological measures. For Rhyming, measures that involved discrimination of rhymes were chosen over those that involved producing rhymes. For Other Phonological, measures that involved identification, discrimination, or manipulation of phonemes were chosen over those that dealt with non-word reading fluency or syllabic segmentation. All measures included are reported in **Table 1**.

Potential Moderating Variables

These 13 studies were then carefully coded for the following study design features, which are reported in **Tables 2, 3**.

1. *Total number of hours in music intervention.*
2. *Type of Control Intervention.* Control interventions included: phonological control, non-auditory control (sports or art), less intensive music control, and no-treatment control. In studies that included more than one intervention or control group, only the group that fully met the requirements outlined in the Inclusion and Exclusion Criteria section was included.
3. *Level of Random Assignment employed.* The following types of assignment were indicated: Student Random (children were enrolled in the study and assigned randomly to intervention and control groups); Random or Non-random Assignment by Class (multiple classes within a school were enrolled in the study and full classes were assigned to the intervention randomly or non-randomly); Random or Non-random Assignment by School (multiple schools were enrolled in the study and participants were assigned to the intervention randomly or non-randomly).
4. *Components of Music Training.* The following component categories of musical activities were coded in a binary manner (i.e., we coded whether or not the intervention included each component): Phonology in Musical context; Gross motor Movement/Kinesthetic activities; Rhythm; Musical Instruments; Rhyming; Clapping/Marching; Visual representations of musical concepts (i.e., visual portrayals of high vs. low pitch or short vs. long sounds); Singing; and Musical notation.
5. *Mean age of participants.*
6. *Subject population (Typically or Atypically developing).*
7. *Was Socio-economic status (SES) reported and controlled for across groups?*
8. *Was IQ controlled reported and controlled for across groups?*

TABLE 1 | Study characteristics.

Study, Year, Journal	Language	Mean age of participants	Reading fluency outcome measure	Phonological outcome measures	
				Rhyming	Other phonological measures
(Bolduc and Lefebvre, 2012), Creative Education	French	4.9 (N = 54)			Phonological awareness measure (PAM; Armand and Montésinos-Gelet, 2001)
(Cogo-Moreira et al., 2013), PLoS ONE	Portuguese (Brazil)	9.2 (N = 235)	Accuracy of Word reading (custom)		Test of phonological awareness (Capovilla and Capovilla, 1998)
(Degé and Schwarzer, 2011), Frontiers in Psychology	German	5.8 (N = 27)			Phonological awareness—total from Bielefelder screening (Jansen et al., 2002)
(Gromko, 2005), Journal of Research in Music Education	English (US)	5.5 (N = 103)	DIBELS letter-naming fluency (Good and Kaminski, 2002)		DIBELS phoneme-segmentation fluency
(Herrera et al., 2011), Psychology of Music	Spanish	4.5 (N = 29)		Rhyme oddity task (custom)	Initial phoneme oddity task (custom)
(Herrera et al., 2011), Psychology of Music	Tamazight	4.7 (N = 27)		Rhyme oddity task (custom)	Initial phoneme oddity task (custom)
(Moreno et al., 2009), Cerebral Cortex	Portuguese (Portugal)	8.3 (N = 32)	Reading inconsistent words (from Portuguese European reading battery, Succena and Castro, 2010)		
(Moreno et al., 2011), Music Perception	English (Canada)	5.3 (N = 60)		Rhyming (from WJ-III, Woodcock et al., 2001)	
(Moritz et al., 2013), Reading and Writing	English (US)	5.6 (N = 30)		Rhyming Discrimination from Phonological awareness test (PAT; Robertson and Salter, 1997)	Isolation of initial phonemes from PAT
(Myant et al., 2008), Educational and Cognitive Psychology	English (UK)	4.3 (N = 59)		Rhyme test from Phonological Assessment Battery (PhAB; Frederickson et al., 1997)	Alliteration test from PhAB
(Register, 2004), Journal of Music Therapy	English (US)	5.5 (N = 43)	Letter-naming fluency from DIBELS (Good and Kaminski, 2001)		Initial sounds fluency from DIBELS
(Thomson et al., 2013), Reading and Writing	English (UK)	9.3 (N = 21)	TOWRE (Torgesen et al., 1999)	Rhyme test from PhAB	Spoonerisms from PhAB
(Yazejian and Peisner-Feinberg, 2009), NHSA Dialog	English (US)	4.4 (N = 181)		Rhyming from Early Phonological Awareness Profile (EPAP; Dickinson and Chaney, 1997)	Phoneme deletion from EPAP

Study information, primary language of participants, age, and outcome measures of studies included in the meta-analyses.

TABLE 2 | Training components.

Study	Total hours of training	Training components							
		Phonology in music context	Movement/ Kinesthetic	Rhythm	Instruments	Rhyming	Clapping/ Marching	Visual representations	Singing Musical notation
Bolduc and Lefebvre, 2012	6.67	✓			✓	✓			✓
Cogo-Moreira et al., 2013	50		✓	✓	✓				✓
Degé and Schwarzer, 2011	16.67		✓	✓	✓				✓
Gromko, 2005	6.5		✓	✓	✓			✓	✓
(Herrera et al., 2011), Spanish	16	✓				✓	✓	✓	✓
(Herrera et al., 2011), Tamazight	16	✓				✓	✓	✓	✓
Moreno et al., 2009	60	✓	✓	✓	✓	✓	✓		✓
Moreno et al., 2011	40			✓					✓
Moritz et al., 2013	90		✓	✓					✓
Myant et al., 2008	17.5		✓	✓			✓		✓
Register, 2004	8.5	✓	✓		✓				✓
Thomson et al., 2013	3	✓		✓			✓	✓	
Yazejian and Peisner-Feinberg, 2009	26		✓	✓	✓	✓		✓	✓

Hours of music training and components of the music intervention for each study.

Statistical Analysis

Effect Size Calculation

For each outcome and measure, a single effect size was computed in the following manner, where ES = effect size:

$$ES = \frac{(\text{Posttest MeanTx} - \text{Pretest MeanTx}) - (\text{Posttest MeanControl} - \text{Pretest MeanControl})}{\text{Pooled Pretest SD}}$$

$$\text{Pooled Pretest SD} = \sqrt{\frac{\text{Pretest SD}_{Tx}^2 * (N_{Tx} - 1) + \text{Pretest SD}_{Control}^2 * (N_{Control} - 1)}{N_{Tx} + N_{Control} - 2}}$$

Data-analysis

Meta-analysis was performed using the open-source statistical software package R (R Core Team, 2015), and employing the “metafor” package (Viechtbauer, 2010). Heterogeneity was computed as I^2 = residual heterogeneity divided by unaccounted variability, and H^2 = unaccounted variability divided by sampling variability (Higgins and Thompson, 2002). Meta-analysis was carried out using two different approaches: random effects model for the separate analysis of each of the three outcome types (Reading Fluency, Rhyming, and Other Phonological outcomes), and mixed effects model for the moderator analysis. Mixed effects was also used for the broader All Phonological Outcomes category since it included non-independent samples from studies that included both Rhyming and Other Phonological Outcomes. Moderator analysis was used to test influence of age, control intervention type, and number of training hours on the efficacy of music

interventions. Given the relatively small number of studies included in the meta-analysis, it was not possible to test additional moderators for each component of training and level of random assignment.

RESULTS

Characteristics of the Studies Included

Publication information, language, age of participants, and outcomes measured are reported in **Table 1**. Participants ranged in mean age from 4.53 to 9.33 years, with a weighted average mean of 6.25. Participants identified with a wide range of native languages (English, Portuguese, German, French, Spanish, and Tamazight). The components of music training are reported in **Table 2** and varied greatly across studies; total hours of training ranged from 3 (Thomson et al., 2013) to 90 (Moritz et al., 2013). Many studies included singing ($k = 12$), rhythm ($k = 9$), instruments ($k = 7$), movement/kinesthetics ($k = 8$), and less than half used Phonology in music context ($k = 6$), rhyming ($k = 5$), clapping/marching ($k = 5$), visual representations of musical concepts ($k = 5$), and only $k = 3$ included music notation.

TABLE 3 | Study controls.

Study	Subject population	Control for IQ	Control for SES	Type of assignment	Control interventions
Bolduc and Lefebvre, 2012	Typical	Yes	SES not reported	Random assignment by class	Phonological control
Cogo-Moreira et al., 2013	Atypical	Yes	SES not reported	Random assignment by school	No-treatment control
Degé and Schwarzer, 2011	Typical	Yes	Yes	Student random	Non-auditory control (sports)
Gromko, 2005	Typical	IQ not reported	No	Non-random assignment by school	No-treatment control
Herrera et al., 2011, Spanish	Typical	Yes	SES not reported	Student random	Phonological control
Herrera et al., 2011, Tamazight	Typical	Yes	SES not reported	Student random	Phonological control
Moreno et al., 2009	Typical	Yes	Yes	Student random	Non-auditory control (art)
Moreno et al., 2011	Typical	Yes	Yes	Student random	Non-auditory control (art)
Moritz et al., 2013	Typical	Yes	No	Non-random assignment by school	Less intensive music control
Myant et al., 2008	Typical	IQ not reported	Yes	Non-random assignment by school	No-treatment control
Register, 2004	Typical	IQ not reported	Yes	Non-random assignment by class	No-treatment control
Thomson et al., 2013	Atypical	Yes	SES not reported	Student random	No-treatment control
Yazejian and Peisner-Feinberg, 2009	Typical	IQ not reported	Yes	Random assignment by class	No-treatment control

This table reports population, IQ, SES, type of assignment, and control interventions for each study.

Several aspects of control factors in the study design are reported in **Table 3**. All but two studies (Cogo-Moreira et al., 2013; Thomson et al., 2013) were conducted on a typically developing children. IQ was reported as equivalent across groups in $k = 9$ studies, and SES was reported as equivalent across groups in only $k = 6$ studies. Many different types of group assignment were found, and only $k = 6$ studies used “true” student random assignment. The remaining studies assigned pre-existing classes (or schools) to different treatment conditions. Control interventions included $k = 3$ studies in which the control group received phonological training, $k = 3$ studies with non-auditory control activities such as art or sports, $k = 6$

studies with no special extra-curricular activities (no-treatment control), and one study where the control group also received music lessons but to a much lesser extent (“less intensive music” control).

Effect Sizes

Means, standard deviations, pre- and post-training, N 's per group, and the computed effect sizes are reported in **Table 4**. Given that this meta-analysis was designed to investigate (1) how music training affects different types of reading-related measures; and (2) how selected aspects of study design (age of participants, hours of training, and type of control intervention)

TABLE 4 | Effect sizes.

Study	Measure	N music	Music group pre-training Mean (SD)	Music group post-training Mean (SD)	N control	Control group pre-training Mean (SD)	Control group post-training Mean (SD)	Standardized effect size
READING FLUENCY DATA								
Cogo-Moreira et al., 2013	Accuracy of word reading	114	9.45 (11.33)	16.14 (15.59)	121	11.22 (14.37)	16.5 (15.33)	0.11
Gromko, 2005	DIBELS letter-naming task	43	33.42 (15.48)	42.63 (15.22)	60	36.27 (18.87)	44.1 (15.63)	0.08
Moreno et al., 2009	Inconsistent word reading	16	41.73 (16.38)	71.35 (13.25)	16	45.83 (17.48)	56.77 (17.53)	1.07
Register, 2004	DIBELS letter naming fluency	22	12.18 (10.58)	20.23 (14.27)	21	17.48 (16.74)	25.38 (17.65)	0.01
Herrera et al., 2011	Word reading	9	49.67 (12.44)	52.44 (11.26)	12	48 (15.97)	48.25 (17.27)	0.17
OTHER PHONOLOGICAL DATA								
Bolduc and Lefebvre, 2012	Phonological Awareness Measure (PAM)	28	10.5 (2.58)	14.8 (3.65)	26	12 (3.12)	15.4 (3.54)	0.31
Cogo-Moreira et al., 2013	Phonological awareness	114	25.79 (4.96)	27.66 (4.64)	121	23.98 (5.13)	25.18 (5.25)	0.13
Degé and Schwarzer, 2011	Phonological Awareness—Total	13	35.77 (2.35)	38.23 (1.17)	14	35.86 (3.18)	36.07 (2.99)	0.78
Gromko, 2005	DIBELS phoneme-segmentation fluency	43	18.61 (16.26)	44.72 (16.94)	60	25.83 (14.73)	41.55 (14.5)	0.67
(Herrera et al., 2011), Spanish	Initial sound	15	42.69 (22.5)	60.14 (12.5)	14	45.8 (17.76)	60.5 (12.63)	0.13
(Herrera et al., 2011), Tamazight	Initial sound	17	42.44 (10.2)	51.99 (8.67)	10	39.72 (12.18)	55.14 (7.9)	−0.52
Moritz et al., 2013	PAT isolation initial	15	7.5 (2.15)	9.93 (0.27)	15	6.57 (2.3)	9.15 (1.21)	−0.07
Myant et al., 2008	Alliteration	28	1.82 (2.58)	3.35 (3.35)	31	0.26 (0.58)	1.11 (1.6)	0.37
Register, 2004	DIBELS initial sounds fluency	22	6 (6.62)	14.27 (8.47)	21	9.52 (6.41)	15.71 (8.04)	0.31
Thomson et al., 2013,	PhAB spoonerisms	9	14.11 (6.54)	17.44 (7.38)	12	14.17 (7.21)	14.83 (6.93)	0.37
Yazejian and Peisner-Feinberg, 2009	Phoneme deletion	111	10.35 (4.19)	12.32 (2.88)	70	8.99 (4.68)	12.03 (3.27)	−0.24
RHYMING DATA								
(Herrera et al., 2011), Spanish	Rhyme oddity	15	42.08 (11.97)	56.64 (6.82)	14	40.56 (14.49)	52.49 (10.94)	0.19
(Herrera et al., 2011), Tamazight	Rhyme oddity	17	46.68 (8.8)	64.65 (9.12)	10	42.92 (11.4)	57.36 (10.27)	0.35
Moreno et al., 2011	Rhyming	30	9.2 (2.9)	11 (3.7)	30	8.6 (3.9)	10 (4.3)	0.11
Moritz et al., 2013	PAT rhyming discrimination	15	7.53 (2.1)	9.86 (0.36)	15	8.64 (1.39)	8.77 (1.54)	1.20
Myant et al., 2008	Rhyme	28	3.86 (2.92)	6.77 (3)	31	3 (2.67)	6.04 (2.93)	−0.05
Thomson et al., 2013	PhAB rhyme	9	16.78 (2.28)	18.78 (2.28)	12	14.08 (5.45)	15.08 (5.87)	0.22
Yazejian and Peisner-Feinberg, 2009	Rhyme recognition	111	3.52 (2.89)	6.05 (3.74)	70	2.76 (2.58)	5.21 (3.85)	0.02

Means and SD's for each group, and effect sizes, are listed for each study (grouped by outcome category type).

would moderate outcomes, the choice to limit the moderator analysis to these three moderator variables was also constrained by the statistical power of conducting meta-regression on only a small number of studies that met the criteria. Thus, meta-analyses were computed separately on reading fluency and phonological awareness, and moderator analyses tested the influence of each of the abovementioned factors on the outcomes.

Meta-analysis Results for Phonological Awareness

Due to the non-independence of the studies that reported both types of phonological awareness outcomes (Rhyming and Other

Phonological) in the same sample, mixed effects analysis was employed to test overall Phonological Awareness. This analysis on All Phonological Awareness ($k = 18$) revealed an effect size of 0.20 (95% CI [0.04, 0.36], $p = 0.01$), showing small but significant gains of music training on phonological skills, shown in the forest plot in **Figure 1**. The test for Heterogeneity [$Q_{(df=17)} = 28.8$, $p = 0.04$] was significant, indicating potential influence of other factors. To investigate these factors and their relation with moderators, phonological outcomes were then further broken down into two separate categories corresponding to Rhyming and Other Phonological outcomes (see Methods section for more information on how measures/outcomes were chosen).

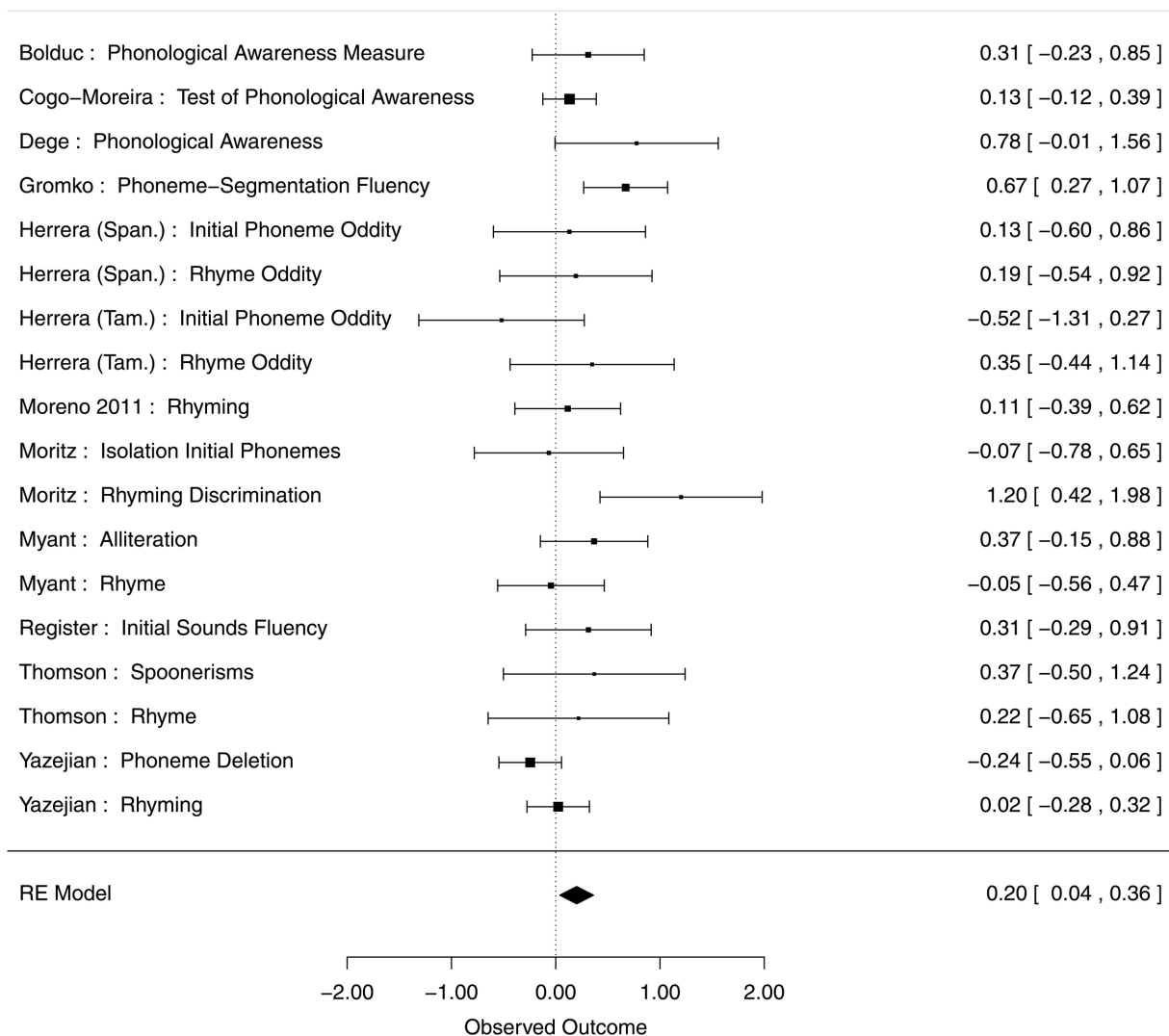


FIGURE 1 | Influence of music training on Phonological Awareness outcomes. The forest plot shows weighted effect sizes for music vs. control group on all phonological outcomes, in each study, and across studies. Confidence intervals are given in brackets.

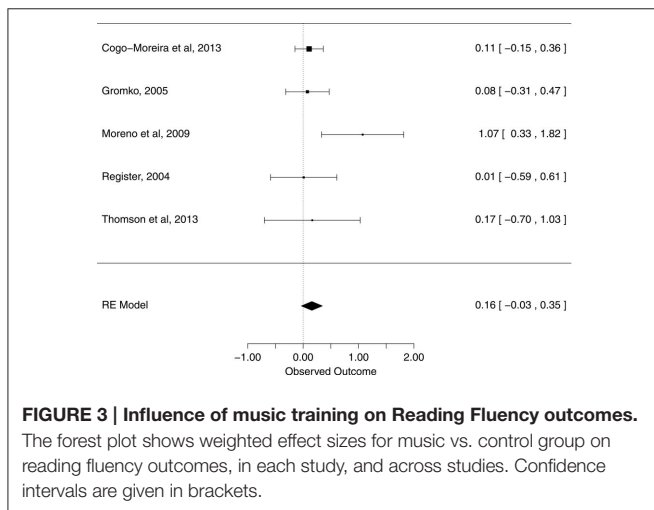
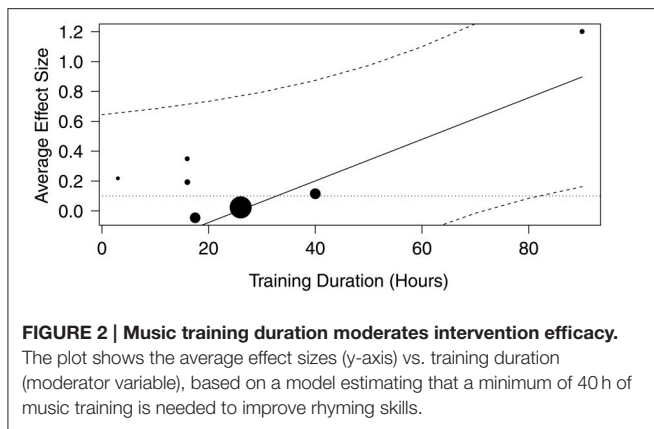
Rhyming Outcomes

Random-effects analysis on the subset of rhyming outcomes ($k = 7$ studies) yielded a weighted average effect size of 0.18 (95% CI [-0.06, 0.42]), which was non-significant at $p = 0.14$. A mixed effects analysis then revealed no significant influence of age ($p = 0.31$) or control intervention type ($p = 0.75$) on the results, but a significant influence ($p = 0.04$) of training hours on rhyming outcomes. These results suggest that an increase in the length of training by 1 h corresponds to an increase of 0.01 (95% CI [0, 0.03]) in the effectiveness of music intervention on rhyming outcomes. The results of this model were then used to predict values of effectiveness given different amounts of training hours. Using the range of values from across all studies from the entire meta-analysis, (3–90 h), and assuming a constant age (5 years) and constant control intervention type, the model predicts that at least 40 h of training are needed to have a significant

effect on Rhyming outcomes, as shown in **Figure 2**. These results should be interpreted with caution, given that the study showing the strongest positive relationship between hours of training and rhyming outcomes (Moritz et al., 2013) had only 15 participants in each group.

Other Phonological Outcomes

Random effects analyses on Other Phonological outcomes ($k = 11$), yielded an average effect size of 0.20 (95% CI [-0.03, 0.42]), which weakly trended toward significance ($p = 0.08$). A mixed effects analysis revealed no significant influence of age ($p = 0.24$), control group type ($p = 0.34$), or training hours ($p = 0.09$) on the model. Heterogeneity was moderate ($I^2 = 40.2\%$; $H^2 = 1.67$) but residual heterogeneity did not reach significance [$QE(df=7) = 11.89, p = 0.10$].



Meta-analysis Results for Reading Fluency

Random effects analysis on the five studies that included Reading Fluency outcomes showed a weighted average effect size of 0.16 (95% CI [-0.03, 0.35], $p = 0.10$), thus showing only a weak trend toward significance of music intervention on reading fluency. Results are shown in **Figure 3**. Heterogeneity was low ($I^2 = 0\%$; $H^2 = 1$), and given the small number of studies ($k = 5$), moderator analysis was not pursued.

Test for Publication Bias

The Rank Correlation Test for Funnel Plot Asymmetry indicated no publication bias for either Reading Fluency (Kendall's tau = 0.60, $p = 0.23$) or Phonological Awareness (Kendall's tau = 0.18, $p = 0.33$).

DISCUSSION

The current meta-analysis was carried out to assess the impact of music intervention on reading-related skills in children, and adds to the literature by specifically highlighting effects of music training transferring to reading-related skills when non-musical reading training is held constant. Results of the meta-analysis on the broad category of Phonological Awareness outcomes

suggest modest gains (a small effect size of $d = 0.20$) for music vs. control groups. This finding is in line with a number of other studies showing better phonological awareness skills in musicians compared to their non-musician peers (Forgeard et al., 2008; Zuk et al., 2013b), and also converges with work showing correlations between music aptitude and phonological skills in children (Lamb and Gregory, 1993; Anvari et al., 2002; Peynircioglu et al., 2002; Dellatolas et al., 2009; Tierney and Kraus, 2013a).

When broken down into subcategories (Rhyming and Other Phonological outcomes), moderator analysis revealed that the effectiveness of music intervention on Rhyming outcomes was dependent on the number of training hours. Total music intervention training hours ranged between 3 and 90 h in the studies included here, and the model estimated that at least 40 h are needed to improve Rhyming skills. To put this number in perspective, other work (e.g., Hambrick et al., 2014) has shown that thousands of hours are typically involved in reaching adult levels of musical expertise. Consideration of how children's music training improves rhyming skills must assess the possibility that results could merely reflect the inclusion of greater rhyming practice within the music interventions relative to the control conditions. Indeed, early childhood music education in group settings typically include activities such as singing and chanting rhyming lyrics. However, several aspects of the studies that support the positive transfer effect for rhyming outcomes suggest that this effect cannot be entirely attributed to this explanation. First, it should be noted that the study with the strongest positive relationship between rhyming outcomes and hours of training (Moritz et al., 2013) reported no rhyming-related training activities, and rather emphasized rhythmic aspects of musical training. Furthermore, of the seven studies with rhyming outcomes, only four were coded as including any report of rhyming training (see **Table 2**). These results, taken together with reports of robust associations between musical rhythm skills and rhyme awareness, in both children with typical development and reading disabilities (e.g., Huss et al., 2011), suggest that other aspects of musical training may impact rhyming skills. Future work is needed to make more definitive conclusions regarding whether intensive rhythm training can improve rhyming and phonological skills in general, given the links between rhythm and reading skills in the literature (e.g., Strait et al., 2011).

The separate meta-analysis on eleven datasets with Other Phonological Outcomes was inconclusive: the effect size was small ($d = 0.2$) and only trended toward significance, with no moderators (age, control intervention type, or training hours) reaching significance. This pattern of results could potentially be due to variability in the many different types of phonological tasks that were included in this category (i.e., Initial Phoneme Oddity, Alliteration, Spoonerisms and others; see **Table 1**) or even to the wide variety of native languages spoken by participants. Further study is needed to determine if certain phonological skills are more susceptible to a positive transfer from music training than others.

The effect size for the separate meta-analysis assessing the impact of music training on reading fluency outcomes was also small ($d = 0.16$) and did not reach significance: moderator

analysis was precluded due to only having five studies in this category. However, it should be pointed out that two of the studies (Cogo-Moreira et al., 2013; Thomson et al., 2013) were on children with reading disabilities, and while there are solid theoretical reasons (see Overy, 2003; Tierney and Kraus, 2013a) to believe that music training could improve reading skills in struggling readers, the intensity of the intervention would likely be an important factor in such attempts.

Moreover, previous meta-analyses with different parameters than the present study have found both a non-significant effect of music on reading skills (Butzlaff, 2000) and significant effects (Standley, 2008). The present study extends these results by including data from additional studies published between 2008 and 2014, and by limiting the scope of studies included to a more rigorously defined comparison, for which reading instruction is controlled across groups. The data quality and variability of study outcomes and confidence intervals are comparable to studies included in other meta-analyses on literacy education (e.g., Lonigan and Shanahan, 2009) and this heterogeneity should be taken into account in the interpretation (as discussed below).

It is interesting to note that a previous meta-analysis on literacy development found medium-to-strong effects of phonological awareness training on reading skills (yet longer term studies produced only small effects), and that phonological awareness was a necessary, but not sufficient condition for reading (Bus and van IJzendoorn, 1999). One could hypothesize that music skills share more variance with phonological skills (due to their auditory bases) than with reading fluency skills, and thus music training may have larger effects on phonological awareness than on reading. Nonetheless, it is also possible that music training could impact reading fluency via a more gradual pathway: beginning more generally by improving auditory discrimination, then affecting rhyming skills and using them to bootstrap further phonological awareness. More intensive training may be needed for these improvements to occur at a level that produces measurable improvements in reading fluency across heterogeneous participant populations.

Overall, the findings of the current meta-analyses are somewhat inconclusive with regards to the hypothesized impact of music education on reading-related skills. The literature search revealed a large amount of variability in outcomes studied, content and intensity of music training, native language of participants, type of subject populations (typically developing vs. reading disordered) and age of participants. In addition, some

of the study designs in the set of studies included in this meta-analysis are laden with potential biases that make it difficult to draw broader conclusions from the findings (see **Table 3**). These inconsistencies include variability in control group activities, lack of information about IQ differences or equivalence across groups; and only 6 studies of 12 reported controlling for socio-economic status across groups. Importantly, most of the studies were quasi-experimental and did not use random assignment to create treatment and control groups. In the case of studies that compared a class (or school) receiving the intervention vs. another control class or school, it is possible that other differences in teacher/student dynamics and educational environment differed across the groups (and therefore either diminished or exaggerated the gains in music training). Although we were able to code and report many of the above characteristics, there were too few studies included in the total meta-analysis to allow a sufficiently powered moderator analysis that would effectively shed light on whether these study characteristics were linked with different trends of results. Thus, the limitations of the present meta-analysis are the heterogeneity of approaches and study designs used, and that the dataset was too underpowered to test all of the potentially influential moderator variables that were coded. Nevertheless, it is interesting to note that all three of the studies (Moreno et al., 2009, 2011; Degé and Schwarzer, 2011) in which SES and IQ were equivalent, and student random assignment was used, also showed large effect sizes on at least one reading-related outcome, indicating a robustness of music training efficacy for improving reading-related skills under methodologically sound circumstances. The quality and breadth of all studies included in the present meta-analyses also provides complementary information to results of a prior meta-analysis on the impact of music on reading skills (i.e., Standley, 2008) in which aspects of the music training may have confounded the findings (e.g., some studies included in their meta-analysis included contrasts where both groups received different types of music training and whether a given group got more music training was unclear). Suggestions for creating a standard of implementation steps for reducing heterogeneity and bias are summarized in **Table 5**.

Moreover, the small effects of music on reading-related outcomes observed in this meta-analysis stand in contrast to the robust results seen in the correlational literature reporting (broadly defined) linguistic advantages in musician children (Magne et al., 2006; Chobert et al., 2011) and adults who had

TABLE 5 | Future directions for studying the impact of music education on reading skills.

Factors to control	Questions of interest to test in future studies
<ul style="list-style-type: none"> • IQ and socio-economic status • Control intervention content • Type and duration of music training • Guidelines for typical and atypical development • Random assignment to experimental groups 	<ul style="list-style-type: none"> • What are the effects of different components of interventions (rhythm, pitch; instruments vs. singing; phonological activities in musical context, etc.) on training efficacy? • What degree of music-driven gains in phonological awareness are needed to impact reading fluency? • What are the mechanisms underlying improvement: such as attention, motivation, (e.g., OPERA hypothesis; Patel, 2011), speech prosody sensitivity, and/or working memory? • How are changes in brain function and structure associated with music-training-driven improvements? • How do individual differences predict response to training? Is there a subset of children that stands to benefit the most from music training?

musical training as children (Skoe and Kraus, 2012). One key difference is that the correlational studies tend to include children who have already had several years of individual instrumental instruction, whereas the intervention studies included here have shorter and less intense music training, and all were conducted in a group rather than individual setting. It could be that the music training in a group setting is less demanding and therefore less likely to make a large impact in terms of transferring to language skills (see OPERA hypothesis for a theoretically driven set of criteria for plasticity; Patel, 2011). Nevertheless, Hyde et al. (2009) showed neural plasticity and improvements in auditory and motor tasks, along with structural brain changes in auditory and motor areas, after 15 months of music training on an instrument. Furthermore, other experimental studies administering group-setting music training to participants randomly assigned to a music group (vs. a non-music control group, i.e., Moreno et al., 2009; Chobert et al., 2014) also found transfer to language perception skills; thus, individual instruction does not appear to be a pre-requisite for music-training-driven improvements in language skills. However, less is known about whether individual lessons and intensive instruction on an instrument are needed to improve reading-related skills.

The literature review encompassed by the present study revealed two somewhat opposing trends: on the one hand, an approach that favors *the contextual use of music* as a fun and motivational context to teach reading and other skills (Standley and Hughes, 1997; Standley, 2008; Darrow, 2009); and on the other hand, an *auditory neuro-development framework* that attributes music-training-related language gains primarily to auditory neural plasticity (Kraus and Chandrasekaran, 2010; Patel, 2011). In the “contextual” approach, phonological awareness and other literacy skills are taught in a musical context: for example, one intervention was described as teaching “literacy skills such as rhyming, letter sounds, vocabulary, or decoding sounds that were accompanied by a chant or song; children’s storybooks that were either read or sung or accompanied by the students on musical instruments as they recognized a previously identified vocabulary word; rearrangement of storybook parts with students asked to put the story pages in order and to retell the story in their own words” (Darrow, 2009, p. 14). Use of nursery rhymes is common and constitutes the foundation of one of the intervention curricula described in a study in the meta-analysis (Bolduc and Lefebvre, 2012). A number of studies have specifically targeted literacy skills within the music training, with musical activities designed to increase print awareness (Standley and Hughes, 1997); letter-naming, letter-sound correspondence, and word building (Register, 2004); and decoding (Register et al., 2007). Interestingly, in many of the *contextual* studies, music is thought of as a positive reinforcer of reading-related exercises, and little mention is made of the auditory system or its physiological underpinnings.

In contrast, the auditory neurodevelopment framework posits that music training strengthens basic auditory and speech processing, which in turn influence phonological perception and reading skills. These gains have been described as domain-general improvements in auditory brain mechanisms underlying temporal and frequency resolution, auditory processing, and

phonological awareness (Tierney and Kraus, 2013a). Experience-based plasticity of brain networks involved in language acquisition is a plausible explanation for the putative transfer of music training to language and literacy skills (reviewed in Kraus and Chandrasekaran, 2010). Randomized study designs conducted with neuro-imaging methods have shown that music lessons (in typically developing children) enhance neural responses to voice-onset-times and syllable durations (Chobert et al., 2014), detection of pitch variations in speech (Moreno et al., 2009), speech segmentation skills (François and Schön, 2011), and discrimination of consonants (Kraus et al., 2014b). Moreover, an association between brain responses to syllables (using the complex Auditory Brainstem Response method) and degree of active engagement (i.e., better classroom participation and attendance) in a music program suggests that the amount of training and level of engagement is an important factor in music-training-driven plasticity (Kraus et al., 2014a).

Another important aspect of the neurodevelopmental framework, thus far not definitively investigated in the literature, is that individual differences in innate (or pre-existing) musical traits may differentially affect music-training-driven plasticity and transfer to language skills. The extant literature does suggest that the relationship between language and music skills varies with different levels of music aptitude (Banai and Ahissar, 2013) and that pre-existing genetic differences likely account for some variation in level of music achievement attained (reviewed in Schellenberg, 2015). Given that individual differences in music abilities can predict some aspects of linguistic competence, even in non-musician children (Strait et al., 2011; Woodruff Carr et al., 2014; Gordon et al., 2015b), taking these individual differences into account could potentially provide a significant path to predicting response to music intervention. In this vein, Seither-Preisler et al. (2014) propose a fascinating neurocognitive model of competence development that would account for the interaction between pre-dispositions and intervention efficacy by modeling plasticity and anatomical influences on music development. They found that the size of the right Heschl’s Gyrus significantly predicted variance in the amount of time that children spent practicing their instruments; the authors interpret this finding as evidence that this particular neurophysiological morphology interacts with motivational factors that determine the amount of time/effort devoted to music. More generally, it is theoretically conceivable that a subset of children has a particular brain architecture that pre-disposes them to faster musical growth and more efficient transfer to language skills; while others may have neural substrates that respond better to other types of language interventions (e.g., phonological only). Continued investigation of these and other hypotheses regarding individual differences may turn out to reduce heterogeneity of findings in future individual studies and meta-analyses on the topic of music-training-driven changes in neural and cognitive activity.

The mixed results obtained in the current meta-analysis could instead signify possible limitations of music training for literacy skills in children. Such an interpretation could be regarded in accordance with previous accounts of modularity of some aspects of language and music (Peretz, 2006). For instance, Peretz et al.

(2015) argue that studies showing “neural overlap” of music and language in brain areas do not necessarily indicate that the same neuronal populations within a given brain area are active for both musical and speech processing. Moreover, it is important to bear in mind that small or non-robust effects of transfer from training to another skill are not unusual in the context of the larger literature on skills transfer. Many of the same methodological challenges (i.e., control group selection) encountered in the current meta-analysis are cited as prevalent issues for the skill-learning field much beyond music and language (Green et al., 2014). To this point, Green and Bavelier (2008) state “in the field of skill learning, transfer of learning from the trained task to even other very similar tasks is generally the exception rather than the rule.” Bransford and Schwartz (1999) suggest that the difficulty in finding consistent results of skills transfer stems in part from the idea that assessments of current knowledge generally do not capture the dynamics of the learning process. In the current meta-analysis, evidence that music training (that in some cases involves rhyming materials) has impacted performance on a standardized test of pre-reading skills (that has different surface features, cues, and demand characteristics) has crossed a substantial hurdle in establishing skills transfer; thus, even small gains should not be considered trivial.

To develop a full picture of the extent of transfer from music experiences to language skills and the possible applicability of the neuro-developmental framework, more work is also needed on the underlying mechanisms of music-related improvements in language when they are reported (either in individual studies or future meta-analyses). These effects could potentially be due to all-around, general acoustic perception/auditory processing skills (affecting perception of pitch, timing, and spectral characteristics); or, the benefits may be only specific to certain aspects of phonology such as fine-tuned detection of voice-onset-time (Zuk et al., 2013b), or perception of prosodic patterns on the supra-syllabic level. Indeed, a growing number of studies have linked speech rhythm sensitivity to early literacy skills. Sensitivity to stress patterns in spoken language are correlated with emerging reading skills in early readers (ages 5–7; Holliman et al., 2008; Goswami et al., 2010), and predict later reading development (Holliman et al., 2010). Struggling readers are also more likely to show weaknesses in perception of speech rhythm (Holliman et al., 2012) and musical rhythm (Huss et al., 2011; Flaunagacco et al., 2014). The temporal sampling theory (Goswami, 2011), along with work on neural oscillations involved in speech comprehension (Luo and Poeppel, 2007; Abrams et al., 2008; Hickok, 2012) converge in their explanation of a temporal scaffolding created by low-frequency stress patterns that facilitates acquisition and comprehension of higher-frequency (e.g., phonetic) information in the speech signal. These mechanisms may be shared by musical rhythm skills (Gordon et al., 2011; Hausen et al., 2013; Hickok et al., 2015; Morillon and Schroeder, 2015). Recent work translating related concepts of rhythm entrainment from dynamic attending theory to speech perception (Schön and Tillmann, 2015) suggest that even short-term rhythmic stimulation can impact phonological processing. A general deficit in these mechanisms of rhythm sensitivity could hinder acquisition of language and literacy

skills (e.g., Leong et al., 2011; Power et al., 2013); individual differences in rhythm sensitivity could possibly mediate response to treatment, and should be taken into account. Likewise, the role of auditory working memory in music-training-driven plasticity is not yet well-understood (Kraus et al., 2012; Ramachandra et al., 2012; Tierney and Kraus, 2013b) and should be accounted for in future intervention studies. **Table 5** summarizes potential questions to be addressed in future work.

The present meta-analysis contributes to the literature by examining the influence of music training on reading-related skills while also constraining the amount of reading instruction received across groups and modeling potentially important moderators (age, hours of training and type of control intervention). The findings yielded modest gains in phonological awareness (mainly in rhyming skills) for music vs. control interventions, but the small subset of studies examining reading fluency skills found no significant aggregate improvements in music vs. control groups. The literature review synthesized results from previous work suggesting potential benefits of music training on non-musical academic skills (e.g., Patel, 2011), supported by some evidence for a transfer from music training to rhyming and phonological awareness skills yielded by the present meta-analysis. This approach has also laid some groundwork for exploring specific aspects of the relationship between reading and music, which may take place in part through enhancement to perception of rhyming. This finding converges with the hypothesis that music supports phonological awareness; further study is needed to determine if intensive and long-term music training can enhance reading fluency via improvements to auditory skills, phonological awareness, and rhyming in particular. Given the limitations discussed here of the work included in this meta-analysis and the potential factors to address (summarized in **Table 5**), further investigation of a positive transfer from music education to reading-related skills is warranted. These investigations should eventually be considered in light of current trends in educational policy to cut funding for arts education (Kratus, 2007), such as when music lessons are eliminated in order to increase instructional time and resources for core subjects.

To draw definitive conclusions on a causal link from music to literacy and possible mediating mechanisms, there is abundant room for further progress in using longitudinal studies to address both the study design factors and the potential moderators of music-training-driven plasticity in reading-related skills. Brain imaging methods may reveal mechanisms underlying this plasticity, and can potentially be exploited to establish innovative approaches for predicting individual differences in response to music training. Recent work linking rhythmic processing to speech sound sensitivity and literacy skills suggests candidate mechanisms for improving reading skills via music education, and warrant further investigation in the context of using music training to remediate reading disabilities in school-age children. Future longitudinal studies incorporating both behavioral reading-related outcomes and measures of neural plasticity in typically developing and struggling readers are also needed in order to assess the viability of the neuro-developmental framework for music interventions.

ACKNOWLEDGMENTS

This project was funded in part through NIH award R01DC007694 to BM. The RedCAP system, made possible through UL1 TR000445 from NCATS/NIH to the Vanderbilt Institute for Clinical and Translational Research, was used for coding and secure data storage. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors would like to gratefully acknowledge Sandra Wilson and Gloria Han for insightful comments on the approach, McKenzie Miller for assistance with coding and data entry, Alison Williams for formatting assistance, and Rita Pfeiffer and two reviewers for

feedback and helpful comments on the manuscript. We would also like to thank Noreen Yazejian, Cathy Moritz, Georgios Papadelis, Sylvain Moreno, San Luis Castro, Mireille Besson, Franziska Degé, Hugo Cogo-Moreira, Dena Register, and Wendy Armstrong for providing data and/or additional information about their studies.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01777>

REFERENCES

- Abrams, D. A., Nicol, T., Zecker, S., and Kraus, N. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J. Neurosci.* 28, 3958–3965. doi: 10.1523/JNEUROSCI.0187-08.2008
- Anvari, S. H., Trainor, L. J., Woodside, J., and Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *J. Exp. Child Psychol.* 83, 111–130. doi: 10.1016/S0022-0965(02)00124-8
- Armand, F., and Montésinos-Gelet, I. (2001). *Apprentissage de la Lecture et de l'Écriture en Milieux Pluriethniques: Études des Contextes Langagiers et du Degré d'Automatisation des Processus en Lecture* (Organisme subventionnaire: Immigration et métropoles). Software programmer: Michel Bastien.
- Banai, K., and Ahissar, M. (2013). Musical experience, auditory perception and reading-related skills in children. *PLoS ONE* 8:e75876. doi: 10.1371/journal.pone.0075876
- Bhide, A., Power, A., and Goswami, U. (2013). A rhythmic musical intervention for poor readers: a comparison of efficacy with a letter-based intervention. *Mind Brain Educ.* 7, 113–123. doi: 10.1111/mbe.12016
- Bidelman, G. M., and Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J. Neurosci.* 35, 1240–1249. doi: 10.1523/JNEUROSCI.3292-14.2015
- Bidelman, G. M., Weiss, M. W., Moreno, S., and Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* 40, 2662–2673. doi: 10.1111/ejn.12627
- Bolduc, J. (2009). Effects of a music programme on kindergartners' phonological awareness skills. *Int. J. Music Educ.* 27, 37–47. doi: 10.1177/0255761408099063
- *Bolduc, J., and Lefebvre, P. (2012). Using nursery rhymes to foster phonological and musical processing skills in kindergartners. *Creative Educ.* 3, 495–502. doi: 10.4236/ce.2012.34075
- Bransford, J. D., and Schwartz, D. L. (1999). Rethinking transfer: a simple proposal with multiple implications. *Rev. Res. Educ.* 24, 61–100.
- Brod, G., and Opitz, B. (2012). Does it really matter? Separating the effects of musical training on syntax acquisition. *Front. Psychol.* 3:543. doi: 10.3389/fpsyg.2012.00543
- Bus, A. G., and van IJzendoorn, M. H. (1999). Phonological awareness and early reading: a meta-analysis of experimental training studies. *J. Educ. Psychol.* 91:403. doi: 10.1037/0022-0663.91.3.403
- Butera, I. M. (2015). From notes to vowels: neural correlations between musical training and speech processing. *J. Neurosci.* 35, 8379–8381. doi: 10.1523/JNEUROSCI.1102-15.2015
- Butzlaff, R. (2000). Can music be used to teach reading? *J. Aesthetic Educ.* 34, 167–178. doi: 10.2307/3333642
- Capovilla, A., and Capovilla, F. (1998). Prova de consciência fonológica: desenvolvimento de dez habilidades da pré-escola a segunda série [Phonological Awareness Test: development of ten abilities from preschool to third grade]. *Temas Desenvol.* 7, 14–20.
- Chobert, J., Francois, C., Velay, J. L., and Besson, M. (2014). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cereb. Cortex* 24, 956–967. doi: 10.1093/cercor/bhs377
- Chobert, J., Marie, C., Francois, C., Schön, D., and Besson, M. (2011). Enhanced passive and active processing of syllables in musician children. *J. Cogn. Neurosci.* 23, 3874–3887. doi: 10.1162/jocn_a.00088
- *Cogo-Moreira, H., Brandão de Ávila, C. R., Ploubidis, G. B., and Mari, J. D. J. (2013). Effectiveness of music education for the improvement of reading skills and academic achievement in young poor readers: a pragmatic cluster-randomized, controlled clinical trial. *PLoS ONE* 8:e59984. doi: 10.1371/journal.pone.0059984
- Darrow, A.-A. (2009). Enhancing literacy in the second grade: five related studies using the register music/reading curriculum. *Appl. Res. Music Educ.* 27, 12–26. doi: 10.1177/8755123308330044
- *Degé, F., and Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Front. Psychol.* 2:124. doi: 10.3389/fpsyg.2011.00124
- Dellatolas, G., Watier, L., Le Normand, M. T., Lubart, T., and Chevrie-Muller, C. (2009). Rhythm reproduction in kindergarten, reading performance at second grade, and developmental dyslexia theories. *Arch. Clin. Neuropsychol.* 24, 555–563. doi: 10.1093/arclin/acp044
- Dickinson, D., and Chaney, C. (1997). *Early Phonological Awareness Profile*. Newton, MA: Education Development Center.
- Douglas, S., and Willatts, P. (1994). The relationship between musical ability and literacy skills. *J. Res. Read.* 17, 99–107. doi: 10.1111/j.1467-9817.1994.tb00057.x
- Evans, S., Meekings, S., Nuttall, H. E., Jasmin, K. M., Boebinger, D., Adank, P., et al. (2014). Does musical enrichment enhance the neural coding of syllables? Neuroscientific interventions and the importance of behavioral data. *Front. Hum. Neurosci.* 8:964. doi: 10.3389/fnhum.2014.00964
- Flaugnacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., Tilli, S., et al. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Front. Hum. Neurosci.* 8:392. doi: 10.3389/fnhum.2014.00392
- Forgeard, M., Schlaug, G., Norton, A., Rosam, C., and Iyengar, U. (2008). The relations between music and phonological processing in normal-reading children and children with dyslexia. *Music Percept.* 25, 383–390. doi: 10.1525/mp.2008.25.4.383
- François, C., and Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cereb. Cortex* 21, 2357–2365. doi: 10.1093/cercor/bhr022
- Frederickson, N., Frith, U., and Reason, R. (1997). *Phonological Assessment Battery: Standardised Edition*. Windsor: NFER-Nelson.
- Good, R., and Kaminski, R. (2001). *Dynamic Indicators of Basic Early Literacy Skills, 5th Edn*. Eugene, OR: Institute for the Development of Educational Achievement.
- Good, R., and Kaminski, R. (2002). *Dynamic Indicators of Basic Early Literacy Skills*. Eugene, OR: Institute for the Development of Educational Achievement.
- Gordon, R. L., Jacobs, M. S., Schuele, C. M., and McAuley, J. D. (2015a). Perspectives on the rhythm-grammar link and its implications for typical

- and atypical language development. *Ann. N.Y. Acad. Sci.* 1337, 16–25. doi: 10.1111/nyas.12683
- Gordon, R. L., Magne, C. L., and Large, E. W. (2011). EEG correlates of song prosody: a new look at the relationship between linguistic and musical rhythm. *Front. Psychol.* 2:352. doi: 10.3389/fpsyg.2011.00352
- Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., and Devin McAuley, J. (2015b). Musical rhythm discrimination explains individual differences in grammar skills in children. *Dev. Sci.* 18, 635–644. doi: 10.1111/desc.12230
- Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci. (Regul. Ed.)* 15, 3–10. doi: 10.1016/j.tics.2010.10.001
- Goswami, U., Gerson, D., and Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Read. Writ.* 23, 995–1019. doi: 10.1007/s11145-009-9186-6
- Goswami, U., Huss, M., Mead, N., Fosker, T., and Verney, J. P. (2013). Perception of patterns of musical beat distribution in phonological developmental dyslexia: significant longitudinal relations with word reading and reading comprehension. *Cortex* 49, 1363–1376. doi: 10.1016/j.cortex.2012.05.005
- Green, C. S., and Bavelier, D. (2008). Exercising your brain: a review of human brain plasticity and training-induced learning. *Psychol. Aging* 23, 692–701. doi: 10.1037/a0014345
- Green, C. S., Strobach, T., and Schubert, T. (2014). On methodological standards in training and transfer experiments. *Psychol. Res.* 78, 756–772. doi: 10.1007/s00426-013-0535-3
- *Gromko, J. E. (2005). The effect of music instruction on phonemic awareness in beginning readers. *J. Res. Music Educ.* 53, 199–209. doi: 10.1177/002242940505300302
- Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., and Campitelli, G. (2014). Deliberate practice: is that all it takes to become an expert? *Intelligence* 45, 34–45. doi: 10.1016/j.intell.2013.04.001
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* 42, 377–381. doi: 10.1016/j.jbi.2008.08.010
- Hausen, M., Torppa, R., Salmela, V. R., Vainio, M., and Särkämö, T. (2013). Music and speech prosody: a common rhythm. *Front. Psychol.* 4:566. doi: 10.3389/fpsyg.2013.00566
- *Herrera, L., Lorenzo, O., Defior, S., Fernandez-Smith, G., and Costa-Giomi, E. (2011). Effects of phonological and musical training on the reading readiness of native- and foreign-Spanish-speaking children. *Psychol. Music* 39, 68–81. doi: 10.1177/0305735610361995
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* 13, 135–145. doi: 10.1038/nrn3158
- Hickok, G., Farahbod, H., and Saberi, K. (2015). The rhythm of perception: entrainment to acoustic rhythms induces subsequent perceptual oscillation. *Psychol. Sci.* 26, 1006–1013. doi: 10.1177/0956797615576533
- Higgins, J. P., and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Stat. Med.* 21, 1539–1558. doi: 10.1002/sim.1186
- Holliman, A. J., Wood, C., and Sheehy, K. (2008). Sensitivity to speech rhythm explains individual differences in reading ability independently of phonological awareness. *Br. J. Dev. Psychol.* 26, 357–367. doi: 10.1348/026151007X241623
- Holliman, A. J., Wood, C., and Sheehy, K. (2010). Does speech rhythm sensitivity predict children's reading ability 1 year later? *J. Educ. Psychol.* 102, 356–366. doi: 10.1037/a0018049
- Holliman, A. J., Wood, C., and Sheehy, K. (2012). A cross-sectional study of prosodic sensitivity and reading difficulties. *J. Res. Read.* 35, 32–48. doi: 10.1111/j.1467-9817.2010.01459.x
- Huss, M., Verney, J. P., Fosker, T., Mead, N., and Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. *Cortex* 47, 674–689. doi: 10.1016/j.cortex.2010.07.010
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., et al. (2009). Musical training shapes structural brain development. *J. Neurosci.* 29, 3019–3025. doi: 10.1523/JNEUROSCI.5118-08.2009
- Jansen, H., Mannhaupt, G., Marx, H., and Skowronek, H. (2002). *Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten*. Göttingen: Hogrefe.
- Kratz, J. (2007). Music education at the tipping point. *Music Educ. J.* 94, 42–48. doi: 10.1177/002743210709400209
- Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605. doi: 10.1038/nrn2882
- Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., et al. (2014a). Auditory learning through active engagement with sound: biological impact of community music lessons in at-risk children. *Front. Neurosci.* 8:351. doi: 10.3389/fnins.2014.00351
- Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., et al. (2014b). Music enrichment programs improve the neural encoding of speech in at-risk children. *J. Neurosci.* 34, 11913–11918. doi: 10.1523/JNEUROSCI.1881-14.2014
- Kraus, N., Strait, D. L., and Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory. *Ann. N.Y. Acad. Sci.* 1252, 100–107. doi: 10.1111/j.1749-6632.2012.06463.x
- Lamb, S. J., and Gregory, A. H. (1993). The relationship between music and reading in beginning readers. *Educ. Psychol.* 13, 19–27. doi: 10.1080/0144341930130103
- Leong, V., and Goswami, U. (2014). Assessment of rhythmic entrainment at multiple timescales in dyslexia: evidence for disruption to syllable timing. *Hear. Res.* 308, 141–161. doi: 10.1016/j.heares.2013.07.015
- Leong, V., Hamalainen, J., Soltesz, F., and Goswami, U. (2011). Rise time perception and detection of syllable stress in adults with developmental dyslexia. *J. Mem. Lang.* 64, 59–73. doi: 10.1016/j.jml.2010.09.003
- Lipsey, M. W., and Wilson, D. B. (2001). *Practical Meta-analysis*. Thousand Oaks, CA: Sage Publications.
- Lonigan, C. J., and Shanahan, T. (2009). *Developing Early Literacy: Report of the National Early Literacy Panel. Executive Summary. A Scientific Synthesis of Early Literacy Development and Implications for Intervention*. Washington, DC: National Institute for Literacy.
- Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004
- 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., Magne, C., and Besson, M. (2011). Musicians and the metric structure of words. *J. Cogn. Neurosci.* 23, 294–305. doi: 10.1162/jocn.2010.21413
- Melby-Lervag, M., Lyster, S. A., and Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *Psychol. Bull.* 138, 322–352. doi: 10.1037/a0026744
- Milovanov, R., Huotilainen, M., Esquef, P. A., Alku, P., Välimäki, V., and Tervaniemi, M. (2009). The role of musical aptitude and language skills in preattentive duration processing in school-aged children. *Neurosci. Lett.* 460, 161–165. doi: 10.1016/j.neulet.2009.05.063
- *Moreno, S., Friesen, D., and Bialystok, E. (2011). Effect of music training on promoting preliterate skills: preliminary causal evidence. *Music Percept.* 29, 165–172. doi: 10.1525/mp.2011.29.2.165
- *Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex* 19, 712–723. doi: 10.1093/cercor/bhn120
- Morillon, B., and Schroeder, C. E. (2015). Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. *Ann. N.Y. Acad. Sci.* 1337, 26–31. doi: 10.1111/nyas.12629
- *Moritz, C., Yampolsky, S., Papadelis, G., Thomson, J., and Wolf, M. (2013). Links between early rhythm skills, musical training, and phonological awareness. *Read. Writ.* 26, 739–769. doi: 10.1007/s11145-012-9389-0
- *Myant, M., Armstrong, W., and Healy, N. (2008). Can music make a difference? A small scale longitudinal study into the effects of music instruction in nursery on later reading ability. *Educ. Child Psychol.* 25, 83–100.
- Overly, K. (2003). Dyslexia and music: from timing deficits to musical intervention. *Ann. N.Y. Acad. Sci.* 999, 497–505. doi: 10.1196/annals.1284.060
- Parbery-Clark, A., Anderson, S., and Kraus, N. (2013). Musicians change their tune: how hearing loss alters the neural code. *Hear. Res.* 302, 121–131. doi: 10.1016/j.heares.2013.03.009
- Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., and Kraus, N. (2011). Musical experience and the aging auditory system: implications for

- cognitive abilities and hearing speech in noise. *PLoS ONE* 6:e18082. doi: 10.1371/journal.pone.0018082
- Patel, A. D. (2008). *Music, Language, and the Brain*. New York, NY: Oxford University Press.
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychol.* 2:142. doi: 10.3389/fpsyg.2011.00142
- Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear. Res.* 308, 98–108. doi: 10.1016/j.heares.2013.08.011
- 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., Vuvan, D., Lagrois, M. É., and Armony, J. L. (2015). Neural overlap in processing music and speech. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140090. doi: 10.1098/rstb.2014.0090
- Peynircioglu, Z. F., Durgunoglu, A. Y., and Üney-Küsefoglu, B. (2002). Phonological awareness and musical aptitude. *J. Res. Read.* 25, 68–80. doi: 10.1111/1467-9817.00159
- Power, A. J., Mead, N., Barnes, L., and Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Front. Hum. Neurosci.* 7:777. doi: 10.3389/fnhum.2013.00777
- Ramachandra, V., Meighan, C., and Gradzki, J. (2012). The impact of musical training on the phonological memory and the central executive: a brief report. *N. Am. J. Psychol.* 14, 541–548.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.
- Register, D. (2001). The effects of an early intervention music curriculum on prereading/writing. *J. Music Ther.* 38, 239–248. doi: 10.1093/jmt/38.3.239
- *Register, D. (2004). The effects of live music groups versus an educational children's television program on the emergent literacy of young children. *J. Music Ther.* 41, 2–27. doi: 10.1093/jmt/41.1.2
- Register, D., Darrow, A.-A., Standley, J., and Swedberg, O. (2007). The use of music to enhance reading skills of second grade students and students with reading disabilities. *J. Music Ther.* 44, 23–37. doi: 10.1093/jmt/44.1.23
- Robertson, C., and Salter, W. (1997). *The Phonological Awareness Test*. East Moline, IL: LinguSystems.
- Schellenberg, E. G. (2015). Music training and speech perception: a gene-environment interaction. *Ann. N.Y. Acad. Sci.* 1337, 170–177. doi: 10.1111/nyas.12627
- Schlaggar, B. L., and McCandliss, B. D. (2007). Development of neural systems for reading. *Annu. Rev. Neurosci.* 30, 475–503. doi: 10.1146/annurev.neuro.28.061604.135645
- Schön, D., and Tillmann, B. (2015). Short- and long-term rhythmic interventions: perspectives for language rehabilitation. *Ann. N.Y. Acad. Sci.* 1337, 32–39. doi: 10.1111/nyas.12635
- Seither-Preisler, A., Parncutt, R., and Schneider, P. (2014). Size and synchronization of auditory cortex promotes musical, literacy, and attentional skills in children. *J. Neurosci.* 34, 10937–10949. doi: 10.1523/JNEUROSCI.5315-13.2014
- 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.
- Skoe, E., and Kraus, N. (2012). A little goes a long way: how the adult brain is shaped by musical training in childhood. *J. Neurosci.* 32, 11507–11510. doi: 10.1523/JNEUROSCI.1949-12.2012
- Slevc, L. R., and Miyake, A. (2006). Individual differences in second-language proficiency: does musical ability matter? *Psychol. Sci.* 17, 675–681. doi: 10.1111/j.1467-9280.2006.01765.x
- Standley, J. M. (2008). Does music instruction help children learn to read? Evidence of a meta-analysis. *Appl. Res. Music Educ.* 27, 17–32. doi: 10.1177/8755123308322270
- Standley, J. M., and Hughes, J. E. (1997). Evaluation of an early intervention music curriculum for enhancing prereading/writing skills. *Music Ther. Perspect.* 15, 79–85. doi: 10.1093/mtp/15.2.79
- Strait, D. L., Hornickel, J., and Kraus, N. (2011). Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behav. Brain Funct.* 7:44. doi: 10.1186/1744-9081-7-44
- Succena, A., and Castro, S. (2010). *Bateria de Avaliação da Leitura em Português Europeu*. Lisboa: CEGOC.
- *Thomson, J. M., Leong, V., and Goswami, U. (2013). Auditory processing interventions and developmental dyslexia: a comparison of phonemic and rhythmic approaches. *Read. Writ.* 26, 139–161. doi: 10.1007/s11455-012-9359-6
- Tierney, A., and Kraus, N. (2013a). Music training for the development of reading skills. *Prog. Brain Res.* 207, 209–241. doi: 10.1016/B978-0-444-63327-9.00008-4
- Tierney, A. T., and Kraus, N. (2013b). The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain Lang.* 124, 225–231. doi: 10.1016/j.bandl.2012.12.014
- Torgesen, J., Wagner, R., and Rashotte, C. (1999). *Test of Word Reading Efficiency (TOWRE)*. Austin, TX: ProEd.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 36, 1–48. doi: 10.18637/jss.v036.i03
- Walker, K. M., Hall, S. E., Klein, R. M., and Phillips, D. P. (2006). Development of perceptual correlates of reading performance. *Brain Res.* 1124, 126–141. doi: 10.1016/j.brainres.2006.09.080
- 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
- Woodcock, R. W., Mather, N., and McGrew, K. S. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities Examiner's Manual*. Itasca, IL: Riverside
- Woodruff Carr, K., White-Schwoch, T., Tierney, A. T., Strait, D. L., and Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14559–14564. doi: 10.1073/pnas.1406219111
- *Yazejian, N., and Peisner-Feinberg, E. S. (2009). Effects of a preschool music and movement curriculum on children's language skills. *NHSA Dialog.* 12, 327–341. doi: 10.1080/15240750903075255
- Zuk, J., Andrade, P. E., Andrade, O. V., Gardiner, M., and Gaab, N. (2013a). Musical, language, and reading abilities in early Portuguese readers. *Front. Psychol.* 4:288. doi: 10.3389/fpsyg.2013.00288
- Zuk, J., Ozernov-Palchik, O., Kim, H., Lakshminarayanan, K., Gabrieli, J. D., Tallal, P., et al. (2013b). Enhanced syllable discrimination thresholds in musicians. *PLoS ONE* 8:e80546. doi: 10.1371/journal.pone.0080546

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 © 2015 Gordon, Fehd and McCandliss. 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.

*Studies included in the meta-analyses are indicated with asterisks.



Sound frequency affects speech emotion perception: results from congenital amusia

Sydney L. Lolli, Ari D. Lewenstein, Julian Basurto, Sean Winnik and Psyche Loui*

Department of Psychology, Program in Neuroscience and Behavior, Wesleyan University, Middletown, CT, USA

OPEN ACCESS

Edited by:

Edward W. Large,
University of Connecticut, USA

Reviewed by:

Erin E. Hannon,
University of Nevada, Las Vegas, USA
Sébastien Paquette,
University of Montréal, Canada

*Correspondence:

Psyche Loui,
Department of Psychology, Program
in Neuroscience and Behavior,
Wesleyan University, 207 High Street,
Middletown, CT 06459, USA
ploui@wesleyan.edu

Specialty section:

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

Received: 08 April 2015

Accepted: 20 August 2015

Published: 08 September 2015

Citation:

Lolli SL, Lewenstein AD, Basurto J,
Winnik S and Loui P (2015) Sound
frequency affects speech emotion
perception: results from
congenital amusia.
Front. Psychol. 6:1340.
doi: 10.3389/fpsyg.2015.01340

Congenital amusics, or “tone-deaf” individuals, show difficulty in perceiving and producing small pitch differences. While amusia has marked effects on music perception, its impact on speech perception is less clear. Here we test the hypothesis that individual differences in pitch perception affect judgment of emotion in speech, by applying low-pass filters to spoken statements of emotional speech. A norming study was first conducted on Mechanical Turk to ensure that the intended emotions from the Macquarie Battery for Evaluation of Prosody were reliably identifiable by US English speakers. The most reliably identified emotional speech samples were used in Experiment 1, in which subjects performed a psychophysical pitch discrimination task, and an emotion identification task under low-pass and unfiltered speech conditions. Results showed a significant correlation between pitch-discrimination threshold and emotion identification accuracy for low-pass filtered speech, with amusics (defined here as those with a pitch discrimination threshold >16 Hz) performing worse than controls. This relationship with pitch discrimination was not seen in unfiltered speech conditions. Given the dissociation between low-pass filtered and unfiltered speech conditions, we inferred that amusics may be compensating for poorer pitch perception by using speech cues that are filtered out in this manipulation. To assess this potential compensation, Experiment 2 was conducted using high-pass filtered speech samples intended to isolate non-pitch cues. No significant correlation was found between pitch discrimination and emotion identification accuracy for high-pass filtered speech. Results from these experiments suggest an influence of low frequency information in identifying emotional content of speech.

Keywords: amusia, tone-deafness, pitch, filtering, speech, emotion, frequency

Introduction

Pitch is a perceptual attribute of sound that allows us to order sounds on a frequency-related scale. It is an integral component of auditory processing, including music and language. Across all spoken languages, pitch is one of several cues used to convey emotional prosody, and in some language (tone languages) pitch is also used to convey meaning in words. Understanding how pitch perception affects our interpretation of speech is essential to fully comprehend the ways in which we communicate emotion through language.

Amusic, or “tone-deaf” individuals, are limited in their ability to perceive and produce pitch (Peretz et al., 2002; Hyde and Peretz, 2004; Vuvan et al., 2015). Though amusia is traditionally

thought of as a music-specific disorder, studies have shown that it may also affect perception of speech. In common-practice Western music, pitches typically vary by a minimum of one semitone. In language, intonation patterns that help us discriminate between statements and questions are characterized by pitch differences that range from 5 to 12 semitones, and occur primarily at the conclusion of a speech fragment (Hutchins et al., 2010). By contrast, pitch changes that reflect prosody in emotional speech lie somewhere in between one and five semitones, and occur over the course of a speech fragment, suggesting that pitch variations in emotion expression are harder to detect than question-statement differences (Dowling and Harwood, 1986).

Consistent with this hypothesis, Hutchins et al. (2010) showed that when asked to discriminate between statements and questions, amusics performed as well as controls. However, when asked to judge whether the same stimuli ended with a rising or falling contour, amusics were significantly less accurate and consistent, suggesting a deficit of pitch awareness in amusics. Though amusic subjects self-reported no difficulties during day-to-day speech processing, Jiang et al. (2012) found that amusics' brain activity was not reliably elicited in response to pitch changes of one semitone in speech [this is in contrast to some early processing of small pitch changes without conscious awareness in music (Peretz et al., 2005, 2009)]. Also, (Nguyen et al., 2009) observed some decreases in sensitivity to pitch inflections found in a tonal language among amusic non-tonal language speakers (Nguyen et al., 2009). Although results from amusics are task-dependent and do overlap with non-amusic controls, studies generally show that amusics have some impairments in speech intonation processing, extending the effects of the disorder beyond music. Other studies have shown that amusics self-report difficulty detecting certain nuances in speech, such as sarcasm, and that they struggle to judge emotional content of speech as accurately as non-amusics (Thompson et al., 2012). In addition, individuals with amusia-like deficiencies report difficulty in determining emotion solely from speech, and may rely more on facial expressions and gestures than control subjects do (Thompson et al., 2012). Though there are other cues in emotional communication that are available to amusics, limitations in the ability to perceive pitch clearly contribute to deficiencies in emotional speech perception.

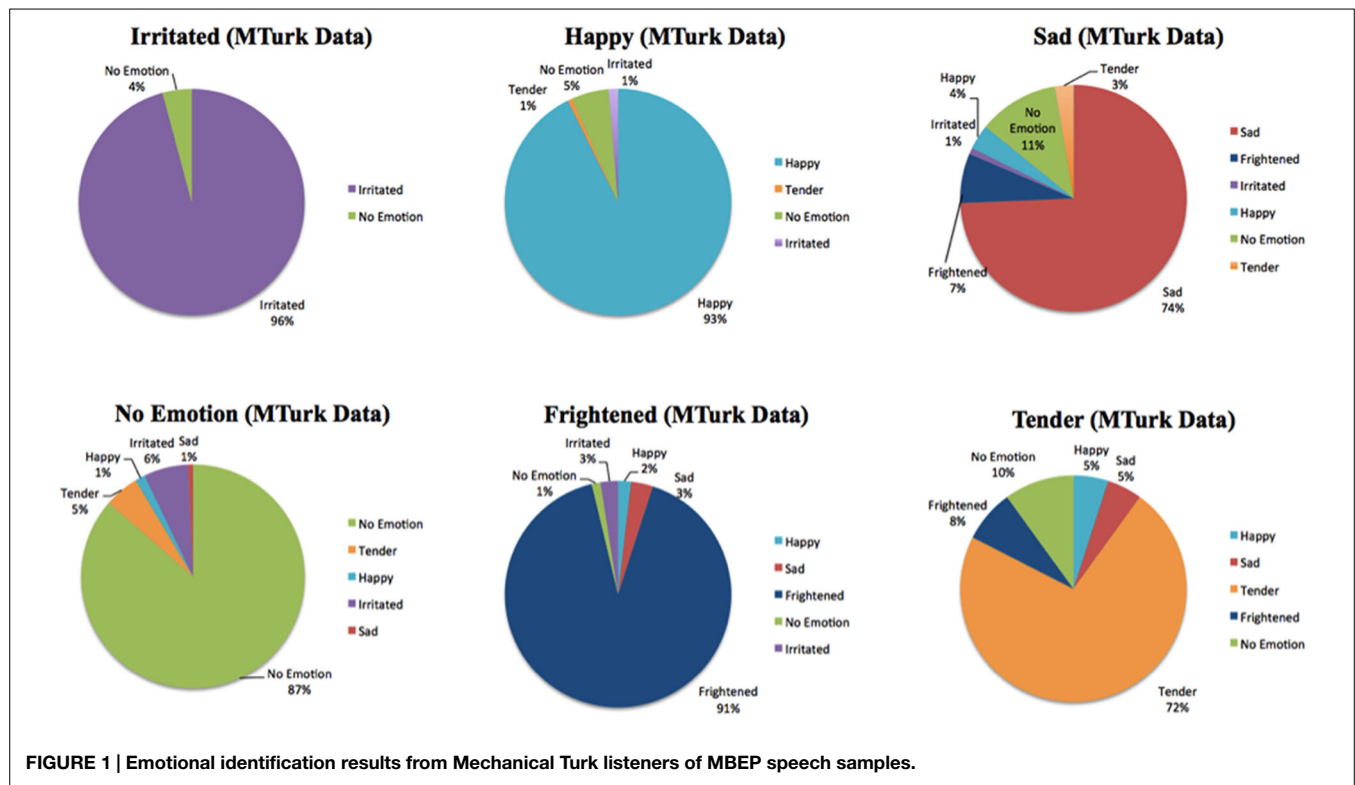
It has been hypothesized that deficiencies may only be noticeable when amusics are presented with very subtly different stimuli. Liu et al. (2012) presented statement-question discrimination tasks to Mandarin speakers, under conditions of natural speech and gliding tone analogs. Amusics were worse at discriminating gliding tone sequences, and had significantly higher thresholds than controls in detecting pitch changes as well as pitch change directions. However, amusics and controls performed similarly in tasks involving multiple acoustic cues, suggesting that instead of using fine-grained pitch differences to interpret meaning, individuals with pitch perception deficits might have relied on some non-pitch cues. In another study, Liu et al. (2010) presented similar statement-question discrimination tasks under the conditions of natural speech, gliding tones, and non-sense speech analogs. Amusics performed significantly worse than non-amusic control participants in discrimination

under all three conditions, suggesting deficiencies not only in samples with isolated pitch contour, but also in natural speech. Liu et al. (2015) again examined this link between amusia and speech processing in Mandarin speakers using speech samples with normal or flattened fundamental frequency contours. Amusics showed reduced speech comprehension when listening to flattened samples in quiet and noisy conditions, while controls only showed reduced speech comprehension in noisy conditions, suggesting that amusics experience speech comprehension difficulties in everyday listening conditions, with deficits extending to impaired segmental processing, rather than being limited to pitch processing.

Our study aims to analyze the extent of impairment in more nuanced areas of speech, namely emotional recognition. It has been suggested that individuals may compensate for poor pitch perception by relying more heavily on alternative cues within speech to infer emotional content, such as stress and emphasis (Hutchins et al., 2010). Speech segments that express five emotions (happy, sad, irritated, fearful, tender) and no emotion are presented as both filtered and non-filtered stimuli to participants. Rather than focusing exclusively on amusic populations, our goal is to test how individual differences in pitch perception can impact the processing of emotional prosody.

Frequency filtering methods are often used in tests that diagnose deficits in auditory perception, in order to simulate subtle differences in music and speech content (Patel et al., 1998; Ayotte et al., 2002; Bhargava and Başkent, 2012; O'Beirne et al., 2012). Low-pass filters may be used to examine speech intelligibility independently or in conjunction with other auditory disturbances (Horwitz et al., 2002; Bhargava and Başkent, 2012). The majority of speech prosody cues are preserved, while speech intelligibility is lost, with a sharply sloped low-pass filter around 500 Hz (Knoll et al., 2009; Guellaï et al., 2014). In our first experiment we applied a low-pass filter that attenuates frequencies above 500 Hz to disrupt intelligibility while still maintaining the fundamental frequency of speech sounds, which gives rise to their pitch contour. In our second experiment, we applied a high-pass filter in order to retain cues other than pitch contour, such as accents and sibilants, which may provide emotional cues. High-pass filters have been used in previous studies, but rarely in amusic populations. Our filter attenuated frequencies below 4800 Hz, providing the listener with minimal pitch contour while preserving rhythmic structure and sibilants.

Natural speech contains many cues that amusics can perceive, prompting them to report predominantly normal speech perception. Studies suggest that amusics who do not report deficiencies in everyday speech may more heavily weigh tempo, mode, and linguistic content in processing emotional significance (Peretz et al., 1998; Gosselin et al., 2015). Low-pass and high-pass filtered speech, in contrast to natural, unfiltered speech, contain less information to factor into individuals' interpretation of emotional content. We hypothesize that there will be a negative correlation between pitch discrimination thresholds and accuracy in emotional identification under low-pass conditions, i.e., that individuals with poorer pitch perception skills are less able to use low-frequency speech cues



to identify emotional prosody. We also hypothesize that unlike low-pass filtering, high-pass filtering speech samples will not affect emotional identification disproportionately for poor pitch perceivers.

Norming Study

The Macquarie Battery for Evaluation of Prosody (MBEP) has been used in previous experiments to assess the effects of amusia on emotional prosody perception (Thompson et al., 2012). The Macquarie database was created from semantically neutral statements (e.g., “The broom is in the closet and the pen is in the drawer”), read by four male and four female actors to represent no emotion and five different emotions (happy, sad, tender, irritated, and frightened). The statements are 14 syllables long, and the emotions were chosen for the variety of acoustic cues that they offer. In total, the database included 96 recorded statements. The statements in the MBEP were recorded in Australia, and thus are recorded with an Australian accent. We performed a norming study on Amazon Mechanical Turk to ensure that American subjects would be able to properly identify emotion in Australian-accented speech.

Methods

Ninety-six statements from MBEP were presented as separate, single-question surveys on Amazon’s Mechanical Turk, and subjects were allowed 1 min to listen and respond by identifying the emotion. Subjects were paid \$0.05 per question. Each of the 96 statements in the database received 10 responses from users in the United States.

Results

Results from the norming study are shown in **Figure 1**. Subjects performed well above chance levels in all emotional categories, confirming that American subjects were able to identify emotion in Australian-accented speech.

Discussion

Listeners were reliably successful at identifying the intended emotion from MBEP speech samples. “Irritated” was the most commonly correctly identified emotion, while “tender” was the least commonly correctly identified emotion. Based on listeners’ responses, statements in which respondents chose the target emotion less than 50% (chance level = 16.7%) of the time were excluded from use in the study. Tender statements were more likely than other emotions to be excluded, as they were typically more difficult to identify. Twelve statements from the set were excluded from use, resulting in 84 speech samples in the rest of the study. These 84 speech samples included 16 of Happy, Frightened, Irritated, and No Emotion, 14 Sad samples, and six Tender samples.

Experiment 1: Low-Pass Filter

Materials and Methods

Participants

Forty participants (21 women and 19 men) aged 18–22 from an introductory psychology course at Wesleyan University participated in exchange for course credit. All participants gave informed consent as approved by the Psychology Ethics Board of Wesleyan University. Participants reported no hearing

impairment, neurological disorders, or psychiatric disorders. Twenty-five of the forty participants reported musical training with varying instruments for lengths of time ranging from 6 months to 13 years. Across participants with previous musical training, an average of 6.5 years of training was reported. All subjects took the Montreal Battery of Evaluation of Amusia (MBEA) as well as the pitch discrimination test. Pitch discrimination thresholds, as identified by the pitch discrimination task (described below), ranged from 1.5 to 48 Hz (mean = 10.5 Hz). Nine subjects were considered amusic based on their inability to identify differences in pitch greater than 16 Hz apart (at 500 Hz) in the pitch discrimination task (amusic mean = 23.2 Hz, SD = 10.4 Hz; control mean = 6.8 Hz, SD = 3.9 Hz). Fifteen subjects were considered amusic based on their scores on the MBEA contour subtest (fewer than 23 correct responses out of 31 possible). Four subjects failed both the pitch discrimination threshold test and the MBEA. While the MBEA and pitch discrimination test both measure aspects of musical perception, especially pitch perception, MBEA is broader and also measures attention and working memory. Here we rely on the pitch discrimination test because we are interested more specifically in pitch discrimination aspects of musical function, rather than the attention and working memory components.

Materials

Several tests were administered to assess musical ability and training: the contour subtest of the a pitch discrimination threshold test MBEA, a questionnaire on demographic information and musical training, and the Shipley Institute of Living Scale (Shipley, 1940), used as a non-verbal IQ control task as it has been shown to be a predictor of WAIS-IQ scores (Paulson and Lin, 1970). Amusia was measured using the contour subtest of the MBEA (Peretz et al., 2003) and a pitch discrimination task. In the contour subtest, two brief melodies are presented that are either identical or differ to varying degrees in pitch contour. The pitch discrimination threshold test (Loui et al., 2008) determines the smallest pitch interval that participants are able to distinguish by presenting a series of two tones and asking whether the second tone is higher or lower in pitch than the first. The test uses a three-up one-down staircase procedure to find the threshold range of pitch perception. The questionnaire administered to the participants included questions about the following: sex, date of birth, first languages, and history of hearing impairment, neurological disorders, or psychological disorders. The questionnaire also included questions on participants' musical training history. If the subject responded that they had trained on an instrument, he or she was asked to share the length of training, age of onset, and the instrument(s) trained on.

A behavioral test was then administered using 84 non-filtered and 84 low-pass filtered speech samples from the MBER, chosen from the norming study reported above. The non-filtered trial condition consisted of natural (unfiltered) speech samples directly from the database, excepting 12 samples that Mechanical Turk workers did not reliably identify with above 50% accuracy. The low-pass filtered trial condition consisted of frequency-filtered versions of the same 84 speech samples, filtering out frequencies above 500 Hz. Filtering was done in Logic X with the plugin

"Channel EQ" (Q factor = 0.75, slope = 48 dB/Octave). This low-pass filtered condition was intended to eliminate formants and other high-frequency cues from the speech samples, while preserving the pitch contour of the speech samples. See **Figure 2** for spectrogram representations of unfiltered (**Figure 2A**) and low-pass filtered (**Figure 2B**) speech samples.

Procedure

Participants were individually administered the tests as stated above in a laboratory setting with minimal noise interference. Stimuli were presented through Sennheiser 280 HD Pro headphones connected to a desktop iMac computer at a comfortable volume for the subject. The experiment was created using Max/MSP and the two trial blocks were presented in a randomized order, with the aim of balancing out any potential order effects of the blocks. All subjects were equally likely to start on unfiltered and filtered speech. The speech samples within each trial block were also presented in a randomized order. Subjects used the mouse to choose one of the six emotion categories listed from among six options: Happy, Sad, Irritated, Frightened, Tender, and No emotion.

Data Analysis

Data were exported from the experiment in Max/MSP to Excel and SPSS for analysis. Pitch discrimination thresholds were log-transformed (log base 10) to achieve normal distribution.

Results

Log pitch discrimination threshold was significantly correlated with emotional identification accuracy in the low-pass filtered condition [$r(38) = -0.38$, $p = 0.015$; **Figure 3A**] but not in the unfiltered speech condition [$r(38) = 0.04$, n.s.; **Figure 3B**].

Amusics (as identified by pitch discrimination thresholds) performed worse than controls in the filtered condition [$t(38) = -3.13$, $p = 0.003$], but not in the unfiltered speech condition [$t(38) = -0.58$, n.s.; **Figure 3C**]. When amusics were identified using the contour subtest of the MBEA, their performance in the low-pass filtered condition was still below that of controls (amusics mean = 62%, SD = 16%; controls mean = 70%, SD = 12%); however the difference was only marginally significant [$t(38) = 1.7$, $p = 0.09$]. Amusics identified using the MBEA contour test did not differ in performance from controls in the unfiltered speech condition [amusics mean = 84%, SD = 10%, controls mean = 81%, SD = 9%, $t(38) = 1.14$, $p = 0.26$].

When holding musical training constant in a partial correlation, accuracy under low-pass conditions was still correlated with pitch discrimination threshold [$r(37) = -0.35$, $p = 0.028$] and unfiltered speech condition accuracy remained uncorrelated [$r(37) = -0.04$, n.s.]. These results confirm that even when controlling for musical training, pitch perception was significantly correlated with emotional identification accuracy under low-pass filtered but not under unfiltered speech conditions. When controlling for Shipley Abstraction scores, the correlations hold at $r(37) = -0.38$, $p = 0.018$ for the low-pass condition, and $r(37) = -0.04$, n.s. for the unfiltered speech

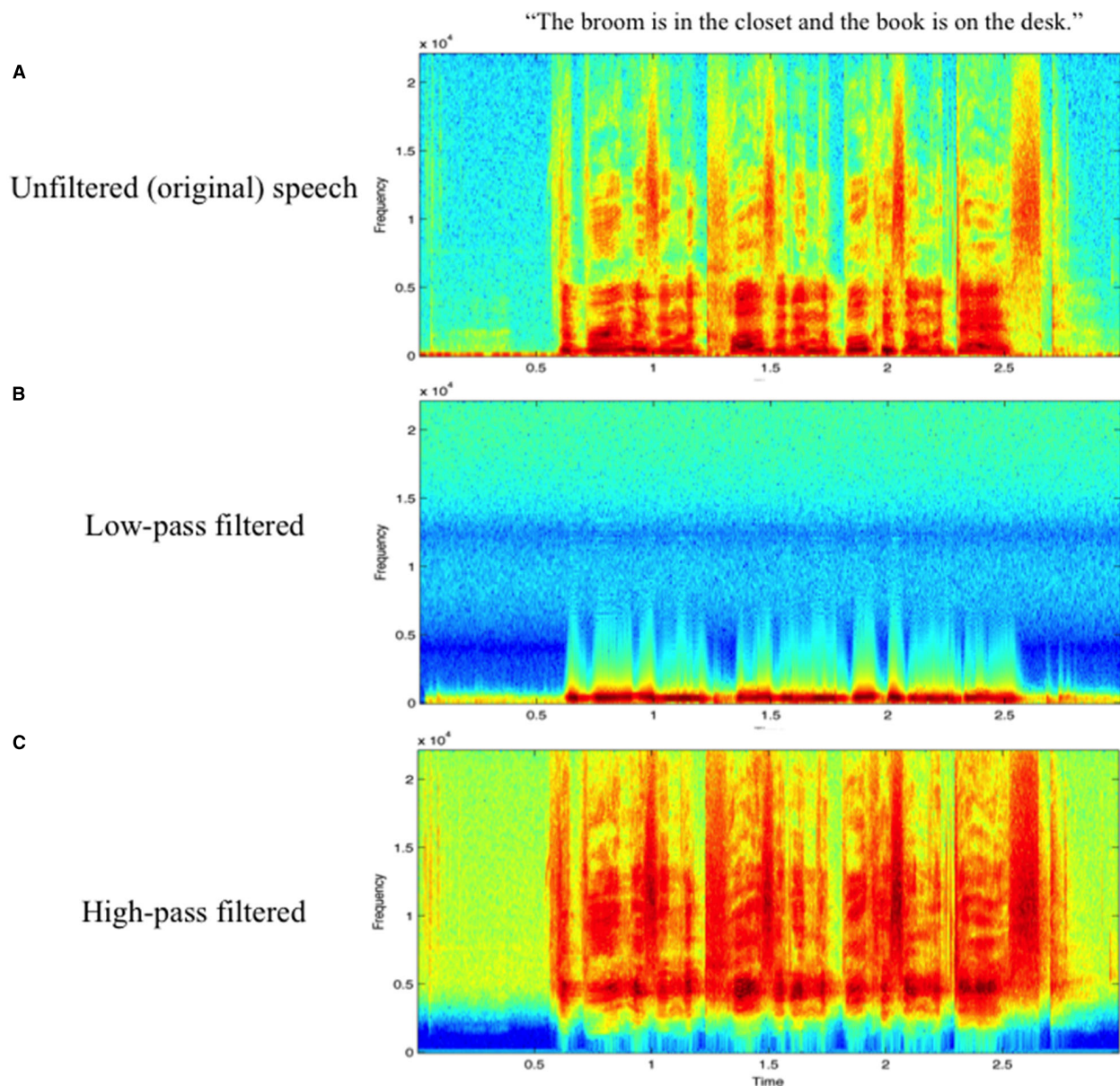
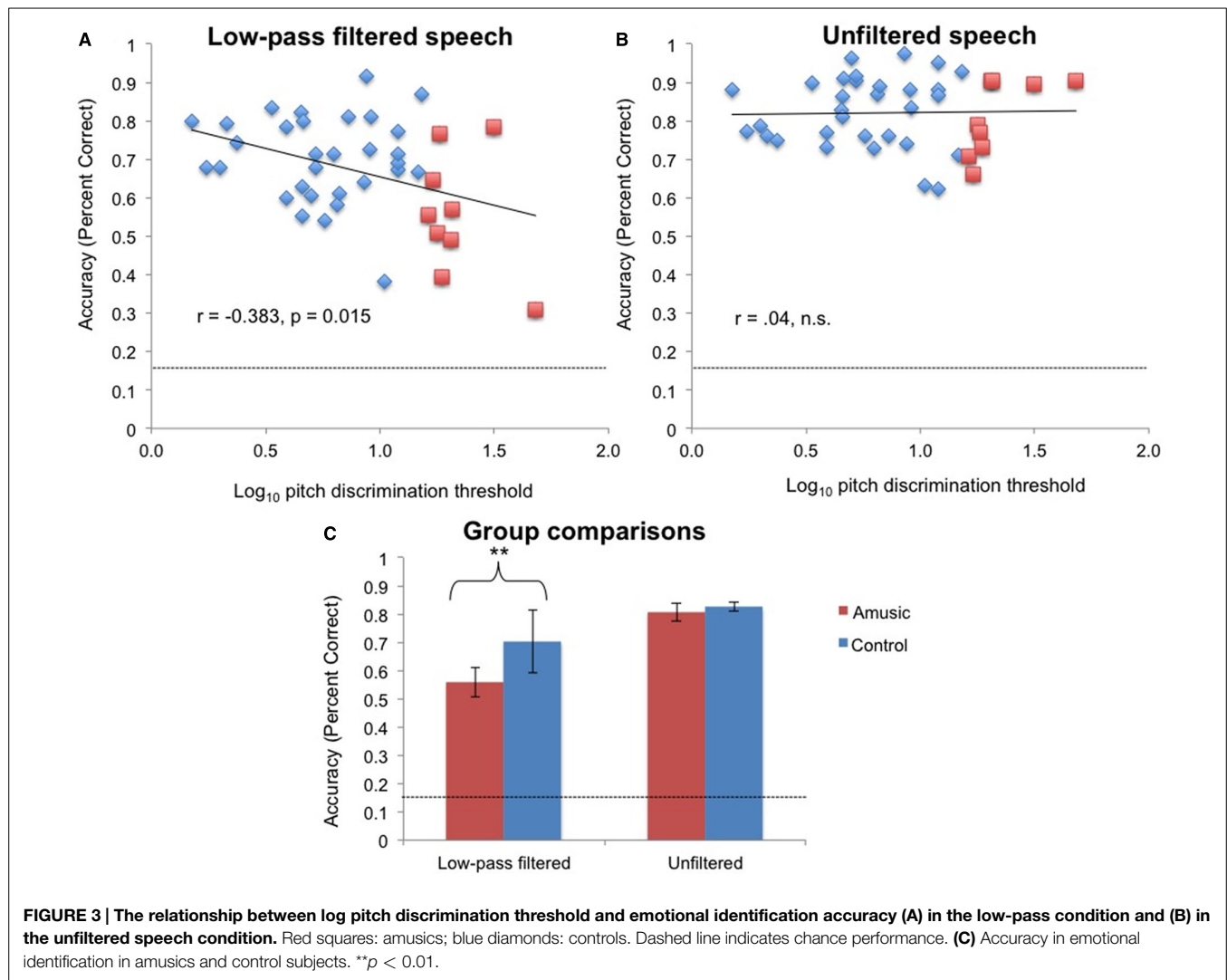


FIGURE 2 | Spectrograms of a representative speech sample in (A) unfiltered, (B) low-pass filtered, and (C) high-pass filtered conditions.

condition. Using both Shipley scores and musical training as control variables, accuracy in the filtered condition remained correlated with pitch discrimination scores [$r(36) = -0.35$, $p = 0.033$] and unfiltered speech accuracy remained uncorrelated [$r(36) = 0.03$, n.s.].

As subjects were randomly assigned to begin the experiment with the low-pass filtered block ($n = 19$, 6 amusics) or the unfiltered block ($n = 21$, 3 amusics), it was possible for block order to have influenced results: specifically, experience with the unfiltered speech condition could have helped a subject's subsequent performance on the low-pass condition. A follow-up analysis was conducted to assess the effects of block order on performance in the low-pass filtered condition. Order was incorporated as a variable in a between-subject ANOVA. A

two-way ANOVA on the dependent variable of accuracy in the low-pass condition, with the factors of group (amusics vs. controls) and block order (low-pass first vs. unfiltered speech first) showed a significant main effect of amusia [$F(1,36) = 5.5$, $p = 0.025$] and a significant main effect of block order [$F(1,36) = 7.3$, $p = 0.01$], as well as a significant interaction between amusia and block order [$F(1,36) = 4.8$, $p = 0.034$]. In addition to confirming that amusics performed worse at emotional identification in low-pass filtered speech, this result suggests that subjects learned to identify emotions via prosody throughout the course of the experiment: those who started with unfiltered speech subsequently performed better on the low-pass filtered condition, compared to those who started on the low-pass filtered condition, presumably because subjects learned during the



unfiltered speech condition to listen for pitch as an emotional cue. Interestingly, the significant interaction between group and block order shows that the amusics who started on the low-pass condition performed worse than the amusics who started on the natural speech condition, who were indistinguishable in performance from controls. This interaction suggests that learning throughout the experiment may occur even more in amusics than in controls.

Scores on the MBEA showed no significant correlation with emotional identification accuracy in the low-pass filtered condition [$r(38) = 0.18, n.s.$]. MBEA was not correlated with emotional identification accuracy under unfiltered speech conditions [$r(38) = -0.04, n.s.$]. Amusics (as identified by MBEA score) did not perform significantly differently between the filtered condition [$t(37) = -0.33, n.s.$] and the unfiltered speech condition [$t(38) = 1.20, n.s.$].

Discussion

Results show a robust association between pitch perception ability and accuracy of emotional identification in speech in

the low-pass filtered conditions, but not in unfiltered speech. Amusic individuals, identified as those who have poor pitch perception abilities, are impaired in identifying the emotional content of speech when high-frequency cues are removed from the speech. These individual differences are uniquely related to pitch discrimination abilities, and are not explained by differences in general IQ or musical training.

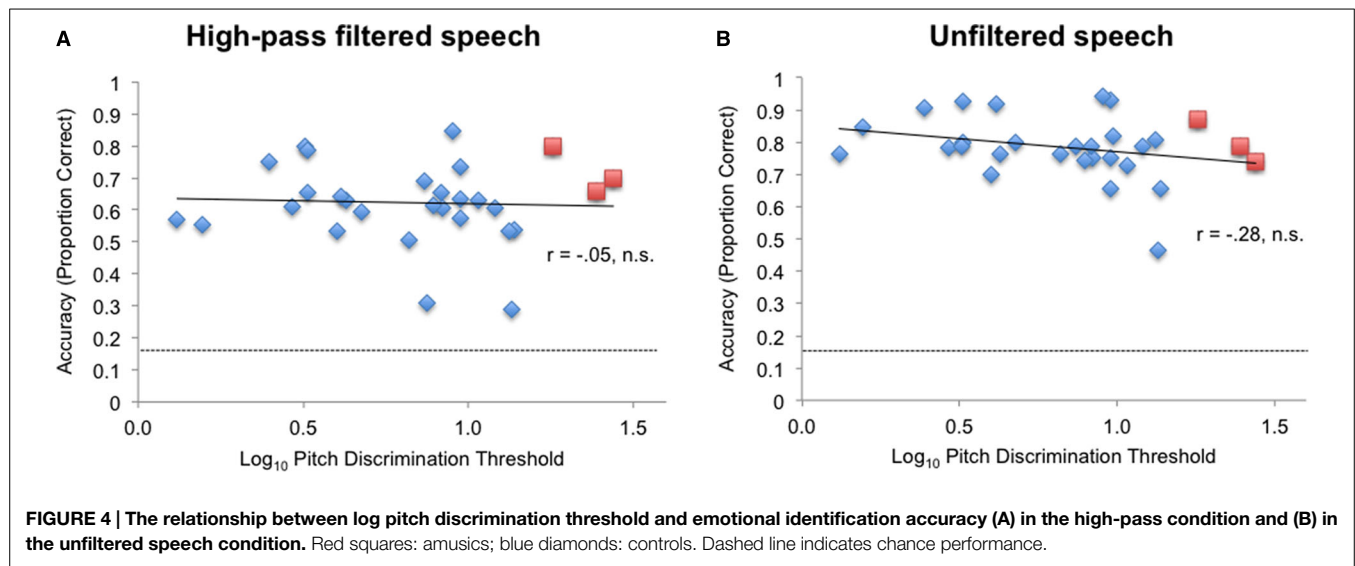
Given the dissociation between low-pass filtered and unfiltered speech conditions, we inferred that amusics may be compensating for poorer pitch perception by using speech cues that are filtered out in the former manipulation. To assess this potential compensation, a second experiment was conducted, using high-pass filtered speech samples intended to isolate non-pitch cues.

Experiment 2: High-Pass Filter

Materials and Methods

Participants

Twenty-nine participants (17 women and 12 men) aged 18–28 from an introductory psychology course at Wesleyan University participated in exchange for course credit. Participants reported



no hearing impairment, neurological disorders, or psychiatric disorders. Twenty-one of the 27 participants reported musical training with varying instruments for lengths of time ranging from 1 to 11 years. Among participants with previous musical training, an average of 5.6 years of training was reported. All subjects took the Montreal Battery as well as the pitch discrimination test. Pitch discrimination thresholds, as identified by the pitch discrimination task (described below), ranged from 1.3 to 27.5 Hz (mean = 10.5 Hz). Three participants were considered amusic based on their inability to identify differences in pitch greater than 16 Hz apart (at 500 Hz) in the pitch discrimination task (amusic mean = 26 Hz, SD = 2.1 Hz; control mean = 7.8 Hz, SD = 4.3 Hz). Twelve participants were considered amusic based on their scores on the MBEA contour subtest (fewer than 23 correct responses out of 31 possible). Three participants failed both the pitch discrimination and the MBEA tests.

Materials

The tests used to assess musical ability and training and the Shipley Institute of Living Scale were the same as administered in Experiment 1. A behavioral test of emotional identification was then administered using the same 84 unfiltered (original) speech samples from the MBEP (the same unfiltered speech samples used in Experiment 1, chosen from the norming study reported above), and 84 new high-pass filtered speech samples generated for this experiment. Filtering was done in Logic X with the plugin “Channel EQ” (Q factor = 0.75, slope = 48 dB/Octave). The frequency cutoff for high-pass filtering was chosen at 4800 Hz (i.e., frequencies lower than 4800 Hz were attenuated) to eliminate cues such as pitch contour and the majority of formant frequencies, while preserving other cues such as speech rate, stress patterns, and rhythm.

Procedure

Stimuli were presented through Sennheiser 280 HD Pro headphones connected to a desktop iMac computer at a comfortable volume for the subject. The main experiment was

created using Max/MSP and the two trial blocks were presented in a randomized order to the participant. The speech samples within each trial block were also presented in a randomized order. Subjects used the mouse to choose one of the six emotion categories as in Experiment 1.

Data Analysis

As in Experiment 1, data were exported from the experiment in Max/MSP to Excel and SPSS for analysis. Pitch discrimination thresholds were log-transformed (log base 10) to achieve normal distribution.

Results

As shown in **Figures 4A,B**, pitch discrimination threshold was not significantly correlated with accuracy under high-pass conditions [$r(27) = -0.05, n.s.$], or with accuracy under unfiltered speech conditions [$r(27) = -0.28, n.s.$]. MBEA was also not significantly correlated with overall accuracy of subjects under unfiltered speech conditions or under high-pass conditions.

While it appears that the high-pass filtering manipulation on the speech samples did not result in the same sensitivity to pitch discrimination differences compared to the low-pass filtered speech in Experiment 1, an additional possibility was that differences between the two experiments resulted from using different subjects between the two experiments, i.e., a sampling difference, which is potentially a confound especially since there were only three subjects who met the pitch-discrimination threshold criterion for amusia within the sample of Experiment 2. In a follow-up analysis to test the equivalence of samples between Experiments 1 and 2, we chose a subset of subjects from among our subjects in Experiment 1 who were matched for pitch discrimination thresholds, Shipley scores, and musical training to our subjects in Experiment 2, thereby repeating our analysis with only 3 amusics. A significant negative correlation was still observed between log pitch discrimination threshold and accuracy in the low-pass filtered speech condition, even within

this reduced subset of the Experiment 1 sample [$r(27) = -0.37$, $t(27) = 2.07$, $p = 0.048$]. This confirms that the samples of amusic and control subjects are comparable between the two experiments, and that the difference in data pattern between Experiments 1 and 2 is due to our experimental manipulations of the speech samples rather than to sampling differences between the experiments.

Discussion

Results showed no significant relationship between emotional identification accuracy and individual differences of pitch discrimination, in either the unfiltered speech or the high-pass filtered speech conditions. Although only three of the 29 subjects in this experiment showed pitch discrimination thresholds that exceeded the cutoff for amusia, a continuum of individual differences in pitch discrimination was captured in the present sample. High-pass filtering the speech samples did not result in any positive relationship between emotional identification and pitch discrimination, suggesting that individuals with poor pitch perception were not systematically using high-frequency information in speech as a potential source of compensatory cues toward emotional identification. Importantly, results were not explained by sampling differences between Experiments 2 and 1, as a matched subset of data from Experiment 1 replicated the negative correlation in the low-pass filtered condition that was not observed in the high-pass filtered condition in Experiment 2.

General Discussion

Results showed a significant negative correlation between pitch discrimination thresholds and emotional identification for low-pass filtered speech, but not high-pass filtered or unfiltered speech. Subjects with poor pitch perception, especially amusics, performed worse than their counterparts in identifying emotions from speech, but only when the speech was low-pass filtered. Amusics were defined here as those with a pitch discrimination threshold of >16 Hz, resulting in nine identified amusics in Experiment 1 and three subjects identified as amusics in Experiment 2. The behavioral dissociation between low-pass and unfiltered speech conditions suggests that low frequency energy bands in speech carry important emotional content, to which amusics are less sensitive.

In the low-pass filtered condition, the observed correlation between emotional identification accuracy and individual differences in pitch discrimination threshold was significant even after controlling for IQ and musical training. This finding suggests that individual differences in pitch perception can exist above and beyond differences in cognitive capacity and musical training, and can have far-reaching consequences that generalize to domains of life beyond musical ability. However, unlike previous reports (Thompson et al., 2012), we did not observe a significant relationship between emotional identification accuracy and pitch discrimination threshold in unfiltered speech. While further work is needed to explain the differences in experiment design that might give rise to

our different findings, the observed dissociation from the current study between low-pass filtered and unfiltered speech conditions supports the hypothesis that amusics could have been compensating for their poorer pitch perception in low frequency sounds by using other cues in the speech stimuli. However, the high-pass filtering manipulation (Experiment 2) did not reveal more reliance on high frequency speech cues among poorer pitch perceivers. This may suggest that frequencies above 4800 Hz (the chosen cutoff for high-pass filtering in Experiment 2) were also not the primary source of the compensatory information in speech that amusics might be using to approach the task of emotional identification. Alternately, both groups were using other cues in speech, not captured in the filters used in these studies, to accomplish the task of emotional identification.

Pitch discrimination thresholds were used to define amusia in these experiments rather than the MBEA, as the latter focuses more on melodic discrimination than on individual differences in pitch discrimination *per se*. While amusic participants performed worse in low-pass trials, accuracy for all participants was well above the chance level of 16%. This finding implies that while the fundamental frequency (below 500 Hz) provides some prosodic information such as pitch contour, cues that exist in the range of frequencies between 500 and 4800 Hz may provide further prosodic cues. These midrange frequencies may have been used for emotion recognition in music, in light of recent findings that amusics are able to show normal recognition of musical emotions (Gosselin et al., 2015). Results are also consistent with recent reports showing that amusia is limited to resolved harmonics (Cousineau et al., 2015). Given these results, examining specific frequency bands for prosodic cues may reveal more in the future about the cues that amusics could be using to identify emotions, and to understand speech and music in communication more generally.

Insight into several additional questions may lead to a more complete model explaining this relationship between pitch discrimination and emotional identification. It remains to be determined if there is a causal link between poor pitch perception and poor emotional recognition, or if a third underlying process leads to both deficiencies, as posited by the musical protolanguage hypothesis (Thompson et al., 2012). Poor pitch perception is associated with multiple behavioral and neural differences, such as differences in neural connectivity (Loui et al., 2009), pitch awareness (Loui et al., 2008; Peretz et al., 2009), learning ability (Loui and Schlaug, 2012), and working memory (Williamson and Stewart, 2010), and different contributions of one factor or another may further affect prosodic recognition.

In that regard, one factor that may affect prosodic recognition is learning differences, which was addressed in a follow-up analysis looking at order effects. This showed a significant interaction between amusia and block order: amusics who started the experiment by listening to low-pass filtered speech performed worse than other amusics who started on unfiltered speech. This interaction suggests that learning throughout the experiment may occur even more in amusics than in controls. While more studies are needed to address this possibility in the future, learning could

potentially be one of the compensatory mechanisms that amusics use to approach the task of emotional identification when pitch perception is impaired.

Given that a significant correlation between pitch discrimination ability and emotional recognition accuracy was found only when high frequency bands were removed, the data suggest that higher frequency information must have played a role in accurate recognition. Further studies may benefit from examining whether these trends are present among all amusics, or whether in-group distinctions can be made between different amusic individuals. Amusia may be a complex class of disorders with subtle disabilities that are currently categorized under a single category. Related symptoms of amusia, such as rhythmic disabilities, poor singing ability, and deficiencies in musical memory, may be examined to determine if these types of disabilities also correlate with deficiencies in recognition of emotional prosody. By investigating emotional identification in speech by individuals with various musical difficulties, future results may contribute further to the debate on the origins of music and language.

References

- Ayotte, J., Peretz, I., and Hyde, K. (2002). Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain* 125, 238–251. doi: 10.1093/brain/awf028
- Bhargava, P., and Başkent, D. (2012). Effects of low-pass filtering on intelligibility of periodically interrupted speech. *J. Acoust. Soc. Am.* 131, EL87–EL92. doi: 10.1121/1.3670000
- Cousineau, M., Oxenham, A. J., and Peretz, I. (2015). Congenital amusia: a cognitive disorder limited to resolved harmonics and with no peripheral basis. *Neuropsychologia* 66, 293–301. doi: 10.1016/j.neuropsychologia.2014.11.031
- Dowling, W. J., and Harwood, D. (1986). *Music Cognition*. New York: Academic Press.
- Gosselin, N., Paquette, S., and Peretz, I. (2015). Sensitivity to musical emotions in congenital amusia. *Cortex* 71, 171–182. doi: 10.1016/j.cortex.2015.06.022
- Guellai, B., Langus, A., and Nespor, M. (2014). Prosody in the hands of the speaker. *Front. Psychol.* 5:700. doi: 10.3389/fpsyg.2014.00700
- Horwitz, A. R., Dubno, J. R., and Ahlstrom, J. B. (2002). Recognition of low-pass-filtered consonants in noise with normal and impaired high-frequency hearing. *J. Acoust. Soc. Am.* 111, 409–416. doi: 10.1121/1.1427357
- Hutchins, S., Gosselin, N., and Peretz, I. (2010). Identification of changes along a continuum of speech intonation is impaired in congenital amusia. *Front. Psychol.* 1:236. doi: 10.3389/fpsyg.2010.00236
- Hyde, K. L., and Peretz, I. (2004). Brains that are out of tune but in time. *Psychol. Sci.* 15, 356–360. doi: 10.1111/j.0956-7976.2004.00683.x
- Jiang, C., Hamm, J. P., Kim, V., Kirk, I., Chen, X., and Yang, Y. (2012). Amusia results in abnormal brain activity following inappropriate intonation during speech comprehension. *PLoS ONE* 7:e41411. doi: 10.1371/journal.pone.0041411
- Knoll, M. A., Uther, M., and Costall, A. (2009). Effects of low-pass filtering on the judgment of vocal affect in speech directed to infants, adults and foreigners. *Speech Commun.* 51, 210–216. doi: 10.1016/j.specom.2008.08.001
- Liu, F., Jiang, C., Thompson, W., Xu, Y., Yang, Y., and Stewart, L. (2012). The mechanism of speech processing in congenital amusia: evidence from Mandarin speakers. *PLoS ONE* 7:e30374. doi: 10.1371/journal.pone.0030374
- Liu, F., Jiang, C., Wang, B., Xu, Y., and Patel, A. D. (2015). A music perception disorder (congenital amusia) influences speech comprehension. *Neuropsychologia* 66, 111–118. doi: 10.1016/j.neuropsychologia.2014.11.001
- Liu, F., Patel, A. D., Fourcin, A., and Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain* 133, 1682–1693. doi: 10.1093/brain/awq089
- Loui, P., Alsop, D., and Schlaug, G. (2009). Tone deafness: a new disconnection syndrome? *J. Neurosci.* 29, 10215–10220. doi: 10.1523/JNEUROSCI.1701-09.2009
- Loui, P., Guenther, F. H., Mathys, C., and Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Curr. Biol.* 18, R331–R332. doi: 10.1016/j.cub.2008.02.045
- Loui, P., and Schlaug, G. (2012). Impaired learning of event frequencies in tone deafness. *Ann. N. Y. Acad. Sci.* 1252, 354–360. doi: 10.1111/j.1749-6632.2011.06401.x
- Nguyen, S., Tillmann, B., Gosselin, N., and Peretz, I. (2009). Tonal language processing in congenital amusia. *Neurosci. Music* 1169, 490–493. doi: 10.1111/j.1749-6632.2009.04855.x
- O’Beirne, G. A., McGaffin, A. J., and Rickard, N. A. (2012). Development of an adaptive low-pass filtered speech test for the identification of auditory processing disorders. *Int. J. Pediatr. Otorhinolaryngol.* 76, 777–782. doi: 10.1016/j.ijporl.2012.02.039
- Patel, A. D., Peretz, I., Tramo, M., and Labreque, R. (1998). Processing prosodic and musical patterns: a neuropsychological investigation. *Brain Lang.* 61, 123–144. doi: 10.1006/brln.1997.1862
- Paulson, M. J., and Lin, T. (1970). Predicting WAIS IQ from Shipley-Hartford scores. *J. Clin. Psychol.* 26, 453–461.
- Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., et al. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron* 33, 185–191. doi: 10.1016/S0896-6273(01)00580-3
- Peretz, I., Brattico, E., Jarvenpää, M., and Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain* 132, 1277–1286. doi: 10.1093/brain/awp055
- Peretz, I., Brattico, E., and Tervaniemi, M. (2005). Abnormal electrical brain responses to pitch in congenital amusia. *Ann. Neurol.* 58, 478–482. doi: 10.1002/ana.20606
- Peretz, I., Champod, S., and Hyde, K. (2003). Varieties of musical disorders: the Montreal Battery of Evaluation of Amusia. *Ann. N. Y. Acad. Sci.* 999, 58–75. doi: 10.1196/annals.1284.006
- Peretz, I., Gagnon, L., and Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. *Cognition* 68, 111–141. doi: 10.1016/S0010-0277(98)00043-2
- Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.* 9, 371–377. doi: 10.1080/00223980.1940.9917704

Conclusion

The present study investigated the relationship between pitch perception and emotional identification in speech. Using a battery of speech that was spoken with different emotional prosody, we showed that poor pitch perception is correlated with lower accuracy in emotional identification tasks, but only for low-pass filtered speech, and not for high-pass filtered or unfiltered speech. The relationship between pitch discrimination and emotional identification accuracy is not explained by differences in IQ and musical training. Future research should be focused toward identifying which speech cues are used by amusics in order to compensate for impaired pitch perception.

Acknowledgments

Supported by startup funds from Wesleyan University and grants from the Grammy Foundation and the Imagination Institute to PL. We thank W. Thompson and two reviewers for helpful comments at different stages of this project.

- Thompson, W. F., Marin, M. M., and Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19027–19032. doi: 10.1073/pnas.1210344109
- Vuvan, D. T., Nunes-Silva, M., and Peretz, I. (2015). Meta-analytic evidence for the non-modularity of pitch processing in congenital amusia. *Cortex* 69, 186–200. doi: 10.1016/j.cortex.2015.05.002
- Williamson, V. J., and Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory* 18, 657–669.

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 © 2015 Lolli, Lewenstein, Basurto, Winnik and Loui. 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.



Cross-domain processing of musical and vocal emotions in cochlear implant users

Alexandre Lehmann^{1,2,3*} and Sébastien Paquette^{2,3}

¹ Department of Otolaryngology Head and Neck Surgery, McGill University, Montreal, QC, Canada, ² International Laboratory for Brain, Music and Sound Research, Center for Research on Brain, Language and Music, Montreal, QC, Canada,

³ Department of Psychology, University of Montreal, Montreal, QC, Canada

Keywords: cross-domain processing, emotion, music, voice, cochlear implant, brain plasticity, neural overlap

Music and voice bear many similarities and share neural resources to some extent. Experience dependent plasticity provides a window into the neural overlap between these two domains. Here, we suggest that research on auditory deprived individuals whose hearing has been bionically restored offers a unique insight into the functional and structural overlap between music and voice. Studying how basic emotions (happiness, sadness, and fear) are perceived in auditory stimuli constitutes a favorable terrain for such an endeavor. We outline a possible neuro-behavioral approach to study the effect of plasticity on cross-domain processing of musical and vocal emotions, using cochlear implant users as a model of reversible sensory deprivation and comparing them to normal-hearing individuals. We discuss the implications of such developments on the current understanding of cross-domain neural overlap.

OPEN ACCESS

Edited by:

McNeel Gordon Jantzen,
Western Washington University, USA

Reviewed by:

Takako Fujioka,
Stanford University, USA

*Correspondence:

Alexandre Lehmann,
alexandre.lehmann@mcgill.ca

Specialty section:

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

Received: 01 July 2015

Accepted: 10 September 2015

Published: 24 September 2015

Citation:

Lehmann A and Paquette S (2015)
Cross-domain processing of musical
and vocal emotions in cochlear
implant users. *Front. Neurosci.* 9:343.
doi: 10.3389/fnins.2015.00343

Cross-domain Neural Overlap and Plasticity

Our musical and vocal perception abilities have such a close relationship that some authors suggested that the former originated from the latter or vice-versa (Honing et al., 2015; Peretz et al., 2015). To what extent do music and voice share functional and structural networks and at which stage of auditory processing they are differentiated are open questions. Functional magnetic resonance imaging (fMRI) studies show the co-activation of brain regions with possibly distinct underlying neural populations (Peretz et al., 2015). Research on expert populations has suggested reciprocal interactions between neural circuits associated with the domains of music and voice (Patel, 2011; White-Schwoch et al., 2013; summarized by Paquette and Mignault Goulet, 2014). Indeed studies have shown that musicians have enhanced speech processing capacity, which is reflected in both cortical and subcortical neural measures (Bidelman et al., 2011, 2014; Parbery-Clark et al., 2012). Musicians can be used as a model of learning-induced plasticity to investigate how such cross-domain transfer effects unfold over time (Strait and Kraus, 2014; Strait et al., 2014). Here we argue that, sensory deprivation offers a complementary model to shed light on the plastic reorganization of brain networks involved in particular functions.

Temporary Deafened Individuals offer a Unique Insight into Auditory Neural Plasticity

Cochlear implants (CI) are bionic devices that can restore the sense of hearing in profoundly deaf individuals. We argue that cochlear implant users offer a promising model to study the mechanisms of cross-domain plasticity because they undergo different trajectories

of auditory development: deafness of various origins results in a variable period of auditory deprivation followed by surgical restoration of auditory input and an intense rehabilitation period, yielding variable individual auditory outcomes.

Signal transmitted from the implant to the auditory nerve is impoverished compared to natural hearing. Critically, the access to pitch cues is impaired, reduced to a small number of frequency bands. As a result, cochlear implant users can potentially perceive speech relatively well in a quiet setting, but understanding it in noise, or accurately perceiving music is very challenging since both tasks rely on pitch information (Gfeller et al., 2007). Perception is not only affected by the impoverished auditory input, but also by neural re-organization following auditory deprivation, from the periphery to the cortex. In absence of auditory input, auditory nerve fibers start to degenerate and the auditory cortex can be recruited by visual and somatosensory systems (Collignon et al., 2011; Lazzouni and Lepore, 2014). Such plastic changes can prevent the auditory cortex from fully recovering its initial function after the auditory input is restored via an implant (Lee et al., 2001; Bavelier and Hirshorn, 2010; Sandmann et al., 2012; Sharma et al., 2015).

To date, little is known of the neural correlates of music and voice processing in cochlear implants and the extent to which those processes overlap. Only one study has performed a direct comparison of the neural correlates of speech and music perception in CI users. Using positron emission tomography (PET), Limb et al. (2009), reported increased activation and greater cortical recruitment in implant recipients compared to normal hearing controls, during both speech and music listening. This effect was stronger for speech—for which CI users are more proficient than music—and suggest a link between auditory performance and degree of auditory cortical activation.

Emotion as a Cross-domain Terrain of Choice to Study Neural Overlap

An important part of our social interaction relies on accurate emotion perception. In normal-hearing individuals, evidence from neuropsychology suggest the existence of an auditory emotional neural pathway, distinct from auditory perception, that might be shared across musical and vocal domains and have both cortical and subcortical components (Peretz, 2011). A systematic comparison of the vocal and musical domains suggests a close acoustical relationship for emotional expression, with similar emotion-specific acoustic cues patterns (Juslin and Laukka, 2003). Several of those patterns relate to the pitch dimension, such as prosody for voice (variations in the pitch contour) and melody for music. The perception of pitch is severely degraded in cochlear implant users, thus limiting their access to those important cues, but other non-pitch based cues can also convey emotions (Gabrielsson and Lindström, 2010). It was recently demonstrated in amusics (individuals with a lifelong pitch perception deficit; Peretz, 2013) that non-pitch based cues (e.g., tempo, pulse clarity) can be used to identify musical emotions (Gosselin et al., 2015). These cues are available to some extent to CI users (Kong et al., 2004;

Looi et al., 2012), and should allow them a certain degree of emotional perception. CI users have a documented deficit in both vocal and musical emotion recognition; emotional categories and dimensions are not uniformly impaired. They can recognize some categories of emotion in voice or music above chance, but not as well as normal hearing controls (Hopman et al., 2012; Nakata et al., 2012; Volkova et al., 2013; Wang et al., 2013). They have difficulty perceiving arousal of musical excerpts but not valence (Ambert-Dahan et al., 2015). These differences could be due to the relatively spared abilities of CI users to perceive temporal variations, while having an impaired pitch perception. They could also reflect differences in the complexity of stimuli employed and how they are handled by speech-optimized processors, suggesting that *ad-hoc* stimuli are required to accurately compare the two domains. This could explain why no study has yet directly compared emotion processing in CI users across the domains of music and voice. To date, there is very little neuro-imaging evidence building up on the aforementioned behavioral findings. Only one study evaluated the impact of two implant processing strategies on the perception of prosody (Agrawal et al., 2013) and demonstrated that electroencephalography (EEG) is a useful tool to reveal differences between strategies coding specific features.

Toward a Study of Cross-domain Processing of Musical and Vocal Emotions in Cochlear Implant Users

A large part of the research on auditory affective processing has been conducted on prosody utilizing words or sentences spoken with various emotional expressions and complex musical pieces expressing varying degrees of emotion. It is not possible to directly compare those results between music and voice because of many confounding variables; factors such as speech semantics, length, harmony, and context are likely to recruit different neural networks. We argue that a necessary first step to study cross-domain processing of musical and vocal emotions is to use an experimental paradigm that moves away from the fairly complex sounds used in the existing literature, using stimuli that enable a controlled comparison between the domains of music and voice. A possible approach would be to use the most primitive affect expressions (primal interjections close to those of babies and animals) in each domain: non-speech vocalizations and brief mono-instrumental musical excerpts.

In the vocal domain, non-speech vocalizations (e.g., screams, laughter) depicting basic emotions that are minimally conventionalized, relatively universal and fundamental to spontaneous human communication (Scherer, 1986), could be used. Stimuli like the Montreal Affective Voices (Belin et al., 2008), consisting of short vocal interjections on the vowel /a/ expressing basic emotions, represent the most primitive form of emotion in their domain. They have minimal semantic information and minimal interaction with linguistic processes (Bestelmeyer et al., 2010). Compared to speech prosody, vocalizations are treated preferentially in the brain (Pell et al., 2015). When it comes to music, finding the most basic emotions

and avoiding interaction with other processes require stepping away from conventional structure (limited by mode or tempo), reducing the length of the stimuli and reducing its emotional complexity. Stimuli like the Musical Emotional Bursts (Paquette et al., 2013) could be used for comparison, they consist of a few spontaneous notes on a clarinet or violin expressing basic musical emotions, they are minimally conventionalized and represent the most primitive form of emotion in their domain. They are all the more similar to vocal stimuli because they use continuous pitch instrument (e.g., the violin which offers a seamless progression between notes, giving the stimuli a quasi-vocal quality), whereas most studies have used discrete pitch instruments (e.g., the piano where one key corresponds to one pitch), which further hinders the direct comparison with vocal stimuli.

These highly similar vocal and musical stimuli seem well-suited to study cross-domain overlap in any population and their primitive quality could be extremely useful to study plasticity in CI users.

A second step would be to pair a well-controlled behavioral paradigm using those stimuli (allowing a direct comparison of musical and vocal domains) with a neuro-imaging modality that is acceptable for use with cochlear implants. Except for a few recent exceptions, implants are not MR-compatible. High-density EEG (Gilley et al., 2010; Zhang et al., 2011; Timm et al., 2014) and PET-scan (Okazawa et al., 1996; Limb et al., 2009; Lazard et al., 2010) have both been used successfully in cochlear implant users. Both methods have drawbacks; EEG recordings are contaminated by massive electrical artifacts from the implant and PET requires the injection of a radioactive isotope. Emerging as a promising brain-imaging modality for CI research is functional near-infrared spectroscopy (fNIRS). fNIRS has been successfully used to study the response to auditory stimuli in cochlear implant users (Sevy et al., 2010) and emotion-related activation in the general population (Herrmann et al., 2003; Plichta et al., 2011). This non-invasive technique measures blood oxygenation level differences using infrared light and is therefore unaffected by electrical artifacts. It is portable and has a better temporal resolution than functional MRI (Villringer and Chance, 1997). Conversely it has a worse spatial resolution and cannot access subcortical sources such as the limbic system (Köchel et al., 2011).

The proposed neuro-behavioral approach would be well-suited to study the effect of plasticity on cross-domain processing of musical and vocal emotions, using cochlear implant users as a model of reversible sensory deprivation and comparing them to normal-hearing individuals. The effect of multiple regressors

could be assessed by recruiting a heterogeneous cohort of individuals spanning the continuum of factors known to affect plasticity such as the duration of auditory deprivation or the age at implantation (Lazard et al., 2012).

This would represent a stepping-stone to ask further questions of interest regarding the effect of plasticity on cross-domain neural overlap. From a basic science perspective, the rationale is to understand a complex system by reverse-engineering its dysfunctions. What are the structural and functional overlaps between music and voice processing after implantation? Would the reduction of auditory cortical resources, together with the fact that music and vocal signals are more similar after being processed by the device, favor an increased neural overlap between domains? Conversely, would any remaining overlap break-down in favor of a more segregated re-organization guided by the non-pitch based, domain relevant cues?

Characterizing those mechanisms can inform novel clinical approaches, possibly through individualized rehabilitation and brain stimulation. For instance, if good performers (CI users with good speech scores) make use of overlapping structures in an optimal fashion compared to poor performers, can we boost residual neural processes in the latter group? It has been suggested that musical training can improve speech outcomes in this population (Patel, 2014), but what stages of the auditory pathway are best candidates for a cross-domain shaping of function and/or structure? Auditory features found to maximize activity of brain networks processing musical and vocal emotions in CI users could be made more salient in device processors.

Cross-domain research on cochlear implant users not only offers a unique insight into auditory neural plasticity, but also has practical implications for patients' rehabilitation, implant design, and programming. We believe that highly comparable stimuli are needed to carry out such studies, together with an optimal imaging technique within a paradigm fine enough to reveal subtle behavioral and neural differences. Such scientific undertaking can further our understanding of how our brain processes vocal and musical emotions and how such cross-domain processing is affected by plasticity. Furthermore, such studies could provide objective measures to support the use of music in the rehabilitation of various disorders.

Acknowledgments

This work was supported by a CRBLM research incubator award (funded by the Fonds de Recherche Nature et Technologies and Société et Culture) to AL and by a graduate scholarship from the Canadian Institutes of Health Research to SP.

References

- Agrawal, D., Thorne, J. D., Viola, F. C., Timm, L., Debener, S., Buchner, A., et al. (2013). Electrophysiological responses to emotional prosody perception in cochlear implant users. *NeuroImage Clin.* 2, 229–238. doi: 10.1016/j.nicl.2013.01.001
- Ambert-Dahan, E., Giraud, A. L., Sterkers, O., and Samson, S. (2015). Judgment of musical emotions after cochlear implantation in adults with progressive deafness. *Neuro Front. Psychol.* 6:181. doi: 10.3389/fpsyg.2015.00181
- Bavelier, D., and Hirshorn, E. A. (2010). I see where you're hearing: how cross-modal plasticity may exploit homologous brain structures. *Nat. Neurosci.* 13:1309. doi: 10.1038/nn1110-1309

- 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
- Bestelmeyer, P. E., Rouger, J., DeBruine, L. M., and Belin, P. (2010). Auditory adaptation in vocal affect perception. *Cognition* 117, 217–223. doi: 10.1016/j.cognition.2010.08.008
- Bidelman, G. M., Gandour, J. T., and Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* 23, 425–434. doi: 10.1162/jocn.2009.21362
- Bidelman, G. M., Weiss, M. W., Moreno, S., and Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* 40, 2662–2673. doi: 10.1111/ejn.12627
- Collignon, O., Vandewalle, G., Voss, P., Albouy, G., Charbonneau, G., Lassonde, M., et al. (2011). Functional specialization for auditory-spatial processing in the occipital cortex of congenitally blind humans. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4435–4440. doi: 10.1073/pnas.1013928108
- Gabrielsson, A., and Lindström, E. (2010). The role of structure in the musical expression of emotions. *Handb. Music Emot. Theory Res. Appl.* 367–400. doi: 10.1093/acprof:oso/9780199230143.003.0014
- Gfeller, K., Turner, C., Oleson, J., Zhang, X., Gantz, B., Froman, R., et al. (2007). Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear Hear.* 28, 412–423. doi: 10.1097/AUD.0b013e3180479318
- Gilley, P. M., Sharma, A., Mitchell, T. V., and Dorman, M. F. (2010). The influence of a sensitive period for auditory-visual integration in children with cochlear implants. *Restor. Neurol. Neurosci.* 28, 207–218. doi: 10.3233/RNN-2010-0525
- Gosselin, N., Paquette, S., and Peretz, I. (2015). Sensitivity to musical emotions in congenital amusia. *Cortex* 71, 171–182. doi: 10.1016/j.cortex.2015.06.022
- Herrmann, M. J., Ehlis, A.-C., and Fallgatter, A. J. (2003). Prefrontal activation through task requirements of emotional induction measured with NIRS. *Biol. Psychol.* 64, 255–263. doi: 10.1016/S0301-0511(03)00095-4
- Honing, H., Cate, C., ten, 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
- Hopyan, T., Peretz, I., Chan, L. P., Papsin, B. C., and Gordon, K. A. (2012). Children using cochlear implants capitalize on acoustical hearing for music perception. *Front. Audit. Cogn. Neurosci.* 3:425. doi: 10.3389/fpsyg.2012.00425
- 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
- Köchel, A., Plichta, M. M., Schäfer, A., Leutgeb, V., Scharmüller, W., Fallgatter, A. J., et al. (2011). Affective perception and imagery: a NIRS study. *Int. J. Psychophysiol.* 80, 192–197. doi: 10.1016/j.ijpsycho.2011.03.006
- Kong, Y.-Y., Cruz, R., Jones, J. A., and Zeng, F.-G. (2004). Music perception with temporal cues in acoustic and electric hearing. *Ear Hear.* 25, 173–185. doi: 10.1097/01.AUD.0000120365.97792.2F
- Lazard, D. S., Lee, H. J., Gaebler, M., Kell, C. A., Truy, E., and Giraud, A. L. (2010). Phonological processing in post-lingual deafness and cochlear implant outcome. *Neuroimage* 49, 3443–3451. doi: 10.1016/j.neuroimage.2009.11.013
- Lazard, D. S., Vincent, C., Venail, F., Van de Heyning, P., Truy, E., Sterkers, O., et al. (2012). Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: a new conceptual model over time. *PLoS ONE* 7:e48739. doi: 10.1371/journal.pone.0048739
- Lazzouni, L., and Lepore, F. (2014). Compensatory plasticity: time matters. *Front. Hum. Neurosci.* 8:340. doi: 10.3389/fnhum.2014.00340
- Lee, D. S., Lee, J. S., Oh, S. H., Kim, S.-K., Kim, J.-W., Chung, J.-K., et al. (2001). Deafness: Cross-modal plasticity and cochlear implants. *Nature* 409, 149–150. doi: 10.1038/35051653
- Lim, C. J., Molloy, A. T., Jiradejvong, P., and Braun, A. R. (2009). Auditory cortical activity during cochlear implant-mediated perception of spoken language, melody, and rhythm. *J. Assoc. Res. Otolaryngol.* 11, 133–143. doi: 10.1007/s10162-009-0184-9
- 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
- Nakata, T., Trehub, S. E., and Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. *J. Acoust. Soc. Am.* 131, 1307. doi: 10.1121/1.3672697
- Okazawa, H., Naito, Y., Yonckura, Y., Sadato, N., Hirano, S., Nishizawa, S., et al. (1996). Cochlear implant efficiency in pre- and postlingually deaf subjects A study with H215O and PET. *Brain* 119, 1297–1306. doi: 10.1093/brain/119.4.1297
- Paquette, S., and Mignault Goulet, G. (2014). Lifetime benefits of musical training. *Front. Neurosci.* 8:89. doi: 10.3389/fnins.2014.00089
- Paquette, S., Peretz, I., and Belin, P. (2013). The “Musical Emotional Bursts”: a validated set of musical affect bursts to investigate auditory affective processing. *Emot. Sci.* 4:509. doi: 10.3389/fpsyg.2013.00509
- Parbery-Clark, A., Tierney, A., Strait, D. L., and Kraus, N. (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience* 219, 111–119. doi: 10.1016/j.neuroscience.2012.05.042
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Audit. Cogn. Neurosci.* 2:142. doi: 10.3389/fpsyg.2011.00142
- Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear. Res.* 308, 98–108. doi: 10.1016/j.heares.2013.08.011
- Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., and Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biol. Psychol.* 111, 14–25. doi: 10.1016/j.biopsycho.2015.08.008
- 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, 20140090–20140090. doi: 10.1098/rstb.2014.0090
- Peretz, I. (2011). “Towards a neurobiology of musical emotions,” in *Handbook of Music and Emotion: Theory, Research, Applications*, eds P. N. Juslin and J. A. Sloboda (Oxford: Oxford University Press), 99–126.
- Peretz, I. (2013). “The biological foundations of music: insights from congenital Amusia,” in *The Psychology of Music 3rd Edn.* ed D. Deutsch (Academic Press), 551–564.
- Plichta, M. M., Gerdes, A. B. M., Alpers, G. W., Harnisch, W., Brill, S., Wieser, M. J., et al. (2011). Auditory cortex activation is modulated by emotion: a functional near-infrared spectroscopy (fNIRS) study. *Neuroimage* 55, 1200–1207. doi: 10.1016/j.neuroimage.2011.01.011
- Sandmann, P., Dillier, N., Eichele, T., Meyer, M., Kegel, A., Pascual-Marqui, R. D., et al. (2012). Visual activation of auditory cortex reflects maladaptive plasticity in cochlear implant users. *Brain J. Neurol.* 135, 555–568. doi: 10.1093/brain/awr329
- Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychol. Bull.* 99, 143–165. doi: 10.1037/0033-2909.99.2.143
- Sevy, A. B. G., Bortfeld, H., Huppert, T. J., Beauchamp, M. S., Tonini, R. E., and Oghalai, J. S. (2010). Neuroimaging with near-infrared spectroscopy demonstrates speech-evoked activity in the auditory cortex of deaf children following cochlear implantation. *Hear. Res.* 270, 39–47. doi: 10.1016/j.heares.2010.09.010
- Sharma, A., Campbell, J., and Cardon, G. (2015). Developmental and cross-modal plasticity in deafness: evidence from the P1 and N1 event related potentials in cochlear implanted children. *Int. J. Psychophysiol.* 95, 135–144. doi: 10.1016/j.ijpsycho.2014.04.007
- Strait, D. L., and Kraus, N. (2014). Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hear. Res.* 308, 109–121. doi: 10.1016/j.heares.2013.08.004
- Strait, D. L., O'Connell, S., Parbery-Clark, A., and Kraus, N. (2014). Musicians' enhanced neural differentiation of speech sounds arises early in life: developmental evidence from ages 3 to 30. *Cereb. Cortex* 24, 2512–2521. doi: 10.1093/cercor/bht103
- Timm, L., Vuust, P., Brattico, E., Agrawal, D., Debener, S., Büchner, A., et al. (2014). Residual neural processing of musical sound features in adult cochlear implant users. *Front. Hum. Neurosci.* 8:181. doi: 10.3389/fnhum.2014.00181
- Villringer, A., and Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends Neurosci.* 20, 435–442. doi: 10.1016/S0166-2236(97)01132-6

- Volkova, A., Trehub, S. E., Schellenberg, E. G., Papsin, B. C., and Gordon, K. A. (2013). Children with bilateral cochlear implants identify emotion in speech and music. *Cochlear Implants Int.* 14, 80–91. doi: 10.1179/1754762812Y.0000000004
- Wang, D. J., Trehub, S. E., Volkova, A., and van Lieshout, P. (2013). Child implant users' imitation of happy- and sad-sounding speech. *Front. Psychol.* 4:351. doi: 10.3389/fpsyg.2013.00351
- White-Schwoch, T., Carr, K. W., Anderson, S., Strait, D. L., and Kraus, N. (2013). Older adults benefit from music training early in life: biological evidence for long-term training-driven plasticity. *J. Neurosci.* 33, 17667–17674. doi: 10.1523/JNEUROSCI.2560-13.2013
- Zhang, F., Hammer, T., Banks, H.-L., Benson, C., Xiang, J., and Fu, Q.-J. (2011). Mismatch negativity and adaptation measures of the late auditory

evoked potential in cochlear implant users. *Hear. Res.* 275, 17–29. doi: 10.1016/j.heares.2010.11.007

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 © 2015 Lehmann and Paquette. 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.



Music and literature: are there shared empathy and predictive mechanisms underlying their affective impact?

Diana Omigie*

Music Department, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany

OPEN ACCESS

Edited by:

McNeel G. Jantzen,
Western Washington University, USA

Reviewed by:

Bradley W. Vines,
Nielsen, USA
Stefan Koelsch,
Freie Universität Berlin, Germany

*Correspondence:

Diana Omigie,
Music Department,
Max Planck Institute for Empirical
Aesthetics, Grüneburgweg 14,
60322 Frankfurt am Main, Germany
diana.omigie@aesthetics.mpg.de

Specialty section:

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

Received: 09 March 2015

Accepted: 05 August 2015

Published: 24 August 2015

Citation:

Omigie D (2015) Music and literature:
are there shared empathy and
predictive mechanisms underlying
their affective impact?
Front. Psychol. 6:1250.
doi: 10.3389/fpsyg.2015.01250

It has been suggested that music and language had a shared evolutionary precursor before becoming mainly responsible for the communication of emotive and referential meaning respectively. However, emphasis on potential differences between music and language may discourage a consideration of the commonalities that music and literature share. Indeed, one possibility is that common mechanisms underlie their affective impact, and the current paper carefully reviews relevant neuroscientific findings to examine such a prospect. First and foremost, it will be demonstrated that considerable evidence of a common role of empathy and predictive processes now exists for the two domains. However, it will also be noted that an important open question remains: namely, whether the mechanisms underlying the subjective experience of uncertainty differ between the two domains with respect to recruitment of phylogenetically ancient emotion areas. It will be concluded that a comparative approach may not only help to reveal general mechanisms underlying our responses to music and literature, but may also help us better understand any idiosyncrasies in their capacity for affective impact.

Keywords: music, literature, emotions, esthetic, empathy, theory of mind, tension, active inference

Introduction

The creation, performance, and consumption of music and literary works are preoccupations present in all cultures. Music and literature (in its original form of storytelling) have ancient origins and do not only lie at the heart of religious and cultural practices and narratives, but also provide widely popular leisure activities in everyday life. A number of psychological theories, some building on ideas dating as far back as classical antiquity (Aristotle, 1961; Longinus, 1965), have sought to account for music and literature's affective capacity independently (Meyer, 1956; Goldman, 2006; Huron, 2006; Zunshine, 2006; Keen, 2007). Others have examined the nature of emotion in response to the arts more generally (Hjort and Lavers, 1997; Robinson, 2007). Most accounts, however, maintain that in spite of any differences in propositional content (Slevc and Patel, 2011), and notwithstanding any prevailing notions regarding evolutionary capacity for emotion (Brown, 2000), both music and literary works share a considerable ability to evoke powerful feelings. Further, evidence that music may evoke semantic representations (Koelsch et al., 2004; Steinbeis and Koelsch, 2008) would seem to temper any strong claims that the two are incomparable in terms of their ability to convey meaning.

Perhaps one of the most important qualities that binds music and literary reading, and differentiates them from a number of other cultural artifacts (such as paintings and sculpture), is that both unfold in time, offering a kind of "narrative" that can be followed (Rabkin, 1973; Maus, 1991; Levinson, 2004). While poetry, like music, may exercise affective impact through

its emphasis on temporal stress and repetition, many different accounts emphasize the impact that even non-versed literary forms (e.g., short stories and novels) can have (Oatley, 1994; Hogan, 2010). Thus, given the timeless and universal appeal of music and literary storytelling, not to mention the claims from a range of theoretical accounts, it seems relevant to explore the emerging neuroscience research for any evidence of an overlap in underlying affective mechanisms. Recent research into music and literature has investigated a breadth of issues varying in their degree of domain specificity (e.g., Bohrn et al., 2012), and a comprehensive review of all possible links that can be made between the two art forms would require a longer format. Accordingly, the current perspective article focuses solely on what are considered key lines of investigation that have seen significant interest in both domains: namely, music and literature's invocation of empathy and predictive processes and the potential role these mechanisms may play in emotion induction.

Inferring and Sharing Emotions

The notion of empathy-like mechanisms being involved in literary response is one that dates as far back as Aristotle's *Poetics* (Aristotle, 1961). Similarly, as early as in the eighteenth century, it was suggested that engagement of children with music is especially valuable in teaching emotions and a good social attitude. Today, in neuroscience and psychology, empathy may be broadly defined as the ability to infer and share emotional experiences (Gallese, 2003). It is held to comprise two different components: a cognitive and an emotional one that encompass the notion of perspective-taking and shared visceral feeling respectively (Shamay-Tsoory et al., 2009). Importantly, while the former is related to the notion of Theory of Mind (TOM) and mentalizing (Frith, 1999), the latter is seen as coinciding with the notion of emotional contagion (Juslin and Västfjäll, 2008; Juslin, 2013), with evidence of a double dissociation observable in the neuroscience literature (Eslinger, 1998; Shamay-Tsoory et al., 2003; Schulte-Ruther et al., 2007).

Fiction, the form of literature seen in short stories and novels, has been described as a kind of simulation of the social world (Mar and Oatley, 2008) and it has been suggested that its invocation of social situations not only explains readers' tendency to mentalize during reading (Gygax et al., 2003) but also to feel emotions themselves (Cupchik et al., 1998; Miall and Kuiken, 2002; Oatley, 2002). Over the years, a large body of neuroimaging studies has focused on the neural correlates of text comprehension (see Mar, 2011, for a review), emotion processing in single words (Citron, 2012) and perspective-taking (van Overwalle, 2009). However, the extent to which cognitive or emotional empathy could be directly linked to the affective impact of literature remained limited. In recent years, however, it is becoming apparent that as text stimuli are rendered increasingly story-like or as feelings and emotions play a larger role in them (in other words, as text stimuli begin to resemble fiction or narrative literature), increasingly recruited are not only those areas involved in TOM processing [e.g., ventromedial prefrontal cortex (vmPFC) and temporoparietal junction (TPJ)] but also limbic or emotion areas like the amygdala, thalamus, and orbitofrontal cortex (OFC; Wallentin et al., 2011). Indeed, in line with the notion of emotion induction occurring

as a consequence of perspective-taking during literary reading, TOM areas and structures like the amygdala have been implicated in narrative contexts concerning characters' feelings (Ferstl et al., 2005), in negatively valenced stories (Altmann et al., 2012), in emotional relative to non-emotional sections of *Harry Potter* (Hsu et al., 2014) and when participants heard spoken narratives describing real-life emotional episodes (Nummenmaa et al., 2014). A recent series of studies has provided further compelling evidence that the greater the emotional content of a story, the greater the recruitment of both cognitive and emotional empathy-related structures such as anterior insula and mid cingulate (Altmann et al., 2012; Hsu et al., 2014, 2015a,b,c). Such findings are in line with the so-called *fiction feeling hypothesis* (Jacobs, 2015), which states that greater emotionality in a narrative results in greater feelings of empathy and immersion.

In music, several studies have implicated various limbic and paralimbic structures in the processing of basic emotions, arousal and valence (e.g., happy vs neutral and consonance vs dissonance; see Koelsch, 2014, for a review). However, it may be argued that since music is not itself an emotional object, at least some emotions induced while listening to it must be inferred (Downey et al., 2013). Supporting the notion that musical emotion may be inferred is the evidence that listeners show activation in structures associated with cognitive empathy during music listening. Steinbeis and Koelsch (2009) showed that when music listeners believed they were listening to a piece of music composed by a human rather than a computer, brain areas typically involved in mentalising, such as the medial prefrontal cortex (mPFC), were activated. Further, in the condition known as Behavioral variant frontotemporal dementia, which is associated with a large network of structures including those involved in mentalising, it was shown that the mentalising deficits normally exhibited by these patients also extended to the music domain (Downey et al., 2013). Specifically, patients were impaired in attributing mental states (e.g., dreamy), but not non-mental characteristics (e.g., raindrops) to music, with performance on the former task being more strongly associated with the vmPFC. Recent evidence of the recruitment of the default mode network (DMN) while listeners listened to their preferred music (Kay et al., 2012; Wilkins et al., 2014) also begs the question of the extent to which mentalising processes determine music preferences. The DMN is a network of structures that is preferentially activated when individuals engage in internal tasks like mind wandering and imagining the future. Critically, however, its sharing of a key structure, the mPFC, with the empathizing network, has been used to explain its frequent recruitment during mentalising and empathizing tasks (Gusnard et al., 2001; Li et al., 2014).

In general, while it may seem highly plausible that readers empathize with human characters in a literary work, the notion of music-evoked empathy has tended to be less intuitive. It is therefore worth noting that in addition to the evidence obtained using neuroscience techniques, numerous behavioral and physiological studies continue to provide persuasive support for the role of empathy-related processes during music listening. For instance, it has been reported that the strength of emotions induced in music listeners (self-report and physiology) modulates as a function of perspective-taking with the music performer

(Miu and Baltes, 2012). Further, it has also been suggested that the discrepancy that sometimes exists between *expressed* and *felt* emotions (Gabrielsson, 2002) may be explained by the subjective degree of empathy felt by the listener for the musician (Egermann and Adams, 2013). Finally, it has been argued that the degree of the empathy trait possessed by a music listener may predict their appreciation of sad music (Taruffi and Koelsch, 2014).

Thus, taken together, a growing body of behavioral, physiological and neuroscience research provides support for the longstanding notion that empathy processes may contribute to the intensity of felt emotion during both literary reading and music listening. Social cognition comprises just one aspect of these activities, however, and, as hinted at above, the temporal unfolding of “information” over time in the two domains, may have an important influence on the way they are experienced. Accounts of brain function that emphasize prediction and active inference (Friston, 2010) are particularly relevant to dynamically unfolding activities like music listening and reading. Thus, it is interesting to consider how such accounts are informing the investigation of emotional responses to these activities and what the result of such investigations are showing.

Predicting the Uncertain

In general, both music and language (the building blocks of literature) are comprised of discrete elements that are not combined haphazardly, but according to a set of principles (Patel, 2008). Just as linguistic syntax refers to the rules that guide the way language is constructed, so also has the term *musical syntax* been used to describe the set of principles guiding the combination of musical elements. In the field of cognitive neuroscience, a comparative approach has revealed similar electrophysiological signatures to irregular or unexpected events in the context of music and language (Patel, 2008). Specifically, “mismatch” responses to low probability events (e.g., Koelsch et al., 2001; Omgie et al., 2013) have been associated with longer processing times (Bharucha and Stoeckig, 1986; Omgie et al., 2012) and localized to the left and right inferior frontal gyrus (Maess et al., 2001; Koelsch et al., 2005). At this point it is worth acknowledging that the necessarily short and highly controlled stimuli that have commonly been used to bring about the signature mismatch responses may seem far removed from the rich and complex literary and musical materials experienced in everyday life. However, these mismatch responses have increasingly been interpreted as support for the Bayesian brain hypothesis, which posits that the brain continuously makes active inferences about how events in the environment will unfold (Garrido et al., 2009; Friston, 2010; Gebauer et al., 2012). Critically, growing investigations into the emotional implications of such predictive processes (e.g., Joffily and Coricelli, 2013; Omgie, 2015) raise the possibility that commonly observed electrophysiological responses reflect a broader mechanism underlying our affective responses to a wide range of stimuli.

Recently, attempts to characterize the experience of continuously and actively predicting have moved away from emphasizing correlates of incorrect predictions (as in the

electrophysiological responses described above) to emphasizing the state of *uncertainty* experienced as a given sequence unfolds (e.g., Hansen and Pearce, 2014; Lehne and Koelsch, 2015). In a recent comprehensive account, the concept of *Tension* was held to be relevant to music, literature (where tension is referred to as suspense), and a range of other activities, and was operationalized as an emotional experience, accompanying continuous prediction making, that arises from a state of uncertainty and need for resolution (Lehne and Koelsch, 2015). The concept of *Tension* has long been used in music listening (Madsen and Fredrickson, 1993; Bigand and Parncutt, 1999; Lerdahl and Krumhansl, 2007; Farbood, 2012), where its build-up and relief is held to be made possible by listeners’ having internalized the tonal systems and forms of their culture’s music. Importantly, feelings of tension in music have also long been related to changes in physiological responses, for instance in response to increased harmonic complexity (e.g., Krumhansl, 1997; Steinbeis et al., 2006). However, only recently, have the neural correlates of musical tension been directly examined using neuroimaging methods (Lehne et al., 2014; see Koelsch, 2014). Indeed, while Koelsch et al. (2008) had demonstrated that structures like the amygdala and OFC are involved in the processing of syntactically irregular musical events (that brought about the previously mentioned mismatch responses), it was also of interest to see that such structures may be linked to the subjective feelings of musical tension (Lehne et al., 2014). Specifically, it was shown that that continuous subjective ratings of tension as provided by participants, correlated with unfolding activity in left pars orbitalis, an area associated with both predictive and affective processing. Further, a region-of-interest analysis was able to confirm the role of amygdala in mediating feelings of increasing relative to decreasing tension during music listening (Lehne et al., 2014). Interestingly, a number of other studies have also been able to indirectly associate subcortical and limbic structures with uncertainty and anticipation in music (Salimpoor et al., 2011; Trost et al., 2012). For instance, Trost et al. (2012) described neural activity in response to a “tension” emotion (characterized by high arousal, negative valence, and unpredictability) not only in sensory and motor areas (linked to prediction making; Schubotz, 2007) but also in structures like the parahippocampal gyrus and caudate nucleus.

Suspense, the concept equivalent to musical tension in its induction of feelings of uncertainty and anticipation (Lehne and Koelsch, 2015), is held to constitute a critical component of narrative literature (Zillmann, 1980; Brewer and Lichtenstein, 1982; Comisky and Bryant, 1982; Oatley, 1994). Further, like musical tension, it has been shown to modulate physiological responses (Zillmann et al., 1975). Thus, of considerable interest was whether suspense evoked by literary reading would also activate the limbic and deep subcortical structures associated with musical tension (Koelsch et al., 2008; Lehne et al., 2014). In a first ever attempt to isolate the neural correlates of suspense that emerges as participants read a literary text for the first time, the authors presented participants with a narrative broken up into numerous shorter segments while their haemodynamic responses were measured (Lehne et al., 2015). Participants were required to

rate each segment, following its presentation, for subjective feelings of suspense. Consequently a parametric regressor that summarized these ratings across participants was used to identify suspense-associated brain regions. The findings were interesting in implicating areas that have been associated with predictive processing in a range of contexts (e.g., inferior frontal gyrus and lateral premotor cortex, see Schubotz, 2007). Further, they were interesting in confirming the role, during the reading of literary texts, of brain areas related to mentalising (e.g., mPFC and TPJ). However, in not implicating a role of subcortical limbic structures in literary tension (as was seen in musical tension), the study from Lehne et al. (2015) suggested differences in the nature of musical and literature-induced uncertainty. Specifically, it suggested differences in the extent of these art forms' recruitment of evolutionary ancient parts of the emotion-processing network.

It remains possible that any conclusions that may be drawn from the studies reviewed above will be moderated following future research. Indeed an important limitation of the literary tension study from Lehne et al. (2015) was the interrupted way in which the stimuli were presented, namely, in segments rather than all at once as in the musical tension study. It remains possible that these interruptions compromised the affective power of the narrative stimuli and, consequently, the extent to which limbic structures could be recruited. Indeed, as seen in the research reviewed earlier, several studies have been able to show a link between limbic activity and perceived emotional intensity of literary stimuli (Ferstl et al., 2005; Wallentin et al., 2011; Altmann et al., 2012; Nummenmaa et al., 2014). Further, there is compelling evidence of the recruitment of limbic regions during the processing of literary stimuli that have been rendered more complex using artistic devices. Here, it is important to point out that in addition to those states of uncertainty that arise from following a plot in the many literary genres that employ suspense (e.g., crime novels, thrillers), states of uncertainty may also arise from a writer's use of literary techniques, of which *defamiliarization* is one (van Peer, 1986; Oatley, 1994; Giora et al., 2004). Defamiliarization is defined as the process whereby a writer makes the familiar unfamiliar and has been shown to reduce the overall predictability of a text, while increasing its

perceived esthetic value (Miall and Kuiken, 1994; Hanauer, 1998). In a recent imaging study, evidence was sought of a contribution of defamiliarization to the affective and esthetic perception of written words (Bohrn et al., 2012). Interestingly, it was shown that defamiliarized proverbs, such as "Time eats money" (a variant of *Time is money*) increased activity not just in syntax and semantics related brain areas, but also in limbic structures like the amygdala and medial OFC. Such findings suggest that even if uncertainty in the unfolding of a plot may not implicate the limbic network to the same degree as uncertainty in music, literature's artistic use of language may provide a rich additional source of emotional power.

Conclusion

In sum, the research literature provides an ever-increasing body of support for the notion of a role of empathy processes during both music listening and literary reading. It also suggests an important role of predictive processes during the consumption of such stimuli, although of interest will be to explore the extent to which uncertainty in the two domains is bound (or not) to activity in the core limbic network. In general, it may be concluded that a comparison of research findings from music and literature focused studies will continue to be enlightening, and that particularly important insights will emerge when studies in the two domains use similar concepts. Critically, it may be expected that while observed overlaps may help to explain the common appeal of music and literature as art forms, differences may help to explain any idiosyncrasies in their respective capacities for affective impact.

Acknowledgments

This work was supported by a stipend to DO from the Max Planck Institute for Empirical Aesthetics. The author thanks Melanie Wald-Fuhrmann and Winfried Menninghaus for their critical reading of earlier drafts of the manuscript. The author is also very grateful to the reviewers for their highly insightful comments and recommendations.

References

- Altmann, U., Bohrn, I. C., Lubrich, O., Menninghaus, W., and Jacobs, A. M. (2012). The power of emotional valence—from cognitive to affective processes in reading. *Front. Hum. Neurosci.* 6:192. doi: 10.3389/fnhum.2012.00192
- Aristotle. (1961). *Poetics*, Butcher S. H., trans, Fergusson, F., intro. New York: Hill & Wang.
- Bharucha, J. J., and Stoeckig, K. (1986). Reaction time and musical expectancy. *J. Exp. Psychol. Hum. Percept. Perform.* 12, 403–410. doi: 10.1037/0096-1523.12.4.403
- Bigand, E., and Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychol. Res.* 62, 237–254. doi: 10.1007/s004260050053
- Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W., and Jacobs, A. M. (2012). Old proverbs in new skins—an fMRI study on defamiliarization. *Front. Psychol.* 3:204. doi: 10.3389/fpsyg.2012.00204
- Brewer, W. F., and Lichtenstein, E. H. (1982). Stories are to entertain: a structural-affect theory of stories. *J. Pragmat.* 6, 473–486. doi: 10.1016/0378-2166(82)90021-2
- Brown, S. (2000). "The 'musilanguage' model of music evolution," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 271–300.
- Citron, F. M. (2012). Neural correlates of written emotion word processing: a review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain Lang.* 122, 211–226. doi: 10.1016/j.bandl.2011.12.007
- Comisky, P., and Bryant, J. (1982). Factors involved in generating suspense. *Hum. Commun. Res.* 9, 49–58. doi: 10.1111/j.1468-2958.1982.tb00682.x
- Cupchik, G. C., Oatley, K., and Vorderer, P. (1998). Emotional effects of reading excerpts from short stories by James Joyce. *Poetics* 25, 363–377. doi: 10.1016/S0304-422X(98)90007-9
- Downey, L. E., Blezat, A., Nicholas, J., Omar, R., Golden, H. L., Mahoney, C. J., et al. (2013). Mentalising music in frontotemporal dementia. *Cortex* 49, 1844–1855. doi: 10.1016/j.cortex.2012.09.011
- Egermann, H., and Adams, S. (2013). Empathy and emotional contagion as a link between recognised and felt emotions in music listening. *Music. Percept.* 31, 139–156. doi: 10.1525/mp.2013.31.2.139
- Eslinger, P. J. (1998). Neurological and neuropsychological bases of empathy. *Euro. Neurol.* 39, 193–199. doi: 10.1159/00007933
- Farbood, M. (2012). A parametric, temporal model of musical tension. *Music. Percept.* 29, 387–428. doi: 10.1525/mp.2012.29.4.387
- Ferstl, E. C., Rinck, M., and von Cramon, D. Y. (2005). Emotional and Temporal Aspects of Situation Model Processing during Text Comprehension:

- an Event-Related fMRI Study. *J. Cogn. Neurosci.* 17, 724–739. doi: 10.1162/0898929053747658
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Frith, C. C. (1999). Interacting minds—A biological basis. *Science* 286, 1692–1695. doi: 10.1126/science.286.5445.1692
- Gabrielsson, A. (2002). Emotion perceived and emotion felt same or different? *Music. Sci.* 2001–2002, 123–147.
- Gallese, V. (2003). The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology* 36, 171–180. doi: 10.1159/000072786
- Garrido, M. I., Kilner, J. M., Kiebel, S. J., and Friston, K. J. (2009). Dynamic causal modeling of the response to frequency deviants. *J. Neurophysiol.* 101, 2620–2631. doi: 10.1152/jn.90291.2008
- Gebauer, L., Kringelbach, M. L., and Vuust, P. (2012). The ever-changing cycles of musical pleasure: the role of dopamine and anticipation. *Psychomusicology* 22, 152–167. doi: 10.1037/a0031126
- Giora, R., Fein, O., Kronrod, A., Elnatan, I., Shuval, N., and Zur, A. (2004). Weapons of mass distraction: optimal innovation and pleasure ratings. *Metaphor Symbol.* 19, 115–141. doi: 10.1207/s15327868ms1902_2
- Goldman, A. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. New York: Oxford University Press. doi: 10.1093/0195138929.001.0001
- Gusnard, D. A., Akbudak, E., Shulman, G. L., and Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4259–4264. doi: 10.1073/pnas.071043098
- Gygax, P., Oakhill, J., and Garnham, A. (2003). The representation of characters' emotional responses: do readers infer specific emotions? *Cogn. Emot.* 17, 413–428. doi: 10.1080/02699930244000048
- Hanauer, D. (1998). The genre-specific hypothesis of reading: reading poetry and encyclopedic items. *Poetics* 26, 63–80. doi: 10.1016/S0304-422X(98)00011-4
- Hansen, N. C., and Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Front. Psychol.* 5:1052. doi: 10.3389/fpsyg.2014.01052
- Hjort, M., and Lavers, S. (1997). *Emotion and the Arts*. Oxford: Oxford University Press.
- Hogan, P. C. (2010). Fiction and Feelings: on the place of literature in the study of Emotion. *Emot. Rev.* 2, 184–195. doi: 10.1177/1754073909352874
- Hsu, C. T., Conrad, M., and Jacobs, A. M. (2014). Fiction feelings in Harry Potter: haemodynamic response in the mid-cingulate cortex correlates with immersive reading experience. *Neuroimage* 25, 1356–1361. doi: 10.1097/wnr.0000000000000272
- Hsu, C. T., Jacobs, A. M., and Conrad, M. (2015a). Can Harry Potter still put a spell on us in a second language? An fMRI study on reading emotion-laden literature in late bilinguals. *Cortex* 63, 282–295. doi: 10.1016/j.cortex.2014.09.002
- Hsu, C. T., Jacobs, A. M., Citron, F., and Conrad, M. (2015b). The emotion potential of words and passages in reading Harry Potter: an fMRI study. *Brain Lang.* 142, 96–114. doi: 10.1016/j.bandl.2015.01.011
- Hsu, C. T., Jacobs, A. M., Altmann, U., and Conrad, M. (2015c). The magical activation of left amygdala when reading Harry Potter: an FMRI study on how descriptions of supra-natural events entertain and enchant. *PLoS ONE* 10:e0118179. doi: 10.1371/journal.pone.0118179
- Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.
- Jacobs, A. M. (2015). “Towards a neurocognitive poetics model of literary reading,” in *Cognitive Neuroscience of Natural Language Use*, ed. R. M. Willems (Cambridge: Cambridge University Press), 135–159.
- Joffily, M., and Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLoS Comp. Biol.* 9:e1003094. doi: 10.1371/journal.pcbi.1003094
- Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Phys. Life. Rev.* 10, 235–266. doi: 10.1016/j.plrev.2013.05.008
- Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 559–575. doi: 10.1017/S0140525X08005293
- Kay, B. P., Meng, X., DiFrancesco, M. W., Holland, S. K., and Szarflaski, J. P. (2012). Moderating effects of music on resting state networks. *Brain Res.* 477, 53–64. doi: 10.1016/j.brainres.2012.01.064
- Keen, S. (2007). *Empathy and the Novel*. New York: Oxford University Press. doi: 10.1093/acprof:oso/9780195175769.001.0001
- Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666
- Koelsch, S., Fritz, T., and Schlaug, G. (2008). Amygdala activity can be modulated by unexpected chord functions during music listening. *Neuroimage* 3, 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., Gunter, T. C., Schröger, E., Tervaniemi, M., Sammler, D., and Friederici, A. D. (2001). Differentiating ERAN and MMN: an ERP study. *Neuroimage* 12, 1385–1389. doi: 10.1097/00001756-200105250-00019
- Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., and Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nat. Neurosci.* 7, 302–307. doi: 10.1038/nn1197
- Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Can. J. Exp. Psychol.* 51, 336–353. doi: 10.1037/1196-1961.51.4.336
- Lehne, M., Engel, P., Rohrmeier, M., Menninghaus, W., Jacobs, A. M., and Koelsch, S. (2015). Reading a suspenseful literary text activates brain areas related to social cognition and predictive inference. *PLoS ONE* 10:e0124550. doi: 10.1371/journal.pone.0124550
- Lehne, M., and Koelsch, S. (2015). Toward a general psychological model of tension and suspense. *Front. Psychol.* 6:79. doi: 10.3389/fpsyg.2015.00079
- Lehne, M., Rohrmeier, M., and Koelsch, S. (2014). Tension-related activity in the orbitofrontal cortex and amygdala: an fMRI study with music. *Soc. Cogn. Affect. Neurosci.* 9, 1515–1523. doi: 10.1093/scan/nst141
- Lerdahl, F., and Krumhansl, C. L. (2007). Modeling tonal tension. *Music. Percept.* 24, 329–366. doi: 10.1525/mp.2007.24.4.329
- Levinson, J. (2004). Music as narrative and music as drama. *Mind Lang.* 19, 428–441. doi: 10.1111/j.0268-1064.2004.00267.x
- Li, W., Mai, X., and Liu, C. (2014). The default mode network and social understanding of others: what do brain connectivity studies tell us. *Front. Hum. Neurosci.* 8:74. doi: 10.3389/fnhum.2014.00074
- Longinus (1965). “On the sublime,” in Aristotle/Horace/Longinus, *Classical Literary Criticism*, trans. T. S. Dorsch (Baltimore: Penguin Books), 97–158.
- Madsen, C. K., and Fredrickson, W. E. (1993). The experience of musical tension: a replication of Nielsen's research using the continuous response digital interface. *J. Music Ther.* 30, 46–63. doi: 10.1093/jmt/30.1.46
- Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001). Musical syntax is processed in the area of Broca: an MEG-study. *Nat. Neurosci.* 4, 540–545. doi: 10.1038/87502
- Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Ann. Rev. Psychol.* 62, 103–134. doi: 10.1146/annurev-psych-120709-145406
- Mar, R. A., and Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspect. Psychol. Sci.* 3, 173–192. doi: 10.1111/j.1745-6924.2008.00073.x
- Maus, F. (1991). Music as narrative. *Indiana Theory Rev.* 12, 1–34.
- Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago: Chicago University Press.
- Miall, D. S., and Kuiken, D. (1994). Foregrounding, defamiliarization, and affect—response to literary stories. *Poetics* 22, 389–407. doi: 10.1016/0304-422X(94)00011-5
- Miall, D. S., and Kuiken, D. (2002). A feeling for fiction: becoming what we behold. *Poetics* 30, 221–241. doi: 10.1016/S0304-422X(02)00011-6
- Miu, A. C., and Baltes, F. R. (2012). Empathy manipulation impacts music-induced emotions: a psychophysiological study on opera. *PLoS ONE* 7:e30618. doi: 10.1371/journal.pone.0030618
- Nummenmaa, L., Saarimäki, H., Glerean, E., Gotsopoulos, A., Jaaskelainen, I. P., Hari, R., et al. (2014). Emotional speech synchronizes brains across listeners and engages large-scale dynamic brain networks. *Neuroimage* 102, 498–509. doi: 10.1016/j.neuroimage.2014.07.063
- Oatley, K. (1994). A taxonomy of the emotions of literary response and a theory of identification in fictional narrative. *Poetics* 23, 53–74. doi: 10.1016/0304-422X(94)P4296-S
- Oatley, K. (2002). “Emotions and the story worlds of fiction,” in *Narrative Impact: Social and Cognitive Foundations*, eds M. C. Green, J. J. Strange, and T. C. Brock (Mahwah, NJ: Lawrence Erlbaum Associates), 39–69.

- Omigie, D. (2015). Dopamine and epistemic curiosity in music listening. *Cogn. Neurosci.* 14, 1–2. doi: 10.1080/17588928.2015.1051013
- Omigie, D., Pearce, M. T., and Stewart, L. (2012). Tracking of pitch probabilities in congenital amusia. *Neuropsychologia* 50, 1483–1493. doi: 10.1016/j.neuropsychologia.2012.02.034
- Omigie, D., Pearce, M. T., Williamson, V. J., and Stewart, L. (2013). Electrophysiological correlates of melodic processing in congenital amusia. *Neuropsychologia* 51, 1749–1762. doi: 10.1016/j.neuropsychologia.2013.05.010
- Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford: Oxford University Press.
- Rabkin, E. S. (1973). *Narrative Suspense: "When Slim Turned Sideways. . ."* Ann Arbor, MI: University of Michigan Press.
- Robinson, J. (2007). *Deeper than Reason. Emotion and its Role in Literature, Music and Art*. Oxford: Oxford University Press.
- Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat. Neurosci.* 14, 257–262. doi: 10.1038/nn.2726
- Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends. Cogn. Sci.* 11, 211–218. doi: 10.1016/j.tics.2007.02.006
- Schulte-Ruther, M., Markowitsch, H. J., Fink, G. R., and Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy. *J. Cogn. Neurosci.* 19, 1354–1372. doi: 10.1162/jocn.2007.19.8.1354
- Shamay-Tsoory, S. G., Aharon-Peretz, J., and Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain* 132, 617–627. doi: 10.1093/brain/awn279
- Shamay-Tsoory, S. G., Tomer, R., Berger, B. D., and Aharon-Peretz, J. (2003). Characterization of empathy deficits following prefrontal brain damage: the role of the right ventromedial prefrontal cortex. *J. Cogn. Neurosci.* 15, 324–337. doi: 10.1162/089892903321593063
- Slevc, L. R., and Patel, A. D. (2011). Meaning in music and language: three key differences. Comment on "Towards a neural basis of processing musical semantics" by Stefan Koelsch. *Phys. Life Rev.* 8, 110–111. doi: 10.1016/j.plrev.2011.05.003
- 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
- Steinbeis, N., and Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cereb. Cortex* 18, 1169–1178. doi: 10.1093/cercor/bhm149
- Steinbeis, N., and Koelsch, S. (2009). Understanding the intentions behind man-made products elicits neural activity in areas dedicated to mental state attribution. *Cereb. Cortex* 19, 619–623. doi: 10.1093/cercor/bhn110
- 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
- Trost, W., Ethofer, T., Zentner, M., and Vuilleumier, P. (2012). Mapping aesthetic musical emotions in the brain. *Cereb. Cortex* 22, 2769–2783. doi: 10.1093/cercor/bhr353
- van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* 30, 829–858. doi: 10.1002/hbm.20547
- van Peer, W. (1986). *Stylistics and Psychology: Investigations of Foregrounding*. London: Croom Helm.
- Wallentin, M., Nielsen, A. H., Vuust, P., Dohn, A., Roerpfstoft, A., and Lund, T. E. (2011). Amygdala and heart rate variability responses from listening to emotionally intense parts of a story. *Neuroimage* 58, 963–973. doi: 10.1016/j.neuroimage.2011.06.077
- Wilkins, R. W., Hodges, D. A., Laurienti, P. J., Stehen, M., and Burdette, J. H. (2014). Network science and the effects of music preference on functional brain connectivity: from Beethoven to Eminem. *Sci. Rep.* 4, 6130. doi: 10.1038/srep06130
- Zillmann, D. (1980). "Anatomy of suspense," in *The Entertainment Functions of Television*, ed. P. H. Tannenbaum (Hillsdale, NJ: Lawrence Erlbaum Associates). 133–163.
- Zillmann, D., Hay, T. A., and Bryant, J. (1975). The effect of suspense and its resolution on the appreciation of dramatic presentations. *J. Res. Personal.* 9, 307–323. doi: 10.1016/0092-6566(75)90005-7
- Zunshine, J. (2006). *Why We Read Fiction: Theory of Mind and the Novel*. Columbus, OH: Ohio State.

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

Copyright © 2015 Omigie. 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.

