

frontiers

RESEARCH TOPICS

THE RELATIONSHIP BETWEEN MUSIC AND LANGUAGE

Hosted by
Lutz Jäncke



frontiers in
PSYCHOLOGY



frontiers

FRONTIERS COPYRIGHT STATEMENT

© Copyright 2007-2012
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, as well as all content on this site is the exclusive property of Frontiers. Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Articles and other user-contributed materials may be downloaded and reproduced subject to any copyright or other notices. No financial payment or reward may be given for any such reproduction except to the author(s) of the article concerned.

As author or other contributor you grant permission 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.

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714

ISBN 978-2-88919-054-6

DOI 10.3389/978-2-88919-054-6

ABOUT FRONTIERS

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

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.

All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

THE RELATIONSHIP BETWEEN MUSIC AND LANGUAGE

Hosted By:

Lutz Jäncke, University of Zurich, Switzerland

Traditionally, music and language have been treated as different psychological faculties. This duality is reflected in older theories about the lateralization of speech and music in that speech functions were thought to be localized on the left and music functions on the right hemisphere. But with the advent of modern brain imaging techniques and the improvement of neurophysiological measures to investigate brain functions an entirely new view on the neural and psychological underpinnings of music and speech has evolved. The main point of convergence in the findings of these new studies is that music and speech functions have many aspects in common and that several neural modules are similarly involved in speech and music. There is also emerging evidence that speech functions can benefit from music functions and vice versa. This new research field has accumulated a lot of new information and it is therefore timely to bring together the work of those researchers who have been most visible, productive, and inspiring in this field and to ask them to present their new work or provide a summary of their laboratory's work.

Table of Contents

- 05 *The Relationship between Music and Language***
Lutz Jäncke
- 07 *The Effect of a Music Program on Phonological Awareness in Preschoolers***
Franziska Degé and Gudrun Schwarzer
- 14 *Native Experience with a Tone Language Enhances Pitch Discrimination and the Timing of Neural Responses to Pitch Change***
Ryan J. Giuliano, Peter Q. Pfordresher, Emily M. Stanley, Shalini Narayana and Nicole Y. Y. Wicha
- 26 *EEG Correlates of Song Prosody: A New Look at the Relationship between Linguistic and Musical Rhythm***
Reyna L. Gordon, Cyrille L. Magne and Edward W. Large
- 39 *Effects of Practice and Experience on the Arcuate Fasciculus: Comparing Singers, Instrumentalists, and Non-Musicians***
Gus F. Halwani, Psyche Loui, Theodor Rüber and Gottfried Schlaug
- 48 *The Influence of Task-Irrelevant Music on Language Processing: Syntactic and Semantic Structures***
Lisianne Hoch, Benedicte Poulin-Charronnat and Barbara Tillmann
- 58 *Relating Pitch Awareness to Phonemic Awareness in Children: Implications for Tone-Deafness and Dyslexia***
Psyche Loui, Kenneth Kroog, Jennifer Zuk, Ellen Winner and Gottfried Schlaug
- 63 *Preserved Statistical Learning of Tonal and Linguistic Material in Congenital Amusia***
Diana Omigie and Lauren Stewart
- 74 *Processing of Voiced and Unvoiced Acoustic Stimuli in Musicians***
Cyrill Guy Martin Ott, Nicolas Langer, Mathias S. Oechslin, Martin Meyer and Lutz Jäncke
- 84 *Tone Language Fluency Impairs Pitch Discrimination***
Isabelle Peretz, Sébastien Nguyen and Stéphanie Cummings
- 89 *Music and Language Processing Share Behavioral and Cerebral Features***
Clara E. James
- 90 *Musical Expertise and Statistical Learning of Musical and Linguistic Structures***
Daniele Schön and Clément François
- 99 *Can You Hear Me Now? Musical Training Shapes Functional Brain Networks for Selective Auditory Attention and Hearing Speech in Noise***
Dana L. Strait and Nina Kraus

- 109** ***Congenital Amusia (or Tone-Deafness) Interferes with Pitch Processing in Tone Languages***
Barbara Tillmann, Denis Burnham, Sebastien Nguyen, Nicolas Grimault, Nathalie Gosselin and Isabelle Peretz
- 124** ***Non-Invasive Brain Stimulation Enhances the Effects of Melodic Intonation Therapy***
Bradley W. Vines, Andrea C. Norton and Gottfried Schlaug
- 134** ***How Singing Works***
Katrin Schulze
- 136** ***Disorders of Pitch Production in Tone Deafness***
Simone Dalla Bella, Magdalena Berkowska and Jakub Sowiński
- 147** ***Transfer of Training between Music and Speech: Common Processing, Attention, and Memory***
Mireille Besson, Julie Chobert and Céline Marie
- 159** ***Implicit Memory in Music and Language***
Marc Ettlinger, Elizabeth H. Margulis and Patrick C. M. Wong
- 169** ***Toward a Neural Basis of Music Perception – A Review and Updated Model***
Stefan Koelsch
- 189** ***The Interplay between Musical and Linguistic Aptitudes: A Review***
Riia Milovanov and Mari Tervaniemi
- 195** ***Why Would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis.***
Aniruddh D. Patel
- 209** ***Neurophysiological Influence of Musical Training on Speech Perception***
Antoine J. Shahin



The relationship between music and language

Lutz Jäncke*

Division Neuropsychology, Institute of Psychology, University of Zurich, Zurich, Switzerland

*Correspondence: l.jaencke@psychologie.uzh.ch

Traditionally, music and language have been treated as different psychological faculties. This duality is reflected in older theories about the lateralization of speech and music in that speech functions were thought to be localized in the left and music functions in the right-hemisphere of the brain. For example, the landmark paper of Bever and Chiarello (1974) emphasized the different roles of both hemispheres in processing music and language information, with the left hemisphere considered more specialized for propositional, analytic, and serial processing and the right-hemisphere more specialized for appositional, holistic, and synthetic relations. This view has been challenged in recent years mainly because of the advent of modern brain imaging techniques and the improvement in neurophysiological measures to investigate brain functions. Using these innovative approaches, an entirely new view on the neural and psychological underpinnings of music and speech has evolved. The findings of these more recent studies show that music and speech functions have many aspects in common and that several neural modules are similarly involved in speech and music (Tallal and Gaab, 2006). There is also emerging evidence that speech functions can benefit from music functions and vice versa. This field of research has accumulated a lot of new information and it is therefore timely to bring together the work of those researchers who have been most visible, productive, and inspiring in this field.

This special issue comprises a collection of 20 review and research papers that focus on the specific relationship between music and language. Of these 20 papers 12 are research papers that report entirely new findings supporting the close relationship between music and language functions. Two papers report findings demonstrating that phonological awareness, which is pivotal for reading and writing skills, is closely related to pitch awareness and musical expertise (Dege and Schwarzer, 2011; Loui et al., 2011). Dege and colleagues even show that pre-schoolers can benefit from a program of musical training to increase their phonological awareness.

Three research papers focus on the relationship between tonal language expertise and musical pitch perception skills and on whether pitch-processing deficits might influence tonal language perception. Giuliano et al. (2011) demonstrated Mandarin speakers are highly sensitive to small pitch changes and interval distances, a sensitivity that was absent in the control group. Using ERPs obtained during the pitch and interval perception tasks, their study reveals earlier ERP responses in Mandarin speakers compared with controls to these pitch changes relative to no-change trials. In their elegant paper, Peretz et al. (2011) report that native speakers of a tone language, in which pitch contributes to word meaning, are impaired in the discrimination of falling pitches in tone sequences as compared with speakers of a non-tone language. Taken together, these two studies illustrate the cross-domain influence of language experience on the perception of pitch, suggesting that the native use of tonal pitch contours in language leads to a general enhancement in the acuity of pitch representations. Tillmann et al. (2011) examined

whether subjects suffering from congenital amusia also demonstrate impairments of pitch-processing in speech, specifically the pitch changes used to contrast lexical tones in tonal languages. Their study revealed that the performance of congenital amusics was inferior to that of controls for all materials including the Mandarin language, this therefore suggesting a domain-general pitch-processing deficit.

Five research papers sought to examine interactions either between musical expertise and language functions or whether an interaction between musical and language functions is beneficial for phonetic perception. Ott et al. (2011) demonstrate that professional musicians process unvoiced stimuli (irrespective of whether these stimuli are speech or non-speech stimuli) differently than controls, this suggesting that early phonetic processing is differently organized depending on musical expertise. Strait and Kraus (2011) report perceptual advantages in musicians for hearing and neural encoding of speech in background noise. They also argue that musicians possess a neural proficiency for selectively engaging and sustaining auditory attention to language and that music thus represents a potential benefit for auditory training. Gordon et al. (2011) examined the interaction between linguistic stress and musical meter and established that alignment of linguistic stress and musical meter in song enhances musical beat tracking and comprehension of lyrics. Their study thus supports the notion of a strong relationship between linguistic and musical rhythm in songs. Hoch et al. (2011) investigated the effect of a musical chord's tonal function on syntactic and semantic processing and conclude that neural and psychological resources of music and language processing strongly overlap. The fifth paper of this group (Omigie and Stewart, 2011) demonstrates that the difficulties amusic individuals have with real-world music cannot be accounted for by an inability to internalize lower-order statistical regularities but may arise from other factors. Although there are still some differences between music and speech-processing, there thus is growing evidence that speech and music processing strongly overlap.

Halwani et al. (2011) examined whether the arcuate fasciculus, a prominent white-matter tract connecting temporal and frontal brain regions, is anatomically different between singers, instrumentalists, and non-musicians. They showed that long-term vocal-motor training might lead to an increase in volume and microstructural complexity (as indexed by fractional anisotropy measures) of the arcuate fasciculus in singers. Most likely, these anatomical changes reflect the necessity in singers of strongly linking together frontal and temporal brain regions. Typically, these regions are also involved in the control of many speech functions. The beneficial impact of music on speech functions has also been demonstrated by Vines et al. (2011) in their research paper. They examined whether the melodic intonation therapy (MIT) in Broca's aphasics can be improved by simultaneously applying anodal transcranial direct current stimulation (tDCS). In fact, they showed that the combination of right-hemisphere anodal-tDCS with MIT speeded up recovery from post-stroke aphasia.

In addition to these 12 research papers there are 8 review and opinion papers that highlight the tight link between music and language. Patel (2011) proposes the so-called OPERA hypothesis with which he explains why music is beneficial for many language functions. The acronym OPERA stands for five conditions which might drive plasticity in speech-processing networks (Overlap: anatomical overlap in the brain networks that process acoustic features used in both music and speech; Precision: music places higher demands on these shared networks than does speech; Emotion: the musical activities that engage this network elicit strong positive emotion; Repetition: the musical activities that engage this network are frequently repeated; Attention: the musical activities that engage this network are associated with focused attention). According to the OPERA hypothesis, when these conditions are met, neural plasticity drives the networks in question to function with higher precision than needed for ordinary speech communication. While Patel's paper is more an opinion paper that puts musical expertise into a broader context, the seven other reviews more or less emphasize specific aspects of the current literature on music and language. Ettlinger et al. (2011) emphasize the specific role of implicitly acquired knowledge, implicit memory, and their associated neural structures in the acquisition of linguistic or musical grammar. Milovanov and Tervaniemi (2011) underscore the beneficial influence of musical aptitude on the acquisition linguistic skills as for example in acquiring a second language. Bella et al. (2011) summarize findings of the existing literature concerning normal singing and poor-pitch singing and suggest that pitch imitation may be selectively inaccurate in the music domain without being affected in speech, thus supporting the separability of

mechanisms subserving pitch production in music and language. In their extensive review of the literature, Besson et al. (2011) discuss the transfer effects from music to speech by specifically focusing on the musical expertise in musicians. Shahin (2011) article reviews neurophysiological evidence supporting an influence of musical training on speech perception at the sensory level, and the question is discussed whether such transfer could facilitate speech perception in individuals with hearing loss. This review also explains the basic neurophysiological measures used in the neurophysiological studies of speech and music perception. The comprehensive review by Koelsch (2011) summarizes findings from neurophysiology and brain imaging on music and language processing and integrates these findings into a broader "neurocognitive model of music perception." Specific emphasis is placed on the comparison of musical syntax and their similarities and differences to language syntax. Schon and Francois (2011) present a review in which they focus on a series of electrophysiological studies that investigated speech segmentation and the extraction of linguistic versus musical information. They demonstrated that musical expertise facilitates the learning of both linguistic and musical structures. A further point is that electrophysiological measures are often more sensitive for identifying music-related differences than behavioral measures.

Taken together, this special issue provides a comprehensive summary of the current knowledge on the tight relationship between music and language functions. Thus, musical training may aid in the prevention, rehabilitation, and remediation of a wide range of language, listening, and learning impairments. On the other hand, this body of evidence might shed new light on how the human brain uses shared network capabilities to generate and control different functions.

REFERENCES

- Bella, S. D., Berkowska, M., and Sowinski, J. (2011). Disorders of pitch production in tone deafness. *Front. Psychol.* 2:164. doi: 10.3389/fpsyg.2011.00164
- 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
- Bever, T. G., and Chiarello, R. J. (1974). Cerebral dominance in musicians and nonmusicians. *Science* 185, 537–539.
- Dege, 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
- Ettlinger, M., Margulis, E. H., and Wong, P. C. (2011). Implicit memory in music and language. *Front. Psychol.* 2:211. doi: 10.3389/fpsyg.2011.00211
- Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., and Wicha, N. Y. (2011). Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Front. Psychol.* 2:146. doi: 10.3389/fpsyg.2011.00146
- 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
- Halwani, G. F., Loui, P., Ruber, 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
- 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
- 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
- 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
- Milovanov, R., and Tervaniemi, M. (2011). The interplay between musical and linguistic aptitudes: a review. *Front. Psychol.* 2:321. doi: 10.3389/fpsyg.2011.00321
- Omigie, D., and Stewart, L. (2011). Preserved statistical learning of tonal and linguistic material in congenital amusia. *Front. Psychol.* 2:109. doi: 10.3389/fpsyg.2011.00109
- Ott, C. G., Langer, N., Oechslin, M., Meyer, M., and Jancke, L. (2011). Processing of voiced and unvoiced acoustic stimuli in musicians. *Front. Psychol.* 2:195. doi: 10.3389/fpsyg.2011.00195
- 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
- Peretz, I., Nguyen, S., and Cummings, S. (2011). Tone language fluency impairs pitch discrimination. *Front. Psychol.* 2:145. doi: 10.3389/fpsyg.2011.00145
- Schon, D., and Francois, C. (2011). Musical expertise and statistical learning of musical and linguistic structures. *Front. Psychol.* 2:167. doi: 10.3389/fpsyg.2011.00167
- Shahin, A. J. (2011). Neurophysiological influence of musical training on speech perception. *Front. Psychol.* 2:126. doi: 10.3389/fpsyg.2011.00126
- Strait, D. L., and Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Front. Psychol.* 2:113. doi: 10.3389/fpsyg.2011.00113
- Tallal, P., and Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends Neurosci.* 29, 382–390.
- Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., and Peretz, I. (2011). Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. *Front. Psychol.* 2:120. doi: 10.3389/fpsyg.2011.00120
- Vines, B. W., Norton, A. C., and Schlaug, G. (2011). Non-invasive brain stimulation enhances the effects of melodic intonation therapy. *Front. Psychol.* 2:230. doi: 10.3389/fpsyg.2011.00230

Received: 05 April 2012; accepted: 07 April 2012; published online: 27 April 2012.

Citation: Jäncke L (2012) The relationship between music and language. *Front. Psychology* 3:123. doi: 10.3389/fpsyg.2012.00123

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

Copyright © 2012 Jäncke. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.



The effect of a music program on phonological awareness in preschoolers

Franziska Degé* and Gudrun Schwarzer

Developmental Psychology, Justus-Liebig-University, Giessen, Germany

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Urs Maurer, University of Zurich, Switzerland
Glenn Schellenberg, University of Toronto, Canada

***Correspondence:**

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

The present experiment investigated the effect of a music program on phonological awareness in preschoolers. In particular, the effects of a music program and a phonological skills program on phonological awareness were compared. If language and music share basic processing mechanisms, the effect of both programs on enhancing phonological awareness should be similar. Forty-one preschoolers (22 boys) were randomly assigned to a phonological skills program, a music program, and a control group that received sports training (from which no effect was expected). Preschoolers were trained for 10 min on a daily basis over a period of 20 weeks. In a pretest, no differences were found between the three groups in regard to age, gender, intelligence, socioeconomic status, and phonological awareness. Children in the phonological skills group and the music group showed significant increases in phonological awareness from pre- to post-test. The children in the sports group did not show a significant increase from pre- to post-test. The enhancement of phonological awareness was basically driven by positive effects of the music program and the phonological skills program on phonological awareness for large phonological units. The data suggests that phonological awareness can be trained with a phonological skills program as well as a music program. These results can be interpreted as evidence of a shared sound category learning mechanism for language and music at preschool age.

Keywords: music, music program, music training, music and language, sound category learning, phonological awareness, phonological skills, preschoolers

INTRODUCTION

Language and music are specific to humans and share several characteristics such as the use of the auditory domain as the input path and the organization of discrete perceptual elements into structured sequences (Patel, 2003). McMullen and Saffran (2004) postulated that language and music share relevant processing mechanisms, especially in childhood. One shared mechanism in childhood is the sound category learning mechanism (McMullen and Saffran, 2004). Consequently, a relationship between language sound categories such as phonemes (phonological awareness) and musical sound categories such as notes should be evident. To test this relation, the present study applied an experimental design to investigate the effect of a music program on phonological awareness in preschoolers. Moreover, the current study compared the effect of a music program and the effect of a phonological skills program on phonological awareness. If the “shared sound category learning mechanism hypothesis” (Patel, 2008) is correct, the effects of the two programs on phonological awareness should be comparable.

Empirical evidence has supported McMullen and Saffran’s (2004) idea that music and language have a common basis in the early years of development. Within the first interactions between adults and infants, adults use infant-directed forms of language and music. Infants prefer these forms of infant-directed speech and songs from an early age onward (Trainor, 1996; Masataka, 1999). Moreover, infant-directed speech is often referred to as a

type of musical speech (Fernald, 1991). This is in accordance with Koelsch and Siebel’s (2005) assumption that the early developing brain processes language as a type of music.

Although musical and linguistic sound systems have different organizations of pitch and timbre, both sound systems face the challenge of extracting a small number of categories that are meaningful from a flow of acoustically variable signals. Moreover, the mind must separate variation within a category, e.g., induced by variable speakers, from variation that constitutes a change of category. These challenges are likely solved by a shared mechanism (Patel, 2008). This “shared sound category learning mechanism hypothesis” predicts comparable individual differences in language and musical abilities. Accordingly, the categorical building blocks of language (e.g., phonemes) should be related to the categorical building blocks of music (e.g., pitches/notes).

This hypothesis is supported by Lamb and Gregory (1993), who revealed a relationship between pitch discrimination and phonemic awareness in 4- and 5-year-old children, suggesting a general relation between musical abilities and phonological awareness. Phonological awareness describes the ability to analyze and manipulate language on two levels. On the word level, phonological awareness refers to 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 repeatedly been shown that phonological awareness

is an important predictor of later reading ability (Pratt and Brady, 1988; Bruck, 1992). Subsequently, Anvari et al. (2002) recruited a larger sample of 4- and 5-year-olds and investigated the relation between musical aptitude and the development of phonological awareness as well as the relation between musical aptitude and the development of reading. In both age groups, musical aptitude was correlated with phonological awareness. The authors concluded that the auditory processing mechanisms necessary for musical perception were related to the auditory processing mechanism necessary for phonological awareness. This relationship was further investigated with respect to the effects of music training on phonological awareness. Bolduc (2009) compared the effect of two music programs on kindergarten children. One music program employed musical activities to increase interest in reading and writing in preschoolers with special needs, whereas the other music program was primarily designed to enhance musical abilities. The music program that focused on enhancing reading and writing was more efficient in enhancing phonological awareness than the other music program. Because there was not a control group without music training, it remains unclear whether both programs significantly enhanced the phonological awareness of the children. Additionally, it is not clear whether the advancement in phonological awareness was due to the music program. In a quasi-experiment, Gromko (2005) investigated the effect of music training on phonological awareness (phoneme segmentation fluency in particular). Children in the treatment kindergarten received music training for 4 months. Children in the control kindergarten received no treatment. Gromko (2005) revealed significantly greater gains in phoneme segmentation fluency in the treatment kindergarten children than in the control kindergarten children. However, because of the pseudo-random assignment of the preschoolers to the treatment and the control group, it is not possible to interpret the results unequivocally. Children in the treatment group may have systematically differed from children in the control group with respect to confounding variables. Indeed, Gromko (2005) reported differences in socioeconomic status between the groups. Moreover, the control group did not receive an alternative training. Therefore, the significant gain in the treatment group may simply represent an effect of extra attention (Hawthorne effect).

The aim of the present experiment was twofold. Firstly, the effect of a music program on phonological awareness was investigated. Previous studies demonstrated a relation between musical abilities as well as music training and phonological awareness. However, because of the correlational and quasi-experimental approaches of these studies, it was not possible to establish the specific causation. Secondly, the current investigation compared the effect of a music program and phonological skills program on phonological awareness to examine the adaptability of the “shared sound category learning mechanism hypothesis.” If this hypothesis is correct, the effect of the music program and the phonological skills program on phonological awareness should be comparable. To our knowledge, no study has addressed this question with such an approach. To this end, an experiment was conducted that randomly assigned preschoolers to a music program, a phonological skills program, and a control group that received sports training (from which no effect was expected). All preschoolers were trained for 10 min on a daily basis for a period of 20 weeks. Phonological

awareness was tested prior to the beginning of training (pretest) and after the completion of training (post-test). This approach allowed inferences of causation and, due to random assignment, ruled out systematic differences between the groups. Moreover, the control group, which received sports training, offered the possibility to control for the effects of retesting, maturation, and attention (i.e., Hawthorne effect). Furthermore, this approach allowed the comparison between the effect of a music program and a program that explicitly addressed phonological awareness.

MATERIALS AND METHODS

PARTICIPANTS

At the beginning of the experiment, the sample consisted of 55 participants. Some children transferred to a different kindergarten during the phase of trainings, and others were not available for the post-test. Five participants were excluded from the analysis because their pretest phonological skills scores were more than 2 SD from the mean. Hence, differences in phonological skills prior to training were minimized. Overall, 75% ($n = 41$) of the original sample was included in the analyses. The dropout rate was not significantly different between the training groups and the control group. The dropouts did not differ significantly in age, gender, socioeconomic status, or intelligence from the remaining sample. The remaining participants were 41 (19 girls) preschoolers from three different kindergartens in Giessen, Germany. At the beginning of the study, the children ranged in age from 5 to 6 years ($M = 5$ years; 9 months, $SD = 4$ months). Participants were randomly assigned to a music program ($n = 13$, six girls), a phonological skills program ($n = 14$, six girls), and a control group that received sports training ($n = 14$, seven girls). All three participating kindergartens had a music group, a phonological skills group, and a control group to rule out kindergarten effects.

MATERIAL

Programs

Children were trained for 10 min on a daily basis for a period of 20 weeks. Thus, participants received a total of 100 sessions. Preschoolers were trained in groups of five to seven children. The programs were implemented by trained research assistants. Each assistant implemented every program (phonological skills, music, sports). The preschoolers were trained at the kindergarten in a quiet room that offered sufficient space for the different tasks. A typical session comprised a short welcome (small talk, attendance check) and two different tasks that were approximately 5 minutes in duration.

The music program was created by the authors. It was based on a well-established program for early music education (Nykrin et al., 2007) and contained joint singing, joint drumming, rhythmic exercises, meter execution, training of rudimentary notation skills, dancing, and playful familiarization with intervals. Typical sessions comprised, for example, the learning of a new song and a listening exercise. At first, all children listened to the song. Then, they attempted to sing along with the CD. Finally, the group sang the song without the CD. The second task involved listening to music recordings and subsequently identifying the tempo or musical instruments on the recording. Other typical sessions involved familiarization with different instruments and joint drumming.

The children were given the opportunity to test several instruments that were scattered across the floor on little “islands” and when indicated, they switched to the next “instrument isle.” Joint drumming activities involved synchronization with a given beat as well as the creation of new beats by the children that had to be learned by the other group members. Another typical session comprised dancing to certain musical themes or synchronization of particular body parts to the music. A second task was joint singing of familiar songs and drumming while singing.

The phonological skills program was a well-established program that was specifically designed to train phonological awareness (Küspert and Schneider, 2003). It contained listening exercises, rhyming exercises, phoneme recognition exercises, syllable exercises, and the introduction of the concepts “word” and “sentence.” Typical early sessions contained listening tasks that involved everyday sounds and rhyming with animal names. Children closed their eyes and guessed the sounds that the trainer produced or chose words that rhymed with an animal name such as cat. Other commonly used tasks were clapping the syllables of the animal, plant, or object displayed on a picture card as well as guessing the last phoneme of a word given by the trainer. Typical sessions at the end of the trainings phase included more difficult tasks such as the identification of words in sentences.

The sports training was created by the authors and contained exercises to train balance, physical strength, endurance, coordination, fine-motor abilities, body perception, and relaxation. Typical sports sessions contained relaxation exercises and well-directed throwing. For relaxation training, children imagined that they were an air mattress losing all air, and the throwing exercises included throwing balls into boxes and buckets of different sizes. Another typical session involved playing football in atypical body positions such as walking on their hands and feet with their belly up (crayfish football) and mastering a path with closed eyes while relying on the instruction of a teammate. Tug-of-war games and balancing objects on different body parts were also typical.

Measures

Control variables such as age, gender, intelligence, and socioeconomic status as well as phonological awareness were assessed.

Using a questionnaire designed by the authors, details about parents’ education was assessed as a measure of socioeconomic status. For statistical purposes, mothers’ and fathers’ education were coded as a dichotomous variable, with 0 for “no university education” and 1 for “university education,” and a composite variable (i.e., the average of the two scores) was used in the analyses.

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 eight 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.

Phonological awareness was assessed with the following four subtests of the Bielefelder Screening (BISC; Jansen et al., 2002): (a) detection of rhymes, (b) synthesis of phonemes into words, (c) segmentation of words into syllables, and (d) phoneme recognition in words. Each subtest consisted of two to four practice items and

10 test items. In the detection of rhymes task, children were asked whether two words rhyme or do not rhyme (e.g., Do train and rain rhyme?). The synthesis of word task requested the synthesis of the initial sound and the remaining word (e.g., mouse) into one word. Children were asked to segment words by clapping their hands in the segmentation into syllables task. 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. In each subtest, a maximum of 10 points was possible, resulting in a maximum composite score of 40 points. In addition, the following two composite scores were calculated: a phonological awareness score for large (words) phonological units consisting of subtest (a) and (c), and a phonological awareness score for small (phonemes) phonological units consisting of subtest (b) and (d).

PROCEDURE

Prior to the pretest, the informed consent of the parents was attained and the demographic questionnaire was sent to the participants. A pretest that assessed intelligence and phonological awareness was conducted. The intelligence test was performed in groups of five to six children, whereas phonological awareness was assessed in individual sessions, and the assessments were performed on consecutive days. The pretest was followed by 20 weeks of training. Subsequent to training, the post-test was conducted. The post-test involved the individual assessment of phonological awareness. The same test that was utilized in the pretest was applied. At the end of the project, each child received a present and a certificate for participation.

RESULTS

CONTROL VARIABLES

Differences in gender, age, intelligence, and socioeconomic status between the music group, the phonological skills group, and the control group were controlled. The ratio of male to female was not significantly different between the three groups, $\chi^2(2, n = 41) = 0.14, p = 0.93$. Concerning mean age, the children in all three groups were comparable, $F(2, 38) = 1.04, p = 0.36$, see **Table 1** for mean and SD. With respect to intelligence, no significant differences between the music group, the phonological skills group, and the control group were revealed, $F(2, 38) = 0.16, p = 0.85$, see **Table 1** for details.

Children in all three groups were comparable in socioeconomic status, $\chi^2(4, n = 37) = 2.68, p = 0.61$, see **Table 2** for details. Four participating families did not provide details about parents’ education. Taken together, the analyses of the control variables indicated that the three groups can be considered as equal in gender ratio, mean age, mean IQ, and socioeconomic status.

Table 1 | Mean and SD for the control variables age and intelligence.

Control variable	Music program <i>M</i> (SD)	Phonological skills program <i>M</i> (SD)	Control group <i>M</i> (SD)
Age (in months)	70.46 (3.9)	69.14 (4.2)	68.21 (4)
Intelligence	113.92 (12.5)	111.64 (10.3)	113.6 (11.9)

PHONOLOGICAL AWARENESS

At the pretest, the music group, phonological skills group, and control group did not significantly differ in phonological awareness, $F(2, 38) = 0.16$, $p = 0.86$. Furthermore, the compositional scores for large and small phonological units revealed no group differences (all $ps > 0.6$).

Phonological awareness scores were entered into a 3 (group: music, phonological skills, control) \times 2 (condition: pretest vs. post-test) ANOVA with repeated measures on the last factor. This analysis revealed a significant main effect of condition, $F(1, 38) = 16.12$, $p < 0.001$. Phonological awareness improved significantly from pretest to post-test. No significant main effect for group was revealed ($p > 0.5$). Furthermore, the analysis revealed a significant group \times condition interaction, $F(2, 38) = 3.24$, $p = 0.05$, see **Figure 1**. Therefore, two ANOVAs comparing each treatment group to the control group and one ANOVA comparing the two treatment groups were calculated.

Table 2 | Distribution of parents' education as a measure of socioeconomic status within the music group, the phonological skills group, and the control group.

	Music program	Phonological skills program	Control group
No. parent	7	5	5
At least one parent	4	7	9

No. parent = no. parent holds a university degree. At least one parent = at least one parent holds a university degree. Four participating families did not provide details about parents' education.

For the comparison between the phonological skills group and the control group, phonological awareness scores were entered into a 2 (group: phonological skills, control) \times 2 (condition: pretest vs. post-test) ANOVA with repeated measures on the last factor. This analysis revealed a significant group \times condition interaction, $F(1, 26) = 4.84$, $p = 0.04$. For the comparison between the music group and the control group, phonological awareness scores were entered into a 2 (group: music, control) \times 2 (condition: pretest vs. post-test) ANOVA with repeated measures on the last factor. This analysis also revealed a significant group \times condition interaction, $F(1, 25) = 7.91$, $p = 0.009$.

The comparison of the two treatment groups revealed no significant group \times condition interaction, $F(1, 25) = 0.06$, $p = 0.81$. Thus, in both treatment groups, the phonological awareness score improved significantly from pre- to post-test, whereas such an improvement was not found in the control group.

Subsequently, differences in phonological awareness were analyzed in greater detail. The groups' phonological awareness on the phoneme level (small phonological units) and on the word level (larger phonological units) was compared.

For the small phonological units, a 3 (group: music, phonological skills, control) \times 2 (condition: pretest vs. posttest) ANOVA with repeated measures on the last factor revealed a significant main effect of condition, $F(1, 38) = 25.39$, $p < 0.001$. No significant interaction or main effect for group was revealed (all $ps > 0.3$). Similarly, in all three groups, phonological awareness of small phonological units improved significantly from pretest to post-test.

For large phonological units, a 3 (group: music, phonological skills, control) \times 2 (condition: pretest vs. posttest) ANOVA with repeated measures on the last factor revealed a marginal significant main effect of condition, $F(1, 38) = 3.39$, $p = 0.073$.

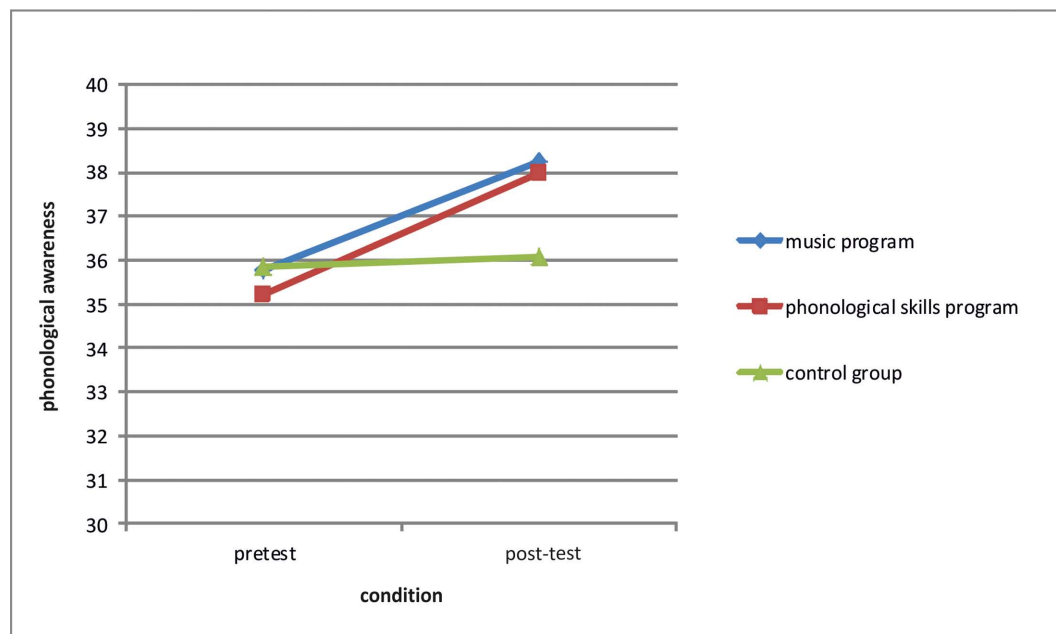


FIGURE 1 | Mean phonological awareness score at pre- and post-test in the music group, the phonological skills group, and the control group.

No significant main effect of group was found ($p > 0.7$). Furthermore, the analysis revealed a significant group \times condition interaction, $F(2, 38) = 4.02$, $p = 0.026$, see **Figure 2**. The individual comparison of each treatment group to the control group yielded significant interactions between group and condition. A 2 (group: phonological skills, control) \times 2 (condition: pretest vs. posttest) ANOVA with repeated measures on the last factor revealed a significant group \times condition interaction, $F(1, 26) = 4.61$, $p = 0.041$, for the comparison between the phonological skills group and the control group. Similarly, for the comparison between the music group and the control group, a 2 (group: music, control) \times 2 (condition: pretest vs. posttest) ANOVA with repeated measures on the last factor revealed a significant group \times condition interaction, $F(1, 25) = 10.68$, $p = 0.003$. Subsequent analyses (2 \times 2 ANOVAs with repeated measures on the last factor for the treatment groups) showed that both treatment groups significantly increased in their phonological awareness of large phonological units from pretest to post-test, $F(1, 25) = 7.17$, $p = 0.009$. No significant main effect for group and no significant group \times condition interaction was revealed ($ps > 0.6$). Thus, in both treatment groups, the phonological awareness of large phonological units score improved significantly from pre- to post-test, whereas such an improvement was not revealed in the control group.

DISCUSSION

The aim of the current experiment was to investigate the effect of a music program on phonological awareness in preschoolers. Additionally, this effect was compared to the effect of a phonological skills program, which was specifically developed to enhance phonological skills.

To this end, preschoolers were randomly assigned to a music program, a phonological skills program, and a control group that received sports training (from which no effect was expected). Participants were trained 10 min daily for 20 weeks. Prior to training, participants were administered a pretest on phonological awareness. Following training, the participants were retested on phonological awareness. At the pretest, no significant differences were found between the treatment groups and the control group with respect to possible confounding variables (gender, age, intelligence, and socioeconomic status). This supported the assumption that the groups did not systematically differ due to random assignment and ruled out the most likely alternative explanations of any observed trainings effect. Additionally, at the pretest, no differences in phonological awareness (overall score, small phonological units, large phonological units) were revealed. Thus, random assignment resulted in three groups that did not systematically differ on the tested variables prior to training.

A positive effect of the music program and the phonological skills program was found on phonological awareness. Preschoolers who participated in the music program or the phonological skills program significantly increased in phonological awareness, whereas such an increase was not found in the control group. Specifically, the treatment and control groups showed differential development in phonological awareness of large phonological units. Both treatment groups improved in phonological awareness of large phonological units, whereas such an improvement was not found in the control group. All three groups showed a similar development in phonological awareness of small phonological units. Thus, the experiment provides evidence of an enhancing effect of the music program on phonological awareness, in particular on phonological awareness of large phonological units.

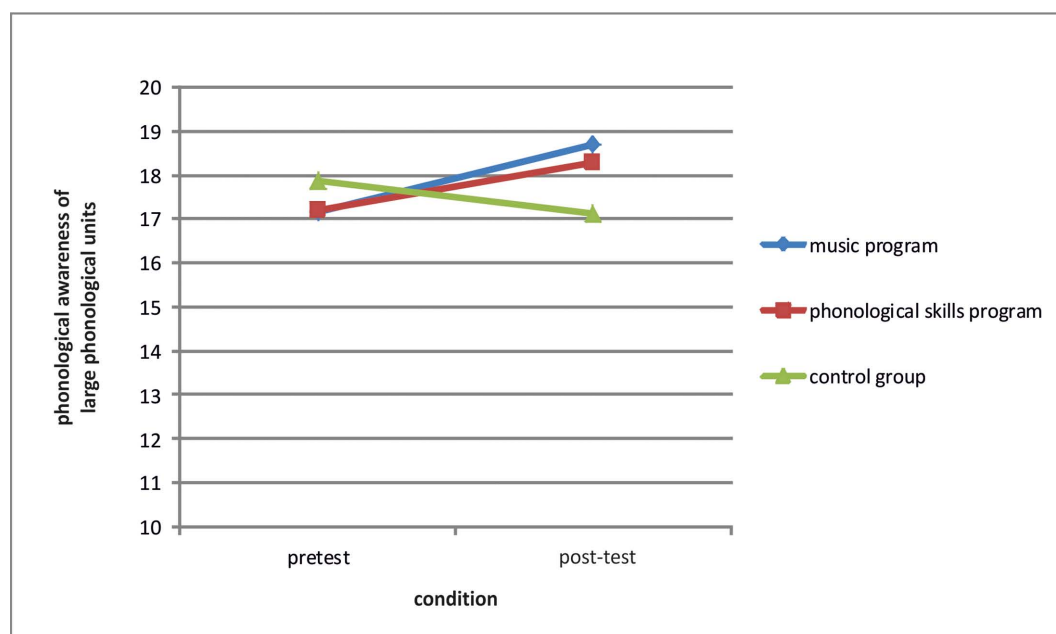


FIGURE 2 | Mean phonological awareness of large phonological units score at pre- and post-test in the music group, the phonological skills group, and the control group.

This result is in accordance with the previous findings of Anvari et al. (2002) as well as Lamb and Gregory (1993), who demonstrated an association between musical aptitude and phonological awareness in 4- and 5-year-old children. Additionally, the present results are in line with the findings of Bolduc (2009), but emphasize the importance of a control group. In the study by Bolduc (2009), it is highly likely that both music programs increased phonological awareness. However, because there was no control group (receiving a non-musical treatment or no treatment), it is impossible to infer the degree of increase in phonological awareness caused by the less effective music program (that was designed particularly to enhance musical abilities). Moreover, the current experiment is in agreement with the results of Gromko (2005), who demonstrated an association between music training and phonological awareness. The present results extended these findings by establishing the specific causation. Because an experimental design was applied, the current results can be interpreted unequivocally in terms of an enhancing effect of the music training program on phonological awareness in preschoolers. Moreover, these results indicate that phonological awareness of large phonological units was more intensely affected by music training and phonological skills training. Thus, the overall enhancement was basically driven by the subtests that assessed phonological awareness of large phonological units. For phonological awareness of small phonological units, the treatment groups and the control group showed a similar development (an increase) over the course of the experiment. This likely represents maturation and/or effects of preschool activities in kindergartens. As preschool activities and maturation affect all preschoolers regardless of the training groups, it seems obvious that all children increased in phonological awareness of small phonological units. Additionally, the most pronounced advancement in phonological awareness of small phonological units occurs when the mapping of graphemes onto phonemes is fostered, which typically takes place in first grade (Marx, 2007). Neither program (music program or phonological skills program) focused on phoneme grapheme mapping, which might partly explain why the three groups developed similarly in phonological awareness of small phonological units. However, the phonological skills addressed small phonological units. Thus, an effect of the phonological skills program might have been expected. Nevertheless, no differences were observed between the groups in phonological awareness of small phonological units. In contrast to the similar development in phonological awareness of small phonological units, the development of large phonological units showed a differential development due to phonological skills and music training. The positive effect of the phonological skills program can be attributed to the extensive training of rhyming, segmenting, and blending. The children in this group explicitly practiced these abilities. The effects of the music training program on large phonological units represent indirect training effects. It is possible that the rhythmical exercises and the combination of rhythm and song lyrics in the children's play songs indirectly trained the ability to rhyme, segment, and blend.

The size of the positive effect of the music program ($d_{\text{corr}} = 0.9$) on phonological awareness of large phonological units was comparable to that of the phonological skills program ($d_{\text{corr}} = 0.6$) on phonological awareness large phonological units. This effect size was similar to effect sizes found in former studies of this phonological skills

program (Schneider et al., 1997). The current results suggest that the music and the phonological skills programs resulted in medium to large effect sizes. Considering that the phonological skills program was explicitly designed to train phonological awareness, it is notable that the music program resulted in a similar effect size. This similarity adds evidence to the "shared sound category learning mechanism hypothesis" for language and music. During the preschool years, the mechanism for sound category learning in music and language seem to partly overlap. Thus, the proposed similarity in neuronal processing of music and language at an early age (McMullen and Saffran, 2004; Koelsch and Siebel, 2005) was supported by the present findings with respect to sound category learning. Future research should test whether the revealed influence occurs in the opposite direction. This would entail investigating effects of phonological awareness programs on pitch discrimination, for example.

A limitation of the experiment is the sample size. The sample was quite small due to the time, effort, and cost-intensive nature of the experiment as well as the initial matching. Future research should attempt to establish causation with a larger sample size. Another important limitation was that no explicit measure of motivation was employed. It is possible that children enjoyed a particular program more than another or that parents influenced children's motivation to participate. Since all activities (phonological skills program, music program, sports training) were designed for preschoolers, it is highly likely that the children enjoyed all programs equally. Nevertheless, this was not explicitly assessed. With respect to parental influence, parents were not blind to the goal (train phonological awareness) but were blind to their child's particular trainings group. However, it is possible that the child reported about the activity in the session and the parents learned their child's group placement. Nevertheless, it remains unclear whether parents practiced phonological awareness at home with the children or attempted to increase children's motivation to participate. Moreover, it is equally likely that parents of the sports training group practiced phonological awareness to prevent deficits or parents of the treatment groups to promote the child's abilities. These issues should be addressed in future research.

Considering the importance of phonological awareness in the emergence of literacy, the results point toward music programs as valuable preparation for learning to read and write. Hence, a music program could either be used as an alternative approach to train phonological awareness or used to complement existing phonological awareness programs. For a successful combination of music programs and existing phonological awareness programs, it may be important to identify the specific musical components of the music program that contributed to the advancement of phonological awareness. Then, the most important components of both training approaches could be combined.

CONCLUSION

The current results indicate that a music program can enhance phonological awareness, in particular phonological awareness of large phonological units (e.g., rhyming, segmenting, and blending). This enhancement is comparable to the effects of a phonological skills program on phonological awareness. Therefore, it seems highly likely that language and music share processing mechanisms, explicitly sound category learning mechanisms.

ACKNOWLEDGMENTS

This research was supported by a grant for educational research (TP9 6050085) from the Justus-Liebig-University Giessen and the graduate school “NeuroAct” (neuronal rep-

resentation and action control) of the University of Gießen and the University of Marburg. 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.
- Bolduc, J. (2009). Effects of a music programme on kindergartners’ phonological awareness skills. *Int. J. Music Educ.* 27, 37–47.
- Bruck, M. (1992). Persistence of dyslexics’ phonological awareness deficits. *Dev. Psychol.* 28, 874–886.
- Fernald, A. (1991). Prosody in speech to children: prelinguistic and linguistic functions. *Ann. Child Dev.* 8, 43–80.
- Gromko, J. E. (2005). The effect of music instruction on phonemic awareness in beginning readers. *J. Res. Music Educ.* 53, 199–209.
- Jansen, H., Mannhaupt, G., Marx, H., and Skowronek, H. (2002). *Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten*. Göttingen: Hogrefe.
- Koelsch, S., and Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends Cogn. Sci. (Regul. Ed.)* 9, 578–584.
- Küspert, P., and Schneider, W. (2003). *Hören, lauschen, lernen Sprachspiele für Kinder im Vorschulalter*, 4 Auflage. Göttingen: Vandenhoeck & Ruprecht.
- Lamb, S. J., and Gregory, A. H. (1993). The relationship between music and reading in beginning readers. *Educ. Psychol.* 13, 19–27.
- Marx, P. (2007). *Lese- und Rechtschreiberwerb*. Paderborn: Schöningh UTB.
- Masataka, N. (1999). Preference for infant-directed singing in 2-day-old hearing infants of deaf parents. *Dev. Psychol.* 35, 1001–1005.
- McMullen, E., and Saffran, J. R. (2004). Music and language: a developmental comparison. *Music Percept.* 21, 289–311.
- Nykrin, R., Grüner, M., and Widmer, M. (eds). (2007). *Musik und Tanz für Kinder*. Mainz: Schott.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681.
- Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford: University Press.
- Pratt, A. C., and Brady, S. (1988). Relation of phonological awareness to reading disability in children and adults. *J. Educ. Psychol.* 80, 319–323.
- Schneider, W., Küspert, P., Roth, E., Vise, M., and Marx, H. (1997). Short- and long-term effects of training phonological awareness in kindergarten: evidence from two German studies. *J. Exp. Child Psychol.* 66, 311–340.
- Trainor, L. J. (1996). Infant preferences for infant-directed versus non-infant-directed play songs and lullabies. *Infant Behav. Dev.* 19, 83–92.
- 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.

Received: 03 March 2011; accepted: 29 May 2011; published online: 20 June 2011.

Citation: Degé F and Schwarzer G (2011) The effect of a music program on phonological awareness in preschoolers. *Front. Psychology* 2:124. doi: 10.3389/fpsyg.2011.00124

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

Copyright © 2011 Degé and Schwarzer. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change

Ryan J. Giuliano¹, Peter Q. Pfordresher², Emily M. Stanley³, Shalini Narayana⁴, and Nicole Y. Y. Wicha^{4,5*}

¹ Department of Psychology, University of Oregon, Eugene, OR, USA

² Department of Psychology, University at Buffalo, State University of New York, Buffalo, NY, USA

³ Department of Psychology, University of Delaware, Newark, DE, USA

⁴ Research Imaging Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

⁵ Department of Biology and Neuroscience Institute, University of Texas at San Antonio, San Antonio, TX, USA

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Martin Meyer, University of Zurich, Switzerland

Psyche Loui, Beth Israel Deaconess Medical Center/Harvard Medical School, USA

Stefan Andreas Elmer, University of Zurich, Zurich, Switzerland

*Correspondence:

Nicole Y. Y. Wicha, Department of Biology – University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, USA.
e-mail: Nicole.Wicha@UTSA.edu

Native tone language experience has been linked with alterations in the production and perception of pitch in language, as well as with the brain response to linguistic and non-linguistic tones. Here we use two experiments to address whether these changes apply to the discrimination of simple pitch changes and pitch intervals. Event related potentials (ERPs) were recorded from native Mandarin speakers and a control group during a same/different task with pairs of pure tones differing only in pitch height, and with pure tone pairs differing only in interval distance. Behaviorally, Mandarin speakers were more accurate than controls at detecting both pitch and interval changes, showing a sensitivity to small pitch changes and interval distances that was absent in the control group. Converging evidence from ERPs obtained during the same tasks revealed an earlier response to change relative to no-change trials in Mandarin speakers, as well as earlier differentiation of trials by change direction relative to controls. These findings illustrate the cross-domain influence of language experience on the perception of pitch, suggesting that the native use of tonal pitch contours in language leads to a general enhancement in the acuity of pitch representations.

Keywords: pitch discrimination, pitch interval discrimination, event related potentials, tone language, auditory perception, auditory cognitive neuroscience, mandarin chinese, neural plasticity

INTRODUCTION

There is increasing evidence that native experience with a tone language changes the perceptual and categorization abilities of listeners for both linguistic and non-linguistic stimuli (i.e., Deutsch, et al., 2006; Krishnan et al., 2009). For example, speakers of Mandarin, where the lexical meaning of a word depends on the overall pitch height as well as change in pitch (melodic contour) in production, are better at discriminating pitch intervals compared to native speakers of English (Pfordresher and Brown, 2009; Hove et al., 2010), a language that uses tone mainly to convey supra-segmental or prosodic information concerning stress or sentence structure (Cruttenden, 1997; Yip, 2002; Gussenhoven, 2004). It is unclear however, whether corresponding temporal dynamics of the neural processes underlying pitch perception actually facilitate this advantage, and whether or not this advantage for tone language speakers transfers to simple non-linguistic stimuli, such as pure tones. With this goal, the current study compares native speakers of a tonal language to speakers of an intonation language with no tone language experience, and measures both accuracy and continuous electrophysiological data on two auditory discrimination tasks.

Languages are classified as either “tonal” or “intonation” based on whether they use pitch to convey word meaning (Yip, 2002). For example, in Mandarin, a Chinese tone language, the word *ma* can mean “mother,” “hemp,” “horse,” or “scold” based on whether a level, rising, falling, or bi-directional (falling then rising) tone is

used, respectively. Acquiring the ability to speak a tone language demands the establishment of fine-grained associations between pitch contours and word meaning, and this appears to lead to general enhancements in pitch processing (Pfordresher and Brown, 2009) and AP abilities (Deutsch, et al., 2009; Deutsch et al., 2006). The use of pitch during language acquisition might exclusively influence pitch processing of speech stimuli, consistent with the view that speech is a specialized, modular ability (Liberman et al., 1967; Liberman and Mattingly, 1985, 1989; Liberman and Whalen, 2000). Alternately, if pitch contrasts in speech are processed by general auditory mechanisms (Bregman et al., 1990), then tone speakers may demonstrate enhanced pitch processing abilities for non-linguistic stimuli, such as simple pure tones.

Recent studies have reported several forms of evidence for this advantage (though see also Burns and Sampat, 1980; Stagray and Downs, 1993; Lee et al., 1996 which have not observed a tone language advantage for the processing of non-speech tones). Higher accuracy has been found among native speakers of an Asian tone language than intonation speakers for discriminating pitch intervals (Pfordresher and Brown, 2009; Hove et al., 2010). Absolute pitch (AP) is more prevalent among Chinese than American music conservatory students (Deutsch et al., 2006). AP abilities have been found to be more accurate in individuals who are very fluent in a tone language, compared to those who are moderately or not at all fluent (Deutsch et al., 2009). Speakers of Japanese (a pitch accented

language) have been found to be better at detecting mistuning of familiar songs than are English speakers (Trehub et al., 2008; but see Schellenberg and Trehub, 2008). Similarly, the reverse relationship has also been documented, where Mandarin speakers with deficits in processing musical pitch are often also worse at identifying and discriminating Mandarin tones than are other speakers (Nan et al., 2010). If indeed tone language experience facilitates the processing of pitch, then additional questions are raised about the specificity of the underlying changes to neural systems implicated in pitch processing, such as whether these changes are limited to the processing of pitch contours similar to the tonal contours in one's native language, or whether these changes apply generally to processes recruited when discriminating between non-linguistic tones differing only in pitch.

In addition to behavioral evidence of a tone language advantage, several studies have demonstrated that native tone language experience leads to changes in the pre-attentive subcortical processing of pitch contours, occurring as early as the auditory brainstem-evoked response. Using the scalp-recorded human frequency following response (FFR), a measure of phase-locked activity in the rostral brainstem, researchers have observed stronger pitch representations, and smoother pitch tracking of Mandarin tonal contours in native speakers of Mandarin Chinese than in English controls (Krishnan et al., 2005, 2009; Swaminathan et al., 2008). Furthermore, this enhanced pitch representation at the brainstem level appears to be driven by pitch information in the auditory stream, specifically for certain dimensions of pitch contours that native tone speakers are familiar with, and not specific to speech (Swaminathan et al., 2008; Krishnan et al., 2009). Musical training may also enhance the extraction of pitch information from Mandarin tones in the brainstem, even among individuals with no knowledge of Mandarin (Wong et al., 2007). These findings suggest that native tone-speaking experience involves plasticity-related changes that optimize which dimensions of the auditory signal are captured at the brainstem and subsequently passed on to pitch-relevant processes at the cortical level, and that these pitch processing advantages may be shared across linguistic and non-linguistic domains.

Although the present study addresses the effect of native tone language experience on the active processing of pitch during a discrimination task, others have shown tone language effects on pre-attentive processing during passive listening tasks. Chandrasekaran et al. (2007a,b) have observed that native tone speakers demonstrate a larger mismatch negativity (MMN) to deviant non-speech homologues of Mandarin pitch contours than English-speaking controls. Even when the deviant non-speech tone was a pitch contour not found in any language, tone language speakers elicited a larger MMN than non-tone language speaking musicians and non-musicians, suggesting that native tone-speaking experience enhances the pre-attentive neural representation of pitch-relevant information (Chandrasekaran et al., 2009). Kaan et al. (2007, 2008) observed an MMN to deviant Thai lexical tones in native speakers of Thai, Mandarin, and English, suggesting all are able to pre-attentively discriminate among tones. Further, they found that perceptual identification training with the tones improved performance on a behavioral discrimination task and decreased MMN amplitude and latency for all groups. Similar MMN latency decrements with

pitch training had been reported previously, although they were conversely observed in conjunction with increases in MMN amplitude (Kraus, et al., 1995; Tremblay et al., 1997). Overall, these findings provide evidence of short- and long-term plasticity-related changes in the auditory system and suggest that native tone language experience will facilitate the timing and magnitude of the brain's response to sound input.

In the few event related potential (ERP) studies that have examined the effects of pitch change more generally, modulations of both early and late ERP components effects have been observed in relation to pitch discrimination. Tervaniemi et al. (2005) found that both early (MMN, N2b) and late ERP components (P300) increased in amplitude to deviant tones in an oddball paradigm with increasing amount of pitch change. Similarly, Peretz et al. (2005) observed increases in N1 amplitude with the increasing amount of pitch change of a deviant tone within a 5-tone sequence. Perhaps most relevant to the current study, training-related increases in P2 and N1c amplitude have been observed during an active-listening task similar to the one employed here, where a standard auditory tone was followed by a second tone of same or differing pitch to which participants must respond indicating whether the two tones were the same or different (Bosnyak et al., 2004). Given that early auditory ERP components are sensitive to modulations in pitch, we predict that a pitch processing advantage for native tone speakers will be reflected in ERPs at early stages of processing (e.g., N1–P2 complex), indexing facilitated perceptual processing.

Herein we present two experiments using ERPs to capture the temporal dynamics of the neural response to pure tones during a simple pitch discrimination task (Experiment 1) and a more complex pitch interval discrimination task (Experiment 2). For both experiments, we first report differences in behavioral performance between native tone language speakers and a control group, followed by a detailed summary of the ERPs of each group for the effect of pitch change and change direction.

EXPERIMENT 1

Experiment 1 follows from Pfordresher and Brown (2009), who found a nominal behavioral advantage in accuracy on a simple pitch discrimination task for tone vs. non-tone speakers. The range of pitch trials used tended toward easily discriminable tones, which is exemplified by the near-perfect performance of several participants in that study and a previous study using an identical task (Pfordresher and Brown, 2007). In those studies (Pfordresher and Brown, 2007, 2009), contrast trials deviated from the base pitch (C5) by 25, 50, 100, 200, 400, 600, and 800 cents, where performance reached high accuracy around 100 cents. Here we used a more challenging range of pitch changes: 7, 13, 25, 50, 100, and 200 cents. It should be noted that accurate performance on this task depends solely on the ability to extract pitch height information from the flat, non-contoured pure tones. Although Asian tone languages undoubtedly place heavy emphasis on pitch contour, research suggests that pitch height is still relevant to tone judgments in Mandarin, as pitch height information is believed to be evaluated separately from, and later integrated with, information about pitch contour (Massaro et al., 1985). Considering that pitch height information is crucial to accurate performance on our task,

we hypothesize that tone language speakers will be better able to discriminate changes in pitch than speakers with no tone language experience, particularly on difficult trials (smaller cent change). In addition to the behavioral advantage, we expect that the ERPs of tone language speakers will reflect enhanced perceptual processing in early sensory components like the N1–P2 complex, and possibly differences at later stages of processing, such as the P300.

METHODS

Participants

Thirty-three right-handed adults^{1,2} grouped as native tone language speakers and a control group were paid for their participation. The Tone group consisted of 17 native speakers of Mandarin Chinese (9 females; mean age = 26.4 years), who were born and raised in China for an average of 22.94 years (range 16–30). They reported using Mandarin for an average of 64% of the day (range 5–90%), all resided in the USA for at least 1 year, and were proficient in English, with an average of 14.71 years of experience in English (range 6–27 years). All were also fluent in at least one other language, including another tone language (i.e., Cantonese or Taiwanese) for 10 of the Tone speakers. The control group consisted of 16 individuals (7 females; mean age = 28.3 years) of various ethnic backgrounds who had no exposure to a tone language. Importantly, the Tone and control groups were not significantly different on years of musical training (tone, $M = 2.30$, control, $M = 1.44$; $p = 0.27$), years of playing an instrument or singing (tone, $M = 3.41$, control, $M = 1.63$; $p = 0.11$), hours listening to music per week (tone, $M = 12.24$, control, $M = 10.44$; $p = 0.60$), and live concerts attended per year (tone, $M = 2.53$, control, $M = 3.19$; $p = 0.60$). No participants in either group reported any private musical training outside of school activities. Additionally, to control for the number of languages spoken as compared to the bilingual Tone speakers, all members of the control group were bilingual in another non-tone language, such that the groups had similar years of experience with a non-native language (tone, $M = 26.35$ years, control, $M = 26.81$ years; $p = 0.80$). Tests of hearing thresholds with an audiometer confirmed that all participants could detect low intensity tones (<35 db) within the frequency range used for testing (500–6000 Hz)³.

¹Fifty-six individuals took part in this study, however in order to closely match the groups, several participants were excluded from data analysis. Four participants were outliers in age (>40 years old), 3 were native speakers of Vietnamese, and 16 others were excluded for an insufficient number of artifact-free trials with correct responses (13), not completing the session (2), or failing the hearing test (1). Two participants in each group did not perform the hearing test due to a technical error, however their performance on the discrimination task was less than one standard deviation from the appropriate group mean, indicating that they had no trouble hearing the stimuli. Note that there is currently no evidence suggesting that hearing sensitivity among normal-hearing populations influences the ability to discriminate pitches (Ayotte et al., 2002).

²Seven people from the control group took part in a parallel fMRI version of this study prior to the ERP session. The time between sessions was on average 20 days. Note that practice effects would be expected to improve performance, yet the control group still underperformed the Tone group in accurately detecting change trials.

³Although years of experience is an incomplete way of characterizing acquired musical ability, a lack of a more valid assessment tool for musical experience has lead researchers in music cognition to routinely use years of training as an indicator of musician/non-musician status (e.g., Bigand and Poulin-Charronnat, 2006).

Stimuli

The stimuli were pure tones created using Matlab software (The Mathworks, Natick, MA, USA) with custom-made scripts, and were presented using E-Prime software (Psychological Software Tools, Inc., Pittsburgh, PA, USA). Tone amplitudes were shaped by an exponential function that resulted in 200 ms attack and 50 ms decay times. The first tone in each pair was always 524 Hz (C5), and was followed by either the identical sound (240 trials) or a tone that was higher (120 trials) or lower (120 trials) in pitch, yielding equiprobability of change and no-change trials. (see Pfordresher and Brown, 2007, 2009). The pitch difference between the first and second note could be 7, 13, 25, 50, 100, or 200 cents (40 trials each), corresponding to approximate frequency differences of 2, 4, 8, 15, 30, and 61 Hz respectively. Like Pfordresher and Brown, we aimed for a geometric series based on subdividing 400 cents into equal intervals (rounding to the nearest cent) to include critical intervals like the semitone and quartertone (50 cents). The order of trials was pseudo-randomized, such that no more than two trials in a row were presented with the same amount or direction of change, with a fixed order presented to all participants.

Procedure

After written consent, participants completed a series of questionnaires requesting information such as demographics, hearing and sight conditions, handedness, language history, and musical experience. For the ERP session, participants sat alone in a sound attenuating, electrically shielded recording chamber and wearing ER-1 insert earphones (Etymotic Research, Inc., Elk Grove Village, IL, USA) to listen to the sound stimuli. Instructions were presented on a computer monitor 4.5 feet from the participant. An experimenter monitored the continuous recording of EEG and a video monitor of the participant throughout the experiment. During a single session, participants performed alternating blocks of pitch discrimination trials and interval discrimination trials (see Experiment 2). After a set of 8 practice trials, participants were presented with 480 pairs of tones across 12 blocks of the pitch discrimination task, interspersed with 16 blocks of interval discrimination trials. During the pitch discrimination trials, participants heard two tones lasting 500 ms each, separated by 500 ms of silence (see **Figure 1**). They judged whether the two tones were the same or different. Participants were asked to fixate their vision on a half-inch white cross on a black screen, which appeared 500 ms before the first tone was heard to orient the participant, and remained on the screen until 500 ms after the end of the last tone to reduce eye movement artifacts in the EEG. This was followed by three fixation crosses that remained on the screen for 3000 ms or until the participant responded (cued delayed-response). Responses before or after the appearance of these three crosses were registered as a null response. Participants responded by pushing one button for “same” and another for “different” with their right hand on a PST serial response box (Psychological Software Tools, Inc., Pittsburgh, PA, USA). At the end of each trial, the screen went blank for 500 ms before the onset of the next trial.

Between each block, participants read a few sentences of an entertaining children’s story (king grisly beard by the Brothers Grimm) at their own pace, to prevent fatigue from the

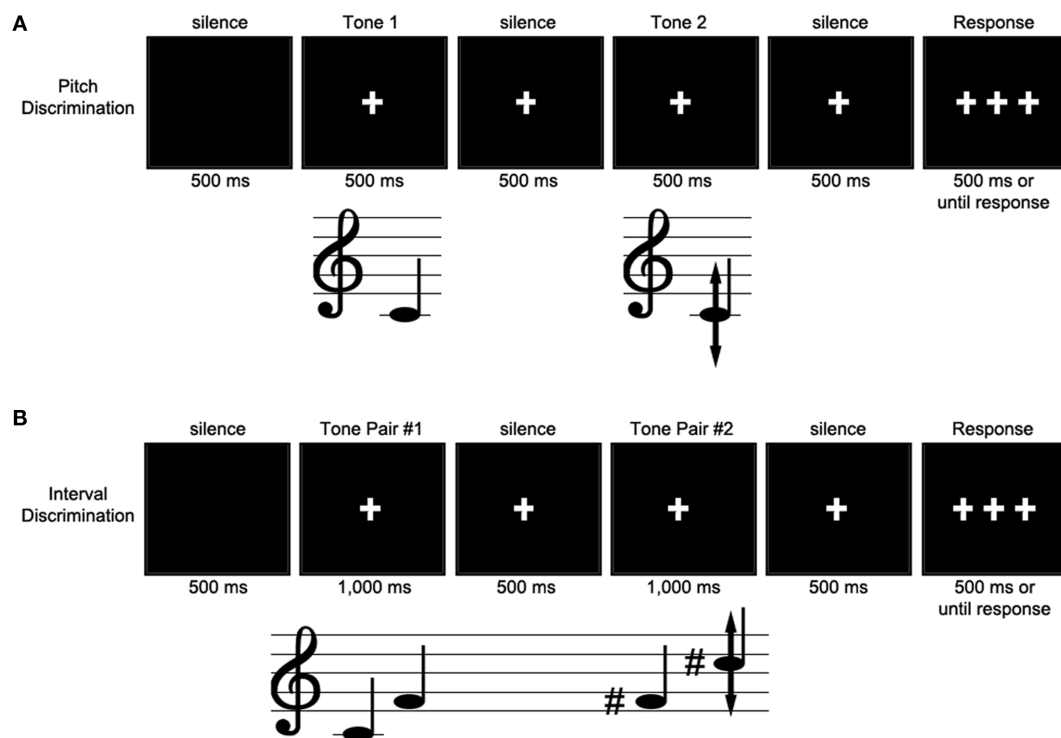


FIGURE 1 | Example trial for Experiments 1 (A) and 2 (B). During the pitch discrimination task, a standard tone was presented, followed 500 ms later by a tone of identical, lower, or higher pitch. Participants indicated their response (same or different) by button press when the three fixation crosses appeared on the screen. ERPs were time-locked to the onset of the second tone.

discrimination task. Participants were also given several breaks. The entire session lasted approximately 2.5 h, with 1.5 h of EEG recording.

ERP recording parameters

Continuous scalp-recorded EEG was acquired using a geodesic array of 26 pre-amplified sintered Ag–AgCl electrodes (BioSemi active electrodes) embedded in a custom electrode cap (Electro-Cap International Inc., using BioSemi electrode holders; see Giuliano and Wicha (2010). Additional electrodes were placed below and at the outer canthi of the left and right eyes to record blinks and eye movement, respectively, and on the left and right mastoid processes to serve as offline reference. Preamplifiers in each electrode were used to reduce induced noise between the electrode and the amplification/digitization system (BioSemi ActiveTwo, BioSemi B.V. Amsterdam), allowing high electrode impedances. Electrode offsets were kept below 35 mV. A first-order analog anti-aliasing filter with a half-power cutoff at 3.6 kHz was applied (see www.biosemi.com). The data were sampled at 512 Hz (2048 Hz with a decimation factor of 1/4) with a bandwidth of DC to 134 Hz, using a fifth order digital sinc filter. Each active electrode was measured online with respect to a common mode sense (CMS) active electrode producing a monopolar (non-differential) channel, and was referenced offline to the average of the left and right mastoids. Data were processed using BrainVision Analyzer 2 (Brain Products GmbH, Munich, Germany). Non-causal Butterworth digital filters were applied with a low cutoff

at 0.1 Hz (12 db/oct) and high cutoff at 30.0 Hz (12 db/oct). The EEG data were segmented in intervals of 1000 ms time-locked to stimulus-onset with a 200 ms prestimulus baseline. Epochs containing blinks, eye movement and excessive artifacts were removed from the data. Artifact rejection thresholds were adjusted for each participant for tests of maximum amplitude to capture blinks, maximum voltage step per ms to capture voltage spikes, minimum amplitude per 50 ms to capture flat lining and saccades, and maximum amplitude difference in 100 ms to capture signal drift. Average waveforms were then calculated for each condition time-locked to the second tone in each pair. Only trials for correctly discriminated tones were included in the grand averages, with change and no-change trials averaged separately. For display purposes only, a high cutoff filter at 15 Hz was applied to the grand average ERPs used in each figure.

RESULTS

Behavioral data

Figure 2 contains a summary of performance on the pitch discrimination task. Performance accuracy was analyzed by the amount and direction of pitch change using hit rate minus false alarm rate, a measure of discrimination accuracy that corrects for response bias (Ayotte et al., 2002). In this case a hit corresponded to the correct detection of a pitch change, while a false alarm corresponded to a response of “different” on a no-change trial. A repeated-measures ANOVA with factors of Group (tone vs. control) \times change amount (7, 13, 25, 50, 100, or 200 cents) \times change

direction (up vs. down) was performed. Overall, performance was better in tone speakers than the control group [group, $F(1, 31) = 7.700$, $p = 0.009$], particularly when discriminating smaller changes in pitch [group \times change amount, $F(5, 155) = 4.405$, $p = 0.014$]. Tone speakers were more accurate than controls on pitch changes of 13 cents ($p = 0.006$), 25 cents ($p = 0.007$), and 50 cents ($p = 0.035$). When collapsing across groups, performance overall improved as the amount of pitch change increased [change amount, $F(5, 155) = 230.224$, $p < 0.001$] and was more accurate on downward than upward pitch changes [change direction, $F(1, 31) = 10.287$, $p = 0.003$] particularly on small pitch change trials [change amount \times change direction, $F(5, 155) = 5.684$, $p = 0.002$]. However, effects of change direction were more pronounced in tone speakers [down, 74.5%; up, 68.5%; $F(1, 31) = 6.533$, $p = 0.016$] than in controls [down, 61.8%; up, 57.0%; $F(1, 31) = 3.957$, $p = 0.056$]. To explore this pattern further we performed planned contrasts by change direction, revealing that tone speakers were more accurate than controls at change amounts of 7 cents ($p = 0.030$), 13 cents ($p = 0.005$), and 25 cents ($p = 0.024$). On upward changes, tone speakers were better at changes of 13 cents ($p = 0.038$), 25 cents ($p = 0.006$), and 50 cents ($p = 0.028$).

Event-related potentials analyses

Grand average ERPs time-locked to the second tone in each pair can be seen in **Figures 3 and 4**. For both groups, the early fronto-central negative deflection (N1) and later central-parietal positivity (P300) were larger in amplitude for change than no-change trials (**Figure 3**; see also footnote 4 regarding contingent negative variation). Upon visual inspection, the ERPs of the control group were characterized by larger N1c and P300 amplitudes to change trials relative to the tone speakers. Taking the direction of pitch change into consideration (**Figure 4**), tone speakers distinguished

between upward and downward pitch changes around 200–400 ms post-stimulus, whereas the control group ERPs to upward and downward changes did not differ until later stages of processing of approximately 400 ms and onward. In the following sections we address results separately for effects of pitch change and change direction, with planned contrasts between groups when a group main effect or interaction with group was present. Mean amplitudes were analyzed in a group \times change (change vs. no-change) ANOVA design including scalp distribution factors of hemisphere (left vs. right) \times laterality (lateral vs. medial) \times anteriority (prefrontal, frontal, central, occipital), such that only 16 of the 26 scalp electrodes were included in the omnibus ERP analyses (left lateral sites, LLPf (prefrontal), LLFr (frontal), LLTe (temporal), LLOc (occipital); left medial sites, LMPf, LMFr, LMCE, LMOc; right lateral sites, RLPf, RLFr, RLTe, RLOc; right medial sites, RMPf, RMFr, RMCE, RMOc). Greenhouse–Geisser corrections were applied to all repeated-measures with more than one degree of freedom to correct for errors of sphericity, and Bonferroni corrections were used for all pair wise comparisons.

Effect of pitch change (change vs. no-change)

Topographic scalp plots of change minus no-change difference waves (**Figure 3B**) suggest that tone speakers differentiated between change and no-change trials at very early stages of processing, specifically at the P50 (50–100 ms). Both groups showed evidence of differentiating between change and no-change trials at the subsequent negativity (N1 complex), although the control group seemed to take longer to resolve this discrepancy, manifesting in a prolonged N1c. At subsequent stages of processing indexed by the P300, both groups showed similar differentiation between change and no-change trials. During the same time window in which the P300 emerged at central-parietal sites, the

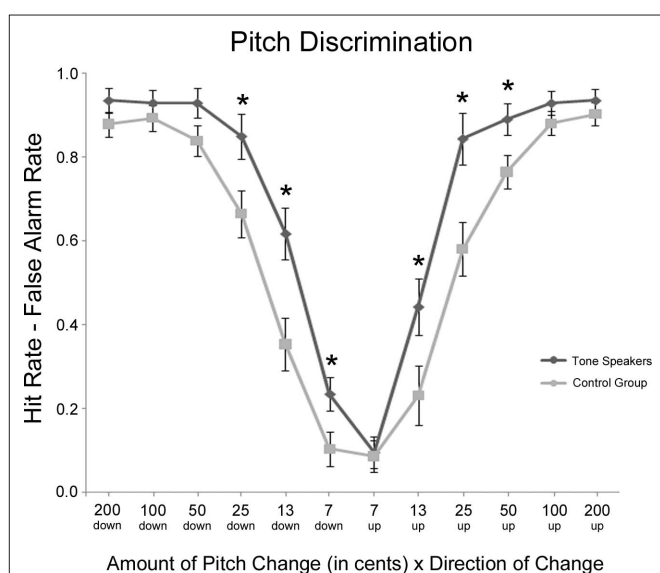


FIGURE 2 | Experiment 1 behavioral performance (hit rate minus false alarm rate) on the pitch discrimination task as a function of the amount and direction of pitch change. Significant group differences between tone speakers and controls are starred ($p < 0.05$).

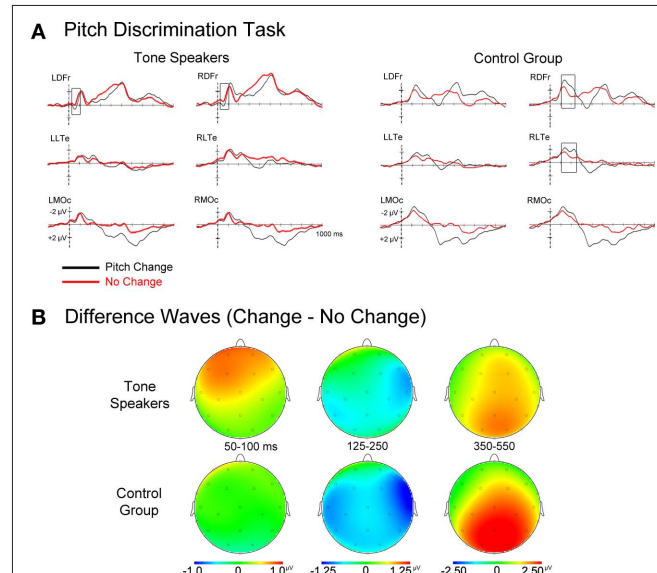


FIGURE 3 | (A) Event related potentials for change and no-change trials time-locked to the second tone of each pair. **(B)** Topographic scalp plots of difference wave activity (change minus no-change trials). Notice the larger P50 for tone speakers, and the later effects for the control group (N1c).

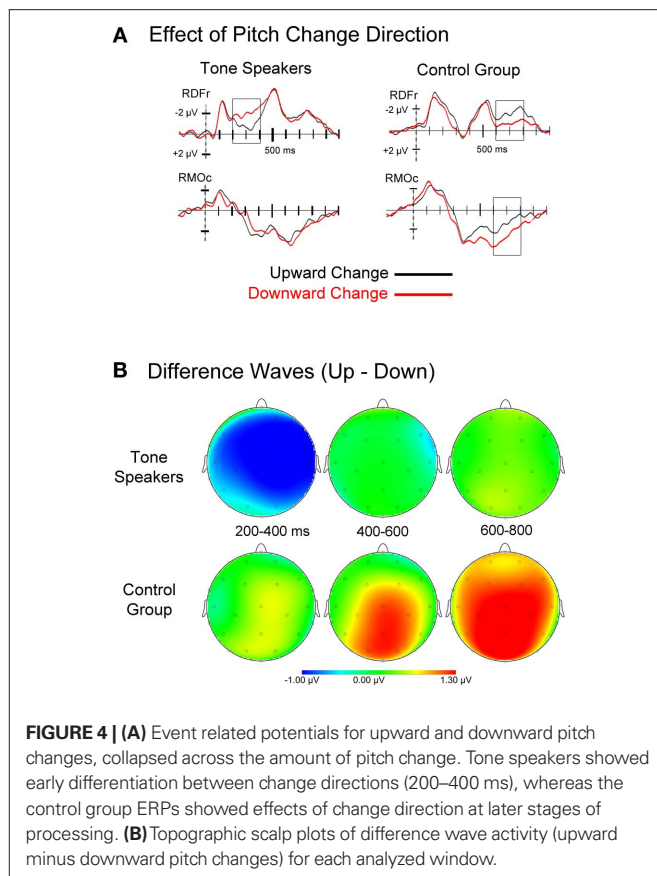


FIGURE 4 | (A) Event related potentials for upward and downward pitch changes, collapsed across the amount of pitch change. Tone speakers showed early differentiation between change directions (200–400 ms), whereas the control group ERPs showed effects of change direction at later stages of processing. **(B)** Topographic scalp plots of difference wave activity (upward minus downward pitch changes) for each analyzed window.

CNV-like⁴ sustained negativity was observed at anterior sites. To capture these effects, mean amplitudes were extracted in time windows of 50–100 ms (P50), 125–250 ms (N1), and 350–550 ms (P300/CNV).

50–100 ms. There was a main effect of pitch change at anterior sites, with larger P50 amplitudes for change than no-change trials at left anterior sites [change \times anteriority, $F(3, 93) = 8.701$, $p = 0.002$; change \times hemisphere \times anteriority, $F(3, 93) = 3.661$, $p = 0.024$; change \times hemisphere \times laterality \times anteriority, $F(3, 93) = 6.450$, $p = 0.001$]. Overall, P50 amplitudes at frontocentral electrodes were larger for tone speakers than the control group [group \times hemisphere \times anteriority, $F(3, 93) = 4.238$, $p = 0.019$; group \times hemisphere \times laterality \times anteriority, $F(3, 93) = 5.081$, $p = 0.006$]. Although there were no significant interactions between group and change, the main effect of change and the differences by group seemed to be carried by larger P50 amplitudes to change than no-change trials in tone speakers only. *Post hoc* analyses confirmed that tone speakers elicited larger P50 amplitudes to change than no-change trials at several bilateral anterior sites (MiPf, LLPf,

LMPf, RMPf, RDFr, LDFr, LMPf, RMPf, and LDCe significance at $p < 0.05$), while this same effect was limited to one electrode in the control group (LLPf).

125–250 ms. A right-lateralized N1 was observed for both groups independent of change condition during this epoch. Similar to the P1, the N1 was larger in amplitude for change than no-change trials [change, $F(1, 31) = 9.608, p = 0.004$], and was most pronounced at medial electrodes over the right hemisphere [change \times hemisphere, $F(1, 31) = 11.568, p = 0.002$; change \times laterality, $F(1, 31) = 6.717, p = 0.014$; change \times anteriority, $F(3, 93) = 7.292, p = 0.004$; change \times hemisphere \times Laterality, $F(1, 31) = 10.714, p = 0.003$; change \times hemisphere \times anteriority, $F(3, 93) = 10.084, p < 0.001$; change \times laterality \times anteriority, $F(3, 93) = 6.706, p = 0.001$; change \times hemisphere \times laterality \times anteriority, $F(3, 93) = 5.069, p = 0.007$]. The distribution of the N1 effect for change vs. no-change trials differed between groups [group \times change \times hemisphere \times laterality, $F(1, 31) = 5.301, p = 0.028$], with a more distributed effect for control subjects (medial sites in the left, $p = 0.022$ and right hemispheres, $p = 0.013$ and at right lateral sites, $p = 0.001$) than tone speakers (right medial electrodes, $p = 0.033$). Additionally, only the control group showed an N1 effect at the right lateral temporal electrode site (RLTe; $p < 0.001$; tone speakers, $p = 0.173$), likely indexing the N1c subcomponent of the auditory N1 potential, with greater amplitude for change than no-change trials.

350–550 ms. Overall, P300 amplitudes were larger for change than no-change trials at frontal, central, and posterior sites [change, $F(1, 31) = 21.056, p < 0.001$; change \times anteriority, $F(3, 93) = 24.576, p < 0.001$], while the CNV was larger for no-change than change trials at medial prefrontal sites [change \times hemisphere, $F(1, 31) = 9.529, p = 0.004$; change \times laterality, $F(1, 31) = 19.223, p < 0.001$; change \times hemisphere \times anteriority, $F(3, 93) = 3.995, p = 0.015$; change \times laterality \times anteriority, $F(3, 93) = 4.674, p = 0.008$]. There was also an overall amplitude difference between groups [group, $F(1, 31) = 5.502, p = 0.026$], with tone speakers eliciting larger negative (CNV) amplitudes than controls at prefrontal and frontal sites, while controls elicited larger positive amplitudes (P300) than tone speakers at posterior sites [group \times laterality \times anteriority, $F(3, 93) = 4.640, p = 0.015$; group \times hemisphere \times laterality \times anteriority, $F(3, 93) = 5.100, p = 0.005$]. Interestingly, tone speakers elicited larger CNV amplitudes than controls on both change and no-change trials, while controls elicited larger P300 amplitudes than tone speakers on change trials only [group \times change \times anteriority, $F(3, 93) = 4.953, p = 0.021$; group \times change \times laterality \times anteriority, $F(3, 93) = 3.002, p = 0.046$].

Direction of change analyses

Visual inspection of the ERPs for upward and downward pitch changes collapsed across amount of change suggests an earlier dissociation between upward and downward change for tone speakers than controls (see **Figure 4**). Mean amplitudes for upward and downward trials were extracted for consecutive epochs of 200–400 ms, 400–600 ms, and 600–800 ms, then subjected to a group \times change direction (up, down) ANOVA.

200–400 ms. Tone speakers elicited larger negative amplitude for downward than upward changes over the right hemisphere [group \times change direction \times laterality, $F(1, 31) = 6.297$, $p = 0.018$; group \times change direction \times hemisphere \times laterality \times anteriority, $F(3,$

⁴Both groups also showed a sustained frontocentral negativity peaking around 500 ms then returning sharply to baseline. Since the task involved a delayed-response upon appearance of a visual cue 500 ms after the target tone, this negativity is likely a contingent negative variation (CNV) reflecting response preparation processes (e.g., Walter et al., 1964; Tecce, 1972; Bender et al., 2004), or an offset potential (Rohrbaugh et al., 1979; Goydke et al., 2004) similar to those observed in auditory studies not requiring a response, though the latter tends to be larger toward the vertex (Picton et al., 1977).

93) = 3.008, $p = 0.043$]. Pairwise comparisons show that this difference was strongest at the lateral frontal and temporal electrodes (RLFr, $p = 0.004$; RLTe, $p = 0.005$), as well as medial prefrontal to central sites (RMPf, $p = 0.053$; RMFr, $p = 0.009$; RMCe, $p = 0.015$). No significant differences were observed for the control group (all p 's > 0.10).

400–600 ms. There were no significant main effects of group or change direction, nor interactions of these factors, in this window.

600–800 ms. Controls elicited larger positive amplitudes to downward than upward changes at medial electrode sites [group \times change direction \times laterality \times anteriority, $F(3, 93) = 3.532$, $p = 0.028$]. Pairwise comparisons revealed that this was significant at frontal ($p = 0.033$), central ($p = 0.006$) and occipital sites ($p = 0.002$). No differences by change direction were observed for tone speakers.

DISCUSSION

The behavioral and electrophysiological results both demonstrate a clear advantage for native tone speakers on this simple pitch discrimination task. Tone speakers were more accurate at discriminating changes in pitch, particularly for smaller change amounts. In addition, tone speakers elicited an earlier difference between change and no-change trials (P50) than was observed in the control group (N1), alluding to faster and more sensitive tone discrimination in speakers of a tone language. Similarly, the ERPs of tone speakers differentiated upward from downward pitch change at an earlier stage of processing (200–400 ms) than controls (600–800 ms), even though the task did not require listeners to indicate the direction of pitch change. ERPs from the control group also showed larger P300 amplitude on change trials compared to tone speakers, indexing a difference in the ability to categorize the stimuli perhaps due to categorical perception requiring more resources for the control group. Finally, subtle differences in scalp distribution between the groups imply that non-tone language speakers employed a more widely distributed neural network to perform the task than tone language speakers.

EXPERIMENT 2

Experiment 1 demonstrated an advantage for tone language speakers on a simple pitch discrimination task. Experiment 2 was designed to test for a similar advantage on a more complex perceptual discrimination task. Experiment 2 required participants to determine if two consecutive pairs of tones were of same or different interval size. The goal was to uncover the electrophysiological correlates for an observed behavioral tone language advantage (Pfordresher and Brown, 2009; Hove et al., 2010), using a similar task and stimuli.

METHODS

Participants

Data from right-handed native tone language speakers ($n = 16$) and controls ($n = 16$) of those who participated in Experiment 1 were used for this experiment.⁵

⁵The sample for Experiment 2 included all participants from Experiment 1, however a single participant was excluded from the Tone group for an insufficient number of artifact-free trials in Experiment 2. The exclusion of this participant did not significantly impact the group matching on music experience, age, education, and other variables.

Stimuli

Stimuli were pure tones as in Experiment 1. During each trial, participants heard two pairs of notes, where the first interval (standard) always comprised C5 (524 Hz) then G5 (784 Hz), resulting in a 700-cent change (where 100 cents = 1 semitone). The second interval started on F#5 (740 Hz) and was immediately followed by a fourth tone, which varied with trial type.

On “no-change” trials (160 total), the fourth tone was C#6, resulting in an equivalent change compared to the previously presented tone pair. On “change” trials (240 total), the fourth tone was either higher or lower in pitch than the no-change note C#6 by 13, 25, 50, 100, 200, or 400 cents (20 trials each). Thus, relative to the standard interval, an “expanded” interval was created by upward pitch changes from C#6, while a “shrunk” interval was formed by downward changes from C#6. Participants were asked to judge the relative size of the two intervals by indicating whether they were “same” or “different” via a button press. The order of trials was pseudo-randomized, such that no more than two trials in a row were presented with the same amount or direction of change, with a fixed order presented to all participants.

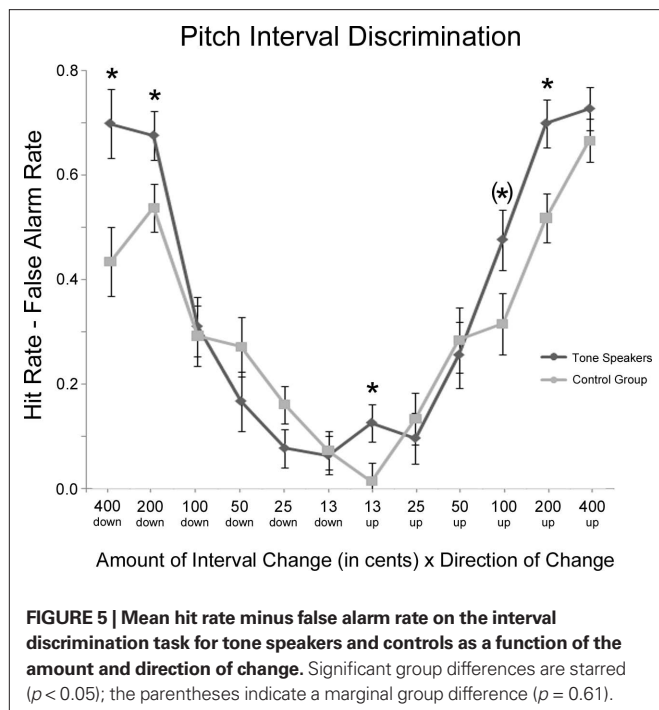
Procedure

The procedure was the same as in Experiment 1. Participants completed a total of 400 trials of interval discrimination across 16 blocks. During each trial (see **Figure 1B**), participants were asked to fixate on a half-inch white cross on a black screen, which appeared 500 ms before the first tone to orient the participant, and remained on the screen until 500 ms after the end of the last tone to reduce eye movement artifacts in the EEG. Participants responded by pushing one button for “same” and another for “different” with their right hand on a response box. At the end of each trial, the screen went blank for 500 ms before the onset of the next trial.

RESULTS

Behavioral data

Figure 5 contains a summary of performance on the interval discrimination task. Response accuracy was analyzed using hit rate minus false alarm rate in a group \times interval distance (13, 25, 50, 100, 200, and 400 cent change) \times change direction (expanded vs. shrunk) repeated-measures ANOVA. For both groups, performance became more accurate as the amount of interval change increased [interval, $F(5, 150) = 131.807$, $p < 0.001$] and was more accurate on expanded than shrunk intervals [interval \times change direction, $F(5, 150) = 2.702$, $p = 0.044$], particularly at large interval distances of 400 cents ($p = 0.001$). Although there was no main effect of group, $F(1, 30) = 1.574$, $p = 0.219$, there was a significant group \times interval interaction, $F(5, 150) = 5.926$, $p = 0.001$, where tone speakers performed better than controls at the largest interval change amounts, specifically 200-cent ($p = 0.012$) and 400-cent change ($p = 0.022$). Furthermore, there was a group \times interval \times change direction interaction, $F(5, 150) = 3.307$, $p = 0.019$, revealing better performance in tone speakers than the control on shrunk intervals of 200 cents ($p = 0.038$) and 400 cents ($p = 0.008$), and on expanded intervals of 13 cents ($p = 0.046$), and 200 cents ($p = 0.009$). Tone speakers were also marginally better on expanded intervals of 100 cents, $p = 0.061$ (note that the apparent difference at 25 cent intervals in **Figure 5** was not significant, $p = 0.101$).



Event-related potentials

Event related potentials were time-locked to the critical changing tone (second tone of the second pair) and analyzed as change relative to no-change (Figure 6) and by change direction (Figure 7). In the following, we describe the results of each analysis in separate sections. As with Experiment 1, all ERP analyses included the electrode factors of hemisphere (left, right) \times laterality (lateral, medial) \times anteriority (prefrontal, frontal, central, occipital), while Greenhouse–Geisser and Bonferroni corrections were applied where appropriate.

Effect of interval change (change vs. no-change)

Event related potentials for all interval no-change and change trials collapsed across change amount, as well as difference waves of change minus no-change activity are shown in Figure 6. Mean amplitudes were extracted in epochs where group ERPs differentiated between change and no-change intervals and subjected to a group \times change (change vs. no-change) ANOVA. The chosen epochs corresponded with an early, right-lateralized N1 effect of change that was only present in tone speakers (125–225 ms), and a later effect of change seen in both groups that was indexed by a P300 and subsequent late positivity (300–500 ms, 500–700 ms).

125–225 ms. Overall, N1 amplitudes were larger for change than no-change trials [change, $F(1, 30) = 14.327$, $p = 0.001$]. However this was mitigated by a marginally significant interaction by group [group \times change, $F(1, 30) = 3.969$, $p = 0.056$], such that this effect was observed in tone speakers ($p < 0.001$) but not in the control group ($p = 0.215$). Pair wise comparisons confirmed that the larger N1 amplitudes for tone speakers on change relative to no-change intervals was observed broadly across the scalp, yet absent in controls.

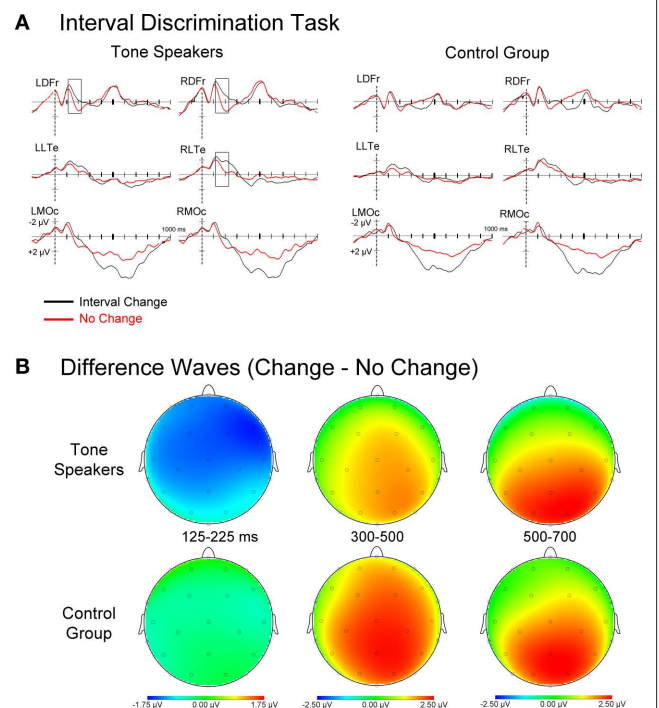


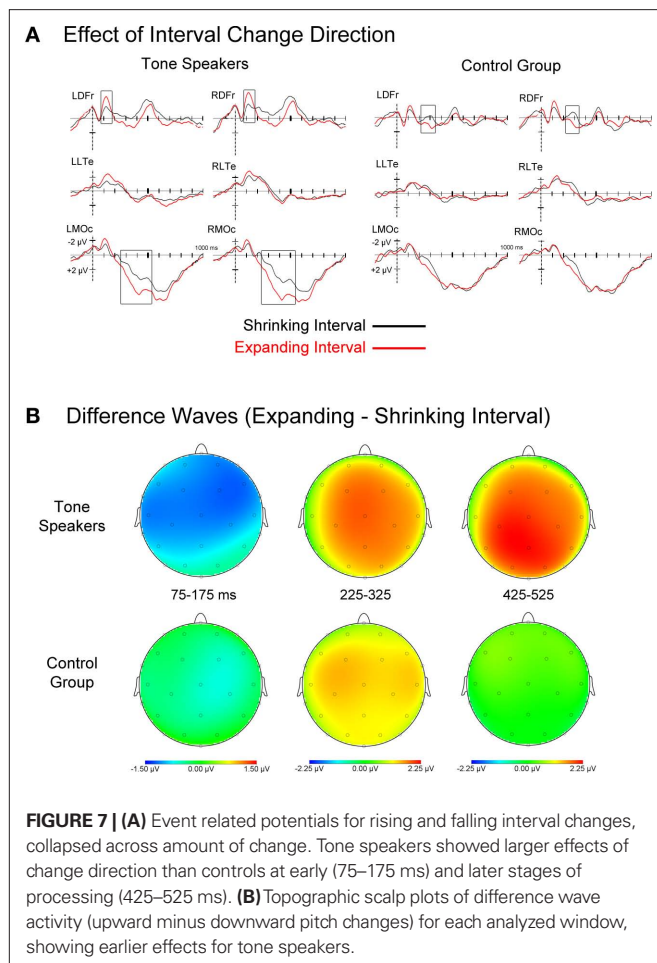
FIGURE 6 | (A) Event related potentials for interval change and no-change trials time-locked to the second tone of the second tone pair. (B) Topographic scalp plots of difference wave activity (change minus no-change trials). Notice the earlier effects of interval change for Tone speakers.

300–500 ms. For both groups, P300 amplitudes were larger for change than no-change trials [change, $F(1, 30) = 19.234$, $p < 0.001$]. However, this effect was more widely distributed in controls than tone speakers [group \times change \times laterality \times anteriority, $F(3, 90) = 3.564$, $p = 0.037$]. This effect was observed at all midline sites, as well as lateral central and occipital sites for the control group, and only at midline frontal, central, and occipital sites for tone speakers.

500–700 ms. Continuing from the previous epoch, the prolonged positivity resulted in larger positive amplitudes for change than no-change trials in both groups [change, $F(1, 30) = 11.541$, $p < 0.001$]. While both groups showed a larger positivity to change than no-change trials at lateral and medial occipital sites, and medial frontal and central sites, this effect extended to lateral central sites in tone speakers only [group \times change \times anteriority, $F(3, 90) = 4.147$, $p = 0.038$; group \times change \times laterality \times anteriority, $F(3, 90) = 5.161$, $p = 0.008$].

Effect of interval change direction

Figure 7 shows ERPs for expanded and shrunk intervals collapsed across amount of interval change, as well as difference waves for rising minus falling intervals. Visual inspection reveals that the ERPs of tone speakers differentiated between expanded and shrunk change around 75 ms and at later stages of processing, while an effect of change direction in non-tone speakers did not emerge until approximately 225 ms. To capture these effects, mean amplitudes



were extracted in windows of 75–175 ms, 225–325 ms, and 425–525 ms, and subjected to a group \times change direction (expanded vs. shrunken) ANOVA.

75–175 ms. Overall, amplitudes in this epoch were more negative for expanded than shrunken intervals [Direction, $F(1, 30) = 7.670$, $p = 0.010$]. There was a trend toward a Group by Change Direction interaction, $F(1, 30) = 3.052$, $p = 0.091$. Planned comparisons indicated larger negative amplitudes to expanded compared to shrunken intervals for Tone speakers ($p = 0.003$), but no difference for controls ($p = 0.475$). Further exploratory comparisons revealed that this effect was significant at frontocentral sites in the left and right hemisphere for Tone speakers, but was absent at all sites for the control group.

225–325 ms. In contrast to the previous epoch, amplitudes were more negative for shrunken than expanded intervals [Direction, $F(1, 30) = 10.527$, $p = 0.003$], and this effect was present in both groups (Tone, $p = 0.039$; controls, $p = 0.021$). No interaction of Group by Change Direction was observed in this window.

425–525 ms. Larger positive amplitudes to expanded than shrunken intervals were observed exclusively in Tone speakers [group \times direction \times laterality, $F(1, 30) = 5.207$, $p = 0.030$]. Planned comparisons revealed the presence of this effect in Tone speakers at

medial prefrontal, frontal, central, and occipital sites. In contrast, no differences by direction of change were observed at any sites in the control group.

DISCUSSION

Like Experiment 1, Experiment 2 shows strong support for a tone language advantage, this time in a more complex perceptual discrimination task. Behaviorally, native tone speakers were more accurate at detecting large changes in interval distance of 200 and 400 cents. In addition, Tone speakers elicited an earlier difference between interval change and no-change trials (125–225 ms) than controls (300–500), alluding to faster and more sensitive tone discrimination in speakers of a tone language. Similarly, the ERPs of Tone speakers showed a difference between expanded and shrunken intervals at an earlier (and later) stage of processing (75–175 ms) than the control group (225–325 ms).

GENERAL DISCUSSION

In this study, native experience speaking a language in which pitch is used to express meaning determined how early the brain responded to differences in pitch for non-speech sounds, and was accompanied by enhanced performance (i.e., greater accuracy in discriminating sounds) on pitch and interval discrimination tasks relative to controls. Not only were native tone speakers more accurate at detecting changes in pitch height and interval distance at various increments, their ERPs demonstrated earlier effects of change vs. no-change, more focal processing of stimulus change, and earlier effects of change direction. These findings support an existing body of research demonstrating the impact of language experience on the way sound is processed in the brain, replicating previous reports of general pitch processing advantages observed in tone language speakers, while adding that these effects are driven by facilitation in the brain response to auditory stimuli at early stages of processing.

The current results demonstrate greater sensitivity to pitch height information in native tone speakers when compared to non-tone language speaking bilinguals, even when listeners were presented with level-contour tones that did not directly resemble linguistic tones used in Mandarin. This is consistent with the view that height and contour of pitch are evaluated separately during the perception of Mandarin lexical tones (Massaro et al., 1985) and coheres with previous findings that native tone-speaking experience leads to general enhancements in pitch processing that are not specific to linguistic input (Deutsch and Henthorn, 2004; Deutsch et al., 2006, 2009). Prior to this study, studies using similar paradigms did not show an advantage for tone language speakers for simple pitch discrimination (Bent et al., 2006; Pfordresher and Brown, 2009). There were some differences between this and prior studies. Whereas the composition of the control non-tone language group in the current study included bilingual speakers (English, plus another non-tonal language), participants in previous studies have primarily been monolingual (English only) speakers. It seems plausible that experience with more than one language can lead to improved performance on a pitch task based on the use of different pitch patterns across languages. For this reason we thought it was important to include a bilingual control group for the bilingual tone speaker group. However, improved

performance based on being bilingual alone would have worked against our finding of a tone language advantage. On the contrary, the current results might be confounded by ethnicity, given that the tone language speakers were all of the same ethnicity, while the non-tone speakers were of different ethnic backgrounds. With our population samples we cannot separate out any role that genetics might play in our findings (c.f., Hove et al., 2010). Nevertheless, there is strong evidence that it is experience and not genetics that plays the more critical role in pitch processing (Crinion et al., 2009; Deutsch et al., 2009). Great care was taken to match the groups on all other characteristics⁶ (see Footnote 1). Finally, the number of trials in the current study was significantly more than those of the other studies, given that a large number of trials are needed for ERP analysis. However, there was no evidence of a practice or habituation effect comparing earlier and later trials in our data. Our results replicate the behavioral tone language advantage for interval discrimination previously reported by Pfordresher and Brown (2009), and suggest that this advantage is specific to larger interval sizes (200- and 400-cent change) for this more challenging perceptual discrimination task. Thus, we provide evidence for a robust advantage across both experiments.

Tone speakers also demonstrated greater sensitivity to the *direction of pitch change* than controls. On pitch discrimination trials, the hit rate of tone speakers was higher for downward than upward changes, particularly at the smallest change amounts (7 and 13 cents). Considering that falling tones occur with greater frequency than rising tones in Mandarin (e.g., Cheng, 1973), this difference may be due to Mandarin speakers' greater experience detecting downward than upward changes in pitch. Thus the cross-domain influence between linguistic and non-linguistic tone processing is influenced by the specific characteristic of the sounds to which a listener is exposed. The same pattern was observed in the ERPs of tone speakers (**Figure 4**), which differentiated downward from upward pitch changes at earlier stages of processing (200–400 ms) than controls (600–800 ms). Although change direction effects were not apparent behaviorally on interval discrimination trials, the more sensitive ERP measures showed earlier and more robust differentiation between expanded and shrunken intervals for tone speakers compared to controls (**Figure 7**).

On the surface, the fact that Mandarin speakers were more sensitive to pitch direction in the current study is similar to previous findings reported by Bent et al. (2006). In that study, Mandarin speakers processed rising and falling non-speech contours differently in an identification task, whereas English speakers were more comparable in their performance across contours. However, unlike the current study, Mandarin speakers in Bent et al. differed only with respect to response bias, not sensitivity, and group differences in that study were only found for non-speech analogs of Mandarin tones (i.e., dynamically changing tones). Bent et al. concluded that linguistic background shapes categorization, but not perception, of auditory signals that resemble speech sounds.

⁶To rule out the possibility that the current results were influenced by musical experience, years of musical training³ was entered as a covariate on ANOVAs of both behavioral and ERP data. For behavioral and ERP analyses in both experiments, there were no main effects of music experience (all p 's > 0.63), nor significant interactions with pitch change (all p 's > 0.71) or change direction (all p 's > 0.23) in any of the time windows analyzed.

The present results have much different implications in that they suggest differences across groups with respect to sensitivity that extend to auditory patterns bearing little resemblance to speech. Moreover, our ERP results indicate that these differences across groups appear early and are related to both perceptual (exogenous, stimulus-driven components such as N1) and categorization (P300) processes. We suspect that both findings fill in part of the continuum from speech to non-speech processing referred to by Bent et al. whereas effects related to response bias may be found for more speech-like analogs, effects related to sensitivity may appear in materials that less resemble speech. Moreover, as mentioned above, effects related to sensitivity as found here may require greater statistical power, and more balanced selection of control participants, to detect.

The *effect of change* (change minus no-change collapsing across pitch change amount) showed early group differences in brain activity at the P50 component, with pitch changes eliciting a larger frontocentral positivity for tone speakers at around 50–100 ms compared to controls (**Figure 3B**), consistent with the distribution of the auditory P50 (e.g., Potts et al., 1998; Korzyukov et al., 2007). Guterman et al. (1992) found that focusing participants' attention on the second click of a click pair increased the P50 amplitude elicited by that click, suggesting that the P50 reflects enhancements in early stimulus processing. This is supported by similar findings of stimulus expectancy and attention effects occurring as early as the P50 (Guterman and Josiassen, 1994; Clementz et al., 2002). The P50 appears to be generated by separate sources in the frontal and temporal lobes (Potts et al., 1998; Grunwald et al., 2003; Korzyukov et al., 2007), where the frontal generator collects and retains information from the temporal lobe generator about the physical parameters of the auditory stimulus (Korzyukov et al., 2007). In light of these findings, the larger P50 amplitudes observed here for tone speakers may suggest that their native experience with a tonal language allows them to extract more information from the auditory stream, in this case information about pitch height. This interpretation is consistent with findings that tone speakers show enhanced processing of pitch in the brainstem (Krishnan et al., 2005, 2009; Swaminathan et al., 2008). In the interval task, although collapsing across interval change amount did not reveal a P50 effect in either group, tone speakers still showed an earlier effect of change than controls indexed by a larger N1-like component (**Figure 6**). Given the greater complexity of the interval discrimination judgments, it is possible that tone speakers' auditory processing advantage is noticeable only at a later cortical stage of processing relative to the simpler pitch discrimination task. We should note that there was evidence of a reduction in the brain response to pitch change in tone speakers at the P300, but this was not replicated in the interval task. Diminished P300 amplitudes in tone speakers may reflect a less resource-intensive category judgment of change or no-change for each pitch, due to pitch discrimination involving more automated processing than in controls (Linding et al., 2003).

CONCLUSION

To our knowledge, this study is the first to demonstrate a tone language pitch processing advantage with simple tones, where accurate performance depends solely upon sensitivity to pitch

height information. Utilizing the temporal precision of ERPs, we provide evidence that the tone language advantage observed on the pitch and interval discrimination tasks was due to an enhancement of early sensory processes, indexed by larger P50 amplitudes to pitch changes and larger N1 amplitudes to the more complex interval changes. Together these findings clearly show that language experience can have a significant impact on an individual's ability to process non-linguistic information, reflecting the subtle but robust cross-domain influence between language and general perception. Native experience with a tone language involves attending to the pitch of spoken syllables in

daily language use from an early age, and this experience appears to enhance the brain's cortical representation of pitch relative to speakers of a non-tone language.

ACKNOWLEDGMENTS

This research was supported by a San Antonio Life Sciences Institute grant #121075, the University of Texas at San Antonio (UTSA) College of Science (Nicole Y. Y. Wicha), and the UTSA Computational Biology Initiative. Nicole Y. Y. Wicha and Ryan J. Giuliano were supported by NICHD/NIGMS HD060435. Peter Q. Pfordresher was supported by NSF grant #BCS-0642592.

REFERENCES

- Ayotte, J., Peretz, I., and Hyde, K. (2002). A group study of adults afflicted with a music-specific disorder. *Brain* 125, 238–251.
- Bender, S., Resch, E., Weisbrod, M., and Oelkers-Ax, R. (2004). Specific task anticipation versus unspecific orienting reaction during early contingent negative variation. *Clin. Neurophysiol.* 115, 1836–1845.
- Bent, T., Bradlow, A. R., and Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 97–103.
- Bigand, E., and Poulin-Charronnat, B. (2006). Are we “experienced listeners?” A review of the musical capacities that do not depend on formal musical training. *Cognition* 100, 100–130.
- Bosnyak, D. J., Eaton, R. A., and Roberts, L. E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40Hz amplitude modulated tones. *Cereb. Cortex* 14, 1088–1099.
- Bregman, A. S., Liao, C., and Levitan, R. (1990). Auditory grouping based on fundamental-frequency and formant peak frequency. *Can. J. Psychol.* 44, 400–413.
- Burns, E. M., and Sampat, K. S. (1980). A note on possible culture-bound effects in frequency discrimination. *J. Acoust. Soc. Am.* 68, 1886–1888.
- Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (2007a). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport* 18, 1963–1967.
- Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (2007b). Mismatch negativity to pitch contours is influenced by language experience. *Brain Res.* 1128, 148–156.
- Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain Lang.* 108, 1–9.
- Cheng, C. C. (1973). A quantitative study of tone in Chinese. *J. Chin. Linguist.* 1, 93–110.
- Clementz, B. A., Barber, S. K., and Dzau, J. R. (2002). Knowledge of stimulus repetition affects the magnitude and spatial distribution of low-frequency event-related brain potentials. *Audiol. Neurotol.* 7, 303–314.
- Crinion, J. T., Green, D. W., Chung, R., Ali, N., Grogan, A., Price, G. R., Mechelli, A., and Price, C. J. (2009). Neuroanatomical markers of speaking Chinese. *Hum. Brain Mapp.* 30, 4108–4115.
- Cruttenden, A. (1997). *Intonation*. Cambridge: Cambridge University Press.
- Deutsch, D., Dooley, K., Henthorn, T., and Head, B. (2009). Absolute pitch among students in an American music conservatory: association with tone language fluency. *J. Acoust. Soc. Am.* 125, 2398–2403.
- Deutsch, D., and Henthorn, T. (2004). Absolute pitch, speech, and tone language: some experiments and a proposed framework. *Music Percept.* 21, 339–356.
- Deutsch, D., Henthorn, T., Marvin, E., and Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: prevalence differences, and evidence for a speech-related critical period. *J. Acoust. Soc. Am.* 119, 719–722.
- Giuliano, R. J., and Wicha, N. Y. Y. (2010). Why the white bear is still there: Electrophysiological evidence for ironic semantic activation during thought suppression. *Brain Res.* 1316, 62–74.
- Goydke, K. N., Altenmüller, E., Moller, J., and Munte, T. F. (2004). Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. *Cogn. Brain Res.* 21, 351–359.
- Grunwald, T., Boutros, N. N., Pezer, N., von Oertzen, J., Fernandez, G., Schaller, C., and Elger, C. E. (2003). Neuronal substrates of sensory gating within the human brain. *Biol. Psychiatry* 53, 511–519.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Guterman, Y., and Josiassen, R. C. (1994). Sensory gating deviance in schizophrenia in the context of task-related effects. *Int. J. Psychophysiol.* 18, 1–12.
- Guterman, Y., Josiassen, R. C., and Bashore, T. R. (1992). Attentional influence on the P50 component of the auditory event-related brain potential. *Int. J. Psychophysiol.* 12, 197–209.
- Hove, M. J., Sutherland, M. E., and Krumhansl, C. L. (2010). Ethnicity effects in relative pitch. *Psychon. Bull. Rev.* 17, 310–316.
- Kaan, E., Barkley, C. M., Bao, M., and Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: an event-related potentials training study. *BMC Neurosci.* 9, 1–17. doi: 10.1186/1471-2202-9-53
- Kaan, E., Wayland, R., Bao, M. Z., and Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: an event-related potential study. *Brain Res.* 1148, 113–122.
- Korzyukov, O., Pflieger, M. E., Wagner, M., Bowyer, S. M., Rosburg, T., Sundaresan, K., Elger, C. E., and Boutros, N. N. (2007). Generators of the intracranial P50 response in auditory sensory gating. *Neuroimage* 35, 814–826.
- Kraus, N., McGee, T., Carrell, T. D., King, C., Tremblay, K., and Nicol, T. (1995). Central auditory-system plasticity associated with speech-discrimination training. *J. Cogn. Neurosci.* 7, 25–32.
- Krishnan, A., Swaminathan, J., and Gandour, J. T. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J. Cogn. Neurosci.* 21, 1092–1105.
- Krishnan, A., Xu, Y. S., Gandour, J., and Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cogn. Brain Res.* 25, 161–168.
- Lee, Y. S., Vakoč, D. A., and Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: a cross-linguistic comparison. *J. Psycholinguist. Res.* 25, 527–542.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461.
- Lieberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech-perception revised. *Cognition* 21, 1–36.
- Lieberman, A. M., and Mattingly, I. G. (1989). A specialization for speech-perception. *Science* 243, 489–494.
- Lieberman, A. M., and Whalen, D. H. (2000). On the relation of speech to language. *Trends Cogn. Sci.* 4, 187–196.
- Linding, M., Zurrón, M., and Diaz, F. (2003). Changes in P300 amplitude during an active standard auditory oddball task. *Biol. Psychol.* 66, 153–167.
- Massaro, D. W., Cohen, M. M., and Tseng, C. Y. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *J. Chin. Linguist.* 13, 267–290.
- Nan, Y., Sun, Y., and Peretz, I. (2010). Congenital amusia in speakers of a tone language: association with lexical tone agnosia. *Brain* 133, 2635–2642.
- Peretz, I., Brattico, E., and Tervaniemi, M. (2005). Abnormal electrical brain responses to pitch in congenital amusia. *Ann. Neurol.* 58, 478–482.
- Pfordresher, P. Q., and Brown, S. (2007). Poor-pitch singing in the absence of “tone deafness”. *Music Percept.* 25, 95–115.
- Pfordresher, P. Q., and Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Atten. Percept. Psychophys.* 71, 1385–1398.
- Picton, T. W., Woods, D. L., Baribeau-Braun, J., and Healey, T. M. G. (1977). Evoked-potential audiometry. *J. Otolaryngol.* 6, 90–119.
- Potts, G. F., Dien, J., Hartry-Speiser, A. L., McDougal, L. M., and Tucker, D. M. (1998). Dense sensor array topography of the event-related potential to task-relevant auditory stimuli. *Electroencephalogr. Clin. Neurophysiol.* 106, 444–456.

- Rohrbaugh, J. W., Sydulko, K., and Lindsley, D. B. (1979). Cortical slow negative waves following non-paired stimuli: effects of modality, intensity and rate of stimulation. *Electroencephalogr. Clin. Neurophysiol.* 46, 416–427.
- Schellenberg, E. G., and Trehub, S. E. (2008). Is there an Asian advantage for pitch memory? *Music Percept.* 25, 241–252.
- Stagray, J. R., and Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and a nontone language. *J. Chin. Linguist.* 21, 143–163.
- Swaminathan, J., Krishnan, A., and Gandour, J. T. (2008). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport* 19, 1163–1167.
- Tecce, J. J. (1972). Contingent negative variation (CNV) and psychological processes in man. *Psychol. Bull.* 77, 73–108.
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., and Schroger, E. (2005). Pitch discrimination accuracy in musicians vs. nonmusicians: an event-related potential and behavioral study. *Exp. Brain Res.* 161, 1–10.
- Trehub, S. E., Schellenberg, E. G., and Nakata, T. (2008). Cross-cultural perspectives on pitch memory. *J. Exp. Child. Psychol.* 100, 40–52.
- Tremblay, K., Kraus, N., Carrell, T. D., and McGee, T. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *J. Acoust. Soc. Am.* 102, 3762–3773.
- Waldo, M. C., and Freedman, R. (1986). Gating of auditory evoked-responses in normal college-students. *Psychiatry Res.* 19, 233–239.
- Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., and Winter, A. L. (1964). Contingent negative variation: an electric sign of sensorimotor association and expectancy in the human brain. *Nature* 203, 380–384.
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.
- Woods, D. L., Alho, K., and Algazi, A. (1993). Intermodal selective attention—evidence for processing in tonotopic auditory fields. *Psychophysiology* 30, 287–295.
- Yip, M. J. W. (2002). *Tone*. Cambridge: Cambridge University 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: 16 February 2011; accepted: 16 June 2011; published online: 03 August 2011.

Citation: Giuliano RJ, Pfordresher PQ, Stanley EM, Narayana S and Wicha NYY (2011) Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Front. Psychology* 2:146. doi: 10.3389/fpsyg.2011.00146

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

Copyright © 2011 Giuliano, Pfordresher, Stanley, Narayana and Wicha. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



EEG correlates of song prosody: a new look at the relationship between linguistic and musical rhythm

Reyna L. Gordon^{1,2*}, Cyrille L. Magne³ and Edward W. Large^{1*}

¹ Music Dynamics Laboratory, Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL, USA

² EEG/ERP Laboratory, Vanderbilt Kennedy Center, Vanderbilt University, Nashville, TN, USA

³ Brain and Language Laboratory, Psychology Department and PhD in Literacy Studies, Middle Tennessee State University, Murfreesboro, TN, USA

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Stefan Debener, University of Oldenburg, Germany
Martin Meyer, University of Zurich, Switzerland

*Correspondence:

Reyna L. Gordon and Edward W. Large, Music Dynamics Laboratory, Center for Complex Systems and Brain Sciences, Florida Atlantic University, 777 Glades Road, Boca Raton, FL, USA.
e-mail: reyna.gordon@vanderbilt.edu;
large@ccs.fau.edu

Song composers incorporate linguistic prosody into their music when setting words to melody, a process called “textsetting.” Composers tend to align the expected stress of the lyrics with strong metrical positions in the music. The present study was designed to explore the idea that temporal alignment helps listeners to better understand song lyrics by directing listeners’ attention to instances where strong syllables occur on strong beats. Three types of textsettings were created by aligning metronome clicks with all, some or none of the strong syllables in sung sentences. Electroencephalographic recordings were taken while participants listened to the sung sentences (primes) and performed a lexical decision task on subsequent words and pseudowords (targets, presented visually). Comparison of misaligned and well-aligned sentences showed that temporal alignment between strong/weak syllables and strong/weak musical beats were associated with modulations of induced beta and evoked gamma power, which have been shown to fluctuate with rhythmic expectancies. Furthermore, targets that followed well-aligned primes elicited greater induced alpha and beta activity, and better lexical decision task performance, compared with targets that followed misaligned and varied sentences. Overall, these findings suggest that alignment of linguistic stress and musical meter in song enhances musical beat tracking and comprehension of lyrics by synchronizing neural activity with strong syllables. This approach may begin to explain the mechanisms underlying the relationship between linguistic and musical rhythm in songs, and how rhythmic attending facilitates learning and recall of song lyrics. Moreover, the observations reported here coincide with a growing number of studies reporting interactions between the linguistic and musical dimensions of song, which likely stem from shared neural resources for processing music and speech.

Keywords: rhythm, EEG, beta/gamma oscillations, song perception, textsetting, prosody, language, music

INTRODUCTION

Song is an ecological model for studying the complex relationship between music and speech, and has been examined in recent studies with cognitive neuroscience methods that have found interactions between various aspects of the linguistic and musical dimensions of songs (Lidji et al., 2009; Gordon et al., 2010; Sammler et al., 2010; Schön et al., 2010). In both music and speech, temporal patterning of events gives rise to hierarchically organized rhythms that are perceived as metrical (Palmer and Hutchins, 2006). Songs are created through the process of *textsetting*, in which the prosodic features of speech are combined with musical melody, uniting their separate metrical structures and forming one rhythmic pattern (Halle and Lerdahl, 1993). Musicologists have long noted that in the process of textsetting, song composers often align linguistically strong syllables with musically strong beats (Kimball, 1996), an observation that has been confirmed in analyses of corpora of English songs (Palmer and Kelly, 1992). In other words, stressed syllables, or syllables that one would expect to be stressed in speech (“strong” syllables), are more likely than unstressed (“weak”) syllables to occur on hierarchically prominent

musical beats. Musicians intuitively emphasize this correspondence, by lengthening well-aligned syllables more than misaligned ones. Furthermore, non-musicians have predictable, consistent intuitions about aligning lyrics to musical rhythms when asked to sing novel lyrics (i.e., new verses) to a familiar melody (Halle and Lerdahl, 1993) or chant lyrics without knowledge of the original melody (Hayes and Kaun, 1996).

It has been proposed that certain goals of music performance, including communication and memorization, may be fulfilled by the temporal alignment of strong syllables with musically strong beats (Palmer and Kelly, 1992). In fact, Johnson et al. (in revision) studied comprehension of sung lyrics in a sentence context by asking listeners to transcribe well-aligned words (strong syllables occurring on strong beats), and misaligned words (weak syllables occurring on strong beats). They found that intelligibility of sung lyrics was better when strong syllables aligned with strong metric positions in the music. Behavioral results from several studies suggest that songs that are well-aligned are also easier to memorize. For instance, Gingold and Abravanel (1987) reported that children had more difficulty learning songs with misaligned textsettings

than songs with well-aligned textsettings. Experiments with adults also point to the efficacy of emphasizing both the linguistic (Wallace and Rubin, 1988; Wallace, 1994) and musical (Purnell-Webb and Speelman, 2008) rhythm during song learning. To summarize, composers, performers, and naïve listeners appear to prefer well-aligned textsettings, which may be associated with more efficient encoding mechanisms.

One explanation for these observations is that attending and comprehension are facilitated when strong syllables occur on strong beats. It thus follows that intelligibility of sung lyrics would be degraded when word stress does not align with strong beats (Johnson et al., in revision). Misaligned words could be perceived as prosodically incongruous sung language, which disrupts attending and hinders comprehension, similar to the way that rhythmic incongruities in speech disrupt intelligibility (Tajima et al., 1997; Field, 2005). The present study was designed to uncover the brain mechanisms underlying the preference for metrical alignment, by testing the hypothesis that well-aligned textsettings help listeners to better understand song lyrics, by directing temporal attention to instances where strong syllables occur on strong beats.

Our approach is based on previous work on perception of speech prosody and musical meter. Several studies have shown that prosody is important for speech segmentation in adults (e.g., Cutler and Norris, 1988; Smith et al., 1989; Mattys et al., 2005) and children (e.g., Jusczyk et al., 1993). It has been proposed that word stress facilitates segmentation of the speech signal by directing temporal attention toward salient events in speech (Shields et al., 1974; Pitt and Samuel, 1990; Quené and Port, 2005): listeners attend more to strong syllables than weak syllables. Rhythmic expectations generated by metrical and syntactic cues also play an important role in predicting word stress (Kelly and Bock, 1988; Pitt and Samuel, 1990). Likewise, when the timing of musical notes is organized to aid the listener in predicting the timing of upcoming events (Schmidt-Kassow et al., 2009), attention is preferentially allocated to strong beats in music (Large and Jones, 1999).

Therefore, it is likely that listeners' temporal attention is focused by well-aligned, predictable textsettings in song, in which strong syllables and strong beats coincide. Cortical rhythms entrain to the temporal structure of acoustic signals (Snyder and Large, 2005; Zanto et al., 2005; Fujioka et al., 2009; Geiser et al., 2009; Nozaradan et al., 2011), and entrainment of neuronal oscillations can function as a mechanism of attentional selection (Lakatos et al., 2008; Stefanics et al., 2010). Entrainment of intrinsic neural processes explains temporal expectancy and perceptual facilitation of events that occur at expected points in time (Large and Jones, 1999; Jones, 2010). The experience of meter is posited to arise from interaction of neural resonances at different frequencies. In light of suggestions of similar mechanisms for perceiving meter in speech and music (Port, 2003; Patel et al., 2006; Patel, 2008; Marie et al., 2011), we sought to probe a possible influence of linguistic rhythm on neural correlates of musical rhythm by investigating prosodically driven changes in these neural responses to sung stimuli.

An experiment was thus designed to explore the idea that alignment of stressed syllables with strong beats in music helps to temporally focus listeners' attention, which in turn facilitates syllable segmentation and comprehension of the lyrics of a song.

Using a cross-modal priming paradigm, listeners were presented with sung sentences (primes) in which all, none, or some of the strong syllables occurred on strong musical beats. Strong beats were determined by a series of isochronous metronome clicks occurring on alternating syllables before and during the sentences. Three kinds of sentence alignment were used: *well-aligned* (regular stress pattern; all strong syllables occur on strong beats), *misaligned* (regular stress pattern; all weak syllables occur on strong beats), *varied* (irregular stress pattern: some strong syllables occur on strong beats, some occur on weak beats). To test the influence of different kinds of sung sentence alignment on a subsequently presented stimulus, each sentence was followed by a target word or pseudoword, presented visually, on which listeners were asked to perform a lexical decision task (i.e., discerning real words from nonsense words).

This approach exploits the ability of a metronome to induce different metrical interpretations (i.e. shifting of strong beats; see Iversen et al., 2009; Geiser et al., 2010; Vlek et al., 2011) on identical melodies, sung with words. An isochronous metronome click train was used to set up metrical expectancies before and during sung sentences, such that the perceived strong musical beats (marked by the clicks) were either aligned or not aligned with the expected linguistic stress of each sung syllable. Alignment was manipulated by shifting the same sung sentence by one quarter note (or shifting the metronome clicks by one quarter note; see Figure 1).

Our first hypothesis was that well-aligned and misaligned syllables would elicit different neural responses. It was expected that strong metronome beats would elicit synchronized rhythmic fluctuations in beta and gamma activity, in light of previous studies suggesting that high frequency EEG/MEG fluctuates with endogenous temporal and metrical expectancies, and persists even in the absence of a tone (Snyder and Large, 2005; Fujioka et al., 2009). If similar mechanisms are at work for perceiving rhythm in speech, then high frequency neural activity may coordinate with strong beats when they align with strong syllables. On the other hand, when a weak syllable is aligned with a strong beat, beat tracking could be disrupted, and therefore typical beta and gamma responses would be attenuated or suppressed. If the anti-phase relationship between strong syllables and strong beats in the misaligned sentences is perceived as syncopation, misaligned syllables may elicit decreased beta power, as reported for sensori-motor syncopation (Jantzen et al., 2001; Mayville et al., 2001).

The second hypothesis was that both alignment and regularity facilitate lexical access. If well-aligned and misaligned syllables are attended to differently online, during the sung sentence primes, then the interplay between linguistic and musical rhythm may also have a cumulative effect on attention to and perception of target words and pseudowords. This may influence the lexico-semantic search process, as shown in previous work on perception of spoken language (Schirmer et al., 2002; Dilley and McAuley, 2008). Thus, we predicted that misaligned and varied sentence primes would hinder target processing, and would be associated with poorer task performance [slower reaction times (RTs), more errors], compared to well-aligned sentence primes. It was therefore predicted that misaligned and varied primes would be more difficult to process than well-aligned primes and thus recruit more cognitive resources for their respective targets. These priming effects could stem from

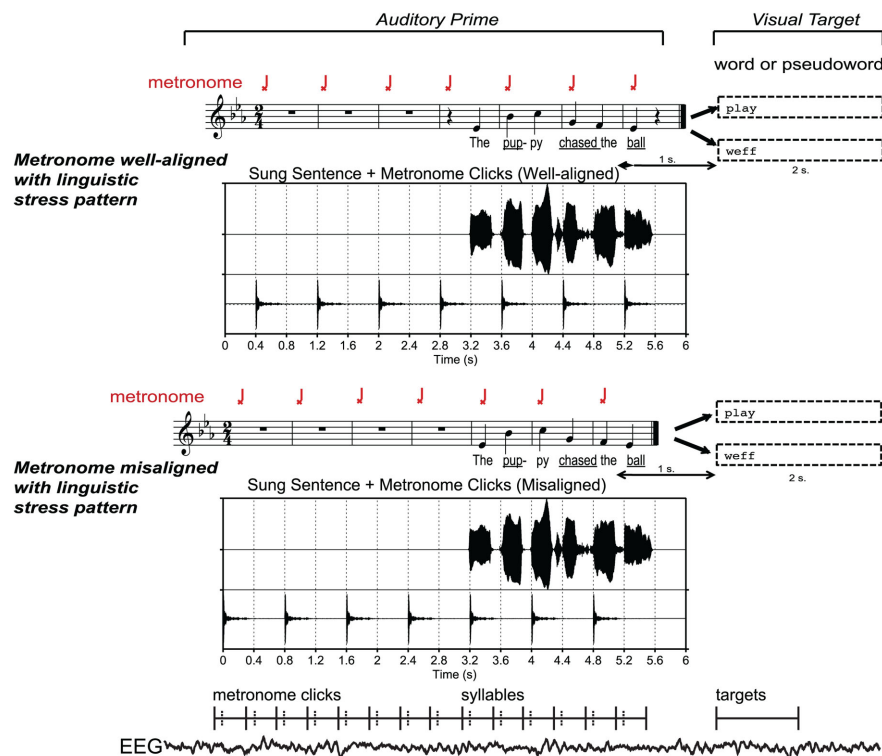


FIGURE 1 | Experimental paradigm and stimuli. Auditory sentence primes were created by aligning a sequence of isochronous metronome clicks with strong syllables in sung sentences (underlined), and shifting by one beat relative to metronome clicks on downbeats. The waveforms show that the same recording of sung sentence can be aligned with metronome clicks on strong syllables to form well-aligned sentences, or with weak syllables to form

misaligned sentences. Targets consisted of words or pseudowords presented visually following each auditory prime. The lower section of the figure gives a schematic view of time windows analyzed for the EEG responses time-locked to strong and weak beats (metronome clicks and syllables) and targets. Bold lines show the boundaries of each latency window, and dashed lines show zero points. See Section “Data Analysis – EEG” for further detail.

differences in domain-general memory mechanisms involved in lexical search (such as changes reported in alpha band activity by Doppelmayr et al., 2005; Grabner et al., 2007; Klimesch et al., 2007) and selective attention to linguistic stimuli, reflected in alpha and beta band suppression (Shahin et al., 2009; Van Elk et al., 2010).

MATERIALS AND METHODS

PARTICIPANTS

Nineteen volunteers participated in this experiment that lasted about 2 h including preparation time. The study was approved by the IRB committees at Florida Atlantic University and Middle Tennessee State University, and all participants gave written informed consent. By self-report, all participants had normal hearing, no known neurological problems, and were right-handed, native speakers of English. Data from three participants was discarded due to excessive ocular artifacts and line noise, thus the final group of subjects consisted of 16 subjects (nine males, mean age = 21.4 years, age range 18–34, mean number years of formal music lessons = 1.9).

STIMULI

Sung sentences

First, a corpus of 144 English sentences was created based on the type of sentences used in a previous study of meter in speech (Kelly

and Bock, 1988). Each sentence contained six syllables that formed three types of metrical patterns: trochaic-regular (strong-weak-strong-weak-strong-weak); iambic-regular (weak-strong-weak-strong-weak-strong); and mixed-irregular (e.g., weak-strong-strong-weak-weak-strong). See **Table 1** for examples of sentences used. The irregular sentences were controlled so that the total number of strong and weak syllables in each syllable position (first through sixth) was balanced over sentences.

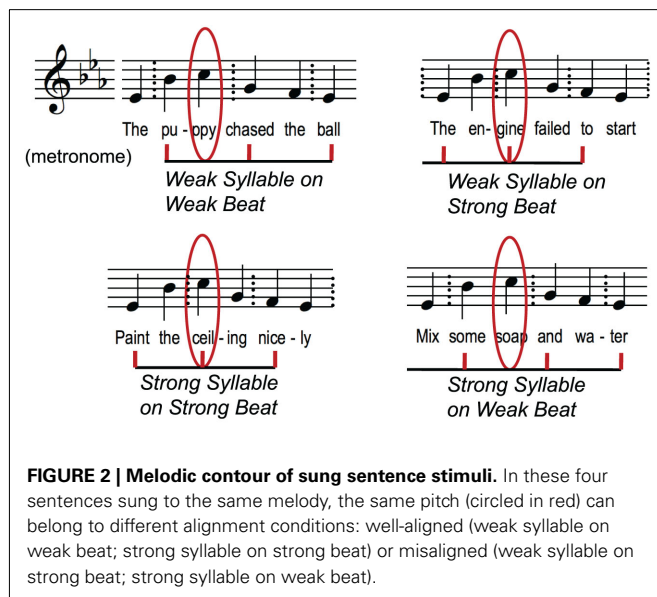
Each sentence from each of the three linguistic conditions was paired with a melody from a corpus of 24 tonal, isochronous pitch sequences composed for this experiment. The first and last notes of each melody were always the tonic, and the second through fifth notes followed an overall contour of rising then falling. Because pitch accents can contribute to the perception of linguistic stress and musical meter (Hannon et al., 2004; Jones, 2009), each melody was used once in each of the six experimental conditions. As shown in **Figure 2**, the peak in melodic contour in a given melody was held constant across experimental conditions and types of syllable alignment (strong or weak syllables on strong or weak beats), so any global effects of metrical alignment could not be attributed to differences in melodic accents across syllables.

Using the MIDI Toolbox (Eerola and Toivainen, 2004), the melodies were exported as MIDI (Musical Instrument Digital Interface) files in preparation for the digital recording session

Table 1 | Experimental design and examples.

Well-aligned prime + target word	
The <u>p</u> ilot <u>f</u> lew the <u>p</u> lane.	+ SKY
Well-aligned prime + target pseudoword	
<u>C</u> lose the <u>o</u> pen <u>w</u> indow.	+ SKULB
Misaligned prime + target word	
<u>C</u> lose the <u>o</u> pen <u>w</u> indow.	+ GLASS
Misaligned prime + target pseudoword	
The <u>p</u> ilot <u>f</u> lew the <u>p</u> lane.	+ GLOIL
Varied prime + target word	
<u>S</u> crub the <u>d</u> irt off your <u>b</u> ike.	+ SOAP
Varied prime + target pseudoword	
<u>S</u> crub the <u>d</u> irt off your <u>b</u> ike.	+ BILM

Strong syllables are underlined, and syllables occurring on strong musical beats are in bold.



with the software GarageBand (Apple, Cupertino, CA, USA). The recording session took place in the IAC chamber in the soundproof booth in the Music Dynamics Lab at FAU. Each sentence was sung by a trained singer (the first author, a soprano) using a metronome set at 150 beats-per-minute, targeting a duration of 400 ms per syllable. The sentences were sung with as little variation as possible in loudness or duration between syllables, regardless of their expected linguistic stress. Vowels in unstressed syllables were not reduced, as they would be in speech; instead, they were pronounced fully, as called for by traditional vocal pedagogy. Each of the 144 sung sentences were edited in Praat (Boersma and Weenink, 2007). Syllable and vowel boundaries were marked by hand for acoustic analyses, by visually inspecting the waveform and spectrogram, and by listening to the segments. The wave files containing the sung sentences were normalized in intensity by equalizing the maximum root-mean-squared (RMS) amplitude across wave files.

Next, acoustic analyses were performed to determine whether strong and weak syllables differed in duration and amplitude. There was no overall difference in RMS amplitude between strong (mean = 0.079, SD = 0.029) and weak (mean = 0.077,

SD = 0.028) syllables [$t(359) = 0.759, p = 0.449$], measured with two-tailed paired t -tests. Next, duration of the first five syllables in each sentence was computed by calculating the inter-vocalic interval (IVI; the interval between the onsets of vowels in adjacent syllables). The analyses showed that strong syllables were slightly shorter¹ (mean = 396 ms, SD = 30) than weak syllables (mean = 401 ms, SD = 29) [$t(359) = -2.458, p = 0.014$].

An isochronous metronome click train (woodblock sound; seven clicks) with an IOI of 800 ms was generated in Praat. The wave file containing the metronome click sounds was normalized in intensity to a level that sounded appropriate when played simultaneously with the normalized sung sentences. Two versions of each sung sentence were created by aligning the metronome click train on alternating sung syllables, from either the vowel onset of the first syllable of each sentence, or 400 ms post-vowel onset (corresponding to the onset of the second syllable). This process created two versions of each sentence, as illustrated in Figure 1, so that precisely the same syllable in each sentence, defined by a single set of boundaries, would be heard with and without a click, across subject lists. Each stimulus was then checked aurally to make sure that clicks were appropriately synchronized with the intended syllables. The alignment of metronome clicks and sung sentences yielded 96 well-aligned sentences, 96 misaligned sentences, and 96 varied sentences. Audio examples of stimuli can also be accessed on the online Supplementary Materials.

Targets

For each of the 144 sentences, a monosyllabic target word that was different from, but semantically related to, the content of the sentence was designated. Words were four or five letters in length, were either nouns or verbs, and part of speech was balanced over conditions. A corpus of 72 monosyllabic pseudowords was selected from the ARC database (Rastle et al., 2002). Pseudowords were orthographically plausible and were also matched in number of letters to the real words. Two experimental lists were created, such that half of the prime sentences in each condition, and in each list, were assigned to pseudoword targets, and half were assigned to (semantically related) real word targets. The same list of pseudowords appeared in both experimental lists but with different prime sentences, and in addition, word frequency was equalized across the two lists (frequency of occurrence in verbal language was derived from the London-Lund Corpus of English Conversation by Brown, 1984).

A Latin Square design ensured that real words and pseudowords would appear with sung sentences with both metronome click patterns across subjects, and no sentence would be heard in more than one condition within the list of stimuli heard by one participant. This counterbalancing created 144 sung sentences \times 2 metronome click patterns \times 2 lists of targets, for a total of 576 different pairs of primes and target, to yield four different experimental lists.

To summarize, an orthogonal manipulation of the factors “Prime Alignment” (well-aligned, misaligned, and varied) and

¹In English, strong syllables typically have longer durations than weak syllables (Ramus et al., 1999); here the IVI for strong syllables are only 5 ms shorter than for weak syllables. This small difference, if perceived, could only bias the results in a direction opposite the hypotheses, i.e., the increased duration would lead to weak syllables being perceived as stressed despite their expected linguistic stress.

“Target Lexicality” (Word, Pseudoword) yielded six experimental conditions listed in **Table 1**, across which the 144 sentences were divided evenly to yield 24 sentences per condition.

PROCEDURE

To familiarize participants with the task, each experimental session began with a block of practice trials consisting of sample stimuli representing each of the experimental conditions; the examples were not heard during the actual experiment. Participants were instructed to fixate on the cross that appeared at the beginning of each trial and to listen to the sung sentence, paying attention to the words and not to worry about the melody or rhythm, and then to decide whether the visual target shown on the screen after the sung sentence was a real word or a nonsense word. They were instructed to respond as quickly and accurately as possible by pressing one of two buttons on the response pad with the index and middle finger of their right hand. Each trial consisted of: a fixation period of 1000 ms, followed by wave file containing the sung sentence prime (6000 ms), and the visual target (2000 ms), which started 1000 ms after the onset of the final beat (see **Figure 1** for trial timing).

Participants then listened through earbuds, adjusted slightly to comfortable volume level for each individual, to the 144 trials from the six experimental conditions, presented in a pseudorandom order, and broken up into three blocks of 48 trials with breaks between each block. They were asked to avoid blinking until a series of X's appeared on the computer screen, at the end of each trial. Response key–finger association and stimulus lists were counter-balanced across participants. The software E-Prime (Psychology Software Tools Inc., Pittsburgh, PA, USA) was used for stimulus presentation and to record behavioral responses [RTs and percent correct (PC)].

EEG DATA ACQUISITION

Electroencephalogram was recorded continuously from 64 Ag/AgCl electrodes embedded in sponges in the Hydrocel Geodesic Sensor Net (EGI, Eugene, OR, USA) placed on the scalp with Cz at the vertex, connected to a NetAmps 300 high-impedance amplifier, using a MacBook Pro computer. Data was referenced online to Cz. All electrical equipment was properly grounded. The sampling frequency was 500 Hz with an anti-aliasing lowpass Butterworth filter of 4 kHz, and impedances were kept below 50 k Ω .

DATA ANALYSIS – BEHAVIORAL (TARGETS)

For each of the six experimental conditions listed in Section “Targets,” PC was computed on all responses. Mean RTs were computed on correct behavioral responses only. Analyses of behavioral data were conducted with e-prime and MATLAB (The MathWorks Inc., Natick, MA, USA) and the statistical software CLEAVE² was used to conduct a 3×2 ANOVA with factors: Alignment (*well-aligned* vs. *misaligned* vs. *varied*), Lexicality (*word* vs. *pseudoword*). The Bonferroni threshold was applied to pairwise *post hoc* tests on significant interactions to correct for multiple comparisons.

²www.nitrc.org/projects/cleave

DATA ANALYSIS – EEG

EEG preprocessing

Electroencephalographic preprocessing was carried out with Net-Station Viewer and Waveform tools. The EEG data was filtered offline with a pass band from 0.5 to 100 Hz. The data were then re-referenced offline to the algebraic average of the left and right mastoid sensors. To analyze responses to the lead-in metronome clicks in the primes, data were epoched at ± 750 ms from click onset (strong beats) and from the midpoint between clicks (weak beats) and then divided into two conditions: strong beats and weak beats. To analyze the EEG recorded during the sung sentences, data were epoched from -450 to $+1050$ ms from the onset of each syllable, which was defined as described in Section “Sung Sentences” (note that long epoch windows are required for wavelet calculation, but statistical analyses are not performed on overlapping time windows of the time–frequency representations (TFRs), as shown in **Figure 1** and described in Cluster Randomization Analysis). Epochs corresponding to prime syllables were divided into the following three conditions: strong syllable on strong beat; weak syllable on strong beat; strong syllable on weak beat. The irregular sentences (*varied* condition) were excluded from the syllable analysis because they became irregular beginning at different syllable positions. To analyze responses to the targets, data were segmented into epochs of -1000 to $+2000$ ms from the onset of the target word/pseudoword. Epochs corresponding to targets were then divided into three conditions: targets following well-aligned primes, misaligned primes, and varied primes. Artifact rejection was performed on epoched data, such that trials containing movement or ocular artifacts, or amplifier saturation were discarded, and likewise, several sensor channels containing excessive noise (in the occipital area) were excluded from further analysis in all subjects. Following artifact rejection, the mean (SD in parentheses) number of remaining trials per participant, in each dataset was as follows. For metronome click dataset: strong beats – 353(68), weak beats – 362(67); for the syllable dataset: Strong Syllable on Strong Beat – 130(20), Weak Syllable on Strong Beat – 130(17), Strong Syllable on Weak Beat 129(16); for the target dataset: Well-aligned – 38(8), Misaligned – 38(8), Irregular 38(8).

Time–frequency representations – evoked and induced

Time–frequency analysis was conducted using the FieldTrip open source toolbox³ (Oostenveld et al., 2011) to calculate both evoked and induced representations of EEG⁴. To obtain the evoked (phase-locked) response, TFRs were computed on the average waveform for each condition and each subject, at each time point and frequency, by convolving the average waveform with a Morlet

³<http://www.ru.nl/neuroimaging/fieldtrip>

⁴Evoked activity is calculated by averaging EEG/MEG activity across trials and then performing a time–frequency transformation on the average waveform (ERP), thus preserving only the oscillations whose phases are time-locked to the stimulus. Induced activity, on the other hand, is calculated by first performing time–frequency transformations on the EEG activity in single trials, and then averaging together the resulting oscillatory power for each single trial, thus preserving the average amplitude envelope (see Tallon-Baudry and Bertrand, 1999 for more information). Through this process, the induced activity includes oscillations that are not necessarily phase-locked and thus would have been eliminated from the evoked during averaging.

wavelet that had a width of six cycles (see Tallon-Baudry et al., 1996, and Herrmann et al., 2005 for more details on wavelet analysis of EEG data). The resulting frequency resolution was $sf = f/6$, and the temporal resolution was $st = 1/sf$, where f is the center frequency for each wavelet, sf is SD in the frequency domain, and st is the SD in the temporal domain. The convolution of wavelets was done between 8 and 50 Hz, with a frequency step of 1 Hz and a time step of 2 ms, between -750 and $+750$ ms from the zero point of metronome click data; between -250 and $+500$ ms for the sung syllables; and from -200 to $+1000$ ms for the targets. Only low gamma band (30–50 Hz) was analyzed, as measuring high gamma band activity would likely have required more trials per condition (to establish an acceptable signal-to-noise ratio, due to heightened effects of small-latency jitter in high frequencies) than were present in this study. To obtain the induced response (phase-invariant), TFRs were computed by convolving single trial data with a Morlet wavelet that had the same parameters as described for the evoked analysis, and averaging power over trials for each subject and condition.

In order to compensate for inter-individual variability in absolute power, a normalization procedure was used for both evoked and induced analyses. For the targets, the spectra for all six conditions were averaged together for each subject (and averaged over time across the entire time period for which power values were computed as described above), to obtain a value for each frequency at each channel, which then served as a baseline. For the metronome clicks and the syllables from the sung sentences, normalization was conducted by using the average power (across epochs) of the response to introductory metronome clicks as a baseline. The power spectra for each condition and each subject were then normalized with respect to these baseline values, resulting in normalized power, expressed as the percent change from the baseline.

Cluster-randomization analysis

Statistical significance between the normalized power spectra in different conditions was tested using cluster-randomization procedure (Maris and Oostenveld, 2007), to identify consistent trends of activity in time–frequency clusters of electrodes. This is a data-driven approach that is particularly useful for spatial localization without having to specify regions of interest *a priori*. This method utilizes planned comparisons between pairs of conditions using cluster-based permutation tests, a solution to the multiple comparisons problem for EEG/MEG data that is prevalent in recent studies of oscillatory activity in language cognition (see Bastiaansen et al., 2010 for a discussion of the advantages of this method). Clustering was performed by identifying power values (for each channel, at each frequency and time point) that showed a similar direction of effect, and then testing the significance of a cluster in a given condition compared to another condition in relation to clusters computed from a random permutation of values drawn from both conditions.

The power spectra were first divided into frequency bands: alpha (8–12 Hz), beta (13–29 Hz), and low gamma (30–50 Hz), and time windows of interest were defined (-100 to 300 ms for the metronome clicks and prime syllables, and 0 – 800 ms for the targets). Frequency bands for the analysis were chosen based

on hypotheses described in the Introduction (primes: beta and gamma; targets: alpha and beta). In each comparison, within a given frequency band⁵, the resulting power values at each time point and channel underwent two-tailed dependent *t*-tests, for each subject. Then, a cluster-level statistic was calculated by taking the sum of the *t*-statistics within every cluster. All data points that did not exceed the preset significance level ($p = 0.025$) were zeroed, and clusters were created by grouping together adjacent non-zero points. The next step was to compute significance probability of the clusters using the (non-parametric) Monte Carlo method⁶. The Monte Carlo significance probability (*p*-value) was determined by calculating the proportion of clusters from random partitions of the data that resulted in a larger test statistic than the clusters on the observed test statistic. *p*-Values < 0.05 were considered significant. Only pairwise comparisons are possible with the clustering method; these planned comparisons are presented in Table 2.

RESULTS

BEHAVIORAL RESULTS: INFLUENCE OF PRIME ALIGNMENT ON TARGET LEXICAL DECISION

In the Reaction Time data (see Table 3), there was a main effect of Lexicality [$F(1,15) = 24.89$, $p < 0.001$]. As predicted,

⁵There is a statistical trade-off involved in specifying whether the clustering procedure would be carried out in a each frequency band without averaging over frequencies, in order to identify only a subset of frequencies showing effects; this was possible on the metronome click data due to the high signal to noise ratio afforded by the number of epochs on the metronome clicks (see EEG Preprocessing for number of epochs kept after artifact rejection). As many fewer epochs were available for each type of prime syllables and targets, the signal-to-noise ratio was less favorable, and therefore, power was averaged within each frequency band to optimize statistical power when carrying out the clustering.

⁶First, the data from both experimental conditions in a given comparison were collected into a single set. Then, random draws were taken from the combined data set to form two subsets, on which the test statistic was computed by summing together the individual *t*-statistics for each data point in the cluster. This process of random partition and computing the test statistic was repeated 2500 times, and those results were collected into a histogram of test statistics, that was then compared to the test statistic of the actual data.

Table 2 | Planned pairwise comparisons for cluster-randomization analysis.

Events	Conditions to be compared	Frequency bands of planned analysis
Metronome	Strong beats (click) vs. weak beats (no click)	Beta, gamma
Syllables	Strong beats: strong vs. weak syllable	Beta, gamma
Syllables	Strong syllables: strong vs. weak beat	Beta, gamma
Targets	All targets for well-aligned vs. misaligned primes	Alpha, beta
Targets	All targets for well-aligned vs. varied primes	Alpha, beta
Targets	All targets for misaligned vs. varied primes	Alpha, beta

participants were always faster at detecting words than pseudowords. There was also a significant interaction of Alignment \times Lexicality [$F(2,30) = 4.12$, $p = 0.0262$]: RT differences were greater on targets following *varied* and *misaligned* primes vs. *well-aligned* primes (each pairwise *post hoc* $p < 0.001$). A follow-up ANOVA was then conducted on the Reaction Time difference (Word minus Pseudoword; see **Figure 3A**). There was a main effect of Alignment [$F(2,30) = 4.12$, $p = 0.0262$]. *Post hoc* tests revealed that *misaligned* led to a larger difference in the RTs ($p = 0.009$) compared to *well-aligned*, and *varied* vs. *well-aligned* also approached significance (*post hoc* $p = 0.0535$). Thus, alignment affects the speed of lexical decision.

In the PC data, performance was nearly at ceiling (see **Table 3**), and there were no main effects. However, the Alignment \times Lexicality interaction was significant [$F(2,30) = 4.11$; $p = 0.0264$]; participants were less accurate for pseudoword than word targets, but only for *varied* primes (pairwise *post hoc* $p < 0.001$). A follow-up one-way ANOVA was then conducted on the difference in PC scores (see **Figure 3B**), showing a main effect of Alignment [$F(2,30) = 4.11$; $p = 0.0264$]. *Post hoc* tests on PC revealed that the difference was significantly larger for *varied* than for *misaligned* ($p = 0.013$) and for *varied* vs. *well-aligned* ($p = 0.028$). Thus, predictability of stress patterns in prime sentences affects the detection of pseudowords.

RESULTS – EEG

Metronome clicks

Time–frequency analyses of the EEG recorded during the lead-in metronome clicks (prior to each sung sentence), averaged together over all the trials and over the clicks within each trial, showed differences in evoked and induced activity between strong beats

(clicks) and weak beats (half-way point between clicks). Two widespread clusters of electrodes (both cluster $ps < 0.001$) showed increased evoked power in the beta band (13–29 Hz) from –100 to 300 ms and in the gamma band (30–50 Hz) from –30 to 122 ms (see **Figure 4A**). Similarly, a cluster of electrodes ($p = 0.044$) over frontal–central regions showed increased induced beta (13–24 Hz) power for strong beats (clicks) vs. decreased power for weak beats, from –100 to 124 ms (see **Figure 4B**).

Syllables in sung sentence primes

Well-aligned syllables were compared to misaligned syllables by comparing strong and weak syllables that occurred on strong beats. Results from the clustering analysis of the prime syllable conditions are described in the following section and clusters significant at the $p < 0.05$ level are reported. Times are given relative to a beat onset time at 0 ms.

A cluster of electrodes (a tendency was found, $p = 0.051$) showed increased induced beta power for strong syllables vs. weak syllables (on strong beats), from –100 to 54 ms, over mostly frontal and central regions (see **Figure 5A**). This result was expected based on Snyder and Large (2005). Thus, similar to the strong beats during the lead-in metronome, strong syllables appear to be associated with increased early-latency high frequency power. This response was not observed when weak syllables occurred on strong beats; instead, the paired comparison showed they were associated with decreased early-latency power.

Next, we considered strong syllables, asking whether responses differed depending on whether they occurred to strong vs. weak beats. Weak and strong beats were compared, holding syllable strength constant (always strong). A significant cluster ($p = 0.035$) of electrodes mostly over left parietal, central, and right fronto-temporal regions (see **Figure 5B**) showed increased evoked gamma power for weak vs. strong beats (on strong syllables), from 116 to 208 ms.

Targets

Contrasting responses to targets preceded by *well-aligned* vs. *misaligned* vs. *varied* primes yielded four significant clusters for induced alpha and beta activity. All conditions showed increased power followed by decreased power relative to baseline (see middle column of **Figure 6**), but contrasting the conditions revealed differences in the degrees of increased and decreased power. Between 238 and 654 ms (cluster $p = 0.006$) from target onset, greater induced alpha power overall was observed for targets preceded by *well-aligned* compared to *varied* sentences, over right hemisphere regions. In a similar latency band (208–650 ms; cluster $p = 0.01$), *well-aligned* primes also led to greater induced alpha power in a

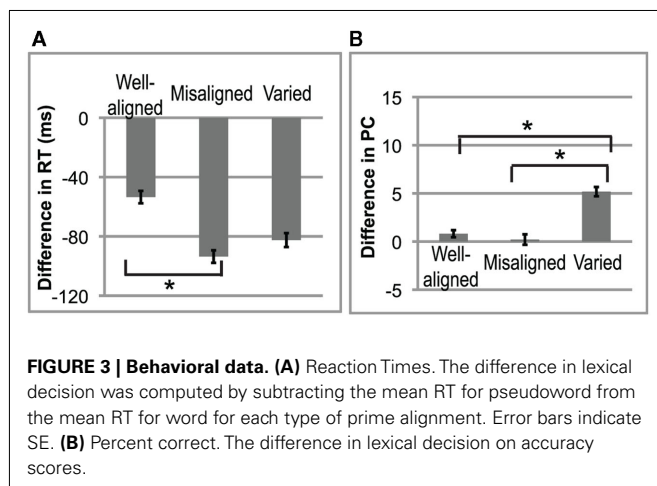
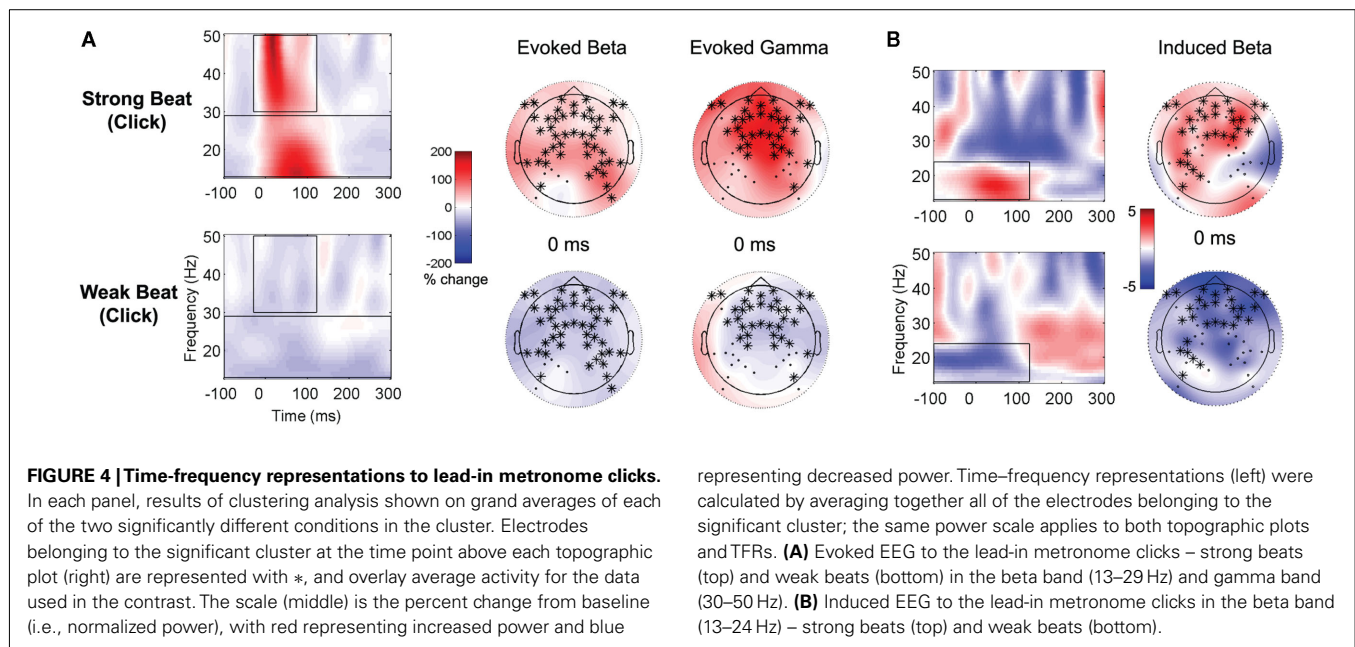


Table 3 | Behavioral data.

Condition	Well-aligned prime + target word	Well-aligned prime + target pseudoword	Misaligned prime + target word	Misaligned prime + target pseudoword	Varied prime + target word	Varied prime + target pseudoword
RTs (ms)	586 (122)	639 (131)	565 (114)	659 (145)	567 (124)	650 (113)
PC	96 (5)	95 (8)	96 (6)	96 (9)	98 (4)	93 (8)

Mean reaction times (RTs) in millisecond and percent correct (PC) for each condition, with SD indicated in parentheses.

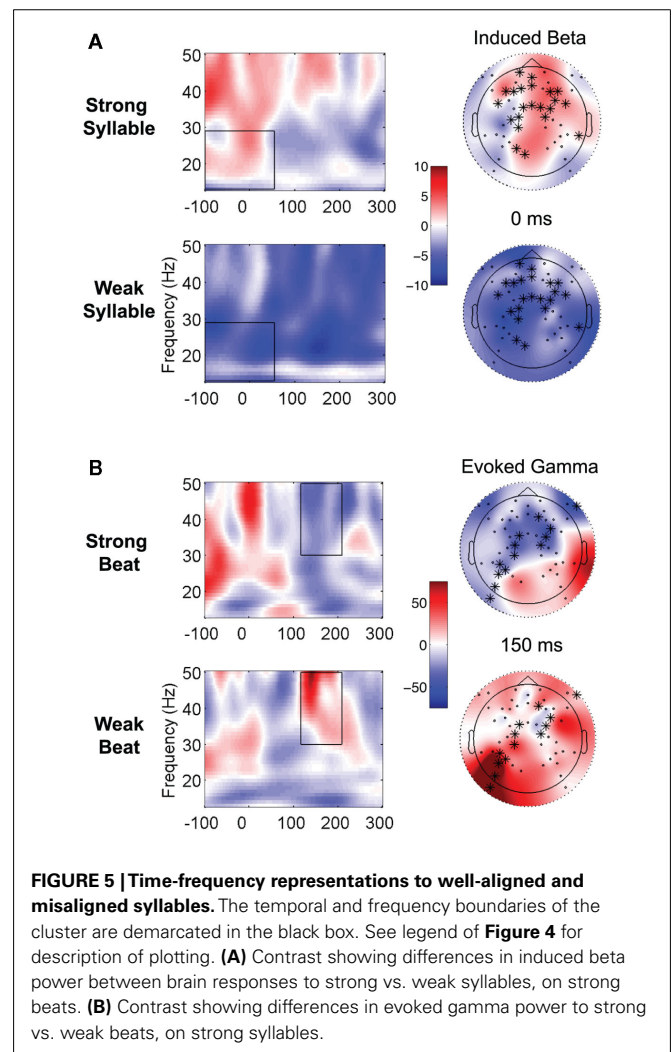


widespread cluster of electrodes, compared to *misaligned* primes (see **Figure 6**, left column).

Next, *well-aligned* primes led to greater early increased induced beta power for targets compared to *misaligned* primes, which led to greater mid-latency decreased power (see **Figure 6**, right column). The activity associated with this cluster of electrodes began in posterior and spread to central regions, stretching from 0 to 526 ms (cluster $p = 0.002$). Finally, *varied* compared to *misaligned* primes led to greater induced beta power, between 228 and 552 ms (cluster $p = 0.03$), starting over right hemisphere regions and then shifting to left hemisphere around 400 ms. Overall, *well-aligned* primes led to greater alpha and beta power on targets, compared to *misaligned* and *varied*. Thus, the pattern of increased followed by decreased alpha and beta power on visual linguistic targets seems to be modulated by the sentence alignment of the preceding prime. No significant clusters were found for the evoked data.

DISCUSSION

In this study, participants listened to sung sentences with different kinds of alignment between linguistic stress and musical meter, and performed a lexical decision task in which they discerned words from pseudowords. Time–frequency analysis of EEG data acquired during the introductory metronome clicks showed expected patterns of increased high frequency power (evoked beta and gamma, and induced beta), and during the sung sentences showed increased induced beta power for well-aligned syllables (strong syllables on strong beats). However, these patterns of brain activity were modified for misaligned syllables. Notably, decreases in induced beta power were associated with the onsets of weak syllables that occurred on strong beats. Moreover, when strong syllables occurred on weak beats, a delayed peak in evoked gamma power was observed. Visual targets primed by misaligned and varied sentences were associated with greater decreases in induced alpha and beta power compared to well-aligned sentences, primarily over central and right regions of the scalp.



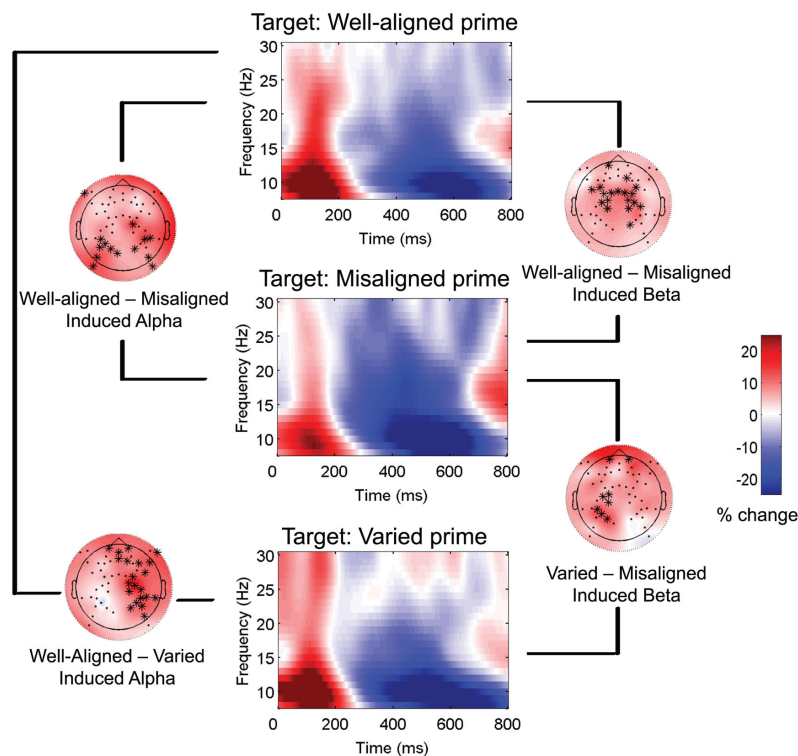


FIGURE 6 | Time-frequency representations to targets: Induced alpha and beta. Brain responses to targets (words and pseudowords combined), primed by three types of sentences: Well-aligned, Misaligned, and Varied. Time-frequency representations (center column) for each condition were

calculated by averaging together all channels. Results of clustering pairs of conditions are shown for induced alpha (left column) and induced beta (right column): topographic plots show difference in scalp power and topography of cluster at 400 ms, between the pair of conditions indicated.

BEAT TRACKING

The lead-in metronome clicks were associated with increased induced beta, evoked beta and evoked gamma power at their onsets, as predicted from the literature on rhythmic tone sequences. The evoked beta and gamma may reflect the perceived onsets and metrical strength of tones (Zanto et al., 2005; Iversen et al., 2009), while the induced beta likely reflects top-down temporal expectancies (Snyder and Large, 2005). By reproducing previous findings, we were able to show that neural attending rhythms are set up during these isochronous metronome clicks prior to the sentence onsets.

EFFECT OF LINGUISTIC AND MUSICAL ALIGNMENT

When strong syllables occurred on strong beats, increased induced beta power was observed relative to weak syllables on strong beats. This finding of increased power at strong beats in the lead-in metronome, and at strong syllables when they occurred on strong beats during the sung sentences, are consistent with the literature showing that the timing of high frequency bursts entrains to the beat in rhythmic tone sequences (e.g., Zanto et al., 2005; Fujioka et al., 2009). Thus, weak syllables, in this contrast, were associated with decreased induced beta power, despite the fact that they occurred on strong musical beats. What is surprising and novel about this finding is that the high frequency response appears not to be entrained to the isochronous metronome clicks during the misaligned sentences. One interpretation of these observations

is that beat tracking is sustained when strong syllables occur on strong beats, whereas it is disturbed when weak syllables occur on strong beats. However, the decreased induced beta power on weak syllables also agrees with results of Mayville et al. (2001) and Jantzen et al. (2001) showing beta desynchronization while participants performed syncopated finger movements compared to synchronized finger movements. Yet if the strong syllables were perceived as syncopated in relation to metronome clicks (strong beats), one might still expect to observe some relative increase in induced beta power on weak syllables that occur on strong beats (Iversen et al., 2009). Instead, beat tracking itself seems to have been disrupted to the point that beta and gamma responses were no longer timelocked to the isochronous metronome, providing an interesting exception to an otherwise consistent phenomenon in studies using tone sequences (Snyder and Large, 2005; Zanto et al., 2005; Fujioka et al., 2009; Iversen et al., 2009).

If beat tracking was disrupted when weak syllables occur on strong beats, would syllable tracking also be disrupted when strong syllables occurred on weak beats? Strong syllables on strong beats were thus contrasted with strong syllables on weak beats. A decrease in evoked gamma power was observed for strong syllables on strong beats at around 150 ms post-click-onset, consistent with the literature showing periodic high frequency synchronization followed by desynchronization during perception of rhythmic patterns (Fujioka et al., 2009). Furthermore, this cluster showed that strong syllables on weak beats were associated with increased

evoked gamma power at around 150 ms. Here both conditions have strong syllables; therefore, syllable tracking seems to be affected by misaligned syllable and beats. This finding suggests that the brain shifts attending rhythms toward strong syllables rather than strong beats during perception of misaligned sung sentences. It remains to be determined whether the addition of singing is specifically responsible for the change in brain response to syncopation, or whether similar results would be observed in more complex syncopated sequences consisting of several instruments, as in orchestral music.

Induced beta activity seems to be involved in predicting the timing of tones in ongoing musical rhythms (Snyder and Large, 2005; Zanto et al., 2005), which are likely related to beta-mediated communication between auditory and motor areas (Fujioka et al., 2009; Fujioka et al., in revision). Evoked beta and gamma fluctuate in response to the presentation of a tone (Snyder and Large, 2005), with evoked beta showing power changes proportional to perceived metrical strength, even in the absence of physical differences in the intensity of tones (Iversen et al., 2009). Evoked gamma responses also discriminate between speech and non-speech sounds within 50 ms from stimulus onset (Palva et al., 2002). Thus, the current findings fit with the literature by differentiating functional roles of induced and evoked EEG. Here induced beta appears to reflect rhythmic anticipation of alignment of strong syllables with strong beats, while evoked EEG for strong syllables appears to be involved in mediating expectations and perception of the onsets of strong syllables.

Taken together, when the syllables and beats are in-phase, the neural response appears to easily entrain to the strong beats/syllables, which occur in alignment. By showing that attentional pulses synchronize to temporally predictable input, this finding is predicted by the dynamic attending theory (Large and Jones, 1999), and corroborated by recent studies on dynamic attending in music (e.g., Snyder and Large, 2005; Geiser et al., 2009, 2010; Nozaradan et al., 2011) and language (e.g., Port, 2003). These observations also coincide with phoneme monitoring data showing that attention is preferentially allocated to strong syllables (Shields et al., 1974; Pitt and Samuel, 1990; Quené and Port, 2005; Wang et al., 2005), which aid in segmenting the speech signal (e.g., Smith et al., 1989). Analyses of speech production and sensorimotor synchronization have also indicated that syllable stress acts as an attractor that captures attention and facilitates production of periodic, perceptually salient events (Kelso et al., 1983; Port, 2003; Lidji et al., 2011). In the present study, the metronome seems to build up expectations that in turn influence the processing of syllable stress in the sung sentences. Given the differences observed here in prime sentences, the subsequently presented targets were then analyzed to uncover potential differences in targets primed by sentences with different types of alignment.

PRIMING EFFECTS ON TARGETS

The brain response to target stimuli was analyzed to investigate whether different types of rhythmic alignment in the primes would affect the perception of visual words and pseudowords. Varied and misaligned sentences, compared to well-aligned sentences, led to lower levels of induced alpha power during presentation of the targets. The decrease in alpha power could be due to increased

top-down attentional demands (Fan et al., 2007; Klimesch et al., 2007) when some or all of the syllables in the prime sentence are misaligned and therefore do not satisfy the type of prosodic expectancies that typically facilitate speech comprehension (e.g., Dilley et al., 2010). Alpha band (de)synchronization is likely related to domain-general mechanisms, such as attention and memory, that play an important role in lexico-semantic processing (Klimesch et al., 1999; Bastiaansen and Hagoort, 2006; Krause et al., 2006). This induced alpha activity seen for the alignment effects on the targets is primarily over the right hemisphere, perhaps related to fMRI data showing right hemispheric lateralization for prosodic aspects of speech (Glasser and Rilling, 2008).

Induced beta suppression (i.e., decreased power) was observed when contrasting targets primed by misaligned vs. well-aligned sentences. This cluster was already significant at the onset of the target stimulus and stretched past 500 ms over central regions, thus beginning earlier and lasting for longer than the induced alpha effects of alignment on the targets for well-aligned vs. misaligned and varied. As seen with the induced alpha in this same pair of conditions, this cluster of beta activity included more electrodes over the right than left hemisphere; however, the difference in lateralization was slight, and only significant at the earlier latencies. The fact that the differences in beta power were still present around the onset of the target stimulus strongly suggests that it reflects an ongoing process of rhythmic cognition in which the brain responds differently to the misaligned primes than the well-aligned primes, over a sustained time period that continues after the offset of the auditory stimulus. Furthermore, another cluster of electrodes showed greater beta suppression for targets primed by misaligned compared to varied primes, suggesting that the proportion of misaligned syllables (all, in the misaligned prime) modulates beta power more than the less predictable alignment and the irregular stress patterns in the varied sentences do.

Induced alpha and beta suppression have been specifically linked to the attention-related modulations required to carry out a semantic task on spoken words (Shahin et al., 2009). In the present study, misaligned sentences may have required greater attentional resources as participants attempted to understand them and search for semantically related targets in the lexicon, leading to the alpha and beta suppression on the targets. Interestingly, listeners' highly focused attentional pulses to the metronome and the lyrics seem to work together in the well-aligned to actually require fewer cognitive resources to understand sentences and conduct lexical search on the targets. That efficient use of periodic fluctuations in well-aligned linguistic and musical rhythms stands in contrast to the less efficient strong beats and stress in opposition, in the misaligned and varied, which seem to lead to heavier overall attentional load and more labored processing of the targets.

Taken together, these results could be interpreted as follows. In the well-aligned sentences, it is easy for participants to attend to strong beats and strong syllables, because they occur together (in-phase). Unified accent structure helps the listener to track beats (induced beta) and strong syllables (evoked gamma) more efficiently, which may facilitate word segmentation in the prime sentences. These well-aligned primes facilitate early stages of processing of targets, reflected in increased alpha and beta power. In the misaligned sentences, strong syllables and strong beats occur in

anti-phase, which may interfere with beat tracking (induced beta) and weaken/delay syllable tracking (evoked gamma). Syncopated structure between linguistic and musical rhythm may hinder segmentation of misaligned words, thus affecting task performance. If intelligibility, and thus clarity of the linguistic content, is weakened by the misalignment, then attentional demands would be greater for the lexical decision task, shown by alpha and beta suppression. Overall, these results of this study suggest that well-aligned textsettings improve rhythmic attending and linguistic comprehension⁷. In fact, Magne et al. (2007) reported that metrical stress violations in French words modulated lexico-semantic processing, even when participants were not attending to the prosodic dimension. The present results may thus be related to other electrophysiological evidence that prosodic expectations are involved in lexico-semantic retrieval at the local (word) level.

Perhaps the most intriguing aspect of the present findings is that induced beta modulations in both the prime syllables and target stimuli differentiate between well-aligned and misaligned conditions. Strong syllables on strong beats are associated with anticipatory beta increased power; increased beta power is also present at the onset of target stimuli preceded by those same well-aligned sentences. Weak syllables on strong beats show anticipatory decreased power, and relatively decreased beta power also characterizes the onset of targets that are preceded by misaligned sentences. The overall increase in beta power for well-aligned sentences may thus carry over into the presentation of its targets. This interpretation would be consistent with results on sensori-motor syncopation (Jantzen et al., 2001; Mayville et al., 2001), suggesting that induced beta suppression on the targets is specifically linked to syncopation between strong syllables and strong beats in misaligned sentences. Non-proximal, or distal, prosodic cues in word sequences can also affect not only perception of subsequently presented, lexically ambiguous target items, but also visual recognition of previously presented words (Dilley and McAuley, 2008; Dilley et al., 2010). In interpreting these effects, the authors discuss the possibility that internal oscillators are involved in building up periodic expectancies that form prosodic context and affect grouping, as in music (Dilley et al., 2010). Future work could utilize song prosody to bias grouping of lexically ambiguous syllables, by manipulating linguistic and musical meter in the “song” modality, thus shedding light on the extent to which metrically driven hierarchical grouping is shared by language and music.

CONCLUSION

We found that beta power increases when strong syllables align with strong beats, supporting the idea that metrical attending is facilitated by alignment; this increase on strong beats disappears when weak syllables are aligned with strong beats, suggesting that misalignment disrupts metrical attending. Overall, these results may provide a first step toward an explanation for the widespread

tendency to align strong syllables and strong beats in songs. Periodic peaks in attentional energy (e.g., as described by Large and Jones, 1999), manifested in neural responses, would be facilitated when linguistic stress and musical beats are in-phase, and disrupted when they are in anti-phase. Typical textsetting behaviors on the part of composers, such as placing stressed syllables on strong metrical positions, and on the part of performers, such as lengthening well-aligned syllables more than misaligned ones (Palmer and Kelly, 1992) could be accounted for by the present findings. Moreover, by showing that alignment modulates lexical decision-making, these results are an encouraging first step in understanding why good textsettings are easier to learn and memorize (Gingold and Abravanel, 1987) and how the rhythmic dimension of lyrics facilitates memorization (Wallace, 1994; Purnell-Webb and Speelman, 2008).

Many further studies on song prosody will need to be undertaken to tease apart these influences; for example, sung words (in a given grammatical category) whose stress is well-aligned or misaligned with the musical meter, could be presented in isolation, or in a consistent word position within sentences, to reduce syntactical processing demands. The comprehension and lexical access of sung words with varying types of alignment and rhythms could be tested directly, while controlling for other linguistic and musical factors that influence intelligibility in songs (Johnson et al., in revision). Caution should be exerted in generalizing these tendencies to different linguistic systems, as certain languages may be more permissible than English toward misalignment in songs (e.g., Spanish: Rodríguez-Vázquez, 2006, and French: Dell and Halle, 2005). In addition to stress-to-beat alignment, alignment is usually achieved through durational accents (Palmer and Kelly, 1992), which of course, are important markers of stress in speech prosody (Ramus et al., 1999) and should thus be a focus of future studies on the neural basis of song prosody. The present work does provide a starting place for using techniques such as metrical strength manipulations in songs, time–frequency analysis of EEG data, and cross-modal priming paradigms, to investigate the neural correlates of textsetting and better understand the interplay between linguistic and musical rhythm in song. Continuing research in this area may shed light on the idea that shared mechanisms for processing linguistic rhythm and musical rhythm contribute to the universal appeal of song.

ACKNOWLEDGMENTS

R. L. Gordon was supported by a Dissertation Fellowship from the American Association for University Women in 2007–2008 and the Daniel B. Newell and Aurel B. Newell Doctoral Fellowship in 2008–2009. E. W. Large was supported by the Air Force Office of Scientific Research (AFOSR FA9550-07-C0095). The authors gratefully acknowledge Steven Bressler, Caroline Palmer, Betty Tuller, and Alexandra Key for helpful comments on earlier versions of the manuscript. Portions of this study were adapted from R. L. Gordon's doctoral dissertation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/Auditory_Cognitive_Neuroscience/10.3389/fpsyg.2011.00352/abstract

⁷While the results do strongly suggest that comprehension is *better* in the well-aligned condition, the present experiment does not allow us to determine whether linguistic comprehension during the well-aligned condition is actually facilitated, or whether comprehension of well-aligned is similar to speech (which does not usually have a musical rhythm), but comprehension in the misaligned and varied conditions is in fact hindered. This is often the challenge in the devising experiments to demonstrate how speech prosody facilitates segmentation.

REFERENCES

- Bastiaansen, M., and Hagoort, P. (2006). "Oscillatory neuronal dynamics during language comprehension," in *Progress in Brain Research*, eds C. Neuper and W. Klimesch (Amsterdam: Elsevier), 179–196.
- Bastiaansen, M., Magyari, L., and Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *J. Cogn. Neurosci.* 22, 1333–1347.
- Boersma, P., and Weenink, D. (2007). *Praat: Doing Phonetics by Computer*. Available at <http://www.praat.org> [accessed February 6, 2007].
- Brown, G. D. A. (1984). A frequency count of 190,000 words in the London-Lund Corpus of English Conversation. *Behav. Res. Methods Instrum. Comput.* 16, 502–532.
- Cutler, A., and Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *J. Exp. Psychol. Hum. Percept. Perform.* 14, 113–121.
- Dell, F., and Halle, J. (2005). "Comparing musical textsetting in French and in English songs," in *Conference Proceedings from Typology of Poetic Forms*, ed. J. L. Aroui, Paris.
- Dilley, L. C., Mattys, S. L., and Vinke, L. (2010). Potent prosody: comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *J. Mem. Lang.* 63, 274–294.
- Dilley, L. C., and McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *J. Mem. Lang.* 59, 294–311.
- Doppelmayr, M., Klimesch, W., Hodlmoser, K., Sauseng, P., and Gruber, W. (2005). Intelligence related upper alpha desynchronization in a semantic memory task. *Brain Res. Bull.* 66, 171–177.
- Eerola, T., and Toivianen, P. (2004). *MIDI Toolbox: MATLAB Tools for Music Research*. Available at <http://www.jyu.fi/musica/miditoolbox/> [accessed June 30, 2005].
- Fan, J., Byrne, J., Worden, M. S., Guise, K. G., Mccandliss, B. D., Fossella, J., and Posner, M. I. (2007). The relation of brain oscillations to attentional networks. *J. Neurosci.* 27, 6197–6206.
- Field, J. (2005). Intelligibility and the listener: the role of lexical stress. *TESOL Q.* 39, 399–423.
- Fujioka, T., Trainor, L. J., Large, E. W., and Ross, B. (2009). Beta and gamma rhythms in human auditory cortex during musical beat processing. *Ann. N. Y. Acad. Sci.* 1169, 89–92.
- Geiser, E., Sandmann, P., Jancke, L., and Meyer, M. (2010). Refinement of metre perception – training increases hierarchical metre processing. *Eur. J. Neurosci.* 32, 1979–1985.
- Geiser, E., Ziegler, E., Jancke, L., and Meyer, M. (2009). Early electrophysiological correlates of meter and rhythm processing in music perception. *Cortex* 45, 93–102.
- Gingold, H., and Abravanel, E. (1987). Music as a mnemonic: the effects of good- and bad-music setting on verbatim recall of short passages by young children. *Psychomusicology* 7, 25–39.
- Glasser, M. F., and Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cereb. Cortex* 18, 2471–2482.
- Gordon, R. L., Schön, D., Magne, C., Astésano, C., and Besson, M. (2010). Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PLoS ONE* 5, e9889. doi:10.1371/journal.pone.0009889
- Grabner, R. H., Brunner, C., Leeb, R., Neuper, C., and Pfurtscheller, G. (2007). Event-related EEG theta and alpha band oscillatory responses during language translation. *Brain Res. Bull.* 72, 57–65.
- Halle, J., and Lerdahl, F. (1993). A generative textsetting model. *Curr. Musicol.* 55, 3–23.
- Hannon, E. E., Snyder, J. S., Eerola, T., and Krumhansl, C. L. (2004). The role of melodic and temporal cues in perceiving musical meter. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 956–974.
- Hayes, B., and Kaun, A. (1996). The role of phonological phrasing in sung and chanted verse. *Linguist. Rev.* 13, 243–303.
- Herrmann, C. S., Grigutsch, M., and Busch, N. A. (2005). "EEG oscillations and wavelet analysis," in *Event-Related Potentials: A Methods Handbook*, ed. T. C. Handy (Cambridge, MA: MIT Press), 229–259.
- Iversen, J. R., Repp, B. H., and Patel, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Ann. N. Y. Acad. Sci.* 1169, 58–73.
- Jantzen, K. J., Fuchs, A., Mayville, J. M., Deecke, L., and Kelso, J. A. (2001). Neuromagnetic activity in alpha and beta bands reflect learning-induced increases in coordinative stability. *Clin. Neurophysiol.* 112, 1685–1697.
- Jones, M. R. (2009). "Musical time," in *Oxford Handbook of Music Psychology*, eds S. Hallam, I. Cross, and M. Thaut (New York: Oxford University Press), 81–92.
- Jones, M. R. (2010). "Attending to sound patterns and the role of entrainment," in *Attention and Time*, eds A. C. Nobre and J. T. Coull (New York: Oxford University Press), 317–330.
- Juszyk, P. W., Cutler, A., and Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Dev.* 64, 675–687.
- Kelly, M. H., and Bock, J. K. (1988). Stress in time. *J. Exp. Psychol. Hum. Percept. Perform.* 14, 389–403.
- Kelso, J. A. S., Tuller, B., and Harris, K. S. (1983). "A 'dynamic pattern' perspective on the control and coordination of movement," in *The Production of Speech*, ed. P. F. Macneilage (New York: Springer-Verlag), 138–173.
- Kimball, C. (1996). *Song: A Guide to Style and Literature*. Washington: Pst Inc.
- Klimesch, W., Doppelmayr, M., Schwaiger, J., Auinger, P., and Winkler, T. (1999). 'Paradoxical' alpha synchronization in a memory task. *Brain Res. Cogn. Brain Res.* 7, 493–501.
- Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res. Rev.* 53, 63–88.
- Krause, C. M., Gronholm, P., Leinonen, A., Laine, M., Sakkinen, A. L., and Soderholm, C. (2006). Modality matters: the effects of stimulus modality on the 4- to 30-Hz brain electric oscillations during a lexical decision task. *Brain Res.* 1110, 182–192.
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113.
- Large, E. W., and Jones, M. R. (1999). The dynamics of attending: how we track time varying events. *Psychol. Rev.* 106, 119–159.
- Lidji, P., Jolicoeur, P., Moreau, P., Kolinsky, R., and Peretz, I. (2009). Integrated preattentive processing of vowel and pitch: a mismatch negativity study. *Ann. N. Y. Acad. Sci.* 1169, 481–484.
- Lidji, P., Palmer, C., Peretz, I., and Morningstar, M. (2011). Listeners feel the beat: entrainment to English and French speech rhythms. *Psychon. Bull. Rev.* doi:10.3758/s13423-011-0163-0
- Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., and Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: behavioral and electrophysiological evidence. *Cereb. Cortex* 17, 2659–2668.
- Marie, C., Magne, C., and Besson, M. (2011). Musicians and the metric structure of words. *J. Cogn. Neurosci.* 23, 294–305.
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190.
- Mattys, S. L., White, L., and Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *J. Exp. Psychol. Gen.* 134, 477–500.
- Mayville, J. M., Fuchs, A., Ding, M., Cheyne, D., Deecke, L., and Kelso, J. A. (2001). Event-related changes in neuromagnetic activity associated with syncopation and synchronization timing tasks. *Hum. Brain Mapp.* 14, 65–80.
- Nozaradan, S., Peretz, I., Missal, M., and Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* 31, 10234–10240.
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 156869.
- Palmer, C., and Hutchins, S. (2006). "What is musical prosody?" in *Psychology of Learning and Motivation*, ed. B. H. Ross (Amsterdam: Elsevier Press), 245–278.
- Palmer, C., and Kelly, M. (1992). Linguistic prosody and musical meter in song. *J. Mem. Lang.* 31, 525–542.
- Palva, S., Palva, J. M., Shtyrov, Y., Kujala, T., Ilmoniemi, R. J., Kaila, K., and Naatanen, R. (2002). Distinct gamma-band evoked responses to speech and non-speech sounds in humans. *J. Neurosci.* 22, RC211.
- Patel, A. D. (2008). *Music, Language, and the Brain*. New York: Oxford University Press.
- Patel, A. D., Iversen, J. R., and Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: the case of British English and French. *J. Acoust. Soc. Am.* 119, 3034–3047.
- Pitt, M. A., and Samuel, A. G. (1990). The use of rhythm in attending to speech. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 564–573.
- Port, R. (2003). Meter and speech. *J. Phon.* 31, 599–611.
- Purnell-Webb, P., and Speelman, C. P. (2008). Effects of music on memory for text. *Percept. Mot. Skills* 106, 927–957.
- Quené, H., and Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica* 62, 1–13.
- Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265–292.
- Rastle, K., Harrington, J., and Coltheart, M. (2002). 358,534 nonwords: the ARC nonword database. *Q. J. Exp. Psychol.* 55A, 1339–1362.

- Rodríguez-Vázquez, R. (2006). The metrics of folk song: a comparative study of text-setting in Spanish and English. *Rhythmica Revista Española de Métrica Comparada* 3–4, 253–281.
- Sammler, D., Baird, A., Valabregue, R., Clement, S., Dupont, S., Belin, P., and Samson, S. (2010). The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. *J. Neurosci.* 30, 3572–3578.
- Schirmer, A., Kotz, S. A., and Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Brain Res. Cogn. Brain Res.* 14, 228–233.
- Schmidt-Kassow, M., Schubotz, R. I., and Kotz, S. A. (2009). Attention and entrainment: P3b varies as a function of temporal predictability. *Neuroreport* 20, 31–36.
- Schön, D., Gordon, R., Campagne, A., Magne, C., Astesano, C., Anton, J. L., and Besson, M. (2010). Similar cerebral networks in language, music and song perception. *Neuroimage* 51, 450–461.
- Shahin, A. J., Picton, T. W., and Miller, L. M. (2009). Brain oscillations during semantic evaluation of speech. *Brain Cogn.* 70, 259–266.
- Shields, J. L., Mchugh, A., and Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *J. Exp. Psychol.* 102, 250–255.
- Smith, M. R., Cutler, A., Butterfield, S., and Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *J. Speech Hear. Res.* 32, 912–920.
- Snyder, J. S., and Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cogn. Brain Res.* 24, 117–126.
- Stefanics, G., Hangya, B., Hernadi, I., Winkler, I., Lakatos, P., and Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *J. Neurosci.* 30, 13578–13585.
- Tajima, K., Port, R., and Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *J. Phon.* 25, 1–24.
- Tallon-Baudry, C., and Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn. Sci. (Regul. Ed.)* 3, 151–162.
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., and Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *J. Neurosci.* 16, 4240–4249.
- Van Elk, M., Van Schie, H. T., Zwaan, R. A., and Bekkering, H. (2010). The functional role of motor activation in language processing: motor cortical oscillations support lexical-semantic retrieval. *Neuroimage* 50, 665–677.
- Vlek, R. J., Schaefer, R. S., Gielen, C. C., Farquhar, J. D., and Desain, P. (2011). Sequenced subjective accents for brain-computer interfaces. *J. Neural Eng.* 8, 036002.
- Wallace, W. T. (1994). Memory for music: effect of melody recall on text. *J. Exp. Psychol.* 20, 1471–1485.
- Wallace, W. T., and Rubin, D. C. (1988). “The Wreck of the Old 97”: a real event remembered in song,” in *Remembering Reconsidering: Ecological and Traditional Approaches to the Study of Memory*, eds U. Niesser and E. Winograd (Cambridge: Cambridge University Press), 283–310.
- Wang, J., Friedman, D., Ritter, W., and Bersick, M. (2005). ERP correlates of involuntary attention capture by prosodic salience in speech. *Psychophysiology* 42, 43–55.
- Zanto, T. P., Large, E. W., Fuchs, A., and Kelso, J. A. S. (2005). Gamma-band responses to perturbed auditory sequences: evidence for synchronization of perceptual processes. *Music Percept.* 22, 535–552.

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: 02 April 2011; accepted: 09 November 2011; published online: 29 November 2011.

Citation: Gordon RL, Magne CL and Large EW (2011) EEG correlates of song prosody: a new look at the relationship between linguistic and musical rhythm. *Front. Psychology* 2:352. doi: 10.3389/fpsyg.2011.00352

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

Copyright © 2011 Gordon, Magne and Large. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits use, distribution, and reproduction in other forums, provided the original authors and source are credited.



Effects of practice and experience on the arcuate fasciculus: comparing singers, instrumentalists, and non-musicians

Gus F. Halwani^{1,2}, Psyche Loui², Theodor Rüber^{2,3} and Gottfried Schlaug^{2*}

¹ Program in Speech and Hearing Bioscience and Technology, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA

² Music and Neuroimaging Laboratory, Department of Neurology, Beth Israel Deaconess Medical Center/Harvard Medical School, Boston, MA, USA

³ Department of Epileptology, Bonn University Hospital, Bonn, Germany

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Mathias S. Oechslin, University of Geneva, Switzerland

Jürgen Hänggi, University of Zurich, Switzerland

*Correspondence:

Gottfried Schlaug, Department of Neurology, Beth Israel Deaconess Medical Center and Harvard Medical School, 330 Brookline Avenue, Palmer 127, Boston, MA 02215, USA.
e-mail: gschlaug@bidmc.harvard.edu

Structure and function of the human brain are affected by training in both linguistic and musical domains. Individuals with intensive vocal musical training provide a useful model for investigating neural adaptations of learning in the vocal-motor domain and can be compared with learning in a more general musical domain. Here we confirm general differences in macrostructure (tract volume) and microstructure (fractional anisotropy, FA) of the arcuate fasciculus (AF), a prominent white-matter tract connecting temporal and frontal brain regions, between singers, instrumentalists, and non-musicians. Both groups of musicians differed from non-musicians in having larger tract volume and higher FA values of the right and left AF. The AF was then subdivided in a dorsal (superior) branch connecting the superior temporal gyrus and the inferior frontal gyrus (STG ↔ IFG), and ventral (inferior) branch connecting the middle temporal gyrus and the inferior frontal gyrus (MTG ↔ IFG). Relative to instrumental musicians, singers had a larger tract volume but lower FA values in the left dorsal AF (STG ↔ IFG), and a similar trend in the left ventral AF (MTG ↔ IFG). This between-group comparison controls for the general effects of musical training, although FA was still higher in singers compared to non-musicians. Both musician groups had higher tract volumes in the right dorsal and ventral tracts compared to non-musicians, but did not show a significant difference between each other. Furthermore, in the singers' group, FA in the left dorsal branch of the AF was inversely correlated with the number of years of participants' vocal training. Our findings suggest that long-term vocal-motor training might lead to an increase in volume and microstructural complexity of specific white-matter tracts connecting regions that are fundamental to sound perception, production, and its feedforward and feedback control which can be differentiated from a more general musician effect.

Keywords: plasticity, white matter, arcuate fasciculus, auditory-motor interactions, tractography, music, singing

INTRODUCTION

In recent years, there has been increased interest in the use of musicians to examine brain adaptation in response to intense and long-term training of musical skills (Trainor et al., 1999; Ross et al., 2003; Bengtsson et al., 2005; Koelsch et al., 2005; Zatorre et al., 2007; Hyde et al., 2009; Moreno et al., 2009; Oechslin et al., 2009; Schlaug et al., 2009a; Wan and Schlaug, 2010b). Although most of the previous work has focused on instrumental training, one type of musical training that may provide additional insights in brain adaptation is training in the voice or auditory-vocal domain (Zarate and Zatorre, 2008; Kleber et al., 2010; Zarate et al., 2010). Singing is an intensive auditory-vocal training which is typically started later in childhood or adolescence compared to instrumental music training. Singing has also received a lot of attention recently, since forms of singing have been shown to have positive effects on various neurological disorders (Wan et al., 2010b).

Regions in the superior temporal lobe, inferior frontal areas, and the associated premotor and motor regions are involved in the feed-forward and feedback control of singing (Pantev et al., 1998; Maess et al., 2001; Levitin and Menon, 2003; Brown et al., 2004; Ozdemir

et al., 2006). This fronto-temporal network of brain regions is connected via the arcuate fasciculus (AF), a prominent white-matter tract which, in its horizontal part, may share some components with the superior longitudinal fasciculus (SLF). The AF has direct fibers connecting the middle and superior temporal gyrus (STG) with inferior frontal regions, but may also have an indirect fiber system connecting the temporal lobe with the inferior parietal lobulus and then the parietal lobulus with frontal lobe regions (Catani et al., 2005; Glasser and Rilling, 2008). It should be noted that these putative "tracts" in fact represent tractography-derived pathways, and thus their fidelity to the anatomy is continuously contested, in particular with regards to whether or not the AF consists of direct fronto-temporal fibers or indirect fibers with synapses and relay stations in the parietal lobule (Catani et al., 2005; Frey et al., 2008; Glasser and Rilling, 2008). The AF and SLF have already been implicated in normal and abnormal vocal-motor activities. Fiber volume in the AF, and regional fractional anisotropy (FA – a measure of the degree of directional preference of water diffusion (Basser, 1995) values along its midpoint, are both lower in tone-deaf individuals (Loui et al., 2009). The phenotypical characteristic of

tone-deaf individuals is that they cannot sing in tune and are not aware of their vocal feedback. The impairment of the AF in tone-deaf subjects supports its presumed role in the feedforward and feedback control of vocal output.

In learning to associate motor actions with sounds and in training an auditory–motor feedback loop connecting the temporal with the frontal lobe, white-matter fiber bundles between these regions might change in size, volume, and composition. One common diffusion tensor imaging (DTI)-derived measure, FA, is a normalized measure expressing the directional diffusion of water protons; the more aligned fibers are within a tract, the higher the tract or regional FA value. FA changes within-group and FA differences between groups can be regarded as a surrogate marker of structural adaptation in the white matter (Lindenberg et al., 2010). Adaptations of the AF have been described as a function of normal development (Barnea-Goraly et al., 2005; Ashtari et al., 2007), and developmental delay has been found to be associated with disordered morphology of the AF, possibly reflecting delayed myelination (Sundaram et al., 2008).

Previous research has identified regions that may control and be strengthened by vocal training (Zarate and Zatorre, 2008; Zarate et al., 2010). In addition to the STG and its reciprocal connections with the inferior premotor and posterior IFG, the middle temporal gyrus (MTG) also has prominent connections with the inferior frontal gyrus (IFG). These two branches can be thought of as two separate branches with possibly slightly different functions in the auditory–motor mapping and auditory–motor control system.

Since the connectivity between STG, MTG, and IFG allows for the planning of complex motor sequences as well as the monitoring and correction of feedback and feedforward motor commands (Guenther et al., 2006), we hypothesized that the connectivity between these regions will be enhanced in musicians as a result of training-dependent or use-dependent plasticity in auditory–motor integration, and perhaps particularly in singers, given their specific auditory–vocal training requiring intense and precise auditory–motor feedforward and feedback control. We examined the connectivity of the AF and its dorsal and ventral branch, which connects the STG, MTG, and IFG, using DTI, an MR imaging technique that enables the visualization and quantitative assessment of white-matter pathways in the brain (Basser et al., 1994, 2000; Makris et al., 1997, 2005; Catani and Thiebaut De Schotten, 2008). Since fine-grained perception and feedback control of pitch and motor actions are both important aspects of musical training, we hypothesized that singers and instrumental musicians would exhibit structural adaptations in the AF relative to non-musicians, although the AF and its subdivisions in both hemispheres might be differentially affected considering the fine-grained sensory–motor mapping of sounds to articulation which we assumed is an important role for the dorsal branch of the AF.

MATERIALS AND METHODS

PARTICIPANTS

Twenty-two musicians were recruited through posting ads on Craigslist, bulletins at local music conservatories, and choral groups in the greater Boston area. Eleven of these individuals were either professional singers or reported to be in training to be professional singers; these singers did have some experience in playing musical instruments, such as piano ($n = 5$), cello (2), trombone (1), trumpet

(1), flute (1), and clarinet (1) but were not actively practicing these musical instruments at the time of this study or in the year prior to enrollment. This group of singers was contrasted with a group of 11 instrumental musicians who did not report singing as a primary musical activity, did not have any formal singing training and were not part of any singing group, but had an equal amount of musical experience as determined by their number of years of musical training. This group's instruments included piano ($n = 4$), violin (4), cello (1), and trombone (2). These two groups of 11 musicians each were contrasted with a third group of 11 participants without any significant instrumental music training (defined as less than 1 year of instrumental music training throughout their life, but not in the year before enrollment in this study), any formal singing training, and no participation in any singing groups. This group of non-musicians was recruited through the same mechanisms as the other groups and was matched to the musician groups with regards to age, gender, and IQ. Singers, instrumental musicians, and non-musicians were similar in the mean age [singers = 25.3 (SD = 2.7) years; instrumental musicians = 27.7 (SD 7.6) years; non-musicians = 27.5 (SD 10.3) years], and gender (5m/6f for instrumentalists, 3m/8f for singers, and 5m/6f for non-musicians). Singers and instrumentalists were matched for number of years in musical training (singers' mean = 15, SD = 5; instrumentalists' mean = 14.3, SD = 9.09) and age of training onset (singers' mean = 6.6, SD = 2.4; instrumentalists' mean = 7.4, SD = 4.4). The three groups were also matched with regards to their IQ as assessed by Shipley's verbal and abstract scaled composite score (Shipley, 1940) (singers' mean = 111, SD = 14, instrumentalists' mean = 114, SD = 10, non-musicians' mean = 111, SD = 13). This study was approved by the Institutional Review Board of the Beth Israel Deaconess Medical Center and all participants provided written informed consent.

IMAGE ACQUISITION

MR images were obtained using a 3-Tesla General Electric scanner. MR sequences included a strongly T1-weighted data set with a voxel resolution of $0.93 \text{ mm} \times 0.93 \text{ mm} \times 1.5 \text{ mm}$. DTI was performed using a diffusion-weighted, single-shot, spin-echo, echo-planar imaging sequence (TE1 = 86.9 ms, relaxation time = 10,000 ms, field of view = 240 mm, matrix size = 94×94 voxels, no skip, NEX = 1, axial acquisition, voxel dimensions = $2.5 \text{ mm} \times 2.5 \text{ mm} \times 2.6 \text{ mm}$). Thirty non-collinear directions with a b value of 1000 s/mm^2 and six volumes with a b value of 0 s/mm^2 were acquired.

DATA PROCESSING

FMRIB's FSL suite was used for pre-processing of raw images (Smith et al., 2004). A 3D affine registration was applied to correct for eddy currents and head motion (Jenkinson and Smith, 2001) followed by BET for brain extraction (Smith, 2002). Eigenvectors and eigenvalues of diffusion tensors were estimated at each voxel and FA was calculated using the *dtifit* function in FSL. Axial diffusivity, a parameter reflecting the principal direction of diffusion in white matter (Song et al., 2002), was estimated using the first eigenvalue of the diffusion tensor, λ_1 . The second and third eigenvalues were averaged and referred to as the radial diffusivity $(\lambda_2 + \lambda_3)/2$ (Basser, 1995; Xue et al., 1999; Basser et al., 2000; Song et al., 2002). Following this, a probability distribution for fiber direction was

calculated for each brain voxel using the *bedpostx* function in FSL (Behrens et al., 2003). To allow for fiber crossings, estimates of two directions per voxel were allowed (Behrens et al., 2007).

TRACTOGRAPHY – ARCUATE FASCICULUS

Regions of interest (ROIs) were drawn by hand, on a single sagittal slice of the native FA image of each subject, according to major anatomical landmarks, shown in **Figure 1**. For each hemisphere, ROIs were drawn on white matter underlying the gray matter in three regions: posterior superior temporal gyrus (pSTG), posterior middle temporal gyrus (pMTG), and posterior inferior frontal gyrus (pIFG). Two separate one-way ANOVAs (with the dependent measure of ROI volume for one, and FA for the other) showed no significant effect of group on either: ROI mean volumes for singers: 124.2 mm^3 (SD = 41.7); instrumentalists' = 136 mm^3 (SD = 39); non-musicians' mean = 124.1 mm^3 (SD = 41.7). FA values: singers' mean = 0.411, SD = 0.067, instrumentalists' mean = 0.422, SD = 0.059, non-musicians' mean = 0.401, SD = 0.062).

Probabilistic tractography was applied to constrain white-matter tracts to the dorsal and ventral branches of the AF in each hemisphere. Tractography was initiated from one seed region to one-way point mask region using the *probtrackx* function in FSL in two separate steps: (1) Tracts traced from the seed region of STG to the waypoint mask of IFG were identified as the dorsal AF. (2) Tracts traced from the seed region of MTG to the waypoint mask of IFG were identified as the ventral AF. These resulting tracts were further masked by setting an intensity threshold of the median intensity values of each tract, and eliminating voxels with intensity values below that threshold. While this provides a way to threshold the tracts similarly across subjects, choosing a cut-off intensity value is somewhat arbitrary, and since there is currently no convention for setting this value (Giorgio et al., 2010), we chose the median intensity value for each tract. To compute tract volume, we used the number of voxels in each tract after applying the cut-off intensity value and multiplied the number of voxels by the voxel size (Loui et al., 2011a,b).

STATISTICAL COMPARISONS – ARCUATE FASCICULUS

For each tract (dorsal AF, ventral AF) in each hemisphere, measures of volume and FA were extracted and compared using a three-way ANOVA with the between-subjects factor of GROUP (with three levels: singers, instrumentalists, and non-musicians) and the within-subject factors of HEMISPHERE (left vs. right) and TRACT (dorsal vs. ventral).

As a control for results from tractography of the AF, we extracted mean FA and volume for non-zero voxels from each subjects' whole-brain FA image, and compared the three groups' means to control for possible whole-brain differences between groups using two one-way ANOVAs (one for FA and one for volume).

RESULTS

TRACT STATISTICS – ARCUATE FASCICULUS

Tract volume was largest in singers, especially in the left hemisphere. This was confirmed using a three-way ANOVA on the dependent variable of tract volume with the between-subjects factor of GROUP (with three levels: singers, instrumentalists, and non-musicians) and the within-subject factors of HEMISPHERE (left vs. right) and TRACT (dorsal vs. ventral). This ANOVA revealed an overall significant effect of GROUP [$F(2,120) = 9.8$, $p < 0.001$], confirming that tract volume was highest in singers, second-highest in instrumentalists, and lowest in non-musicians. Furthermore, the same ANOVA revealed a significant main effect of HEMISPHERE [$F(1,120) = 9.2$, $p = 0.003$], with tract volume being higher in the left than in the right hemisphere. Finally, the interaction between GROUP and HEMISPHERE was significant [$F(2,120) = 4.4$, $p = 0.014$], confirming that the tract volume in the left hemisphere was especially higher among singers relative to the other two groups. These results are summarized in **Figure 2**.

In order to ascertain that tract volume in the left hemisphere was especially higher among singers relative to the other two groups, we performed pairwise *post hoc* comparison for each tract between singers and instrumentalists, singers and non-musicians, and between instrumentalists and non-musicians applying appropriate Bonferroni corrections. The dorsal branch of the left AF showed a significantly higher volume in singers relative to instrumentalists [$t(10) = 3.36$, $p = 0.007$], surviving Bonferroni corrections for two hemispheres and two branches of the AF in each hemisphere. The same branch's volume is also significantly higher in singers relative to non-musicians [$t(10) = 3.76$, $p = 0.004$], also surviving *post hoc* corrections for the four branches of the AF.

Tract FA was lower in singers than in instrumentalists, especially in the dorsal branch of the left hemisphere. This was tested using a three-way ANOVA on the dependent variable of FA with the between-subjects factor of GROUP and the within-subjects factors of HEMISPHERE and TRACT. The main effect of GROUP was significant [$F(2,120) = 8.174$, $p < 0.001$], confirming that FA was highest in instrumentalists, followed by singers and then by non-musicians, although this finding did not survive strict correction

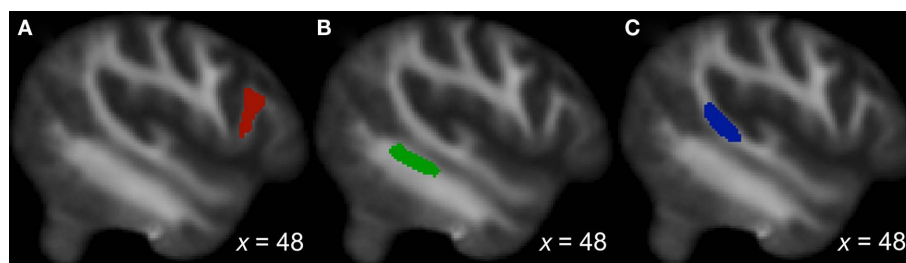
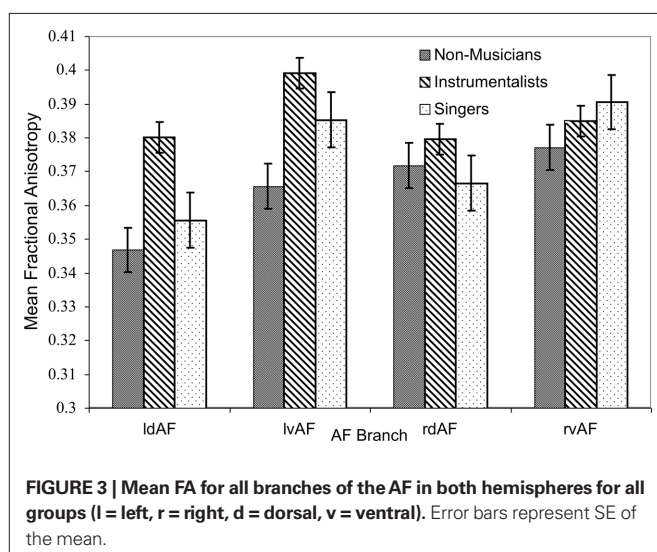
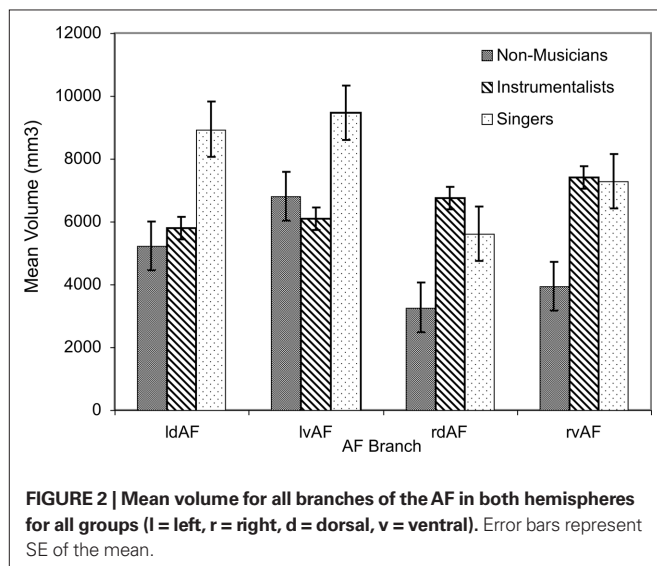


FIGURE 1 | Region of interest locations displayed on the FA images: (A) IFG, (B) MTG, (C) STG.

for multiple tests. The main effect of TRACT was significant [$F(1,120) = 12.49, p = 0.001$], confirming that tract FA was lower in the dorsal branch relative to the ventral branch. Finally, an interaction between GROUP and HEMISPHERE showed a trend toward significance [$F(2,120) = 2.73, p = 0.069$], showing that FA was lower in the left hemisphere of singers compared to both hemispheres of instrumentalists. These results are summarized in **Figure 3**.

To explore the potentially differing contributions of axial and radial diffusivities to the observed FA difference between singers and instrumentalists, we extracted the singers' and instrumentalists' left dorsal AF's mean axial (λ_1) and radial [$(\lambda_2 + \lambda_3)/2$] diffusivities. A direct comparison of these parameters between singers and instrumentalists suggested that the finding of lower FA was strongly due to higher radial diffusivity [$(\lambda_2 + \lambda_3)/2$] in singers relative to instrumentalists [$t(20) = 2.7, p = 0.01$], and not due to differences in axial diffusivity [$t(20) = 1.01, p = 0.33$].



To test whether the between-group differences in the AF could be explained by whole-brain differences in FA or volume, separate one-way ANOVAs were conducted on the dependent variables of FA and volume. Results showed no significant between-group differences in whole-brain FA or volume (all $ps > 0.1$), suggesting that differences in the AF cannot be explained by whole-brain differences.

FA ALONG THE ARCUATE FASCICULUS

Having identified the left dorsal AF as the main tract of interest that showed differences between singers and matched instrumental musician controls, we further sought to identify the region along the AF that shows maximal between-group differences. Visual comparison of the FA maps of tracts identified in singers and instrumentalists (**Figure 4**) suggested that a part of the longitudinal portion of the left dorsal AF showed the strongest FA differences between groups.

To explore this observed difference along the tract, the longitudinal portion of the left dorsal AF was divided into 4 bins of 10 coronal slices each and the mean FA from each bin (mean FA of 10 coronal slices) was extracted for each subject, and compared between singers and instrumentalists using a two-tailed t -test. The comparison revealed significantly lower FA for the singers in bin 3 [$y = -22$ mm to $y = -13$ mm, $t(9) = 4.29, p = 0.002$], around the midpoint of the longitudinal portion of the left dorsal AF. **Figure 5** shows bin-by-bin comparisons between singer and instrumental groups in the longitudinal portion of the left dorsal AF, whereas **Figures 6 and 7** show the entire left dorsal AF tract in axial (**Figure 6**) and sagittal (**Figure 7**) views, highlighting the section that is significantly lower in FA in singers than controls: on average, singers possessed lower FA in the longitudinal portion of the left dorsal AF, and more specifically in bin 3 (from $y = -22$ mm to $y = -13$ mm).

BEHAVIORAL CORRELATES OF FA DIFFERENCES

To explore the relationship between singers and FA values in the left dorsal AF, we tested the hypothesis that the number of years of singing training is predictive of FA value for the left dorsal AF using a linear regression (shown in **Figure 8**). Within the singer group only, the reported number of years of singing training inversely predicted the FA value of each individual's left dorsal AF ($R^2 = -0.387, p = 0.04$). For both FA and volume, a similar relationship was not observed in the left dorsal AF of control musicians ($R^2 = -0.08, p > 0.1$), nor for any other tracts in both singers' and instrumentalists (all $|R^2|s < 0.16$, all $ps > 0.2$).

DISCUSSION

We report a DTI comparison between singers, instrumental musicians, and non-musicians. Our results showed that the AF, a bundle of white-matter fibers that connects regions of the brain known to be involved in sound perception and production as well as the feed-forward and feedback control and vocal output, has a larger volume and higher FA values in musicians (both singers and instrumentalists) relative to non-musicians. Among musicians, singers showed significantly higher volume in the dorsal and ventral branches of the left AF, and lower FA in the left dorsal AF, relative to instrumentalists. In contrast, GROUP had no effect on whole-brain FA measures between the three groups, suggesting that effects could not be explained by whole-brain differences. A bin-by-bin analysis

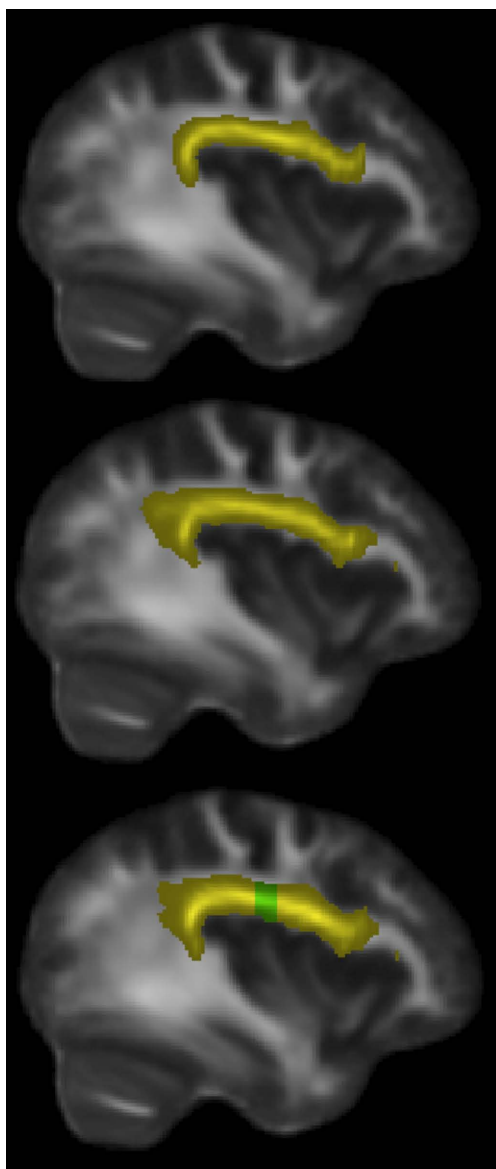


FIGURE 4 | Sagittal slice $x = 38$ mm, showing left dorsal AF in yellow: instrumental controls (top). Singers (middle). All musicians (bottom), with region showing significant FA differences at $y = -22$ to -13 mm, shown in green.

of FA differences in the longitudinal portion of the AF among musicians shows significantly lower FA among singers along the midpoint of the left dorsal AF, relative to instrumentalists.

The finding of higher FA in musicians compared to non-musicians, but lower FA in singers relative to instrumental musicians provides support for auditory–motor training-induced differences in the AF for musicians in general and singers in particular. The lower FA may reflect less aligned fibers or more fiber crossings among singers along the course of the AF. While DTI does not resolve the directionality of fiber tracts, tractography methods make it possible to infer tract volume based on identified tracts that connect several endpoints of gray matter (Giorgio et al., 2010).

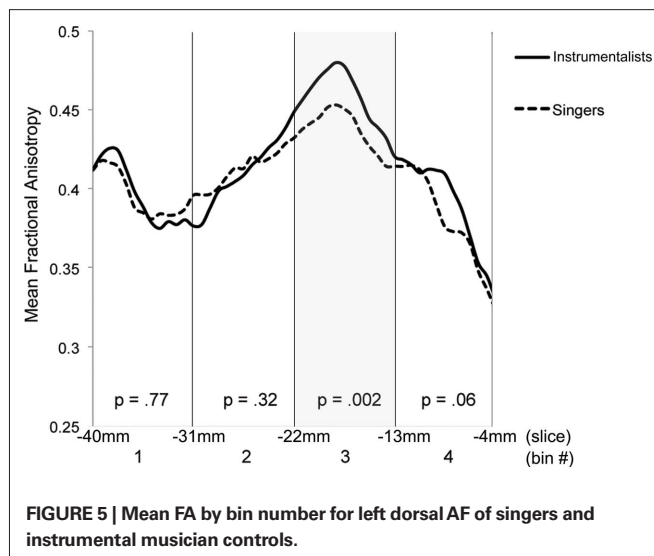


FIGURE 5 | Mean FA by bin number for left dorsal AF of singers and instrumental musician controls.

By using seed regions in the superior and middle temporal gyri, and waypoint masks in the IFG of each hemisphere in each brain, we were able to identify the dorsal and ventral branches of the left and right AF. Both dorsal and ventral branches of the left AF were identified as being larger in singers than in instrumentalists, with stronger effects in the dorsal branch. Interestingly, this effect was only observed in the left hemisphere.

Previous research has shown that individuals with difficulties in pitch perception and production, i.e., tone-deaf individuals, have diminished white-matter volume in their right AF, with pitch perception abilities being correlated with the dorsal branch of the right AF and pitch perception–production mismatch being correlated with the right ventral branch (Loui et al., 2009), suggesting that fine-grained control of pitch perception and production abilities are related to the dorsal branch (i.e., STG ↔ IFG connections). In contrast, both singers and instrumental musicians in this study have expertise in pitch perception and production, as the instrumental group is matched for number of years of musical training. It is only when both the singers and instrumentalists are compared with non-musicians that differences in the right AF are observed. This might suggest that the right AF shows more of a domain general adaptation effect in activities that involve matching sounds with actions independent of whether these actions are articulatory actions or hand actions. The left AF and in particular the dorsal branch of the left AF showed the most profound differences comparing singers with instrumentalists. Although this finding is somewhat surprising, it suggests that the left AF, which already shows an adaptation when individuals acquire language (Barnea-Goraly et al., 2005; Ashtari et al., 2007) and is usually larger and more complex than the right AF (Vernooij et al., 2007; Glasser and Rilling, 2008), might be the structure that adapts the most to the specific requirements of vocal–motor and auditory–motor integration.

Furthermore, it is also possible that the dorsal (STG ↔ IFG) and ventral (MTG ↔ IFG) branches have different functions and therefore show different adaptations. The ventral (MTG–IFG) branch could be a fast and coarse system that does not rely on precise auditory feedback while the dorsal (STG ↔ IFG) branch could be

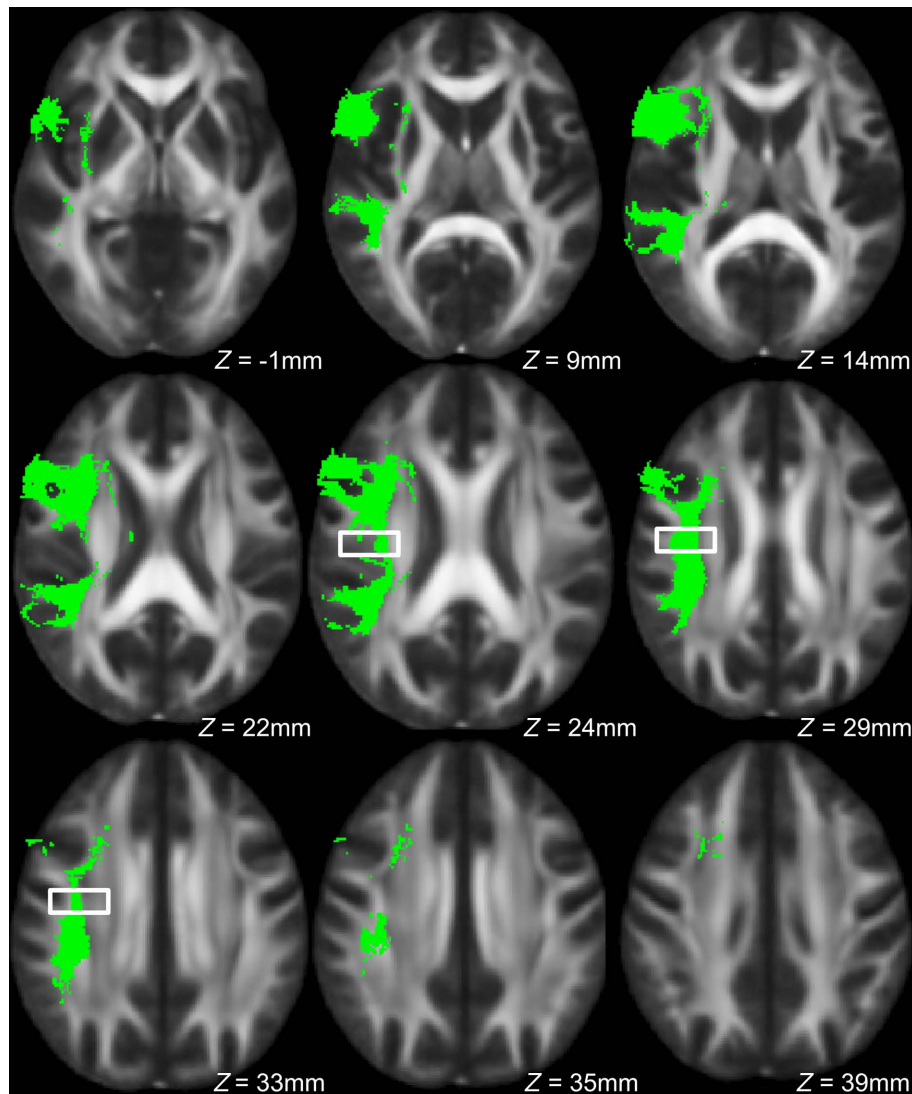


FIGURE 6 | Axial view of left dorsal AF across all subjects in green, with region showing significant between-group FA differences indicated by the box.

a slower, more precise system that is under conscious control and heavily relies on auditory feedback to make adjustments in the auditory–motor mapping of vocal output.

Similar findings of lower FA values in white-matter regions have been reported in other studies comparing experts in a particular domain with non-experts. For example, simultaneous interpreters (individuals who translate a source’s speech in real-time) were reported to have lower FA relative to controls, in subregions of their corpus callosum, as well as in the white matter underlying the left anterior insula and inferior parietal lobe, regions thought to be involved in articulatory control and sensory–motor mapping for speech (Elmer et al., 2010).

Diffusion tensor imaging provides *in vivo* measures of white-matter connectivity through its parameters of FA, axial diffusivity, radial diffusivity, and shape and volume of white-matter tracts. FA is a scalar measure that characterizes the degree of directional preference in diffusion of water, and is affected by the degree of

myelination and axonal coherence. Axial diffusivity is said to be related to the integrity of the axons, whereas radial diffusivity is said to be related to the degree of myelination (Song et al., 2002; Schaechter et al., 2009). However there is much debate about the validity of these associations especially when applied to gray–white-matter boundaries or regions with multiple crossing fibers (Wheeler-Kingshott and Cercignani, 2009).

The finding of lower FA due to higher radial diffusivity in singers, in particular in bin 3 of the tract (around the midpoint of the longitudinal portion of the AF, most proximal to regions of the motor and somatosensory cortices) suggests that the singers’ AFs may be adapted to allow for more connectivity between the primary nodes of the AF (STG/MTG and IFG), the motor/somatosensory system, and/or the inferior parietal lobule via the AF’s putative “indirect pathway” (as described in Catani et al., 2005). Since the FA difference is most strong in bin 3 and bin 3 roughly coincides with the location of the pre- and postcentral gyrus, this suggests

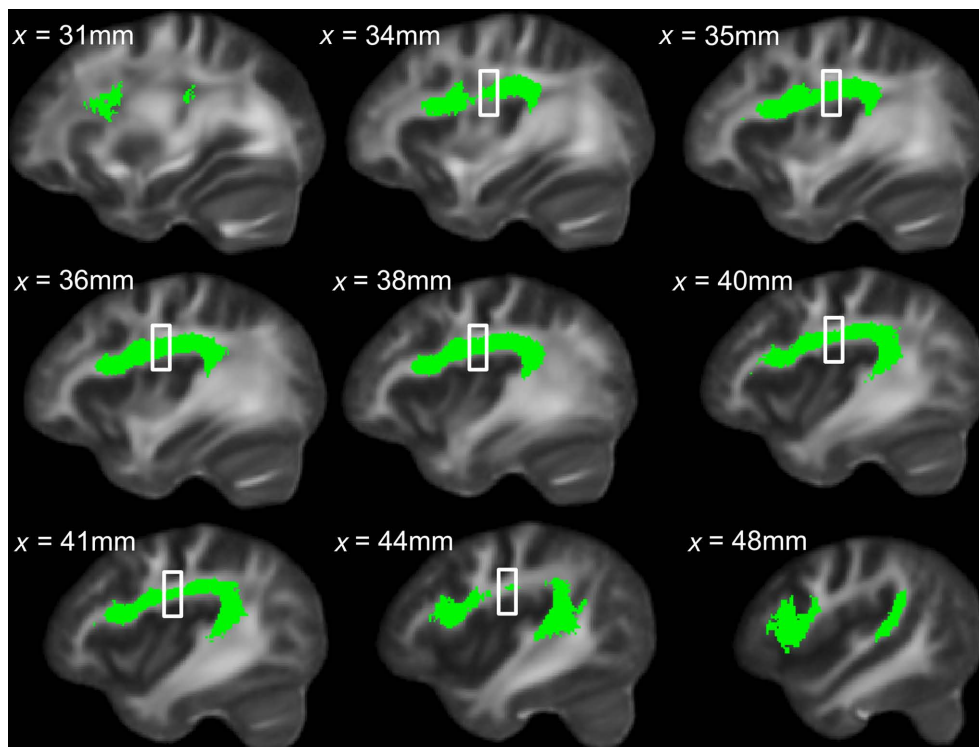


FIGURE 7 | Sagittal view of left dorsal AF across all subjects in green, with region showing significant between-group FA differences indicated by the box.

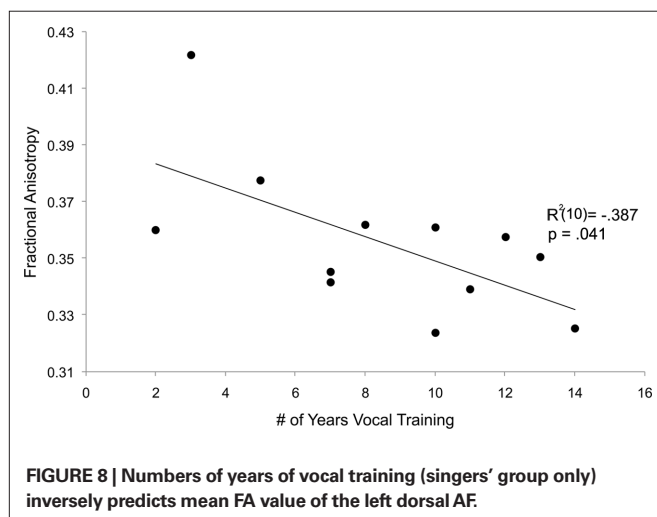


FIGURE 8 | Numbers of years of vocal training (singers' group only) inversely predicts mean FA value of the left dorsal AF.

that at least part of the increase in radial diffusivity in singers may be reflective of less alignment in one direction and possibly more branching in other directions (e.g., perpendicular to the axial plane of the tract, toward the pre- and postcentral gyrus). Differences in white-matter microstructure of this kind (increased radial diffusivity with no difference in axial diffusivity) are most frequently interpreted as indicating differences in myelination (Song et al., 2002; Alexander et al., 2007; Jito et al., 2008), but changes in myelin or axon density due to axonal sprouting are also possible (Dancause et al., 2005; Carmichael, 2006), although the extent to which DTI is sensitive to such phenomena is not known. Nevertheless, enhanced

connectivity of this kind in singers might be an adaptation for increased sensitivity for feedback information from the inferior parts of the somatosensory cortex (homuncular representations of sensory feedback from speech areas, e.g., proprioception of jaws, tongue, and lips), as well as feedforward information provided to the inferior regions of the inferior motor/premotor strip (homuncular representations of the tongue, jaw, lips, and larynx). This interaction of feedback and feedforward information in which the AF may play a critical role has been described in detail in Guenther et al. (2006) for syllable production in speaking and might be similar for intoned syllable production in singing. Functional neuroimaging studies have shown an extended activation after vocal skills training involving inferior motor/premotor and somatosensory regions (Kleber et al., 2010). In contrast to instrumental musicians, who exercise fine non-vocal-motor control while engaging their vocal system minimally during a performance, singers must always monitor their breathing as well as proprioception from their vocal apparatus. This added cognitive demand necessitates stronger connectivity between temporal, inferior frontal, as well as inferior motor/premotor, and inferior somatosensory regions; this may be reflected in differing white-matter architecture in the AF of singers, relative to instrumentalists.

Our results suggest that musical training, particularly vocal training, is associated with structural adaptations in the AF: a tract that is important for linguistic as well as musical functions. In individuals who receive primarily vocal musical training, adaptations are observed especially in the left dorsal AF, which is larger in volume but lower in FA among singers (relative to instrumentalists). This combination of DTI differences may reflect more fiber

crossings in white-matter regions between STG and IFG, possibly reflecting a more complex pattern of connectivity for the AF that may be associated with vocal training as compared to other kinds of musical training.

The present study provides further support for the use of musically trained individuals, especially singers, as a model for structural as well as functional adaptations of the auditory-motor system (Schlaug, 2001) by showing structural differences between the brains of those engaged in specific types of musical training (vocal vs. instrumental). The intensive training that is typical in learning to sing may be translated into neurorehabilitation programs for aphasic patients and other groups of impaired vocal output (Schlaug et al., 2008; Wan and Schlaug, 2010a; Wan et al., 2010a), whose white-matter pathways in the undamaged hemisphere show adaptations in

response to intense intonation-based speech therapy (Schlaug et al., 2009b; Wan et al., 2010b). By characterizing the differences in white-matter connectivity between singers and non-singer controls who are matched in musical training, the present study suggests that intense musical training, particularly vocal-motor training, has an effect on the macro- and microstructural anatomy of vocal-motor networks.

ACKNOWLEDGMENTS

The study was supported by research grants from NIH/NIDCD (RO1 DC009823 and RO1 DC008796). Gus Halwani acknowledges support from a training grant (T32 DC00038) from NIH/NIDCD. Furthermore, support from the Nancy Lurie Marks Family Foundation and from the family of Rosalyn and Richard Slifka is also acknowledged.

REFERENCES

- Alexander, A. L., Lee, J. E., Lazar, M., and Field, A. S. (2007). Diffusion tensor imaging of the brain. *Neurotherapeutics* 4, 316–329.
- Ashtari, M., Cervellione, K. L., Hasan, K. M., Wu, J., McIlree, C., Kester, H., Ardekani, B. A., Roofeh, D., Szeszko, P. R., and Kumra, S. (2007). White matter development during late adolescence in healthy males: a cross-sectional diffusion tensor imaging study. *Neuroimage* 35, 501–510.
- Barnea-Goraly, N., Menon, V., Eckert, M., Tamm, L., Bammer, R., Karchemskiy, A., Dant, C. C., and Reiss, A. L. (2005). White matter development during childhood and adolescence: a cross-sectional diffusion tensor imaging study. *Cereb. Cortex* 15, 1848–1854.
- Basser, P. J. (1995). Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. *NMR Biomed.* 8, 333–344.
- Basser, P. J., Mattiello, J., and Lebihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 66, 259–267.
- Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.* 44, 625–632.
- Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage* 34, 144–155.
- Behrens, T. E., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., Matthews, P. M., Brady, J. M., and Smith, S. M. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* 50, 1077–1088.
- Bengtsson, S. L., Nagy, Z., Skare, S., Forsman, L., Forssberg, H., and Ullen, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nat. Neurosci.* 8, 1148–1150.
- Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., and Parsons, L. M. (2004). The song system of the human brain. *Brain Res. Cogn. Brain Res.* 20, 363–375.
- Carmichael, S. T. (2006). Cellular and molecular mechanisms of neural repair after stroke: making waves. *Ann. Neurol.* 59, 735–742.
- Catani, M., Jones, D. K., and Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16.
- Catani, M., and Thiebaut De Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex* 44, 1105–1132.
- Dancuse, N., Barbay, S., Frost, S. B., Plautz, E. J., Chen, D., Zoubina, E. V., Stowe, A. M., and Nudo, R. J. (2005). Extensive cortical rewiring after brain injury. *J. Neurosci.* 25, 10167–10179.
- Elmer, S., Hanggi, J., Meyer, M., and Jancke, L. (2010). Differential language expertise related to white matter architecture in regions subserving sensory-motor coupling, articulation, and interhemispheric transfer. *Hum. Brain Mapp.* PMID: 21162044. [Epub ahead of print].
- Frey, S., Campbell, J. S., Pike, G. B., and Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J. Neurosci.* 28, 11435–11444.
- Giorgio, A., Santelli, L., Tomassini, V., Bosnell, R., Smith, S., De Stefano, N., and Johansen-Berg, H. (2010). Age-related changes in grey and white matter structure throughout adulthood. *Neuroimage* 51, 943–951.
- Glasser, M. F., and Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cereb. Cortex* 18, 2471–2482.
- Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (2006). Neural modeling and imaging of the cortical interaction underlying syllable production. *Brain Lang.* 96, 280–301.
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., and Schlaug, G. (2009). Music training shapes structural brain development. *J. Neurosci.* 29, 3019–3025.
- Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156.
- Jito, J., Nakasu, S., Ito, R., Fukami, T., Morikawa, S., and Inubushi, T. (2008). Maturation changes in diffusion anisotropy in the rat corpus callosum: comparison with quantitative histological evaluation. *J. Magn. Reson. Imaging* 28, 847–854.
- Kleber, B., Veit, R., Birbaumer, N., Gruzelier, J., and Lotze, M. (2010). The brain of opera singers: experience-dependent changes in functional activation. *Cereb. Cortex* 20, 1144–1152.
- 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.
- Levitin, D. J., and Menon, V. (2003). Musical structure is processed in “language” areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *Neuroimage* 20, 2142–2152.
- Lindenberg, R., Renga, V., Zhu, L. L., Betzler, F., Alsop, D., and Schlaug, G. (2010). Structural integrity of corticospinal motor fibers predicts motor impairment in chronic stroke. *Neurology* 74, 280–287.
- Loui, P., Alsop, D., and Schlaug, G. (2009). Tone deafness: a new disconnection syndrome? *J. Neurosci.* 29, 10215–10220.
- Loui, P., Li, H. C., Hohmann, A., and Schlaug, G. (2011a). Enhanced cortical connectivity in absolute pitch musicians: a model for local hyperconnectivity. *J. Cogn. Neurosci.* 23, 1015–1026.
- Loui, P., Li, H. C., and Schlaug, G. (2011b). White matter integrity in right hemisphere predicts pitch related grammar learning. *Neuroimage* 55, 500–507.
- 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.
- Makris, N., Kennedy, D. N., Mcinerney, S., Sorensen, A. G., Wang, R., Caviness, V. S. Jr., and Pandya, D. N. (2005). Segmentation of subcomponents within the superior longitudinal fascicle in humans: a quantitative, in vivo, DT-MRI study. *Cereb. Cortex* 15, 854–869.
- Makris, N., Worth, A. J., Sorensen, A. G., Papadimitriou, G. M., Wu, O., Reese, T. G., Wedeen, V. J., Davis, T. L., Stakes, J. W., Caviness, V. S., Kaplan, E., Rosen, B. R., Pandya, D. N., and Kennedy, D. N. (1997). Morphometry of in vivo human white matter association pathways with diffusion-weighted magnetic resonance imaging. *Ann. Neurol.* 42, 951–962.
- 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.
- Oechslin, M. S., Imfeld, A., Loenneker, T., Meyer, M., and Jancke, L. (2009). The plasticity of the superior longitudinal fasciculus as a function of musical expertise: a diffusion tensor imaging study. *Front. Hum. Neurosci.* 3:76. doi: 10.3389/fnhum.2009.0076.2009
- Ozdemir, E., Norton, A., and Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage* 33, 628–635.
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature* 392, 811–814.

- Ross, D. A., Olson, I. R., and Gore, J. C. (2003). Cortical plasticity in an early blind musician: an fMRI study. *Magn. Reson. Imaging* 21, 821–828.
- Schaechter, J. D., Fricker, Z. P., Perdue, K. L., Helmer, K. G., Vangel, M. G., Greve, D. N., and Makris, N. (2009). Microstructural status of ipsilesional and contralesional corticospinal tract correlates with motor skill in chronic stroke patients. *Hum. Brain Mapp.* 30, 3461–3474.
- Schlaug, G. (2001). The brain of musicians. A model for functional and structural adaptation. *Ann. N. Y. Acad. Sci.* 930, 281–299.
- Schlaug, G., Forgeard, M., Zhu, L., Norton, A., and Winner, E. (2009a). Training-induced neuroplasticity in young children. *Ann. N. Y. Acad. Sci.* 1169, 205–208.
- Schlaug, G., Marchina, S., and Norton, A. (2009b). Evidence for plasticity in white-matter tracts of patients with chronic Broca's aphasia undergoing intense intonation-based speech therapy. *Ann. N. Y. Acad. Sci.* 1169, 385–394.
- Schlaug, G., Marchina, S., and Norton, A. (2008). From singing to speaking: why singing may lead to recovery of expressive language function in patients with Broca's aphasia. *Music Percept.* 25, 315–323.
- Shipley, W. C. (1940). A self-administered scale for measuring intellectual impairment and deterioration. *J. Psychol.* 9, 371–377.
- Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155.
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., and Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23(Suppl. 1), S208–S219.
- Song, S. K., Sun, S. W., Ramsbottom, M. J., Chang, C., Russell, J., and Cross, A. H. (2002). Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. *Neuroimage* 17, 1429–1436.
- Sundaram, S. K., Sivaswamy, L., Makki, M. I., Behen, M. E., and Chugani, H. T. (2008). Absence of arcuate fasciculus in children with global developmental delay of unknown etiology: a diffusion tensor imaging study. *J. Pediatr.* 152, 250–255.
- Trainor, L. J., Desjardins, R. N., and Rockel, C. (1999). A comparison of contour and interval processing in musicians and non-musicians using event-related potentials. *Aust. J. Psychol.* 51, 147–153.
- Vernooij, M. W., Smits, M., Wielopolski, P. A., Houston, G. C., Krestin, G. P., and Van Der Lugt, A. (2007). Fiber density asymmetry of the arcuate fasciculus in relation to functional hemispheric language lateralization in both right- and left-handed healthy subjects: a combined fMRI and DTI study. *Neuroimage* 35, 1064–1076.
- Wan, C. Y., Rueber, T. R., Hohmann, A., and Schlaug, G. (2010a). The therapeutic effects of singing in neurological disorders. *Music Percept.* 27, 287–295.
- Wan, C. Y., Demaine, K., Zipse, L., and Schlaug, G. (2010b). From music making to speaking: engaging the mirror neuron system in autism. *Brain Res. Bull.* 82, 161–168.
- Wan, C. Y., and Schlaug, G. (2010a). Neural pathways for language in autism: the potential for music-based treatments. *Future Neurol.* 5, 797–805.
- Wan, C. Y., and Schlaug, G. (2010b). Music making as a tool for promoting brain plasticity across the life-span. *Neuroscientist* 16, 566–577.
- Wheeler-Kingshott, C. A., and Cercignani, M. (2009). About “axial” and “radial” diffusivities. *Magn. Reson. Med.* 61, 1255–1260.
- Xue, R., Van Zijl, P. C., Crain, B. J., Solaiyappan, M., and Mori, S. (1999). In vivo three-dimensional reconstruction of rat brain axonal projections by diffusion tensor imaging. *Magn. Reson. Med.* 42, 1123–1127.
- Zarate, J. M., Wood, S., and Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia* 48, 607–618.
- Zarate, J. M., and Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage* 40, 1871–1887.
- Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* 8, 547–558.

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: 17 March 2011; paper pending published: 16 April 2011; accepted: 23 June 2011; published online: 07 July 2011.

Citation: Halwani GF, Loui P, Rüber T and Schlaug G (2011) Effects of practice and experience on the arcuate fasciculus: comparing singers, instrumentalists, and non-musicians. *Front. Psychology* 2:156. doi: 10.3389/fpsyg.2011.00156

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

Copyright © 2011 Halwani, Loui, Rüber and Schlaug. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



The influence of task-irrelevant music on language processing: syntactic and semantic structures

Lisianne Hoch¹, Benedicte Poulin-Charronnat^{1,2} and Barbara Tillmann^{1*}

¹ Lyon Neuroscience Research Center, CNRS UMR5292, INSERM U1028, Université Lyon1, Lyon, France

² Laboratoire d'Etude de l'Apprentissage et du Développement, CNRS-UMR 5022, Université de Bourgogne, Dijon, France

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Eckart Altenmüller, University of Music and Drama Hannover, Germany
Antoine Shahin, The Ohio State University, USA
Patrick Wong, Northwestern University, USA

*Correspondence:

Barbara Tillmann, Team Auditory Cognition and Psychoacoustics, Lyon Neuroscience Research Center, 50 Avenue Tony Garnier, F-69366 Lyon Cedex 07, France.
e-mail: btillmann@olfac.univ-lyon1.fr

Recent research has suggested that music and language processing share neural resources, leading to new hypotheses about interference in the simultaneous processing of these two structures. The present study investigated the effect of a musical chord's tonal function on syntactic processing (Experiment 1) and semantic processing (Experiment 2) using a cross-modal paradigm and controlling for acoustic differences. Participants read sentences and performed a lexical decision task on the last word, which was, syntactically or semantically, expected or unexpected. The simultaneously presented (task-irrelevant) musical sequences ended on either an expected tonic or a less-expected subdominant chord. Experiment 1 revealed interactive effects between music-syntactic and linguistic-syntactic processing. Experiment 2 showed only main effects of both music-syntactic and linguistic-semantic expectations. An additional analysis over the two experiments revealed that linguistic violations interacted with musical violations, though not differently as a function of the type of linguistic violations. The present findings were discussed in light of currently available data on the processing of music as well as of syntax and semantics in language, leading to the hypothesis that resources might be shared for structural integration processes and sequencing.

Keywords: musical expectancy, semantic expectancy, syntactic expectancy, structural integration, cross-modal interactions

INTRODUCTION

Music and language are rule-governed systems. The rules, which organize events (e.g., chords, words) structurally over time, define syntactic principles. Acculturated listeners and speakers have implicit knowledge of musical and linguistic syntax, allowing them to develop expectations about future musical and linguistic events (Besson and Schön, 2003; Patel, 2008). Such structural similarities have encouraged investigations about the domain-specificity or generality of music and language processing.

Evidence of independence between music and language processing mainly comes from neuropsychological studies revealing a double dissociation between music and language deficits for brain-damaged patients (Basso and Capitani, 1985; Peretz et al., 1994, 1997). While few behavioral and neurophysiological studies have reported independent processes (Besson et al., 1998; Bonnel et al., 2001), several behavioral and neurophysiological studies have reported dependent processes for music and language (Patel et al., 1998; Bigand et al., 2001; Koelsch et al., 2001, 2002, 2005; Maess et al., 2001; Tillmann et al., 2003; Poulin-Charronnat et al., 2005; Steinbeis and Koelsch, 2008; Fedorenko et al., 2009; Slevc et al., 2009). Behavioral studies have shown, for example, that musical structures influence syllable (Bigand et al., 2001) and word (Poulin-Charronnat et al., 2005) processing in vocal music. Event-related potential (ERP) studies have shown that (1) the late positivity P600, which reflects syntactic integration, is not language-specific but can be elicited by music-syntactic violations (Patel et al., 1998), and (2) music-syntactic violations can elicit an early right anterior negativity (ERAN), which is comparable to the left-lateralized (early) anterior negativity (i.e., (E)LAN) elicited by linguistic-syntactic

violations (Koelsch et al., 2001). In addition, functional imaging data have revealed that cortical areas involved in linguistic processing (e.g., Broca's area) are also involved in music processing (Maess et al., 2001; Koelsch et al., 2002; Tillmann et al., 2003). Based on these data, Patel (2003) proposed the "shared syntactic integration resource hypothesis" (SSIRH): Music and language share neural resources for processes linked to the structural integration of events (i.e., processing of structural relations between events (chords/words) in working memory). In contrast, the musical and linguistic representations would be stored in distinct neural networks and can be selectively damaged, thus reconciling the double dissociations that had been observed in patients (Basso and Capitani, 1985; Peretz et al., 1994, 1997).

The SSIRH predicts that "tasks which combine linguistic and musical syntactic integration will show interference between the two" (Patel, 2003, p. 679). This hypothesis has been initially supported by ERP studies showing interactive influences between the simultaneous processing of music-syntactic and linguistic-syntactic structures (Koelsch et al., 2005; Steinbeis and Koelsch, 2008). Using a cross-modal paradigm, visually presented sentences were synchronized with auditorily presented chord sequences. As in previous studies investigating either language processing (e.g., Gunter et al., 2000) or music processing (e.g., Koelsch et al., 2001), syntactically incorrect words elicited a LAN, while music-syntactically unexpected chords (i.e., Neapolitan chords containing out-of-key tones) elicited an ERAN. Most importantly, linguistic and musical syntax interacted: A music-syntactically unexpected chord reduced the amplitude of the LAN while a linguistic-syntactically unexpected word reduced the amplitude of the ERAN. Recently (and in parallel

to our present study), a behavioral study measuring reading-times in a self-paced reading paradigm reported cross-modal interactive influences of music-syntactic violations on syntactic complexity processing in garden-path sentences (Slevc et al., 2009). Reading-times showed enhanced garden-path effects (i.e., longer reading-times for syntactically unexpected words) when sentences were presented simultaneously with music-syntactically unexpected, out-of-key chords (compared to expected in-key chords).

These cross-modal experiments used strong musical expectancy violations and introduced unexpected, out-of-key chords or chords with out-of-key tones that created acoustic violations with the preceding context (thus creating a confound with the music-syntactic violation). However, previous research investigating music perception has promoted the need for controlled musical materials with the aim to disentangle the influence of acoustic deviance from that of music-syntactic processing (Tekman and Bharucha, 1998; Bigand et al., 2003, 2006; Koelsch et al., 2007). In particular, musical context effects can be due to knowledge-driven processes (referred to as cognitive priming) or sensory-driven processes (i.e., sensory priming). The former result from the activation of listeners' knowledge of Western musical syntax, the later from the difference in acoustic overlap between prime and target (i.e., harmonic spectra, tone repetition vs. novelty). Out-of-key tones do not only create an expectancy violation based on listeners' tonal knowledge and musical-syntax processing, but create sensory dissonance with the other context tones as well (Terhardt, 1984). These perceptual changes define unexpected events as sensory deviants. The use of unexpected events with out-of-key tones does not allow distinguishing whether context effects are due to violations of musical structures or to violations of sensory features. Relatively subtle tonal manipulations (such as comparing target chords belonging to the same context key, i.e., tonic vs. subdominant) avoid creating any contextual dissonance and thus allow focusing on cognitive priming and musical structure processing.

This kind of experimental control is also relevant for the investigation of interactions between music and language processing, notably to be able to study interactions only due to music-syntactic processing (and not due to acoustic deviance processing). Our Experiment 1 investigated the interaction between simultaneous music-syntactic and linguistic-syntactic processing with more subtle music-syntactic violations than those previously used. For this aim, the music-syntactic manipulation used the in-key subdominant chord as the unexpected chord and the in-key tonic chord as the expected chord. With this musical material, previous studies have shown faster and more accurate processing for the expected tonic chord than for the less-expected subdominant chord (referred to as the tonal function effect). Originally, this tonal function effect was observed with tasks that focused on a perceptual feature of the target chord (e.g., consonance/dissonance judgment, Bigand et al., 2003). More recently, the tonal function effect was observed with tasks that focused on a simultaneously presented linguistic feature, such as sung, spoken, and visual syllables (Bigand et al., 2001; Escoffier and Tillmann, 2008; Hoch and Tillmann, 2010, respectively). The effect of tonal function on syllable processing suggests that listeners process musical structures, even when these structures are task-irrelevant.

In a cross-modal paradigm (i.e., task-relevant visual information presented with a musical background), Escoffier and Tillmann (2008) compared target chord processing in tonal sequences (ending with an expected tonic or a less-expected subdominant chord) and in tonally neutral sequences (i.e., baseline sequences). The comparison to neutral baseline sequences followed a rationale previously used in psycholinguistics to study costs and benefits of contextual expectations (e.g., Jonides and Mack, 1984). In contrast to the tonal sequences (ending on tonic vs. subdominant), which install a tonal center and evoke expectations for the tonic in particular, the baseline sequences do not install a tonal center and thus do not guide listeners' tonal expectations. In Escoffier and Tillmann (2008), the processing of the visual event was faster for the tonic condition than for the baseline condition, but was not slowed down in the subdominant condition, which did not differ from the baseline condition. This comparison to baseline thus suggests that the cross-modal influence of the tonal function on the visual processing is not due to a general distraction or attentional shift (due to the unexpected subdominant), but rather to a benefit of processing thanks to the expected tonic chord.

Our Experiment 1 used a cross-modal paradigm: Sentences were presented visually in synchrony with auditorily presented task-irrelevant chord sequences, which ended on an expected tonic chord or a less-expected subdominant chord. Participants performed a lexical decision task on the final word, which was syntactically expected or unexpected. If previously reported interactive influences between music-syntactic and linguistic-syntactic processing were not due to the acoustic violation created by the strong music-syntactic violation (i.e., out-of-key chords or tones, Koelsch et al., 2005; Steinbeis and Koelsch, 2008), then Experiment 1 should reveal an interaction: The linguistic-syntactic expectancy effect should be modulated by the tonal function of the final chord and the tonal function effect (i.e., the musical expectancy effect) should be modulated by the linguistic-syntactic expectancy. Based on the previously reported tonic facilitation in a cross-modal paradigm (Escoffier and Tillmann, 2008), interfering processes between music-syntactic and linguistic-syntactic processing should be reflected in a reduced or vanished tonic facilitation when simultaneously presented with a syntactically unexpected word. Such interfering processes between music-syntactic and linguistic-syntactic processing would support the hypothesis of shared structural and temporal integration resources, as formulated for syntax in the SSIRH (Patel, 2003).

EXPERIMENT 1: SYNTACTIC EXPECTANCY AND MUSICAL EXPECTANCY

METHOD

Participants

Thirty-two students from the University of Lyon ($M \pm SD = 22 \pm 2.83$ years) participated in Experiment 1. Number of years of formal musical practice ranged from 0 to 12 (3.72 ± 4.11 years; $Mdn = 2$ years). All participants gave informed consent, and none of the participants reported to have an auditory impairment.

Materials

Musical material. Twelve eight-chord sequences from Bigand et al. (2001) were used; half ended on the expected tonic chord and the other half on the less-expected subdominant chord. The

inter-chord distance was set to zero. The first seven chords sounded for 625 ms each (thus a stimulus onset asynchrony of 625 ms), and the final chord sounded for 850 ms (due to the timbre's resonance). Sequences were generated with Cubase 5.1 (Steinberg) and Grand Piano sound samples using Halion software sampler (Steinberg).

Linguistic material. In 24 French sentences of eight syllables (Appendix), the final word of each sentence defined the target, which was monosyllabic and syntactically expected. Aiming for semantically expected target-words, the syntactically expected sentences were constructed on the basis of a pretest. In the pretest, the targets were given as the appropriate final word by at least 60% of the participants (students from the University of Lyon that did not participate in Experiment 1) in a free-completion task. The syntactic expectancy manipulation consisted of a gender violation (also named morphosyntactic violation, see Gunter et al., 2000): The target word was either syntactically expected or unexpected relative to the preceding article (e.g., “Le méchant chien dort dans la_[fem.] **niche**_[fem.] / The nasty dog is sleeping in the_[fem.] **kennel**_[fem.]” vs. “Le méchant chien dort dans le_[masc.] **niche**_[fem.] / The nasty dog is sleeping in the_[masc.] **kennel**_[fem.],” see Appendix). This manipulation was applied to all 24 sentences, leading to 48 experimental sentences.

Pseudo-words were constructed from target-words by altering one letter [e.g., “puche” instead of “poche” (pocket)]. This modification changed only one phoneme and did not violate orthographic or phonological rules of French. None of the pseudo-words were pseudo-homonyms in French.

Each sentence was visually presented, syllable-by-syllable, on the center of the screen. For polysyllabic words, a dash was presented at the end of the syllable to indicate that the word continued on the next screen. The syllable-by-syllable presentation allowed reading at a comfortable pace using the chords' stimulus onset asynchrony of 625 ms. The first seven syllables of each sentence were presented in white (380 ms followed by an inter-syllable-interval of 245 ms), and the target was displayed in red up to the participant's response, with a timeout of 1200 ms.

Audio-visual presentation. The onset of each syllable was synchronized with the onset of each chord so that the target word (or pseudo-word) was synchronized with the onset of the final chord. The 96 sentences (i.e., the 48 sentences ending on either a word or a pseudo-word) were presented with a musical sequence ending on either a tonic or a subdominant chord. The resulting 192 trials were presented in pseudorandom order, where the presentation of the same sentence was separated by at least four other sentences and consecutive presentations of each experimental condition were limited to five repetitions. A different pseudorandom order was created for each participant. The experiment was run with the PsyScope Software (Cohen et al., 1993).

Procedure

Participants were informed that sentences containing eight syllables were presented syllable-by-syllable on successive screens with music in the background. Their task was to read the sentences and to decide as quickly and accurately as possible whether the last element was a word or a pseudo-word (i.e., lexical decision task). Before the presentation of the first syllable, a fixation cross appeared

in the center of the screen. Participants were first trained in the lexical decision task with eight isolated words and pseudo-words that was followed by training with four sentences accompanied by chord sequences. Error feedback was given in the training and experimental phases. A 250-ms noise mask followed each trial to diminish the trace of the preceding musical sequence in the sensory memory buffer and thus its influence on the processing of the following musical sequence (as, for example, in Bigand et al., 2003; Escoffier and Tillmann, 2008).

RESULTS

The mean accuracy was 94 and 97% for words and pseudo-words, respectively. The means of correct response times (RTs) were 588 ms (range: 455–826 ms) and 643 ms (range: 495–829 ms) for words and pseudo-words, respectively. Correct RTs were individually normalized with a mean of 0 and a SD of 1, providing z-scores. For words¹, percentages of correct responses (Table 1) and normalized RTs (Figure 1) were, respectively, analyzed by two 2 × 2 ANOVAs with syntactic expectancy (expected, unexpected) and tonal function (tonic, subdominant) as within-participant factors and either participants (F_1) or target-words (F_2) as a random variable.

For correct responses and normalized RTs, the main effect of syntactic expectancy was significant, $F_1(1, 31) = 27.99, p < 0.001$, $MSE = 13.61$, $F_2(1, 23) = 12.68, p < 0.005$, $MSE = 28.87$, and $F_1(1, 31) = 57.68, p < 0.001$, $MSE = 0.03$, $F_2(1, 23) = 39.52, p < 0.001$, $MSE = 0.03$, respectively. Expected words were processed more accurately and faster than syntactically unexpected words. For the normalized RTs, the two-way interaction between syntactic expectancy and tonal function was significant, $F_1(1, 31) = 6.25, p < 0.05$, $MSE = 0.03$, $F_2(1, 23) = 8.95, p < 0.01$, $MSE = 0.02$. For expected words, a musical expectancy effect was observed: Processing was faster when the target word was presented with a tonic rather than with a subdominant chord $F_1(1, 31) = 6.12, p < 0.05$, $MSE = 0.05$, $F_2(1, 23) = 5.59, p < 0.05$, $MSE = 0.03$. For syntactically unexpected words, this tonic facilitation was not observed (F_1 and $F_2 < 1$). In addition, the syntactic expectancy effect was decreased when the target word was presented together with a subdominant chord, $F_1(1, 31) = 10.78, p < 0.005$, $MSE = 0.03$, $F_2(1, 23) = 8.71, p < 0.01$, $MSE = 0.02$, rather than with a tonic chord, $F_1(1, 31) = 50.96, p < 0.001$, $MSE = 0.03$, $F_2(1, 23) = 38.84, p < 0.001$, $MSE = 0.03$.

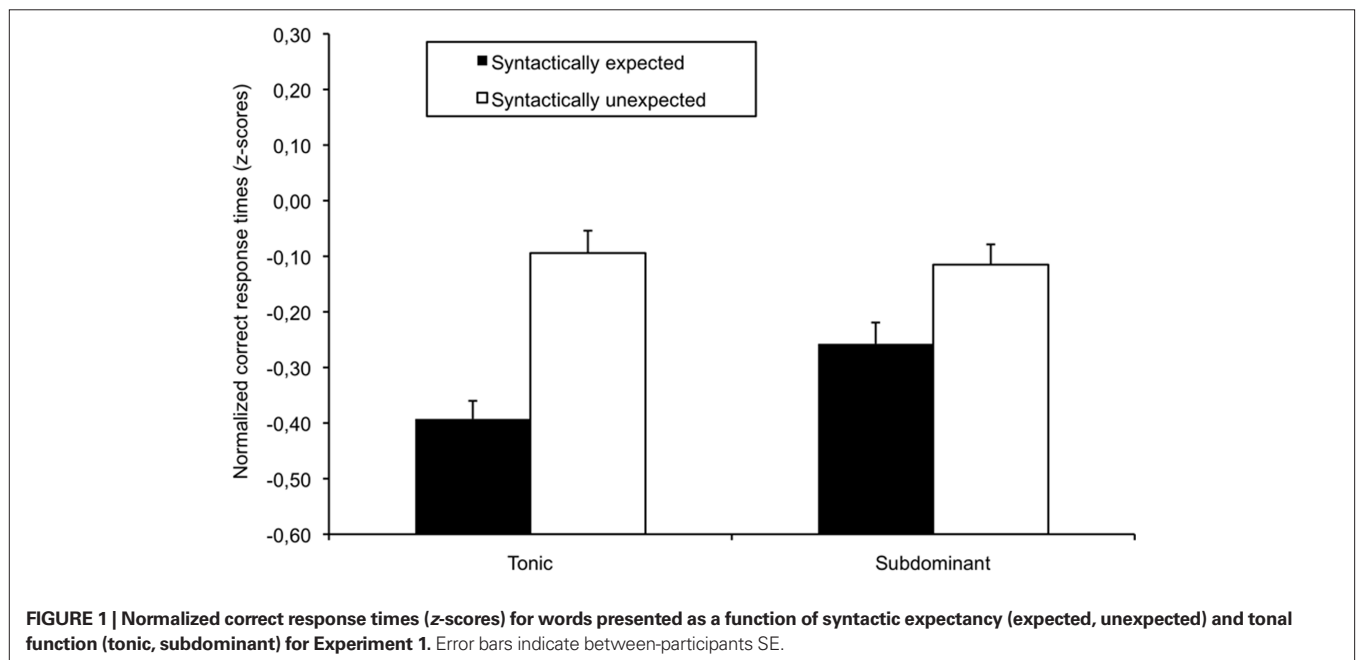
DISCUSSION

In Experiment 1, expected words were processed faster and more accurately than syntactically unexpected words, reflecting the influence of morphosyntactic expectancies (see Colé and Segui, 1994; Friederici et al., 1998; Gunter et al., 2000, for similar results). This syntactic expectancy effect was modulated by the tonal function of simultaneously presented chords: It was reduced for words presented with a less-expected subdominant chord compared to an expected tonic chord. The present interaction also revealed that

¹Analyses for pseudo-words are reported here for completion. For Experiments 1 and 2, respectively, percentages of correct responses and normalized RTs were analyzed by 2 × 2 ANOVAs with linguistic expectancy and tonal function as within-participant factors. No significant effects were observed, except a main effect of linguistic expectancy in Experiment 1. Pseudo-words that matched expected words were processed faster than pseudo-words that matched syntactically unexpected ones, $F_1(1, 31) = 4.49, p < 0.05$, $MSE = 0.04$, $F_2(1, 23) = 6.53, p < 0.05$, $MSE = 0.02$.

Table 1 | Percentages of correct responses (%CR) and correct response times (RTs, raw latencies in ms) presented as a function of Linguistic Expectancies (Syntactic for Experiment 1 and Semantic for Experiment 2) and Tonal Function (Tonic, Subdominant). SE are indicated in brackets.

Tonal function		Experiment 1				Experiment 2			
		Syntactic expectancy				Semantic expectancy			
		Syntactically expected		Syntactically unexpected		Semantically expected		Semantically unexpected	
%CR	Tonic	95.83	(0.75)	92.19	(0.89)	97.01	(0.82)	92.19	(1.45)
	subdominant	95.57	(0.77)	92.32	(1.15)	95.18	(1.38)	90.36	(1.52)
RTs	Tonic	564.91	(16.67)	603.32	(18.18)	572.15	(18.93)	635.30	(17.39)
	subdominant	582.45	(17.43)	601.47	(17.95)	583.46	(18.95)	646.32	(18.40)



the musical expectancy effect was observed only for the expected words, but not for the syntactically unexpected words. The expected words showed a tonic facilitation as previously observed for sung, spoken and visual syllables (Bigand et al., 2001; Escoffier and Tillmann, 2008; Hoch and Tillmann, 2010, respectively) and for expected words in sung sentences (Poulin-Charronnat et al., 2005). In contrast, syntactically unexpected words did not show the tonic facilitation. Escoffier and Tillmann (2008) had shown that in a cross-modal paradigm the relative facilitation between tonic and subdominant chords was due to a benefit of the tonic, rather than to a cost of the subdominant. Given that this facilitation requires the processing of musical structures and tonal functions, the interference observed here with the processing of the unexpected word suggests that musical structure and linguistic syntax processing tap into the same processing resources, thus hindering the otherwise observed tonic benefit.

In contrast to the findings of Experiment 1 that were observed with morphosyntactic agreement manipulations (see also Koelsch et al., 2005; Steinbeis and Koelsch, 2008), Slevc et al. (2009) have

reported a different type of interactive pattern between music-syntactic and linguistic-syntactic processing using a syntactic garden-path manipulation (see also Fedorenko et al., 2009, with object-extracted sentences): Self-paced reading-times were slower for target-words in syntactic garden-path sentences compared to simple sentences, but this effect was greater when presented with an unexpected out-of-key chord (107 ms) than with an expected in-key chord (31 ms). This comparison suggests that the type of linguistic-syntactic manipulation might lead to different interactive patterns with music-syntactic processing. Slevc et al. (2009) and our Experiment 1 also differed in the type of music-syntactic violation tested (i.e., unexpected musical events that were out-of-key vs. in-key). This difference might also influence the type of interactive pattern observed, notably with an additional cost due to strong musical violations that include acoustic violations. This hypothesis would need to be tested in an experimental paradigm that also integrates a baseline condition. However, it is worth noting that in Slevc et al. (2009), the expected words did not show the typically observed musical expectancy effect (i.e., faster processing

for the expected tonic chord, e.g., Bharucha and Stoeckig, 1986): Reading-times of expected words were faster with unexpected out-of-key chords (i.e., 606 ms) than with expected in-key chords (i.e., 639 ms, see Slevc et al., 2009 for a discussion).

Despite differences in interactive data patterns between our Experiment 1 and the study by Slevc et al. (2009), overall findings showed interactive influences between simultaneous music-syntactic and linguistic-syntactic processing, thus suggesting shared processing resources for musical-syntax and linguistic syntax, as proposed by the SSIRH (Patel, 2003).

EXPERIMENT 2: SEMANTIC EXPECTANCY AND MUSICAL EXPECTANCY

To further investigate whether shared resources between music and language processing are restricted to linguistic-syntactic processing or extend to linguistic-semantic processing, previous studies have investigated the simultaneous processing of musical syntax and linguistic semantics (Koelsch et al., 2005; Steinbeis and Koelsch, 2008; Slevc et al., 2009; see also Poulin-Charronnat et al., 2005). In contrast to the consistently observed interactive influences between the simultaneous processing of syntactic structures in music and language, these studies have revealed mixed data patterns. With a cross-modal presentation, the processing of a linguistic-semantic violation was not influenced by the simultaneous processing of a music-syntactic violation (i.e., out-of-key chords or tones, Koelsch et al., 2005; Slevc et al., 2009). However, when participants were required to perform a dual-task on language (visually presented) and music (auditorily presented), interactive influences were observed between music processing and semantic processing (Steinbeis and Koelsch, 2008). A similar interactive pattern was observed for vocal music and the use of more subtle musical expectancy violations in a behavioral priming paradigm (Poulin-Charronnat et al., 2005): The musical expectancy effect (i.e., faster processing for expected tonic chords than for less-expected subdominant chords) was observed only for semantically expected words but vanished for semantically unexpected words. This interaction also revealed that tonal function modulated semantic processing: The semantic expectancy effect (i.e., faster processing for semantically expected words than for unexpected words) was reduced for words sung on the less-expected subdominant chord compared to words sung on the expected tonic chord. In contrast to the studies that did not report interactive influences between musical and semantic structures (Koelsch et al., 2005; Slevc et al., 2009), the interactive influences reported by Steinbeis and Koelsch (2008) and Poulin-Charronnat et al. (2005) suggest the extension of shared resources from syntactic structures to semantic structures.

In view of the previous studies, which revealed mixed data patterns using various materials and tasks, Experiment 2 investigated simultaneous linguistic-semantic and music-syntactic processing with the musical materials and the linguistic task of Experiment 1. The musical and linguistic materials and the linguistic task were comparable to Poulin-Charronnat et al. (2005), but here used in a cross-modal presentation. As in Experiment 1, Experiment 2 used music-syntactic manipulation without acoustic confound, thus contrasting with all the previous studies using a cross-modal presentation (Koelsch et al., 2005; Steinbeis and Koelsch, 2008; Slevc et al., 2009): Chord sequences ended on either the expected

tonic or the less-expected subdominant chord. For the language manipulation, the final word of the sentence was semantically expected or unexpected (as in all the previous studies). Thus, Experiment 2 investigated whether previously observed interactive influences between music and semantic processing (Poulin-Charronnat et al., 2005; Steinbeis and Koelsch, 2008) extends to a cross-modal presentation with subtle musical expectancy violations and with a task focusing on language only. If previously reported interactive influences between music-syntactic and linguistic-semantic processing were not due to the experimental task and/or the form of presentation of the materials (i.e., vocal music in Poulin-Charronnat et al., 2005, and dual-task paradigm in Steinbeis and Koelsch, 2008), then in Experiment 2, the linguistic-semantic expectancy effect should be modulated by the tonal function of the final chord, and the tonal function effect (i.e., the musical expectancy effect) should be modulated by the linguistic-semantic expectancy.

METHOD

Participants

Thirty-two students from the University of Lyon ($M \pm SD = 21 \pm 2.38$ years) participated in Experiment 2. The number of years of formal musical practice ranged from 0 to 14 (2.78 ± 4.32 years; $Mdn = 0$). This mean level of musical practice did not differ significantly from that of the participants in Experiment 1 [$t(31) = 0.90$, $p = 0.38$]. All participants gave informed consent. None of the participants had declared to have an auditory impairment, and none participated in Experiment 1.

Materials and procedure

The musical sequences and the sentences ending on syntactically expected words of Experiment 1 were used. These syntactically expected words were also semantically expected, and they had been chosen, as described in Experiment 1, based on a free-completion test. Sentences were then matched by pair so that the semantically expected target word of one sentence defined the semantically unexpected target word of another sentence (and *vice versa*), resulting in 24 sentences ending on a semantically expected word and 24 sentences ending on a semantically unexpected word [e.g., “Le méchant chien dort dans la **niche** (vs. **tente**)”/“*The nasty dog is sleeping in the **kennel** (vs. **tent**)*”]. These combinations of semantically unexpected target-words were pretested (with 29 new participants, students of the University of Lyon). The semantically unexpected word was never chosen as an appropriate ending among multiple choices (including expected/unexpected words and fillers), except by two participants for one sentence (sentence 22, see Appendix). The procedure was as described for Experiment 1.

RESULTS

The mean accuracy was 94% and 95% for target-words and pseudo-words, respectively. The means of correct RTs were 608 ms (range, 447–879 ms) and 656 ms (range, 533–880 ms) for words and pseudo-words, respectively. As for Experiment 1, correct RTs were individually normalized with a mean of 0 and a SD of 1, providing z-scores (see footnote 1). For words, percentages of correct responses (Table 1) and normalized RTs (Figure 2) were, respectively, analyzed by two 2×2 ANOVAs with semantic expectancy

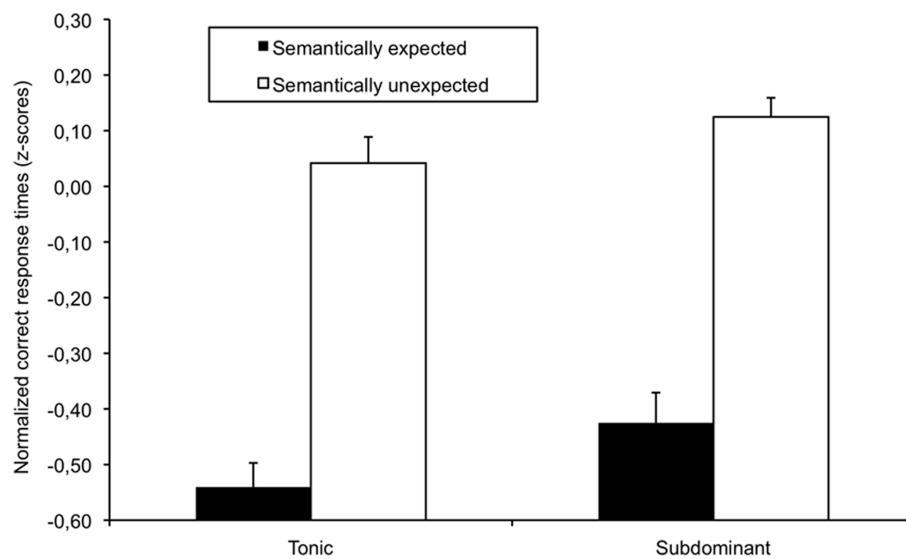


FIGURE 2 | Normalized correct response times (z-scores) for words presented as a function of semantic expectancy (expected, unexpected) and tonal function (tonic, subdominant) for Experiment 2. Error bars indicate between-participants SE.

(expected, unexpected) and tonal function (tonic, subdominant) as within-participant factors and either participants (F_1) or target-words (F_2) as random variables.

For correct responses and normalized RTs, the main effect of semantic expectancy was significant, $F_1(1, 31) = 20.98, p < 0.001$, $MSE = 35.40$, $F_2(1, 23) = 22.31, p < 0.001$, $MSE = 24.30$, and $F_1(1, 31) = 127.88, p < 0.001$, $MSE = 0.08$, $F_2(1, 23) = 91.44, p < 0.001$, $MSE = 0.09$, respectively: Expected words were processed more accurately and faster than semantically unexpected words. The main effect of tonal function was significant for normalized RTs, $F_1(1, 31) = 6.05, p < 0.05$, $MSE = 0.05$, $F_2(1, 23) = 7.71, p < 0.05$, $MSE = 0.03$, and marginally significant for correct responses, $F_1(1, 31) = 3.28, p = 0.08$, $MSE = 32.41$, $F_2(1, 23) = 3.70, p = 0.07$, $MSE = 20.02$, target word processing was more accurate and faster when presented in synchrony with a tonic chord than with a subdominant chord. The two-way interaction between semantic expectancy and tonal function was not significant [$F_1(1, 31) = 0.21, p = 0.65$, $MSE = 0.04$; $F_2(1, 23) = 0.83, p = 0.37$, $MSE = 0.03$ for normalized RTs; and $F_1(1, 31) = 0.000, p = 1.00$, $MSE = 12.88$, $F_2(1, 23) = 0.005, p = 0.94$, $MSE = 19.00$ for percentages of correct responses].

DISCUSSION

Experiment 2 replicated the well-known semantic priming effect (see McNamara, 2005 for a review), with semantically expected words being processed faster and more accurately than unexpected words. In addition, a main effect of tonal function was observed with faster and more accurate processing of words presented in synchrony with an expected tonic chord than with a less-expected subdominant chord. The facilitated visual word processing when simultaneously presented with a tonic chord confirmed the previously described tonic facilitation, as observed for sung, spoken, and visual syllables (Bigand et al., 2001; Escoffier and Tillmann, 2008; Hoch and Tillmann, 2010).

Experiment 2 did not reveal interactive influences between linguistic-semantic and music-syntactic processing. The semantic priming effect did not modulate the musical expectancy effect nor was it modulated by the musical expectancy effect. The absence of interactive influence is consistent with some of the previous data observed with stronger musical expectancy violations (Koelsch et al., 2005; Slevc et al., 2009). As previously argued, it might suggest independent resources between music-syntactic and linguistic-semantic processing (see however the General Discussion here below).

Further insight can be provided by the comparison of the various data patterns of previous experimental studies that used similar materials, but differed in material presentation or tasks. In contrast to the data of Experiment 2, the data by Poulin-Charronnat et al. (2005) suggested shared resources for music and semantic processing: They revealed interactive influences with similar materials and tasks as in Experiment 2, but presented as vocal music. The data by Steinbeis and Koelsch (2008) also suggested shared resources with the observation of interactive influences for the same materials and cross-modal presentation as in Koelsch et al. (2005), but using a dual-task (instead of a single task leading to a data pattern suggesting independent influences). Taken together, these findings suggest that the presentation form of the experimental material (as separate information streams in a cross-modal paradigm or combined in one information stream as in vocal music) and the attentional level required for the processing of each material type (in single vs. dual-tasks) influence whether interactive data patterns are observed between simultaneous music and semantic processing. In particular, interactive patterns might be observed for semantics when the tonal function is part of the task-relevant information stream(s).

It is worth noting that in contrast to the chord material used in the previously cited studies, behavioral and ERPs for sung melodies suggested independent processing of music and semantic

expectancy violations (Besson et al., 1998; Bonnel et al., 2001). However, even for this melodic material, (1) a reduction of the N400 for semantically unexpected words sung on an unexpected tone can be noticed in the reported ERPs (see Besson et al., 1998, **Figure 2**, p. 496, but not significant), and (2) these studies (Besson et al., 1998; Bonnel et al., 2001) used two explicit tasks (i.e., semantic and musical coherence judgments) that might lead participants to separately analyze musical and linguistic information, thus weakening interactive influences.

GENERAL DISCUSSION

Our study manipulated the tonal function of chords together with either syntactic or semantic structures in a cross-modal paradigm. Experiment 1 revealed interactive influences between music-syntactic and linguistic-syntactic processing. The effect of linguistic-syntactic expectancy was reduced when the target word was simultaneously presented with a less-expected subdominant chord, and there was no effect of musical expectancy for syntactically unexpected words. This outcome extends interactive influences previously observed with strong violations of musical expectancies (Koelsch et al., 2005; Steinbeis and Koelsch, 2008; Fedorenko et al., 2009; Slevc et al., 2009) to more subtle violations of musical expectancies. By contrast, Experiment 2 did not reveal interactive influences between music-syntactic and linguistic-semantic processing: Only main effects of tonal function and semantic expectancy were observed. The absence of interactive influences is in agreement with some of the previous studies investigating music-syntactic and linguistic-semantic processing (Koelsch et al., 2005; Slevc et al., 2009), but differs from the interactive influences consistently observed for music-syntactic and linguistic-syntactic processing (Experiment 1; Koelsch et al., 2005; Slevc et al., 2009).

Overall, the findings of our two cross-modal experiments can be interpreted as new support for the SSIRH (Patel, 2003), notably by revealing interactive influences between music-syntactic and linguistic-syntactic processing, but not between music-syntactic and linguistic-semantic processing. More importantly, the controlled construction of our musical material allowed focusing on the influence of musical structure processing, notably by excluding that the observed interference was due to acoustic deviance processing (thus going beyond previous studies using out-of-key events, e.g., Koelsch et al., 2005; Steinbeis and Koelsch, 2008). In particular here, the influence of the musical structures was based on the tonic facilitation (rather than on a subdominant cost), as suggested by previous research including a neutral baseline condition (Escoffier and Tillmann, 2008).

In the following, we propose to discuss our study together with the currently available data on simultaneous musical and linguistic (syntactic or semantic) structure processing, revealing the need for future studies and suggesting an extension of the SSIRH. Simultaneous music and syntactic processing has consistently shown interactive influences (Koelsch et al., 2005; Steinbeis and Koelsch, 2008; Slevc et al., 2009; Fedorenko et al., 2009; Experiment 1), while simultaneous music and semantic processing has either shown interactive influences or not (Poulin-Charronnat et al., 2005; Steinbeis and Koelsch, 2008 vs. Koelsch et al., 2005; Slevc et al., 2009; Experiment 2). To further investigate simultaneous musical

and linguistic structure processing (notably as a function of the type of linguistic expectancies, syntax vs. semantics), we ran an additional analysis combining Experiments 1 and 2: This combined analysis revealed a two-way interaction between musical and linguistic expectancy effects, but this interaction was not significantly modulated by the type of linguistic manipulation (syntactic or semantic)². Interestingly, even though the interaction was not significant in Experiment 2, the effect sizes suggested a similar interactive pattern as in Experiment 1: The influence of tonal function was stronger for expected words ($\eta_p^2 = 0.16$ and $\eta_p^2 = 0.13$ for syntax in Experiment 1, and for semantics in Experiment 2, respectively) than for unexpected words ($\eta_p^2 = 0.00$ and $\eta_p^2 = 0.07$, for syntax and semantics, respectively). It is worth noting that also in Slevc et al. (2009), the mean RT data for semantics mirrored this interactive pattern (see also Poulin-Charronnat et al., 2005), even if not significantly: The mean reading-times of semantically expected and unexpected words differed more strongly when the simultaneously presented chord was the expected in-key chord (i.e., 71 ms) than when it was the unexpected out-of-key chord (i.e., 38 ms, see Slevc et al., 2009, **Table 1**, p. 377).

Beyond the influence of the presentation form and the experimental task (see Discussion of Experiment 2), the rather unstable data pattern for music and semantic processing, which is observed over the currently available studies, might have been based on the type of semantic violation that has been used in all the cited studies, and that contrasts to the expectancy violations applied to syntax. The challenge to aim for comparable violation types and to equate the levels of processing difficulty between the to-be-compared materials is not only encountered by research investigating language and music, but is also well-known in research investigating linguistic syntax and semantics (e.g., Friederici et al., 1993, 2003). In studies investigating language and music processing, up to now, the syntactic expectancy violations consisted of syntactic errors (e.g., gender violations, Koelsch et al., 2005; Steinbeis and Koelsch, 2008; Experiment 1) or syntactic complex sentences (Fedorenko et al., 2009; Slevc et al., 2009), while the semantic expectancy violations consisted of correct, but low-cloze probability words (Koelsch et al., 2005; Poulin-Charronnat et al., 2005; Steinbeis and Koelsch, 2008; Slevc et al., 2009; Experiment 2). This comparison thus points to the potential influence of the strength of the structure manipulations, and it raises the question whether the use of stronger semantic violations (semantic errors, semantically implausible words) or semantically more complex structures might produce more consistent interference with simultaneous music processing. The underlying hypothesis would be that, as previously suggested by Steinbeis and Koelsch (2008), neural resources might be shared for the processing of structure violation and the integration of unexpected events for both musical structures and linguistic-semantic structures. As phrased by Slevc et al. (2009), music and language processing might share “resources for a more general type of processing (e.g., for a process of integrating new information into any type of evolving representation),” p. 375).

²The two-way interaction between musical and linguistic expectancy manipulation was significant, $F(1, 62) = 4.01$, $p < 0.05$, $MSE = 0.04$, but the three-way interaction integrating also the type of linguistic manipulation (syntactic or semantic) was not significant: $F(1, 62) = 1.73$, $p = 0.19$, $MSE = 0.04$.

Research in music cognition and psycholinguistics has suggested that both music and language processing require integrative processes. Each incoming event needs to be integrated on-line into an updated mental representation of the global context, notably to form a coherent and meaningful representation (Friederici, 2001; Jackendoff, 2002; Patel, 2003, 2008; Hagoort, 2005; Tillmann, 2005). For language, structural integration of information over time is necessary for both syntactic and semantic processing (Gibson, 1998; Friederici, 2001; Jackendoff, 2002; Hagoort, 2005). At the sentence level, readers or listeners need to integrate newly incoming information to update their mental representation, and create a coherent and meaningful situational model (e.g., van Dijk and Kintsch, 1983; Kintsch, 1988). This research in music and language cognition, taken together with the observation of interactive influences between music and syntactic processing and between music and semantic processing (at least in some of the studies), led us to propose that structural integration resources might be the key concept of shared resources in the processing of music, syntax, and semantics. The SSIRH might thus be extended from syntactic to more general structural integration resources. This hypothesis of shared structural integration resources needs to be further investigated by testing (1) stronger semantic expectancy violations (e.g., semantically implausible words) or sentences with more complex semantic structures (e.g., semantic garden-path sentences, role assignment violations), which require more complex processes such as a reanalysis and a reinterpretation of the previous information, and (2) predictions for interactive influences beyond music and linguistic structure processing, notably for the simultaneous processing of other structured materials, such as arithmetic, movies, dance, or action sequences (see also Jackendoff, 2009)³. This hypothesis integrates in research revealing the role of inferior frontal cortex (in particular, Broca's area and its right-hemisphere homolog) not only in language processing for syntax and semantics (e.g., Kotz et al., 2002), but also in the processing of musical structures (e.g., Maess et al., 2001; Tillmann et al., 2003) and artificial grammar structures (Petersson et al., 2004, 2010) as well as sequential manipulation

and structuring of notes, syllables, visuo-spatial materials or action sequences (e.g., Gelfand and Bookheimer, 2003; Tettamanti and Weniger, 2006) and the perception and production of temporal sequences (Schubotz et al., 2000; Coull, 2004). Converging evidence has been provided by patient studies (in particular Broca aphasics), showing deficits in language processing as well as musical structure processing (Patel et al., 2008), artificial grammar learning (Christiansen et al., 2010) and the perception of human action (Fazio et al., 2009). As developed by Petersson et al. (2010), the inferior frontal cortex (centered on BA 44 and 45) might thus be a main candidate for the encoding and subsequent processing of discrete elements presented in a sequence, thus providing domain-general neural resources for on-line sequence processing that requires the structural integration of events in an incremental and recursive manner.

CONCLUSION

Recent research has investigated the hypothesis of neural resources shared between music and language processing, in particular differentiating syntax and semantics of the language material. While musical structure processing interacted consistently with linguistic syntax processing, the data patterns observed for musical-syntax and linguistic-semantics were less clear, showing either interactive patterns or not. Our study tested the simultaneous processing of musical syntax with either linguistic syntax (Experiment 1) or semantics (Experiment 2) and the observed data pattern is in agreement with these previous findings. A closer analysis of material, violation types as well as experimental tasks suggests the extended hypothesis of shared structural integration resources that goes beyond syntactic integration. This extended hypothesis now requires future research to further investigate not only music and language processing, but also the behavioral and neural correlates of the simultaneous processing of other material types that require structural and temporal integration (such as action, for example).

ACKNOWLEDGMENTS

Thanks to Pierre Sereau for help in preparing and running pre-tests. This research was supported by the grant ANR-09-BLAN-0310.

³These predictions have been started to be tested in recent works (see Sammler et al., 2010 for action; Hoch and Tillmann, submitted, for maths).

REFERENCES

- Basso, A., and Capitani, E. (1985). Spared musical abilities in a conductor with global aphasia and ideomotor apraxia. *J. Neurol. Neurosurg. Psychiatr.* 48, 407–412.
- Besson, M., Faïta, F., Peretz, I., Bonnel, A. M., and Requin, J. (1998). Singing in the brain: independence of lyrics and tunes. *Psychol. Sci.* 9, 494–498.
- Besson, M., and Schön, D. (2003). "Comparison between music and language," in *The Cognitive Neuroscience of Music*, eds I. Peretz and R. Zatorre (Oxford: Oxford University Press), 269–293.
- Bharucha, J. J., and Stoeckig, K. (1986). Reaction time and musical expectancy: priming of chord. *J. Exp. Psychol. Hum. Percept. Perform.* 12, 403–410.
- Bigand, E., Poulin, B., Tillmann, B., Madurell, F., and D'Adamo, D. (2003). Sensory versus cognitive components in harmonic priming. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 159–171.
- Bigand, E., Tillmann, B., and Poulin-Charronnat, B. (2006). A module for syntactic processing in music? *Trends Cogn. Sci. (Regul. Ed.)* 10, 195–196.
- Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D. A., and Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition* 81, B11–B20.
- Bonnel, A. M., Faïta, F., Peretz, I., and Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: evidence for independent processing. *Percept. Psychophys.* 63, 1201–1213.
- Christiansen, M. H., Kelly, M. L., Shillcock, R. C., and Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition* 116, 382–393.
- Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (1993). Psyscope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behav. Res. Methods Instrum. Comput.* 25, 257–271.
- Colé, P., and Segui, J. (1994). Grammatical incongruity and vocabulary types. *Mem. Cognit.* 22, 387–394.
- Coull, J. T. (2004). fMRI studies of temporal attention: allocating attention within, or towards, time. *Brain Res. Cogn. Brain Res.* 21, 216–226.
- Escoffier, N., and Tillmann, B. (2008). The tonal function of a task-irrelevant chord modulates speed of visual processing. *Cognition* 107, 1070–1080.
- Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A. C., Pozzo, T., Calzolari, F., Granieri, E., and Fadiga, L. (2009). Encoding of human action in Broca's area. *Brain* 132, 1980–1988.
- 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. Cognit.* 37, 1–9.
- Friederici, A. D. (2001). "The neural basis of sentence processing: a neurocognitive model," in *Towards a New Functional Anatomy of Language Cognition*, eds G. Hickok and D. Poeppel (special issue).

- Friederici, A. D., Pfeifer, E., and Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Brain Res. Cogn. Brain Res.* 1, 183–192.
- Friederici, A. D., Rüschemeyer, S.-A., Hahne, A., and Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb. Cortex* 13, 170–177.
- Friederici, A. D., Schriefers, H., and Lindenberger, U. (1998). Differential age effects on semantic and syntactic priming. *Int. J. Behav. Dev.* 22, 813–845.
- Gelfand, J. R., and Bookheimer, S. Y. (2003). Dissociating neural mechanisms of temporal sequencing and processing phonemes. *Neuron* 38, 831–842.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 76, 1–76.
- Gunter, T. C., Friederici, A. D., and Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J. Cogn. Neurosci.* 12, 556–568.
- Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends Cogn. Sci. (Regul. Ed.)* 9, 416–423.
- Hoch, L., and Tillmann, B. (2010). Laterality effects for musical structure processing: a dichotic listening study. *Neuropsychology* 24, 661–666.
- Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. New York: Oxford University Press.
- Jackendoff, R. (2009). Parallels and non-parallels between language and music. *Music Percept.* 26, 195–204.
- Jonides, J., and Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychol. Bull.* 96, 29–44.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychol. Rev.* 95, 163–182.
- Koelsch, S., Gunter, T. C., Schröger, E., Tervaniemi, M., Sammler, D., and Friederici, A. D. (2001). Differentiating ERAN and MMN: an ERP study. *Neuroreport* 12, 1385–1389.
- Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., and Friederici, A. D. (2002). Bach speaks: a cortical “language-network” serves the processing of music. *Neuroimage* 17, 956–966.
- 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.
- Koelsch, S., Jentschke, S., Sammler, D., and Mietschen, D. (2007). Untangling syntactic and sensory processing: an ERP study of music perception. *Psychophysiology* 44, 476–490.
- Kotz, S. A., Cappa, S. F., von Cramon, D. Y., and Friederici, A. D. (2002). Modulation of the lexical-semantic network by auditory semantic priming: an event-related functional MRI study. *Neuroimage* 17, 1761–1772.
- 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, 540–544.
- McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Sussex: Psychology Press.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681.
- Patel, A. D. (2008). *Music, Language, and the Brain*. New York: Oxford University Press.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. J. (1998). Processing syntactic relations in language and music: an event-related potential study. *J. Cogn. Neurosci.* 10, 71–733.
- Patel, A. D., Iversen, J. R., Wassenaar, M., and Hagoort, P. (2008). Musical syntactic processing in agrammatic Broca’s aphasia. *Aphasiology* 22, 776–789.
- Peretz, I., Belleville, S., and Fontaine, S. (1997). Dissociation entre musique et langage après atteinte cérébrale: un nouveau cas d’amusie sans aphasie. *Can. J. Exp. Psychol.* 51, 354–367.
- Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., and Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain* 117, 1283–1301.
- Petersson, K. M., Folia, V., and Hagoort, P. (2010). What artificial grammar learning reveals about the neurobiology of syntax. *Brain Lang.* [Epub ahead of print].
- Petersson, K. M., Forkstam, C., and Ingvar, M. (2004). Artificial syntactic violations activate Broca’s region. *Cogn. Sci.* 28, 383–407.
- Poulin-Charronnat, B., Bigand, E., Madurell, F., and Peereman, R. (2005). Musical structure modulates semantic priming in vocal music. *Cognition* 94, B67–B78.
- Sammler, D., Harding, E. E., D’Ausilio, A., Fadiga, L., and Koelsch, S. (2010). “Music and action: do they share neural resources?” in *11th International Conference of Music Perception and Cognition*, eds S. M. Demorest, S. J. Morrison, and P. S. Campbell (Seattle, WA: Causal Productions), 87.
- Schubotz, R. I., Friederici, A. D., and von Cramon, D. Y. (2000). Time perception and motor timing: a common cortical and subcortical basis revealed by fMRI. *Neuroimage* 11, 1–12.
- 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.
- 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.
- Tekman, H. G., and Bharucha, J. J. (1998). Implicit knowledge versus psychoacoustic in priming of chords. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 252–260.
- Terhardt, E. (1984). The concept of musical consonance: a link between music and psychoacoustics. *Music Percept.* 1, 276–295.
- Tettamanti, M., and Weniger, D. (2006). Broca’s area: a supramodal hierarchical processor? *Cortex* 42, 491–494.
- Tillmann, B. (2005). Implicit investigations of tonal knowledge in nonmusicians listeners. *Ann. N. Y. Acad. Sci.* 1060, 1–11.
- Tillmann, B., Janata, P., and Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Brain Res. Cogn. Brain Res.* 16, 145–161.
- van Dijk, T. A., and Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic 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: 14 February 2011; accepted: 13 May 2011; published online: 06 June 2011.
Citation: Hoch L, Poulin-Charronnat B and Tillmann B (2011) The influence of task-irrelevant music on language processing: syntactic and semantic structures. *Front. Psychology* 2:112. doi: 10.3389/fpsyg.2011.00112
This article was submitted to *Frontiers in Auditory Cognitive Neuroscience*, a specialty of *Frontiers in Psychology*.
Copyright © 2011 Hoch, Poulin-Charronnat and Tillmann. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.

APPENDIX

No.	Context	Expected target word	Syntactically unexpected target word	Semantically unexpected target word
1	Le bon cam-peur dort dans <i>The good camper is sleeping in</i>	la tente <i>the_(fem.) tent</i>	le tente <i>the_(masc.) tent</i>	la niche <i>the_(fem.) kennel</i>
2	Le mé-chant chien dort dans <i>The nasty dog is sleeping in</i>	la niche <i>the_(fem.) kennel</i>	le niche <i>the_(masc.) kennel</i>	la tente <i>the_(fem.) tent</i>
3	Le beau mou-choir est dans <i>The beautiful handkerchief is in</i>	la poche <i>the_(fem.) pocket</i>	le poche <i>the_(masc.) pocket</i>	la tasse <i>the_(fem.) cup</i>
4	Le pe-tit sucre est dans <i>The small piece of sugar is in</i>	la tasse <i>the_(fem.) cup</i>	le tasse <i>the_(masc.) cup</i>	la poche <i>the_(fem.) pocket</i>
5	Le nou-veau train est dans <i>The new train is coming into</i>	la gare <i>the_(fem.) station</i>	le gare <i>the_(masc.) station</i>	la classe <i>the_(fem.) class</i>
6	Le sage é-lève est dans <i>The wise student is in</i>	la classe <i>the_(fem.) class</i>	le classe <i>the_(masc.) class</i>	la gare <i>the_(fem.) station</i>
7	Le jo-vial bu-veur lève <i>The cheerful drinker is rising</i>	le verre <i>the_(masc.) glass</i>	la verre <i>the_(fem.) glass</i>	le sort <i>the_(masc.) fate</i>
8	Le mé-chant sor-cier jette <i>The evil wizard is casting</i>	le sort <i>the_(masc.) fate</i>	la sort <i>the_(fem.) fate</i>	le verre <i>the_(masc.) glass</i>
9	Le gros gâ-teau gonfle dans <i>The big cake is rising in</i>	le four <i>the_(masc.) oven</i>	la four <i>the_(fem.) oven</i>	le trou <i>the_(masc.) hole</i>
10	Le pe-tit rat entre dans <i>The little rat is coming into</i>	le trou <i>the_(masc.) hole</i>	la trou <i>the_(fem.) hole</i>	le four <i>the_(masc.) oven</i>
11	Le sé-rieux doc-teur prend <i>The serious doctor is taking</i>	le pouls <i>the_(masc.) pulse</i>	la pouls <i>the_(fem.) pulse</i>	le nord <i>the_(masc.) north</i>
12	Le jeune ran-don-neur perd <i>The young hiker loses</i>	le nord <i>the_(masc.) north</i>	la nord <i>the_(fem.) north</i>	le pouls <i>the_(masc.) pulse</i>
13	Le vieux bou-lan-ger roule <i>The old baker is rolling</i>	la pâte <i>the_(fem.) dough</i>	le pâte <i>the_(masc.) dough</i>	la feuille <i>the_(fem.) sheet</i>
14	Le jeune é-co-li-er plie <i>The young schoolboy is folding</i>	la feuille <i>the_(fem.) sheet</i>	le feuille <i>the_(masc.) sheet</i>	la pâte <i>the_(fem.) dough</i>
15	Le bon mè-nui-sier coupe <i>The good carpenter is cutting</i>	la planche <i>the_(fem.) board</i>	le planche <i>the_(masc.) board</i>	la viande <i>the_(fem.) meat</i>
16	Le cos-taud bou-cher coupe <i>The strong butcher is cutting</i>	la viande <i>the_(fem.) meat</i>	le viande <i>the_(masc.) meat</i>	la planche <i>the_(fem.) board</i>
17	Le jeune nu-dist est sur <i>The young nudist is on</i>	la plage <i>the_(fem.) beach</i>	le plage <i>the_(masc.) beach</i>	la scène <i>(fem.) stage</i>
18	Le grand chan-teur est sur <i>The famous singer is on</i>	la scène <i>(fem.) stage</i>	le scène <i>(masc.) stage</i>	la plage <i>the_(fem.) beach</i>
19	Le gour-mand che-val mange <i>The greedy horse is eating</i>	le foin <i>the_(masc.) hay</i>	la foin <i>the_(fem.) hay</i>	le joint <i>the_(masc.) gasket</i>
20	Le pe-tit dro-gué fume <i>The little drug addict is smoking</i>	le joint <i>the_(masc.) joint</i>	la joint <i>the_(fem.) joint</i>	le foin <i>the_(masc.) hay</i>
21	Le jeune é-co-li-er lève <i>The young schoolboy is lifting</i>	le doigt <i>the_(masc.) finger</i>	la doigt <i>the_(fem.) finger</i>	le clou <i>the_(masc.) nail</i>
22	Le gauche bri-co-leur plante <i>The clumsy home constructor is planting</i>	le clou <i>the_(masc.) nail</i>	la clou <i>the_(fem.) nail</i>	le doigt <i>the_(masc.) finger</i>
23	Le bel a-vion vole dans <i>The beautiful plane is flying in</i>	le ciel <i>the_(masc.) sky</i>	la ciel <i>the_(fem.) sky</i>	le bus <i>the_(masc.) bus</i>
24	Le jeune é-lève monte dans <i>The young student is getting into</i>	le bus <i>the_(masc.) bus</i>	la bus <i>the_(fem.) bus</i>	le ciel <i>the_(masc.) sky</i>



Relating pitch awareness to phonemic awareness in children: implications for tone-deafness and dyslexia

Psyche Loui^{1*}, Kenneth Kroog², Jennifer Zuk¹, Ellen Winner^{2,3} and Gottfried Schlaug¹

¹ Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

² Department of Psychology, Boston College, Chestnut Hill, MA, USA

³ Project Zero, Harvard Graduate School of Education, Cambridge, MA, USA

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Lutz Jäncke, University of Zurich, Switzerland

Nina Kraus, Northwestern University, USA

Franziska Degé, Justus-Liebig-University, Germany

*Correspondence:

Psyche Loui, Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Avenue, Palmer 127, Boston, MA 02215, USA.

e-mail: ploui@bidmc.harvard.edu

Language and music are complex cognitive and neural functions that rely on awareness of one's own sound productions. Information on the awareness of vocal pitch, and its relation to phonemic awareness which is crucial for learning to read, will be important for understanding the relationship between tone-deafness and developmental language disorders such as dyslexia. Here we show that phonemic awareness skills are positively correlated with pitch perception–production skills in children. Children between the ages of seven and nine were tested on pitch perception and production, phonemic awareness, and IQ. Results showed a significant positive correlation between pitch perception–production and phonemic awareness, suggesting that the relationship between musical and linguistic sound processing is intimately linked to awareness at the level of pitch and phonemes. Since tone-deafness is a pitch-related impairment and dyslexia is a deficit of phonemic awareness, we suggest that dyslexia and tone-deafness may have a shared and/or common neural basis.

Keywords: music, language, development, phonemic awareness, pitch perception, production, tone-deafness, dyslexia

INTRODUCTION

Developmental dyslexia is a disorder in which children with normal intelligence and sensory abilities show learning deficits for reading (Démonet et al., 2004). Children with dyslexia are characterized by an impairment in phonemic awareness, which is the ability to process and manipulate spoken words made up of individual sounds or phonemes (Bradley and Bryant, 1983; Stanovich, 1988; Stahl and Murray, 1994; Snow et al., 1998). In recent years, the relationship between phonemic awareness and musical ability is a topic of increased interest. Studies have demonstrated significant relationships between phonemic awareness and musical sound processing in children as well as adults (Stadler, 1990; Anvari et al., 2002; Jentschke et al., 2005; Jones et al., 2009a; Moreno et al., 2009). One line of research shows that the processing of rhythm and meter in music predicts phonological difficulties in children with dyslexia, suggesting that dyslexia may arise from a difficulty in processing the fast temporal structure of speech sounds (Tallal and Piercy, 1973; Merzenich et al., 1996; Overly, 2003; Benasich et al., 2006; Huss et al., 2011). Since consonants within a language are largely differentiated by temporally rapid sound elements, the ability to process temporally rapid changes is important for phonemic awareness. Support for this theory comes from the finding that auditory training that involves the gradual increase in temporal frequency of tones leads to improvements in auditory processing speed, speech discrimination, and language comprehension (Tallal, 2004; Tallal and Gaab, 2006).

Another line of research has shown links between phonemic awareness and pitch processing, rather than temporal processing, of musical sounds. As different consonants and vowels in speech are characterized by different frequency as well as temporal content, and pitch is the direct perceptual correlate of frequency, pitch

perception and phoneme perception might share the same process of frequency discrimination. Thus, sensitivity to frequency changes may be important for reading ability as well. In previous research, phonemic awareness was shown to be correlated with pitch discrimination, pitch awareness (as assessed using the tonal test of the Gordon Primary Measures of Audiation (PMMA; Gordon, 1979), melody discrimination, and tone manipulation tasks (Lamb and Gregory, 1993; Anvari, et al., 2002; Peynircioglu et al., 2002; Bolduc, 2005; Forgeard et al., 2008a). Additionally, children with specific language impairment were shown to be impaired at pitch discrimination tasks, further suggesting that the link between phonemic awareness and musical ability holds in the domain of pitch (Mengler et al., 2005).

Individuals who are tone-deaf also have impaired pitch processing. Tone-deafness, or congenital amusia, is a deficit of musical ability that according to some reports affects up to 17% of the normal adult population (Kalmus and Fry, 1980; Cuddy et al., 2005; Loui et al., 2009; Peretz et al., 2009). Little is known regarding the developmental trajectory of tone-deafness in children, although the topic is beginning to be explored (Lebrun et al., in press). Adults who are tone-deaf have deficits in pitch production as well as perception, and show a mismatch between their pitch perception and production abilities (Loui et al., 2008). They also show impairments in phonemic awareness (Jones et al. 2009a) as well as impairments in the processing of speech intonation (Liu et al., 2010) and in the understanding of lexical tones in a tonal language (Nan et al., 2010). Taken together, the research suggests a relationship between phonemic awareness and pitch perception. However, the relationship between phonemic awareness and pitch production is yet unclear, especially in young children. Understanding tone-deaf behavior in children – as characterized by deficits in

pitch perception, production, and the relationship between the two – and the possible relationship between tone-deafness and phonemic awareness, is important for characterizing the specificity of difficulties that children with developmental language disorders might face in their auditory environment.

In what follows, we examine the relationship between pitch perception and production, on the one hand, and phonemic awareness, on the other hand, in children between the ages of seven and nine. This age range was chosen because it is the earliest time when measurable deficits in literacy development can be readily identified (Shaywitz and Shaywitz, 2005). We hypothesized that the degree of correlation between perception and production performance (which we call pitch awareness) should be positively associated with level of phonemic awareness.

MATERIALS AND METHODS

PARTICIPANTS

Thirty-two children (16 females, 16 males) were recruited from Grant School No. 7 in Passaic, NJ, USA, and from schools in the greater Boston area. Children's mean age was 7.6 years ($SD = 0.7$ years). Children had a mean of 0.4 years of musical training outside of normal school education (range = 0–3 years, with five children having more than a year of musical training). Median household income was between US \$35,000 and \$50,000.

STIMULI AND PROCEDURE

Prior to testing, parents were contacted to obtain informed consent and to fill out a basic questionnaire. This questionnaire included items asking for the family's contact information as well as the child's date of birth and whether or not they had received any musical training outside of normal schooling. Each child was tested on a battery of tests that assessed phonemic awareness, pitch perception and production, and general intelligence. The pitch perception and production tests were used previously in our lab with tone-deaf adults (Loui, et al., 2008). Phonemic awareness scores were obtained from a combination of existing tests, the Sound Categorization tests (Bradley and Bryant, 1985) and Auditory Analysis tests (Rosner and Simon, 1971).

Phonemic awareness tests

Two phonemic awareness tests were administered: the Sound Categorization Test (Bradley and Bryant, 1985) and the Auditory Analysis Test (Rosner and Simon, 1971). The Sound Categorization Test presents participants with four words per trial, with the instruction to find the word with the last sound that differed (e.g., “fan cat hat mat” – target answer: “fan”), the first sound that differed (e.g., “rot rod rock box” – target answer: “box”), and the middle sound that differed (e.g., “mop hop tap lop” – target answer: “tap”). Two practice trials followed by 10 test trials were administered in each category (last sound, first sound, middle sound), yielding a total of 36 trials (six practice and 30 test).

The Auditory Analysis Test presents participants with two practice items followed by 40 test items. Each item consists of a compound word (e.g., “cowboy”) with instructions to say the word without one of the compounds (e.g., “say *cowboy* without *boy*”). Target answer: “cow”).

Pitch perception and production test

The Pitch Perception and Production Test was created in our lab and reveals mismatched pitch perception and production abilities in the tone-deaf adults (Loui, et al., 2008): while non-tone-deaf controls demonstrate similar levels of performance between pitch perception and production in this test, tone-deaf individuals show varied levels of performance, suggesting that processing streams for perception and production might be dissociated in the auditory system. The test presents participants with pairs of pure tones forming intervals. Within each pair, the first tone has a fundamental frequency of 500 Hz and the second tone ranges in frequency from 267 Hz to 750 Hz. Each tone was 500 ms in duration, with a gap of 500 ms between the two tones. Thirteen tone pairs were presented. After each presentation, the tasks were first to reproduce the tone pair by humming (pitch production test), and then to state whether the second tone was higher or lower than the first (pitch perception test). The order of the two tasks was counterbalanced across participants.

Non-verbal intelligence test

Non-verbal intelligence was assessed using the Kaufman brief intelligence test (KBIT; Kaufman and Kaufman, 1990). This consisted of a series of up to 40 test items. Each test item consists of a puzzle matrix with a single missing piece. Children were asked to choose the item that fit best within the matrix given a choice of several puzzle pieces. The KBIT is a well-established test that has been shown to be correlated with non-verbal intelligence in children of the age of four and above (Kaufman and Kaufman, 1990). This test was administered so that we could control for possible confounds of non-verbal intelligence.

DATA ANALYSIS

Age (years and months) at the time of testing was calculated from date of birth as reported by the parents. Socio-economic status of each child was estimated based on the median household income reported in US census data (<http://2010.census.gov/2010census/data/>) for each child's zip code.

On average, children scored 22 correct out of 40 (74%) on the Sound Categorization test ($SD = 22\%$) and 21 correct out of 40 (54%) on the Auditory Analysis test ($SD = 27\%$). Since there was no *a priori* reason to expect that performance on Sound Categorization and Auditory Analysis tests might show different relationships with pitch perception and production, each child's raw scores (not age-normalized) for Sound Categorization and Auditory Analysis tests were summed to form a single Phonemic Awareness score for statistical comparisons against pitch-awareness measures.

Previous studies have found that the agreement between pitch perception and production is low among adults who are tone-deaf (Loui, et al., 2008). To obtain an unbiased measure of internal consistency of children's perception and production skills, we calculated a single Pitch Perception–Production index for each child using the following procedure. For each trial, the direction of each participant's hummed pitch intervals (i.e., the production task) was first coded as +1 (for rising intervals) or –1 (for falling intervals). Then the verbal response in the perception task was coded for

each trial with $a + 1$ (if they said that the second tone was higher than the first) or $a - 1$ (if they said that the second tone was lower than the first). The raw scores for the production and perception tasks were then correlated to form the Perception–Production index for each participant. A correlation score of 1.0 indicates perfect congruence between the perception and production of interval directions (i.e., the child always produced an ascending interval in the trials where s/he reported perceiving an ascending interval), a score of 0 indicated no correlation between perception and production (i.e., on trials where the child produced an ascending interval, s/he was equally likely to report hearing an ascending or a descending interval), and a correlation of -1.0 indicates opposite performance between perception and production. This correlation was then Fisher-transformed to obtain a z-score for each subject. The Pitch Perception–Production Index indicated whether the participants hummed back the same pitch directions that they perceived from the recordings. This score does not reflect whether each of the perception/production responses was correct, but only whether there was internal consistency between each child's perception and production.

Bivariate pairwise correlations were then performed to evaluate relationships between each pair of variables (Phonemic Awareness score, Pitch Perception–Production index, KBIT score, and Age). In addition, in order to eliminate the possible influences of confounding variables of IQ and age, partial correlations were run between Phonemic Awareness scores and the Pitch Perception–Production index, while controlling for KBIT, age, SES, and musical training.

RESULTS

Pairwise correlations revealed five significant positive correlations: Pitch Perception–Production was correlated with Phonemic Awareness ($r = 0.795, p < 0.001$), KBIT was correlated with Phonemic Awareness ($r = 0.593, p = 0.001$), Pitch Perception–Production was correlated with KBIT ($r = 0.708, p < 0.001$), Pitch Perception–Production was correlated with Age ($r = 0.576, p = 0.003$), and KBIT was correlated with Age ($r = 0.695, p < 0.001$), as shown in Table 1.

Having established multiple significant bivariate relationships between each set of variables, the next step was to determine whether the relationship between Phonemic Awareness and Pitch Perception–Production existed even after controlling for general intelligence (as measured by the KBIT), age, SES, and musical training. A partial correlation was run examining the relationship of the Perception–Production Index and Phonemic Awareness, with age,

KBIT, SES, and number of years of musical training controlled. Partial correlations revealed significant correlations between Pitch Perception–Production and Phonemic Awareness even after partialing out the effects of intelligence, age, SES, and musical training ($r_{\text{partial}} = 0.586, p = 0.008$). The relationship between Phonemic Awareness and Pitch Perception–Production is shown in Figure 1.

Considering the two separate components of the Pitch Perception–Production index, pitch perception-only and production-only scores were not significantly correlated with the phonemic awareness score (correlation between phonemic awareness and perception scores: $r = 0.29, p = 0.17$; between phonemic awareness and production scores: $r = 0.062, p = 0.80$). Perception-only and production-only scores were also not significantly correlated with the Sound Categorization and Auditory Analysis tests that made up the Phonemic Awareness score (correlation between Sound Categorization and pitch perception: $r = 0.15, p = 0.47$; between Sound Categorization and pitch production: $r = 0.11, p = 0.60$; between Auditory Analysis and pitch perception: $r = 0.19, p = 0.40$; between Auditory Analysis and pitch production: $r = -0.077, p = 0.72$). However, the Pitch Perception–Production index was significantly correlated with each of the two Phonemic Awareness tests (correlation between Perception–Production index and Sound Categorization: $r = 0.671, p < 0.001$; between Perception–Production

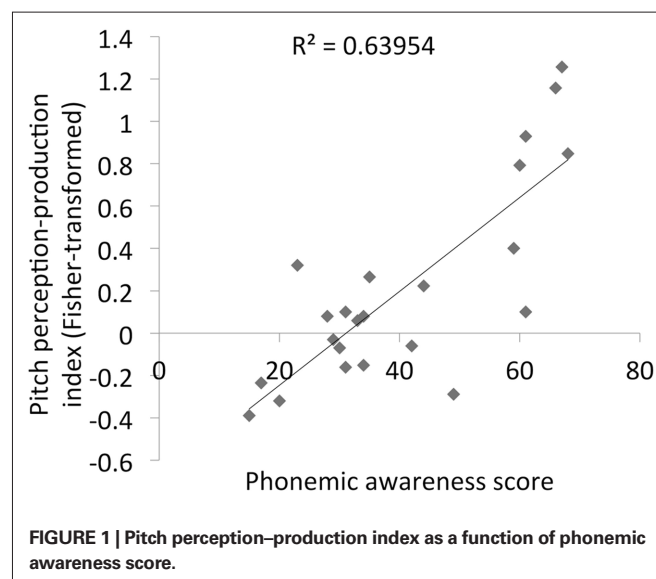


Table 1 | Pairwise correlation coefficients between variables of phonemic awareness scores, pitch perception–production scores, Kaufman brief intelligence test scores, age, musical training, and SES.

	Pitch perception–production	KBIT	Age	Musical training	SES
Phonemic awareness	0.800**	0.593**	0.347	0.582**	0.741**
Pitch perception–production		0.735**	0.576**	0.802**	0.489**
KBIT			0.695**	0.901**	0.742**
Age				0.755**	0.641**
Musical training					0.805**

** $p < 0.01$.

index and Auditory Analysis test: $r = 0.791$, $p < 0.001$), suggesting that it is the combination of pitch perception and production skills, rather than each of the two skills considered separately, that accounts for individual differences in phonemic awareness.

DISCUSSION

This study demonstrates that phonemic awareness is correlated with pitch awareness, as assessed by the degree of agreement between pitch perception and pitch production. These results are independent of age, non-verbal IQ, SES, and musical training. While prior studies have shown that phonemic awareness, a crucial predictor of reading ability, is related to musical ability (Anvari, et al., 2002; Forgeard et al., 2008b; Jones, et al., 2009a), our study is the first to demonstrate an association between phonemic awareness and level of agreement between pitch perception and pitch production. Given that individuals with dyslexia are impaired in phonemic awareness (Stahl and Murray, 1994) and tone-deaf individuals show mismatches between pitch perception and production (Loui, et al., 2008), the current results suggest an intriguing association between dyslexia and tone-deafness and warrant further investigation of a possible shared impairment across these disorders.

The present correlational findings now motivate research into whether phonemic processing and pitch processing are mediated by a shared neural substrate. It has been shown that tone-deafness in adults is characterized by a mismatch between pitch perception and production (Loui, et al., 2008; Liu, et al., 2010) as well as to deficits in phonemic awareness (Jones et al., 2009b). It has also been established that phonemic awareness is deficient in dyslexia (Snow, et al., 1998; Stanovich, 1988; Wood et al., 2005). Taken together, the present results suggest that tone-deafness and dyslexia may be intimately related. Since the task of perceiving and producing pitch intervals requires awareness of one's own voice, the present results suggest that a deficit in awareness of one's own voice may underlie both tone-deafness and dyslexia. Although deficits in pitch awareness are certainly not the sole deficit in dyslexia, a pitch-awareness deficit could be one of many risk factors for language impairment (Leppanen et al., 2002).

The relationship between pitch perception and pitch production has been investigated in previous studies involving auditory feedback of one's own voice in adults (Burnett and Larson, 2002; Pfordresher, 2005). While auditory feedback is a useful way to investigate how one responds to one's own voice, another measure of vocal awareness, especially in the domain of pitch, can be best assessed by recording subjects' perception and production of pitched intervals (pairs of pitches) and comparing each individual's perceptual response and production output for internal consistency. Our previous work has shown that tone-deaf adults lack conscious awareness of their own vocal productions of pitched intervals (Loui, et al., 2008). Adults identified as tone-deaf by the Montreal Battery for Evaluation of Amusia (Peretz et al., 2003) as well as by their responses to psychophysically defined frequency discrimination thresholds (Foxton et al., 2004) demonstrate a paradoxical mismatch between pitch interval perception and production, where they could produce small pitched intervals in the correct direction despite being unable to perceive the direction of the pitched interval according to their own verbal report. Although tone-deafness (or congenital amusia) has been described in adults for decades, the incidence of

tone-deafness in children is unknown, with the first case report of amusia having only recently been reported in one child (Lebrun et al., in press). Patterns of behavior from some of the children in the present study, i.e., children who showed negative correlations between perception and production performance, are similar to the previously reported perception–production mismatch in tone-deaf adults. In that regard, the pitch perception–production tests in the current study might be useful in the future as a sensitive screening test for early identification of tone-deafness in childhood. While it is unclear whether this mismatch behavior is a sign of a developing auditory-motor system, or whether these children are currently tone-deaf, or will develop into tone-deaf adults, the present results show that they exhibit behavioral characteristics of tone-deaf adults and thus may put them into an at-risk category for tone-deafness.

Neuroimaging work has shown that the behavioral pattern of mismatch between pitch interval perception and production is associated with atypical white matter connectivity in the arcuate fasciculus, which connects the superior and middle temporal gyri in the temporal lobe with the posterior inferior frontal gyrus in the frontal lobe (Loui, et al., 2009). Notably, individuals who showed more congruent perception and production abilities had larger inferior branches of the right arcuate fasciculus. Given that arcuate fasciculus volume is correlated between the left and right hemispheres (Wahl et al., 2010) but also shows significant hemispheric asymmetry (Catani et al., 2005; Vernooij et al., 2007), the present results may be a result of structural correspondence between left and right arcuate fasciculi during development, such that children with a well-developed right inferior arcuate fasciculi will also have highly internally consistent pitch perception and production abilities and possess well-developed left inferior arcuate fasciculus, which may be involved in phonemic awareness. Alternatively, the two domains of sound awareness (pitch and phonemes) may involve the same neural structures, providing further support for overlapping structures and/or cognitive resources used in language and music (Besson and Schon, 2001; Slevc et al., 2009; Sammler et al., 2010). In future studies that involve larger groups of children with and without the diagnosis of dyslexia, research may be able to tease apart the behavioral patterns and neural substrates of pitch awareness and phonemic awareness. Determining the extent to which pitch and phoneme processing are connected may help us explore commonalities between tone-deafness and dyslexia, in the hope of designing more refined rehabilitation strategies for dyslexia and perhaps even tone-deafness in the future.

CONCLUSION

The present results demonstrate an association between phonemic awareness and pitch awareness. Since phonemic awareness is a crucial predictor of reading ability and is disrupted in dyslexia, whereas the agreement between pitch perception and production is characteristic of tone-deaf behavior, the present results suggest that dyslexia and tone-deafness are related and may share a common basis.

ACKNOWLEDGMENTS

This study was supported by research grants from NIH (R01 DC009823) and Templeton Foundation. We are grateful to all our participants and to Maureen Kroogfor recruiting participants at Grant School # 7 in Passaic, New Jersey.

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.
- Benasich, A. A., Choudhury, N., Friedman, J. T., Realpe-Bonilla, T., Chojnowska, C., and Gou, Z. (2006). The infant as a prelinguistic model for language learning impairments: predicting from event-related potentials to behavior. *Neuropsychologia* 44, 396–411.
- Besson, M., and Schon, D. (2001). Comparison between language and music. *Ann. N. Y. Acad. Sci.* 930, 232–258.
- Bolduc, J., and Montésinos-Gelet, I. (2005). Pitch processing and phonological awareness. *Psychomusicology* 19, 3–14.
- Bradley, L., and Bryant, P. (1985). *Rhyme and Reason in Reading and Spelling*. Ann Arbor: The University of Michigan Press.
- Bradley, P., and Bryant, L. (1983). Categorizing sounds and learning to read: a causal connection. *Nature* 301, 419–421.
- Burnett, T. A., and Larson, C. R. (2002). Early pitch-shift response is active in both steady and dynamic voice pitch control. *J. Acoust. Soc. Am.* 112(3 Pt 1), 1058–1063.
- Catani, M., Jones, D. K., and Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16.
- Cuddy, L. L., Balkwill, L. L., Peretz, I., and Holden, R. R. (2005). Musical difficulties are rare: a study of “tone deafness” among university students. *Ann. N. Y. Acad. Sci.* 1060, 311–324.
- Démonet, J.-F., Taylor, M. J., and Chaux, Y. (2004). Developmental dyslexia. *Lancet* 363, 1451–1460.
- Forgeard, M., Schlaug, G., Norton, A., Rosam, C., Iyengar, U., and Winner, E. (2008a). The relation between music and phonological processing in normal-reading children and children with dyslexia. *Music Percept.* 25, 383–390.
- Forgeard, M., Winner, E., Norton, A., and Schlaug, G. (2008b). 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
- Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., and Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying “tone deafness”. *Brain* 127(Pt 4), 801–810.
- Gordon, E. E. (1979). *Primary measures of musical audition*. Chicago, IL: G.I.A. Publications.
- 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.
- Jentschke, S., Koelsch, S., and Friederici, A. D. (2005). Investigating the relationship of music and language in children: influences of musical training and language impairment. *Ann. N. Y. Acad. Sci.* 1060, 231–242.
- Jones, J. L., Lucker, J., Zalewski, C., Brewer, C., and Drayna, D. (2009a). Phonological processing in adults with deficits in musical pitch recognition. *J. Commun. Disord.* 42, 226–234.
- Jones, J. L., Zalewski, C., Brewer, C., Lucker, J., and Drayna, D. (2009b). Widespread auditory deficits in tone deafness. *Ear Hear.* 30, 63–72.
- Kalmus, H., and Fry, D. B. (1980). On tone deafness (dysmelodia): frequency, development, genetics and musical background. *Ann. Hum. Genet.* 43, 369–382.
- Kaufman, A. S., and Kaufman, N. L. (1990). *Kaufman Brief Intelligence Test*. Circle Pines, MN: AGS Publishing.
- Lamb, S. J., and Gregory, A. H. (1993). The relationship between music and reading in beginning readers. *Educ. Psychol.* 13, 19–27.
- Lebrun, M. A., Moreau, P., McNally-Gagnon, A., Mignault Goulet, G., and Peretz, I. (in press). Congenital amusia in childhood: a case study. *Cortex*.
- Leppanen, P. H., Richardson, U., Pihko, E., Eklund, K. M., Guttorm, T. K., Aro, M., and Lyytinen, H. (2002). Brain responses to changes in speech sound durations differ between infants with and without familial risk for dyslexia. *Dev. Neuropsychol.* 22, 407–422.
- Liu, F., Patel, A. D., Fourcin, A., and Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain* 133, 1682–1693.
- Loui, P., Alsop, D., and Schlaug, G. (2009). Tone-deafness: a disconnection syndrome? *J. Neurosci.* 29, 10215–10220.
- Loui, P., Guenther, F. H., Mathys, C., and Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Curr. Biol.* 18, R331–R332.
- Mengler, E. D., Hogben, J. H., Michie, P., and Bishop, D. V. (2005). Poor frequency discrimination is related to oral language disorder in children: a psychoacoustic study. *Dyslexia* 11, 155–173.
- Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., and Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science* 271, 77–81.
- 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.
- Nan, Y., Sun, Y., and Peretz, I. (2010). Congenital amusia in speakers of a tone language: association with lexical tone agnosia. *Brain* 133, 2635–2642.
- Overy, K. (2003). Dyslexia and music. From timing deficits to musical intervention. *Ann. N. Y. Acad. Sci.* 999, 497–505.
- Peretz, I., Brattico, E., Jarvenpää, M., and Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain* 132, 1277–1286.
- Peretz, I., Champod, A. S., and Hyde, K. (2003). Varieties of musical disorders. The montreal battery of evaluation of amusia. *Ann. N. Y. Acad. Sci.* 999, 58–75.
- Peynircioglu, Z., Durgunoglu, A., and Oney-Kusefoglu, B. (2002). Phonological awareness and musical aptitude. *J. Res. Read.* 25, 68–80.
- Pfordresher, P. Q. (2005). Auditory feedback in music performance: the role of melodic structure and musical skill. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 1331–1345.
- Rosner, J., and Simon, D. P. (1971). The auditory analysis test: an initial report. *J. Learn. Disabil.* 4, 384.
- Sammler, D., Baird, A., Valabregue, R., Clement, S., Dupont, S., Belin, P., and Samson, S. (2010). The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. *J. Neurosci.* 30, 3572–3578.
- Shaywitz, S. E., and Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biol. Psychiatry* 57, 1301–1309.
- 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.
- Snow, C. E., Burns, M. S., and Griffin, P. (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.
- Stadler, S. (1990). Vocal pitch matching ability in children between four and nine years of age. *Eur. J. High Ability* 1, 33–41.
- Stahl, S. A., and Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *J. Educ. Psychol.* 86, 221–234.
- Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: the phonological-core variable-difference model. *J. Learn. Disabil.* 21, 590–604.
- Tallal, P. (2004). Improving language and literacy is a matter of time. *Nat. Rev. Neurosci.* 5, 721–728.
- Tallal, P., and Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends Neurosci.* 29, 382–390.
- Tallal, P., and Piercy, M. (1973). Developmental aphasia: impaired rate of non-verbal processing as a function of sensory modality. *Neuropsychologia* 11, 389–398.
- Vernooij, M. W., Smits, M., Wielopolski, P. A., Houston, G. C., Krestin, G. P., and van der Lugt, A. (2007). Fiber density asymmetry of the arcuate fasciculus in relation to functional hemispheric language lateralization in both right- and left-handed healthy subjects: a combined fMRI and DTI study. *Neuroimage* 35, 1064–1076.
- Wahl, M., Li, Y.-O., Ng, J., LaHue, S. C., Cooper, S. R., Sherr, E. H., and Mukherjee, P. (2010). Microstructural correlations of white matter tracts in the human brain. *Neuroimage* 51, 531–541.
- Wood, F. B., Hill, D. E., Meyer, M. S., and Flowers, D. L. (2005). Predictive assessment of reading. *Ann. Dyslexia* 55, 193–216.

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: 16 February 2011; accepted: 13 May 2011; published online: 30 May 2011.
 Citation: 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. Psychology* 2:111. doi: 10.3389/fpsyg.2011.00111
 This article was submitted to *Frontiers in Auditory Cognitive Neuroscience*, a specialty of *Frontiers in Psychology*. Copyright © 2011 Loui, Kroog, Zuk, Winner and Schlaug. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



Preserved statistical learning of tonal and linguistic material in congenital amusia

Diana Omigie and Lauren Stewart*

Department of Psychology, Goldsmiths, University of London, London, UK

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Barbara Tillmann, Centre National de La Recherche Scientifique, France
Isabelle Peretz, Université de Montréal, Canada

***Correspondence:**

Lauren Stewart, Department of Psychology, Goldsmiths, University of London, New Cross, London SE14 6NW, UK.
e-mail: l.stewart@gold.ac.uk

Congenital amusia is a lifelong disorder whereby individuals have pervasive difficulties in perceiving and producing music. In contrast, typical individuals display a sophisticated understanding of musical structure, even in the absence of musical training. Previous research has shown that they acquire this knowledge implicitly, through exposure to music's statistical regularities. The present study tested the hypothesis that congenital amusia may result from a failure to internalize statistical regularities – specifically, lower-order transitional probabilities. To explore the specificity of any potential deficits to the musical domain, learning was examined with both tonal and linguistic material. Participants were exposed to structured tonal and linguistic sequences and, in a subsequent test phase, were required to identify items which had been heard in the exposure phase, as distinct from foils comprising elements that had been present during exposure, but presented in a different temporal order. Amusic and control individuals showed comparable learning, for both tonal and linguistic material, even when the tonal stream included pitch intervals around one semitone. However analysis of binary confidence ratings revealed that amusic individuals have less confidence in their abilities and that their performance in learning tasks may not be contingent on explicit knowledge formation or level of awareness to the degree shown in typical individuals. The current findings suggest that the difficulties amusic individuals have with real-world music cannot be accounted for by an inability to internalize lower-order statistical regularities but may arise from other factors.

Keywords: congenital amusia, statistical learning, implicit learning

INTRODUCTION

Research into developmental disorders seeks to explain why basic abilities that are acquired effortlessly by most humans prove difficult for others. The capacity to perceive music is one such example. While the ability to make sense of music is recognized as a fundamental human trait (Blacking, 1995), an estimated 4% of the population have pervasive and lifelong difficulties in this domain (Kalmus and Fry, 1980). Such individuals cannot discriminate between different melodies, recognize music that should be familiar to them without the aid of lyrics, detect anomalous pitches in familiar and unfamiliar melodies or judge dissonance in musical excerpts (Ayotte et al., 2002; Peretz et al., 2003). The presence of this disorder, termed “congenital amusia,” is typically ascertained using a diagnostic tool known as the Montreal battery for the evaluation of amusia (MBEA: Peretz et al., 2003) and cannot be explained by peripheral hearing problems or general cognitive deficits (Ayotte et al., 2002; Peretz et al., 2002).

The lack of facility these individuals have with music urges us to consider how the more typical human capacity to perceive music comes about. Music is composed of a small set of elements which combine to form complex hierarchical structures (Lerdahl and Jackendoff, 1983; Bod, 2002). Despite this level of complexity, it appears that almost all humans possess knowledge of musical structure (Bigand and Poulin-Charronnat, 2006). In individuals without musical training, such knowledge is implicit and cannot be verbalized, but studies which probe listeners' musical expectations find that non-musicians are just as sensitive as formally

trained musicians to many aspects of musical structure (Cuddy and Badertscher, 1987; Hebert, Peretz and Gagnon, 1995; Schellenberg, 1996; Honing and Ladinig, 2009; Marmel and Tillmann, 2009). Importantly, listeners acquire this knowledge in an incidental manner and without any awareness of their doing so, in a process commonly referred to as “implicit learning” (Reber, 1992).

A growing body of work suggests that the implicit learning of the musical knowledge that is possessed by most listeners is acquired via the internalization of statistical regularities (Smith et al., 1994; Tillmann et al., 2000; Tillmann and McAdams, 2004; Jonaitis and Saffran, 2009), which confers sensitivity to several aspects of musical structure that can be demonstrated across a range of musical tasks. These include making subjective ratings on goodness of fit, melodic expectation, and goodness of completion (Krumhansl and Keil, 1982; Cuddy and Badertscher, 1987; Schmuckler, 1989; Brown et al., 1994; Toiviainen and Krumhansl, 2003) as well as demonstrating sensitivity to musical tensions and relaxations in sequences of chords (Bigand et al., 1996; Bigand and Parncutt, 1999).

One paradigm, originating in the language acquisition literature, has been particularly influential in demonstrating the ability of listeners to compute the statistical properties of their auditory environment. Saffran et al. (1996) demonstrated that adult listeners exposed to a nonsense speech language comprised of tri-syllabic units (which, from now on, will be referred to as “words,” following Saffran et al., 1996) were able to discover boundaries between these words by computing the transitional probabilities between adjacent syllables. The authors showed that even though the speech stream

was continuous, with no temporal cues between adjacent words, listeners in a later test phase were able to successfully discriminate between words in the language they had been exposed to versus foils containing the same syllables, which were arranged in a different order (so called “non-words”). In a separate experiment, the authors demonstrated that listeners were also able to discriminate between words and foils in which either the first or third syllable in a word from the language had been substituted with a different syllable (so called “part-words”). Importantly for the present research, in an analogous study using tonal sequences, the authors reported that the learning mechanism by which listeners carried out this sequence segmentation was not confined to linguistic materials. Saffran et al. (1999) presented participants with a continuous tone stream comprised of tone-triplet units (which the authors termed “tone words”) made up of musical notes from the octave above middle C. They showed that after 21 min of exposure, listeners were able to distinguish the tone words they had been exposed to from both non-word and part-word foils. While this paradigm focuses on transitional probabilities between adjacent tone elements, other paradigms have also examined listeners’ sensitivity to transitional probabilities within sequences of harmonic elements (Jonaitis and Saffran, 2009), pitch intervals (Saffran and Griepentrog, 2001), and timbral elements (Tillmann and McAdams, 2004), the statistical learning of non-adjacent dependencies in tonal stimuli (Creel et al., 2004; Kuhn and Dienes, 2005; Gebhart et al., 2009), and the facilitative effect of musical information on language learning (Schön et al., 2008). Taken together, results from these studies show that listeners require only a limited amount of exposure to internalize the statistical properties of a completely novel musical system.

Evidence that musical competencies arise largely from implicit learning of regularities in our musical environment suggests at least two testable hypotheses concerning the nature of musical deficits in congenital amusia. One hypothesis may be that such individuals lack the learning mechanism that permits internalization of regularities from a structured sound stream. The disproportionate difficulties seen with music, as opposed to language, would predict that a faulty learning mechanism would be restricted to tonal, rather than linguistic material. A second hypothesis may be that the learning mechanism is intact, but a difficulty in detection and/or discrimination of small pitch changes is the limiting factor in building up knowledge of musical structure.

The present study addressed these hypotheses, by testing a group of amusic and control participants on their ability to internalize the regularities present in structured linguistic and tonal materials, given equal amounts of exposure. Following the paradigm used by Saffran et al. (1996, 1999) our participants were exposed to streams made up of words comprised of either syllables or tones. Critically, only the statistical properties within the stream served as a reliable cue as to the location of word boundaries. In a subsequent test phase, participants were then required to demonstrate their knowledge of these word boundaries, by distinguishing between words they had heard in the exposure phase and non-words, which were comprised of identical syllables or tones that were arranged in a different temporal order. Two types of tonal material were used. In the first, intervals within the tone sequence exceeded psychophysically measured thresholds across the amusic group (“supra-threshold” condition) while in the second (“sub-threshold” condition) intervals within the

tone sequence were smaller, including a semitone. According to the literature, many amusics have difficulty with the detection and/or discrimination of pitch direction around this level (Foxton et al., 2004; Hyde and Peretz, 2004; Liu et al., 2010).

If general learning mechanisms are compromised, we would predict inferior learning across all conditions in the amusic group. If learning mechanisms are compromised for tonal material only, we would predict inferior learning for both tonal conditions in the amusic group but equivalent learning across both groups with the linguistic material. Finally, if learning mechanisms are intact, for both linguistic and tonal material, but the learning of amusics is limited by a poor sensitivity to pitch change, we would predict equivalent learning in both groups for the linguistic material and for the supra-threshold tonal condition, but inferior learning for the sub-threshold tonal condition.

In addition to recording accuracy rates for the above tasks, we also collected binary confidence judgments on a trial-to-trial basis for the sub-threshold tonal condition. Recent studies have suggested that amusia may be a disorder of awareness, rather than perception, i.e., such individuals can represent pitch changes adequately, but these do not reach conscious awareness, resulting in poor performance on tests which probe musical perception explicitly (Peretz et al., 2009; Hyde et al., 2011). Such a hypothesis would predict that even if amusics and controls show comparable learning, as indicated by equivalent accuracy in identifying words they had been previously exposed to, individuals with amusia may show a bias toward reporting low confidence compared to control individuals.

MATERIALS AND METHODS

PARTICIPANTS

A total of 24 participants (12 amusic, 12 control) took part in this study. All participants were recruited via an online assessment based on the scale and rhythm subtest of the MBEA (Peretz et al., 2003)¹. Each participant took the online test twice and if they consistently achieved a score of 22/30 or less, they were invited to come in to the lab where assessment could take place under controlled conditions. Each participant was administered four MBEA subtests (scale, contour, interval, and rhythm sub-tests) in a sound-attenuated booth in order to confirm the presence or absence of amusia. Previous research had shown that amusia is characterized by poor perception in the pitch-based sub-tests of the MBEA (scale, contour, interval) while only half of them typically show a deficit in the rhythm test (Peretz et al., 2003). Thus we calculated a composite score for the three pitch-based sub-tests, using 65 out of 90 as a cut off score, whereby individuals were classified as amusic if their composite score fell below this value (Peretz et al., 2003; Liu et al., 2010). The amusic and control sample were matched on age, gender, score on the national adult reading test (NART: Nelson and Willison, 1991), Digit span (Wechsler adult intelligence scale, WAIS: Wechsler, 1997), number of years of formal education and number of years of musical education. In addition, two pitch threshold tasks were conducted. A pitch change detection task and a pitch direction discrimination task, both employing a two-alternative forced-choice AXB adaptive tracking procedure with pure tones, were used to assess thresholds for the detection of a simple pitch change and the discrimination of pitch direction

¹www.delosis.com/listening/home.html

respectively (see Liu et al., 2010 for further details). **Table 1** provides background information on the two groups, while **Table 2** provides mean scores on the MBEA sub-tests and pitch thresholds. In addition to performing significantly worse on four sub-tests of the MBEA, the current cohort of amusic individuals differed significantly in their thresholds for the discrimination of pitch direction (Controls: $M = 0.18$, $SD = 0.08$, Range = 0.09–0.33; Amusics: $M = 1.05$, $SD = 1.07$, Range = 0.10–2.97). However the two groups did not differ significantly in thresholds for the detection of a pitch change with only one amusic individual having a threshold above one semitone (Controls: $M = 0.15$, $SD = 0.06$, Range = 0.08–0.26; Amusics: $M = 0.27$, $SD = 0.33$, Range = 0.07–1.29).

STIMULI

Stimuli for the three conditions (linguistic, supra-threshold tonal, and sub-threshold tonal conditions) were based on those used by Saffran et al. (1996, 1999). The linguistic sequences were created

from 11 syllables obtained by pairing the consonants p, t, b, and d with the vowels a, i, and u. Syllabic sounds were excised from the recorded speech of a native English speaker who was required to read aloud a string of words in which the required syllables were inserted. All pitch information was subsequently removed from the syllable sounds using Praat software (Boersma, 2001). Subsequently, the syllable sounds were stretched or compressed (as necessary) to a fixed duration of 280 ms using Audacity software².

Figure 1 shows the frequencies used in both tonal conditions. Following Saffran et al. (1999), the sub-threshold sequences were constructed from eleven tones drawn from the chromatic scale beginning at C4 (261.3 Hz). As in Saffran et al. (1999), all the tones from C4 to B4 were used, excluding A#. The supra-threshold sequences were constructed from a novel scale with unfamiliar interval sizes, obtained by dividing the two-octave span from C4 (261.3 Hz) into 11 evenly log-spaced divisions. Thus, the 11 tones in the sub-threshold condition were generated using the formula: Frequency (Hz) = $261.63 \times 2^{n/12}$, with n referring to the number of steps along the chromatic scale (0–9, 11) while the 11 tones in the supra-threshold condition followed the formula: Frequency (Hz) = $261.63 \times 4^{n/11}$, where n is the number of equal sized steps along the new scale (0–10). Consequently, the tones used in the sub-threshold tonal condition were 261.63, 277.18, 293.66, 311.13, 329.63, 349.23, 369.99, 392.00, 415.30, 440, and 493.88 Hz while those used in the supra-threshold tonal condition were 261.63, 296.77, 336.63, 381.84, 433.13, 491.31, 557.29, 632.14, 717.05, 813.36, and 922.60 Hz. All tones were sine tones generated in Matlab³ with a duration of 330 ms and an envelope rise and fall time of 10 ms on either side.

Language construction

For all conditions (linguistic, supra-threshold tonal, sub-threshold tonal), two languages analogous in statistical structure were prepared to ensure that any potential learning could not be accounted for by idiosyncratic aspects of one language in particular. Both languages were comprised of the same elements

Table 1 | Amusic and control participant characteristics; summary of the two groups in terms of their mean age, gender, years of musical training, and education, NART and total digit span (forward and backward).

Group	Age	Gender	Years of musical training	Years of education	NART	Digit span
AMUSIC						
μ	52.83	5M	0.58	15.92	42.25	22.58
σ	9.65	7F	1.24	1.93	5.69	3.48
CONTROL						
μ	51.08	4M	1.10	16.08	44.55	21.17
σ	8.90	8F	1.82	2.71	3.31	3.27
t-TESTS						
t	0.46		−0.82	−0.17	−1.21	1.02
p	0.65		0.42	0.86	0.24	0.35

F, female; M, male; R, right; L, left; μ , Mean; σ , SD; t , test statistic of the independent samples t-test; p , probability value.

²<http://audacity.sourceforge.net/>

³<http://www.mathworks.com/products/matlab>

Table 2 | Mean scores of the amusic and control groups on sub-tests of the Montreal Battery for the evaluation of amusia (MBEA). A pitch composite score below the cut off of 65 out of 90 was considered to be diagnostic of amusia.

Group	MBEA scale	MBEA contour	MBEA interval	MBEA rhythm	Pitch composite	Detection threshold	Direction threshold
AMUSIC							
μ	19.75	19.58	18.25	24.17	57.58	0.27	1.05
σ	2.26	2.61	2.01	3.13	5.70	0.33	1.07
CONTROL							
μ	27.33	27.42	27.33	28.5	82.08	0.15	0.18
σ	2.35	2.27	2.84	1.31	6.17	0.06	0.08
t-TESTS							
t	−8.06	−7.84	−9.05	−4.42	−10.11	1.24	2.79
p	<0.001	<0.001	<0.001	<0.001	<0.001	0.240	0.020

μ , Mean; σ , SD; t , test statistic of the independent samples t-test; p , probability value. The pitch composite score is the mean score based on the scale, contour, and interval sub-tests of the MBEA.

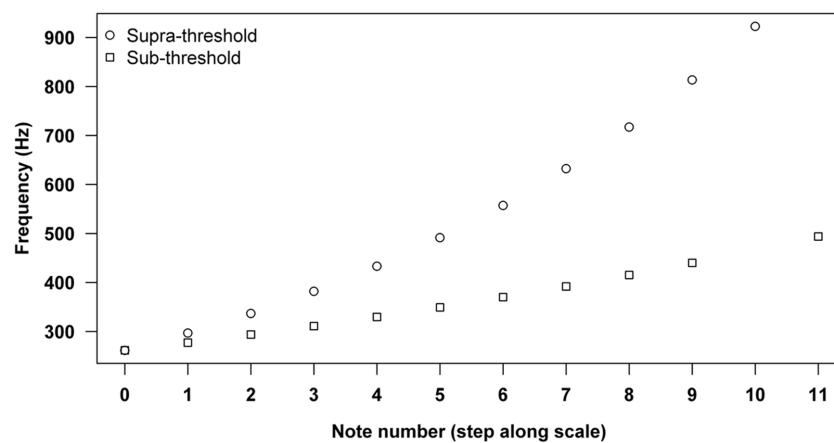


FIGURE 1 | Frequency of tones used in the sub-threshold and supra-threshold tonal conditions. Note numbers 0–9 and 11 in the sub-threshold condition correspond to C4 to A4 and B4 while note numbers 0–10 in the supra-threshold condition correspond to tones from a novel scale obtained by dividing the two-octave span from C4 (261.3 Hz) into 11 evenly log-spaced divisions.

that had been arranged to make different words, and differed only in the transitional probabilities between elements of the words (see Saffran et al., 1999 for further details). For half the participants of each group, language 1 was used in the listening phase, and words from language 2 were used as the non-word foils during the test phase, while the opposite was the case for the remaining participants.

Each language comprised six words. In language 1 of the linguistic condition, the six words used were babupu, bupada, dutaba, patubi, pidabu, and tutibu while in language 2, they were batida, bitada, dutupi, tipuba, tipabu, and tapuba. In the sub-threshold tonal condition, language 1 comprised of six tone words taken from the chromatic scale beginning at C4; ADB, DFE, GG#A, FCF#, D#ED, and CC#D whilst language 2 comprised of a different set of six tone words from the chromatic scale beginning at C4; AC#E, F#G#E, GCD#, C#BA, C#FD, G#BA. To create tone words that were analogous in structure across the two tonal conditions, words in the supra-threshold condition were created by substituting frequencies in the sub-threshold words with frequencies from the novel scale that corresponded in terms of the number of steps from C4. Tone words in the two conditions were identical in pattern and differed only in the actual frequencies, and consequently the size of interval occurring between adjacent tones (Figure 2).

Sequence concatenation

To create each sequence, the six words from the given language were concatenated in random order to create six different blocks containing 18 words each. Concatenation adhered to two strict conditions; that a word could not follow itself and that there were no silent gaps between words. The six blocks created in this way were then further concatenated to create sequences consisting of 432 words (72 tokens of each word). As the sequences in the tonal conditions consisted of units with a duration of 330 ms, these lasted approximately 7 min. The sequences in the linguistic condition, consisting of syllable sounds of 280 ms length, were approximately 6 min long.

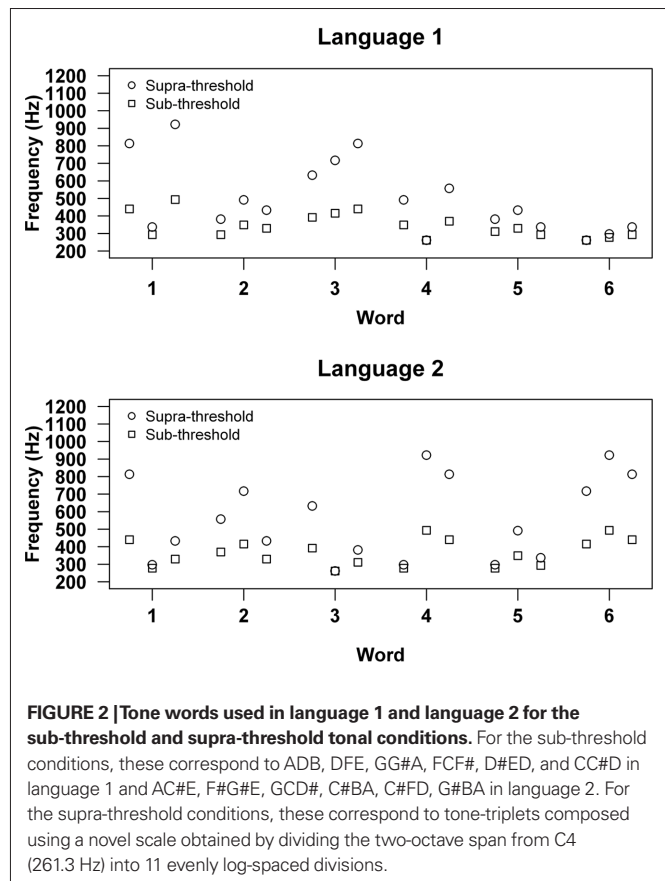


FIGURE 2 | Tone words used in language 1 and language 2 for the sub-threshold and supra-threshold tonal conditions. For the sub-threshold conditions, these correspond to ADB, DFE, GG#A, FCF#, D#ED, and CC#D in language 1 and AC#E, F#G#E, GCD#, C#BA, C#FD, G#BA in language 2. For the supra-threshold conditions, these correspond to tone-triplets composed using a novel scale obtained by dividing the two-octave span from C4 (261.3 Hz) into 11 evenly log-spaced divisions.

PROCEDURE

Participants gave written consent to participate in the experiments, which were approved by the Ethics Committee at Goldsmiths, University of London. All experiments were conducted in a sound-attenuated booth. Sounds for the listening and test phase were

presented through an external sound card (Edirol UA-4FX USB Audio Capture) at a fixed intensity level of 73 dB using Sennheiser headphones HD 202. Programs for stimulus presentation and the collection of data were written in Matlab⁴.

As languages in the sub-threshold and supra-threshold tonal conditions comprised analogous words (but over a different frequency range), it was important to eliminate any potential carryover effects between the conditions. This was achieved by splitting each group in two such that one half of each group was exposed to language 1 of the linguistic condition, language 2 of the supra-threshold tonal condition and language 1 of the sub-threshold tonal condition while the other half of each group was exposed to language 2 of the linguistic condition, language 1 of the supra-threshold tonal condition and language 2 of the sub-threshold condition.

To avoid fatigue effects, testing took place over two sessions. In the first session, participants were run on the supra-threshold tonal condition and on the linguistic condition. The order in which the conditions were presented to participants was counterbalanced for both the amusic and control groups. The linguistic and tonal conditions were separated by a period in which participants carried out a completely unrelated task. The second testing session was carried out on a different day on average 7 months later. In these sessions, participants were run on the sub-threshold tonal condition only.

Exposure lasted approximately 21 min in total for the tonal conditions and 18 min for the linguistic condition. Instructions for all three conditions were identical for the listening phase. Participants were told that they would hear a stream of sounds. They were asked to avoid analyzing the stream but also to refrain from blocking out the sounds as they would be tested on what they had heard afterward. They were then presented with three blocks of one of the sound sequences described previously with the opportunity for a short break between blocks.

Immediately after the exposure phase, the testing phase commenced, starting with three practice trials. Participants were then presented with 36 trials. Each trial comprised two words; one of which they had heard during exposure and another which had the same constituent parts, but which had not appeared in combination during the exposure phase. For all three conditions, the 36 trials were created by exhaustively pairing the six words from both languages such that on each trial participants exposed to opposing languages were expected to select opposing items. Within a trial, words were presented with a 750-ms inter-stimulus interval and there was an inter-trial interval of 5 s during which the participant was required to make their response.

On each trial of the test phase for the conditions run in the first session (the linguistic condition and the supra-threshold tonal condition), the participant's task was to indicate, using the computer keyboard, which word (the first or the second) in the pair they had heard during the exposure phase. In the second session (the sub-threshold tonal condition), participants were additionally required to indicate whether or not they were confident about their decision by responding "confident" or "not confident" immediately after. As this condition required participants to make two responses (compared to one in the previous conditions), responses in this session were entered into the computer by the experimenter so as to avoid inputting error. Two different random orders of the test

trials were generated for each condition and following Saffran et al. (1999) each participant was randomly assigned to one of the two different random orders in each condition.

RESULTS

EVIDENCE OF LEARNING: PERFORMANCE DURING THE TEST PHASE

Figure 3 shows scores for all individuals, by group, across all three conditions. As shown in Table 3, single-sample *t*-tests (all two-tailed) revealed an overall performance that was significantly greater than chance for both groups across all conditions. Independent sample *t*-tests revealed no significant differences between the scores of individuals assigned to alternative orders of the test trials in any of the three conditions (all $p > 0.05$) so data were treated similarly regardless of this factor.

Individual participants' data were entered into a preliminary $2 \times 2 \times 3$ split plot ANOVA with condition (linguistic, supra-threshold tonal, sub-threshold tonal) as a within subject factor and group (amusic, control) and language set (one, two) as between subject factors. The aim of this initial analysis was to observe any effect of the set of languages to which participants were allocated. There were no significant main effects of language set, group, or condition [Language set: $F(1, 20) = 0.03$, $MSe = 5.69$, $p = 0.87$, Group: $F(1, 20) = 0.55$, $MSe = 5.69$, $p = 0.47$; Condition: $F(2, 40) = 2.48$, $MSe = 19.69$, $p = 0.10$], nor were there any significant interactions (all $p > 0.05$).

Given that performance was not differentially affected according to the precise set of languages a participant had been allocated to, scores were collapsed across this factor to increase the power of the main analysis. A 2×3 split plot ANOVA with group (amusic versus control) as a between-subjects factor and condition (linguistic, supra-threshold, sub-threshold) as a within-subjects factor was carried out in order to re-assess the main effects of group and condition. No difference was found between control and amusic subjects: $F(1, 22) = 0.60$, $MSe = 5.19$, $p = 0.45$, or across conditions: $F(2, 44) = 2.39$, $MSe = 20.4$, $p = 0.10$, nor was there a significant interaction between group and condition, suggesting that both groups performed equally well on all conditions: $F(2, 44) = 0.05$, $MSe = 20.4$, $p = 0.95$.

Having employed a within-subjects design in the current study, further analysis investigated the possibility that repeated testing on the same individuals may have resulted in order effects during the first session, where the linguistic condition and the supra-threshold tonal condition conditions were carried within an hour of each other. However, an independent samples *t*-test indicated that participants who carried out the linguistic condition first did not perform any better in the supra-threshold tonal condition ($M = 22.92$) compared with those who carried out the supra-threshold tonal condition first [$M = 23.50$, $t(22) = -0.035$, $p = 0.73$].

Finally, of key interest was whether participants' performance on the two tonal conditions could be accounted for by psychophysically measured pitch detection and pitch discrimination thresholds. Results from correlation analyses with each of the groups treated separately (Table 4), showed no significant relationship between learning and perceptual thresholds.

CONFIDENCE JUDGMENTS

The next stage of analysis evaluated the degree to which participants' confidence in their performance predicted their accuracy in the test phase of the sub-threshold task, as well as the overall

⁴<http://www.mathworks.com/products/matlab>

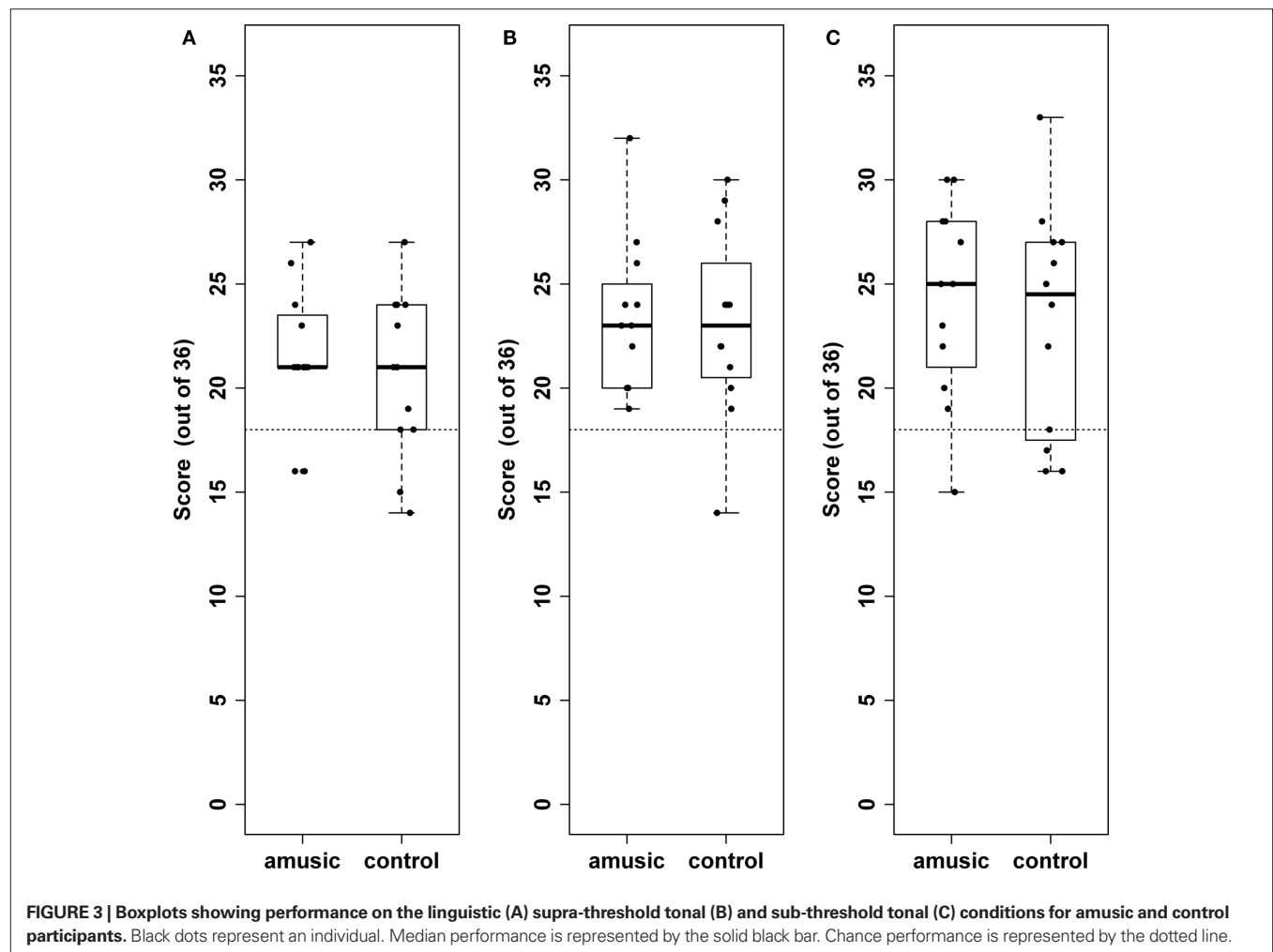


Table 3 | Means, SD, and results for one sample *t*-tests (*df* = 11) against chance performance for amusic and control individuals across all three conditions.

	Linguistic	Supra-threshold tonal	Sub-threshold tonal
AMUSIC			
μ	21.50	23.33	24.33
σ	3.32	3.70	4.68
<i>t</i>	3.66	4.99	3.89
<i>p</i>	0.004	<0.001	0.003
CONTROL			
μ	20.67	23.08	23.25
σ	3.96	4.52	5.48
<i>t</i>	2.33	3.89	3.32
<i>p</i>	0.040	0.003	0.007

μ , Mean; σ , SD; *t*, test statistic of the independent samples *t*-test; *p*, probability value.

Table 4 | Results of Pearson correlations (*df* = 10) between the overall performance of both groups in the two conditions and psychophysically measured pitch direction and discrimination thresholds.

		Supra-threshold tonal	Sub-threshold tonal
AMUSIC			
Pitch detection	<i>r</i>	-0.35	-0.24
	<i>p</i>	0.26	0.46
Pitch direction	<i>r</i>	0.01	-0.36
	<i>p</i>	0.99	0.25
CONTROL			
Pitch detection	<i>r</i>	-0.04	-0.29
	<i>p</i>	0.90	0.35
Pitch direction	<i>r</i>	-0.49	-0.20
	<i>p</i>	0.11	0.53

r, Test statistic of the Pearson's product moment correlation; *p*, probability value.

bias in confidence responses reported by the two groups. Analysis of the confidence ratings given on a trial-by-trial basis was carried out using Signal Detection Theory (Green and Swets, 1966).

Following previous studies employing these methods, a “hit” was considered to be a correct response with high confidence, whereas a “false alarm” was an incorrect response with high

confidence (Kunimoto et al., 2001; Tunney and Shanks, 2003). Using hit and false alarm rates, computed by expressing the number of hits and false alarms as a proportion of correct and incorrect responses respectively, two key variables were extracted for each participant: their “awareness” or ability to judge whether a correct or an incorrect response had been made (the discriminability index, d') and their tendency to favor one response (“confident” versus “not confident”) over the other (the response bias, c). The former, d' , was computed as $d' = z(\text{hit rate}) - z(\text{false alarm rate})$, while the latter, c , was computed as $c = -0.5 [z(\text{hit rate}) + z(\text{false alarm rate})]$ (Macmillan and Creelman, 2001). A higher d' denotes greater awareness compared with a lower one and a d' value significantly greater than zero indicates presence of explicit knowledge. A negative value c denotes a liberal response bias (more likely to report “confident”), and a positive c value denotes a conservative response bias (less likely to report “confident”).

Table 5 shows means and SD of the hit rates, false alarm rates, d' and c for both groups. Although the control group had a higher mean d' , an independent sample t -test revealed no difference between the groups in their ability to discriminate correct responses from incorrect ones [$t(22) = -0.40, p = 0.70$]. Further, neither group had a mean d' significantly greater than zero [amusics: $t(11) = 0.74, p = 0.48$; controls: $t(11) = 1.61, p = 0.14$] suggesting that knowledge acquired was largely implicit and failed to reach full conscious awareness in both groups (Tunney and Shanks, 2003; Dienes and Scott, 2005). The next analysis examined whether there were any differences in response biases (c) between the two groups using an independent samples t -test. This revealed that the amusic group exhibited significantly greater conservatism than the control group when judging their performance [$t(22) = 3.15, p < 0.01$]. In other words, amusic individuals were less likely than controls to give a “confident” response.

Finally, using correlation analyses, it was investigated whether either awareness level (d') or the response bias (c) predicted participants performance, as defined by the number of correct responses out of 36 in the test phase. No relationship was seen between the response bias and performance in either the amusic ($r = -0.46, p = 0.13$) or the control group ($r = -0.07, p = 0.84$). However, results shown in **Figure 4** revealed that while controls who had a greater level of awareness were also more accurate in the test phase ($r = 0.62, p = 0.03$), there was no such relationship in the amusic group ($r = 0.27, p = 0.40$).

Table 5 | Mean hit rates, false alarm rates, and d' and c values for amusic and control participants.

	p(H)	p(FA)	d'	c
AMUSIC				
μ	0.40	0.34	0.15	0.47
σ	0.29	0.23	0.70	0.81
CONTROL				
μ	0.70	0.63	0.25	-0.48
σ	0.21	0.27	0.54	0.65

μ , Mean; σ , SD.

DISCUSSION

The facility that typical individuals have in perceiving music is built upon long term schematic knowledge gained incidentally over a life-time of exposure to the statistical properties of one's own musical culture (Tillmann et al., 2000). Individuals who exhibit pervasive and lifelong difficulties with music may lack such knowledge, either due to inadequate learning mechanisms, or intact learning mechanisms, which are rendered less effective owing to an insensitivity to small pitch changes. The present study aimed to distinguish between these possibilities, as well as to consider the possibility that any potential learning deficits were specific to music, rather than manifested more generally across the auditory domain.

A cohort of amusic individuals and matched controls were given equal opportunity to learn the regularities present within novel tonal and linguistic materials. In all conditions, participants were exposed to structured sequences made up of discrete “words” (tri-syllabic or tone-triplets) that were concatenated in such a manner that the only cues to where words began and ended were the transitional probabilities between adjacent syllables and tones. Following an exposure phase, participants heard pairs of words and identified which word had been present in the exposure phase. Two tonal conditions were used, spanning one and two-octaves respectively, in order to determine whether the use of small intervals could explain any potential lack of learning in the amusic group. Evidence of learning was seen across all three conditions, and equivalently for amusic individuals and controls. Interestingly neither the manipulation of pitch range (and thus interval size), nor the degree of familiarity with the tonal materials (the supra-threshold was constructed using a novel scale) had any effect on the degree of learning, for either group, and measured pitch thresholds did not correlate with the degree of learning shown by either group.

The finding of intact learning for the tonal and linguistic material suggests that difficulties in real-world music perception are unlikely to be explained in terms of a faulty learning mechanism. However, the current study has used non-word foils to assess learning in the test phase and it is possible that the use of part-word foils, which differ by just one syllable or tone, and therefore constitute a more difficult discrimination task in the test phase, would reveal more subtle deficits. Another issue worthy of consideration is that the current design involved multiple testing on the same individuals. While the use of a within-subjects design was deemed more suitable than a between-subjects design (where any observed differences across conditions could not be accounted for solely by the nature of the materials) its use resulted in a scenario whereby participants were no longer naïve to the task demands following the first experiment. It was unlikely that any transfer effects could occur between the first and second sessions, which were carried out several months apart, however we addressed the issue of repeated testing in the first session by counterbalancing the order in which individuals undertook the linguistic and supra-threshold tonal conditions. *Post hoc* analysis of the scores of individuals assigned to one or the other condition first verified the absence of any order effects.

The finding that learning was equivalently good for both tonal conditions is particularly striking, since previous studies have suggested that an insensitivity to pitch may lie at the heart of the disorder (Peretz et al., 2002; Foxton et al., 2004; Hyde and Peretz, 2004). The theory holds that, owing to the prevalence of small

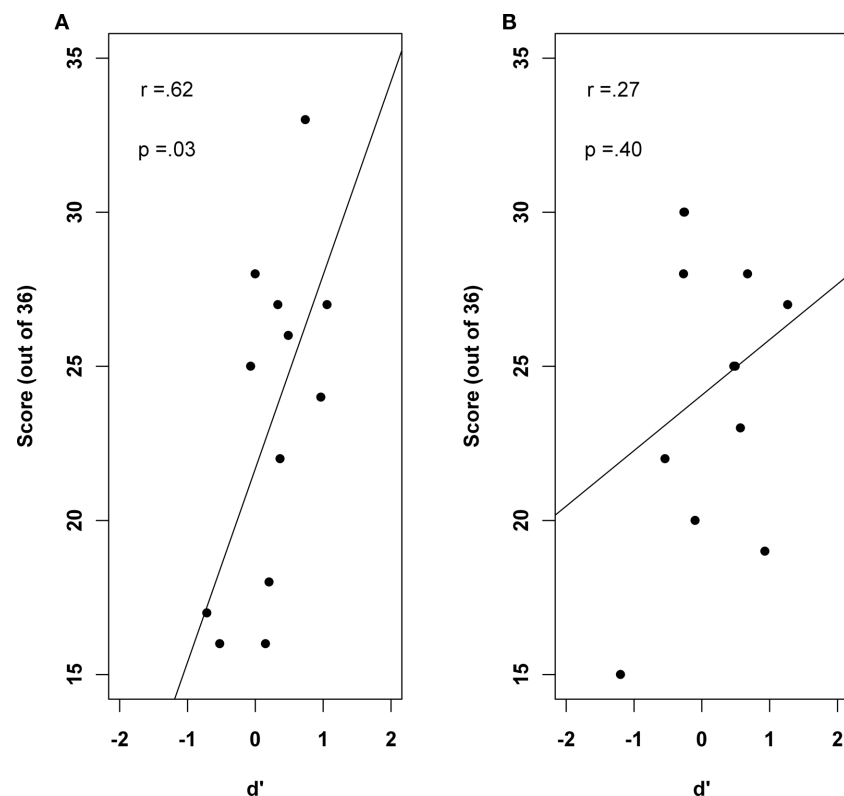


FIGURE 4 | Scatter plot showing the significant correlation between d' and performance for the control group (A) and the null correlation in the amusic group (B).

intervals in Western music (Dowling and Harwood, 1986; Vos and Troost, 1989), an insensitivity to such small intervals would have downstream effects for the acquisition of higher-order music features such as contour (Stewart et al., 2006) and the assimilation of musical scales which is central to the tonal encoding of pitch (Peretz and Hyde, 2003).

In the current cohort, only one amusic individual had a pitch detection threshold greater than a semitone. This might seem surprising given the common notion amusia arises from a fine-grained pitch discrimination problem (Hyde and Peretz, 2004). However a close examination of the literature suggests a mixed picture regarding the issue of pitch sensitivity in amusia (see Stewart, 2011). For instance, in one of the first studies to report fundamental pitch discrimination deficits in a cohort of amusic individuals, participants were required to monitor a sequence of five monotonic piano notes for a possible change in pitch at the fourth note (Hyde and Peretz, 2004). The authors reported that whilst controls were able to detect pitch intervals as small as a quarter of a semitone, amusic individuals were unable to detect a pitch change of a semitone or less. However, results from other studies have suggested that while pitch change detection is indeed worse in amusic individuals, their thresholds may be reduced to below a semitone when perception is assessed using alternative tasks. For instance, Foxton et al. (2004) and Tillmann et al. (2009) presented two pairs of sounds, one of which consisted of identical tones and the other of two tones of different frequencies, and required participants to indicate which of the two pairs differed in pitch. In a more recent study (Liu et al.,

2010), participants were required to indicate whether the first or last tone in a series of three tones contained a pitch glide. The forced-choice nature of these tasks makes them “criterion-free” (Kershaw, 1985; Macmillan and Creelman, 2001) which eliminates the effect of any bias amusic individuals may have toward favoring a more conservative “no change” response. Interestingly, results from these latter studies suggest that even though thresholds for the detection of a pitch change tend to be normal when assessed using forced-choice methods, thresholds for the discrimination of pitch direction are not, with a sizeable subgroup of amusic individuals possessing thresholds of a semitone or more.

This mixed picture regarding pitch thresholds in amusia suggests that performance is highly dependent on the way in which knowledge is probed. Neuro-imaging studies have demonstrated that individuals with amusia unconsciously process pitch deviations which they are unable to report explicitly (Peretz et al., 2009; Hyde et al., 2011). Measuring electrical brain activity using EEG, Peretz et al. (2009) reported that amusic individuals, showed evidence of an early right lateralized negative brain response to notes in melodies from the MBEA which had been mistuned by a quarter of a tone. Importantly, these same amusic individuals demonstrated no awareness of the incongruous mistuned notes in an equivalent behavioral task. Further to this evidence of intact processing in the auditory cortex, a recent fMRI study further confirmed that the brains of amusic individuals are sensitive to extremely fine pitch changes with activation in both the left and right auditory cortices increasing as a function of increasing pitch distance (Hyde et al., 2011). Although

experimental studies in which participants are exposed to artificial grammars provide evidence that implicit learning often gives way to explicit awareness after sufficient exposure (Shanks and St. John, 1994), the confidence ratings collected in the present study indicated that neither the control nor the amusic group were able to acquire full explicit knowledge of the sequences' structure. The finding of normal learning, as measured via largely implicit knowledge corroborates these previous studies, which suggest a level of unconscious knowledge in amusia that is in stark contrast to the explicit knowledge required for successful performance in musical perception tasks such as the MBEA (Peretz et al., 2009).

Analysis of response biases revealed that individuals with amusia were less confident about their performance, though no less accurate than controls. While the groups did not differ from each other in terms of how aware they were of their performance, a striking positive association was observed between awareness and performance in the control group that was not observed in the amusic group. The presence of this relationship in controls is not surprising as increasing awareness indicates an increasing tendency toward explicit knowledge acquisition and it is reasonable for performance in a learning task to correlate with levels of awareness (when unconscious) or explicit knowledge (when conscious). In contrast, the absence of this association in the amusic sample suggests a degree of dissociation whereby the level of awareness demonstrated by an individual does not predict their performance. What this finding suggests is that, in contrast to controls for whom performance in learning tasks may be largely contingent on awareness (Shanks and St. John, 1994), at least some individuals with amusia are able to perform well in the absence of any ability to discriminate when they are making a correct response from when they are making an incorrect one.

A dissociation between performance and explicit knowledge of performance has been frequently reported in the neuropsychological literature, for instance with amnesic patients who often show preserved memory in priming tasks while lacking any explicit memory for the same information (Graf et al., 1984; Knowlton et al., 1992; Reber et al., 2003). Such a dissociation has also been reported by Tillmann et al. (2007) in an individual with acquired amusia, as well as in a group of individuals with congenital amusia (unpublished data), who demonstrate implicit knowledge of musical structure in a harmonic priming task. However, attempts to disentangle the nature of implicit and explicit processing in individuals with congenital amusia suggest that their difficulties may not simply be attributed to a lack of confidence or awareness. In contrast to studies suggesting that amusic individuals possess intact musical knowledge but can not explicitly report it, results from other studies that use indirect tasks and electrophysiological methods (in which neither attention nor knowledge declaration is required) have provided evidence that their deficits include a genuine insensitivity to pitch change (Pfeuty and Peretz, 2010) and tonal aspects of musical structure (Peretz et al., 2009). The failure of control individuals to acquire explicit knowledge in the current task makes it impossible to fully address the question of the extent to which amusic individuals are able to acquire explicit knowledge through incidental listening but the approach of collecting confidence ratings seems promising and future studies using tasks in which controls reliably acquire explicit knowledge may help to resolve this.

Two other possibilities may be considered to explain why amusic individuals have such difficulties in perceiving music, despite showing intact learning for statistical regularities in tonal material. While exposure time was equated across groups in the present study, it could be argued that amusic individuals have had insufficient exposure to real-world music. This seems unlikely, given that young children are able to show knowledge of musical structure despite relatively limited exposure to musical material (Krumhansl and Keil, 1982; Trainor and Trehub, 1994), at least compared with a life-time of incidental exposure in adult amusics. In addition, several amusics have been reported to have engaged in protracted attempts to learn music and to have grown up in a musically enculturated environment (Ayotte et al., 2002; Stewart, 2008). More plausible is the possibility that the internalization of statistical regularities from real-world music may be more complex compared with the first-order transitional probabilities used in the present study. Higher-order transitional probabilities or relational probabilities between non-adjacent tones (Creel et al., 2004; Gebhart et al., 2009) may be more relevant to the acquisition of knowledge required to support an understanding of melodic and harmonic structure compared with the simple first-order transitional probabilities used in the present study (Tillmann et al., 2000; Jonaitis and Saffran, 2009). Future studies might assess the ability of amusic individuals to internalize the rules guiding more complex musical grammars (Loui et al., 2010; Rohrmeier et al., 2011) though it should be borne in mind that performance on artificial grammar learning tasks simulating more complex musical systems may be limited by the short-term memory deficits shown by many individuals with congenital amusia (Gosselin et al., 2009; Williamson and Stewart, 2010; Williamson et al., 2010).

CONCLUSION

In sum, the present study has provided evidence that while individuals with amusia may possess some of the fundamental mechanisms required to build knowledge of musical structure, they lack confidence in their ability and display different patterns of awareness compared with typical individuals. The striking ability of the current cohort to internalize statistical regularities based on a tonal sequence of intervals including the semitone, urges a reconsideration of the view that amusia may emerge owing to an insensitivity to small pitch changes (Peretz and Hyde, 2003). Whether the difficulties that amusic individuals show with real music can be related to a failure to internalize higher-order regularities remains to be investigated. Results from future studies investigating the state of higher-order learning mechanisms and the types of knowledge acquired, will have a critical bearing on the extent to which the condition may be considered a disorder of awareness, rather than perception.

ACKNOWLEDGMENTS

This study was supported by a grant to Lauren Stewart from the Economic and Social Research council (ESRC; RES-061-25-0155) and a joint studentship awarded to Diana Omigie by the ESRC and the Medical Research council (MRC). We would like to thank all the participants for their continued assistance in our studies of amusia, and for their insights.

REFERENCES

- Ayotte, J., Peretz, I., and Hyde, K. L. (2002). Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain* 125, 238–251.
- Bigand, E., and Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychol. Res.* 62, 237–254.
- Bigand, E., Parncutt, R., and Lerdahl, F. (1996). Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Percept. Psychophys.* 58, 125–141.
- Bigand, E., and Poulin-Charronnat, B. (2006). Are we “experienced listeners?” A review of the musical capacities that do not depend on formal musical training. *Cognition* 100, 100–130.
- Blacking, J. (1995). *Music, Culture and Experience*. London: University of Chicago Press.
- Bod, R. (2002). A unified model of structural organization in language and music. *J. Artif. Intell. Res.* 17, 289–308.
- Boersma, P. (2001). Praat: a system for doing phonetics by computer. *Glot Int.* 5, 341–345.
- Brown, H., Butler, D., and Jones, M. R. (1994). Musical and temporal influences on key discovery. *Music Percept.* 11, 371–407.
- Creel, S., Newport, E., and Aslin, R. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1119–1130.
- Cuddy, L. L., and Badertscher, B. (1987). Recovery of the tonal hierarchy—some comparisons across age and levels of musical experience. *Percept. Psychophys.* 41, 609–620.
- Dienes, Z., and Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychol. Res.* 69, 338–351.
- Dowling, W. J., and Harwood, D. L. (1986). *Music Cognition*. Orlando: Academic press.
- Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., and Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying “tone deafness.” *Brain* 127, 801–810.
- Gebhart, A., Newport, E., and Aslin, R. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon. Bull. Rev.* 16, 486–490.
- Gosselin, N., Jolicoeur, P., and Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Ann. N. Y. Acad. Sci.* 1169, 270–272.
- Graf, P., Squire, L. R., and Mandler, G. (1984). The information that amnesic patients do not forget. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 164–178.
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Hebert, S., Peretz, I., and Gagnon, L. (1995). Perceiving the tonal ending of tune excerpts: the roles of pre-existing representation and musical expertise. *Can. J. Exp. Psychol.* 49, 193–209.
- Honing, H., and Ladinig, O. (2009). Exposure influences expressive timing judgments in music. *J. Exp. Psychol. Hum. Learn.* 35, 281–288.
- Hyde, K. L., and Peretz, I. (2004). Brains that are out of tune but in time. *Psychol. Sci.* 15, 356–360.
- Hyde, K. L., Zatorre, R., and Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cereb. Cortex* 21, 292–299.
- Jonaitis, E. M., and Saffran, J. R. (2009). Learning harmony: the role of serial statistics. *Cogn. Sci.* 33, 951–968.
- Kalmus, H., and Fry, D. B. (1980). On tune deafness (dysmelodia) – frequency, development, genetics and musical background. *Ann. Hum. Genet.* 43, 369–382.
- Kershaw, C. D. (1985). Statistical properties of staircase estimates from two interval forced choice experiments. *Br. J. Math. Stat. Psychol.* 38, 35–43.
- Knowlton, B. J., Ramus, S. J., and Squire, L. R. (1992). Intact artificial grammar learning in amnesia – dissociation of classification learning and explicit memory for specific instances. *Psychol. Sci.* 3, 172–179.
- Krumhansl, C. L., and Keil, F. C. (1982). Acquisition of the hierarchy of tonal functions in music. *Mem. Cognit.* 10, 243–251.
- Kuhn, G., and Dienes, Z. (2005). Implicit learning of nonlocal musical rules: implicitly learning more than chunks. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1417–1432.
- Kunimoto, C., Miller, J., and Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Conscious. Cogn.* 10, 294–340.
- Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- Liu, F., Patel, A. D., Fourcin, A., and Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain* 133, 1682–1693.
- Loui, P., Wessel, D. L., and Kam, C. L. H. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Percept.* 27, 377–388.
- Macmillan, N. A., and Creelman, C. D. (2001). *Detection Theory: A User's Guide*. Cambridge: Cambridge University Press.
- Marmel, F., and Tillmann, B. (2009). Tonal priming beyond tonics. *Music Percept.* 26, 211–221.
- Nelson, H. E., and Willison, J. (1991). *National Adult Reading Test Manual*, 2nd Edn. Windsor: NFER-Nelson.
- Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., and Jutras, B. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron* 33, 185–191.
- Peretz, I., Brattico, E., Jarvenpää, M., and Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain* 132, 1277–1286.
- Peretz, I., Champod, A. S., and Hyde, K. L. (2003). Varieties of musical disorders. The Montreal battery of evaluation of amusia. *Ann. N. Y. Acad. Sci.* 999, 58–75.
- Peretz, I., and Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends Cogn. Sci.* 7, 362–367.
- Pfeuty, M., and Peretz, I. (2010). Abnormal pitch-time interference in congenital amusia: evidence from an implicit test. *Atten. Percept. Psychophys.* 72, 763–744.
- Reber, A. S. (1992). The cognitive unconscious – an evolutionary perspective. *Conscious. Cogn.* 1, 93–133.
- Reber, P. J., Martinez, L. A., and Weintraub, S. (2003). Artificial grammar learning in Alzheimer's disease. *Cogn. Affect. Behav. Neurosci.* 3, 145–153.
- Rohrmeier, M., Rebuschat, P., and Cross, I. (2011). Incidental and online learning of melodic structure. *Conscious. Cogn.* 20, 214–222.
- Saffran, J., Johnson, E., Aslin, R., and Newport, E. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.
- Saffran, J. R., and Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: evidence for developmental reorganization. *Dev. Psychol.* 37, 74–85.
- Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996). Word segmentation: the role of distributional cues. *J. Mem. Lang.* 35, 606–621.
- Schellenberg, E. G. (1996). Expectancy in melody: tests of the implication-realisation model. *Cognition* 58, 75–125.
- Schmuckler, M. A. (1989). Expectation in music – investigation of melodic and harmonic processes. *Music Percept.* 7, 109–150.
- Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., and Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition* 106, 975–983.
- Shanks, D. R., and St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behav. Brain Sci.* 17, 367–4470.
- Smith, J. D., Nelson, D. G. K., Grohskopf, L. A., and Appleton, T. (1994). What child is this – what interval was that – familiar tunes and music perception in novice listeners. *Cognition* 52, 23–54.
- Stewart, L. (2008). Fractionating the musical mind: insights from congenital amusia. *Curr. Opin. Neurobiol.* 18, 127–130.
- Stewart, L. (2011). Characterising congenital amusia. *Q. J. Exp. Psychol.* 64, 625–638.
- Stewart, L., von Kriegstein, K., Warren, J. D., and Griffiths, T. D. (2006). Music and the brain: disorders of musical listening. *Brain* 129, 2533–2553.
- Tillmann, B., Bharucha, J. J., and Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychol. Rev.* 107, 885–913.
- Tillmann, B., and McAdams, S. (2004). Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis)similarities. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1131–1142.
- 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.
- Tillmann, B., Schulze, K., and Foxton, J. M. (2009). Congenital amusia: a short-term memory deficit for non-verbal, but not verbal sounds. *Brain Cogn.* 71, 259–264.
- Toivianen, P., and Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: the dynamics of tonality induction. *Perception* 32, 741–766.
- Trainor, L. J., and Trehub, S. E. (1994). Key membership and implied harmony in western tonal music: developmental perspectives. *Percept. Psychophys.* 56, 125–132.
- Tunney, R. J., and Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Mem. Cognit.* 31, 1060–1071.
- Vos, P. G., and Troost, J. M. (1989). Ascending and descending melodic intervals – statistical findings and their perceptual relevance. *Music Percept.* 6, 383–396.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-III (WAIS-III)*. San Antonio, TX: The Psychological Cooperation.
- Williamson, V. J., McDonald, C., Deutsch, D., Griffiths, T. D., and Stewart, L.

(2010). Faster decline of pitch memory over time in congenital amusia. *Adv. Cogn. Psychol.* 6, 15–22.

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.

Received: 14 February 2011; accepted: 13 May 2011; published online: 03 June 2011.

Citation: Omigie D and Stewart L (2011) Preserved statistical learning of tonal and linguistic material in congenital amusia. *Front. Psychology* 2:109. doi: 10.3389/fpsyg.2011.00109

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

Copyright © 2011 Omigie and Stewart. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



Processing of voiced and unvoiced acoustic stimuli in musicians

Cyrill Guy Martin Ott^{1*}, Nicolas Langer¹, Mathias S. Oechslin^{2,3}, Martin Meyer¹ and Lutz Jäncke¹

¹ Division Neuropsychology, Psychological Institute, University of Zurich, Zurich, Switzerland

² Geneva Neuroscience Center, University of Geneva, Geneva, Switzerland

³ Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland

Edited by:

Anne-Lise Giraud, INSERM-Ecole Normale Supérieure, France

Reviewed by:

Mitchell Steinschneider, Albert Einstein College of Medicine, USA
Anne Caclin, INSERM, France

*Correspondence:

Cyrill Guy Martin Ott, Division Neuropsychology, Institute of Psychology, University Zurich, Binzmühlestrasse 14/Box 25, CH-8050 Zurich, Switzerland.
e-mail: c.ott@psychologie.uzh.ch

Past research has shown that musical training induces changes in the processing of supra-segmental aspects of speech, such as pitch and prosody. The aim of the present study was to determine whether musical expertise also leads to an altered neurophysiological processing of sub-segmental information available in the speech signal, in particular the voice-onset-time. Using high-density EEG-recordings we analyzed the neurophysiological responses to voiced and unvoiced consonant-vowel-syllables and noise-analogs in 26 German speaking adult musicians and non-musicians. From the EEG the N1 amplitude of the event-related potential and two microstates from the topographical EEG analysis (one around the N1 amplitude and one immediately preceding the N1 microstate) were calculated to the different stimuli. Similar to earlier studies the N1 amplitude was different to voiced and unvoiced stimuli in non-musicians with larger amplitudes to voiced stimuli. The more refined microstate analysis revealed that the microstate within the N1 time window was shorter to unvoiced stimuli in non-musicians. For musicians there was no difference for the N1 amplitudes and the corresponding microstates between voiced and unvoiced stimuli. In addition, there was a longer very early microstate preceding the microstate at the N1 time window to non-speech stimuli only in musicians. Taken together, our findings suggest that musicians process unvoiced stimuli (irrespective whether these stimuli are speech or non-speech stimuli) differently than controls. We propose that musicians utilize the same network to analyze unvoiced stimuli as for the analysis of voiced stimuli. As a further explanation it is also possible that musicians devote more neurophysiological resources into the analysis of unvoiced segments.

Keywords: musicians, speech, musical expertise, N1, microstates, topographic pattern analysis, transfer effect, experience dependent plasticity

INTRODUCTION

The past 15 years have seen a vast amount of research on music and associated neural processes. One of the most prominent findings is that the intensive practice and training needed to achieve high proficiency in musicianship leads to structural and functional short and long-term alterations in the brain (for a review, see Schlaug, 2001; Münte et al., 2002; Jäncke, 2009). Researchers have therefore become increasingly aware of the value of musicianship as a model for brain plasticity. In view of this, the question arises whether these music-training induced plasticity effects are restricted particularly to structures and functions underlying the perception and production of music or are more widespread in affecting also other domains such as language and speech.

A number of recent studies indicate that the language network is affected by plastic alterations due to musical training (e.g., Thompson et al., 2003; Schön et al., 2004; Moreno and Besson, 2006; Besson et al., 2007; Marques et al., 2007; Parbery-Clark et al., 2009; Gordon et al., 2010; Marie et al., 2010a,b; Colombo et al., 2011). However, most of these studies focused on pitch processing and prosody, nonetheless showing a substantial transfer effect of intensive musical training on auditory language functions. Efficient

analysis of spectral acoustic information in the speech signal is indeed crucial for auditory language perception, but the same is true for temporal cues (Shannon et al., 1995; Davis and Johnsrude, 2007). As musical training involves melodic, supra-segmental and subtle timing issues, the processing of rapidly changing temporal cues in the speech signal and the underlying neural substrates may also be altered in musicians. Studies which have examined musicians and non-musicians in terms of their ability to discriminate segmental phonetic cues mainly focused on the spectral domain and generally have revealed negative results. For example, Hillenbrand et al. (1990) have demonstrated that musicians are not faster or even better in discriminating synthetic speech sounds. A similar finding has been provided by Marie et al. (2010a), who examined whether musical expertise has an effect on the discrimination of tonal and segmental variations in tone language. In addition, they measured event-related potentials (ERPs) and identified that musical expertise did not influence segmental processing as indicated by the amplitudes and latencies of early components like the N1. These authors, however, demonstrated strong influences of musical expertise on the later occurring “cognitive” ERP components (N2, P3a, and P3b).

Anyhow, the speech signal carries not only rapidly changing spectral, but also temporal information on a segmental level. Musical training on the other hand, includes practising the perception of rhythm, meter, and subtle timing issues. Thus, it might be the case that when focusing on the temporal domain, musicians and non-musicians actually do differ with respect to the processing of segmental phonetic cues. One of the most important temporal acoustic cues carrying linguistically and phonetically relevant information is the voice-onset-time (VOT). This mainly refers to voicing differences between stop-consonants in a broad range of languages. VOT is defined as the duration of the delay between release of closure and start of voicing (Lisker and Abramson, 1964). Electrophysiological studies have consistently demonstrated VOT-related auditory evoked potential (AEP) differences in the N1 component in humans and animals (e.g., Sharma and Dorman, 1999; Steinschneider et al., 1999, 2005; Sharma et al., 2000; Trébuchon-Da Fonseca et al., 2005; Zaehle et al., 2007). The N1 component is considered to be particularly important in reflecting the basic encoding of acoustic information of the auditory cortex (Näätänen and Picton, 1987; Picton et al., 1987). Moreover, it has been shown that the electrophysiologic patterns elicited by the processing of consonant-vowel (CV) syllables and corresponding non-speech sounds with similar temporal characteristics are comparable in humans (Liégeois-Chauvel et al., 1999; Zaehle et al., 2007). This indicates a general mechanism for encoding and analyzing successive temporal changes in acoustic signals. Modulations of the N1 component reflect the central auditory representation of speech and non-speech sounds (Ostroff et al., 1998; Sharma et al., 2000; Meyer et al., 2006; Pratt et al., 2007). The present study thus aimed to examine potential differences between musicians and non-musicians in the processing of changing temporal acoustic cues. This was done by recording and comparing scalp AEP's in response to CV-syllables and corresponding noise-analogs with varying VOT's and noise-onset times (NOT), respectively. Based on the findings of Zaehle et al. (2007), which clearly demonstrate that the AEP's N1 component is especially responsive to temporal variations in speech and non-speech stimuli, we focused our ("classical") AEP analysis on amplitude modulations at the vertex electrode in the N1 time window. These modulations were elicited by brief auditory stimuli that varied systematically along an acoustic and a linguistic dimension.

We hypothesized that within the non-musicians group we would find stronger N1 amplitudes for stimuli with short (voiced) compared to long (unvoiced) VOT's/NOT's, as this has previously been shown by Zaehle et al. (2007) and various others (e.g., Simos et al., 1998; Sharma et al., 2000; Steinschneider et al., 2005). On the other hand, if musical expertise has an influence on early phonetic processing in the temporal domain, then musicians should show a different pattern in the corresponding neurophysiological indicators of early phonetic processing within the time window of the N1.

In order to reveal potential differences between musicians and non-musicians in distinct stages of information processing of VOT's/NOT's, we also analyzed the spatial variations of the scalp voltage distribution over time between conditions and groups, an approach known as Topographic Pattern Analysis or evoked potential (EP) mapping. This method searches for time segments of stable map topography that represents functional microstates of the brain. These are assumed to reflect distinct information processing steps

and provide several advantages over the classical ERP analysis, such as experimenter and reference independence (for an overview, see, e.g., Murray et al., 2008; Michel et al., 2009). This kind of analysis benefits mostly from the high topographic resolution that high-density EEG-recordings provide. Using this approach, we searched for stable map topographies before, during, and after the N1 time window. Here, we are interested in examining whether the duration of these maps are different between musicians and non-musicians.

MATERIALS AND METHODS

SUBJECTS

Thirty-one healthy volunteers with no history of neurological pathology and with normal audiological status participated in this study. All participants were native German or Swiss German speakers. One group consisted of 16 musicians (11 women, 5 men, mean age \pm SD of 26.7 ± 10.1 years), with formal training starting at a mean age of 6.02 ± 2.2 years. All musicians were students, music teachers and/or members of an orchestra/band and practiced their instrument daily for 1–6 h. Most of them play several instruments of which the most frequent were piano, violin, and clarinet. The other group comprised 15 non-musicians (11 women, 4 men, mean age \pm SD of 26.07 ± 7.9 years) with no formal musical training and no history of musical instrumental performance. According to the Annett-Handedness-Questionnaire (Annett, 1970), 13 participants within the musician group were consistently right-handed, whereas two were ambidextrous and one was left-handed. In the non-musician group, 11 participants were consistent right-handers, one was ambidextrous and three were left-handed. In order to determine each participant's degree of musical competence, the "Advanced Measures of Music Audiation" by Gordon (1989) was applied. All participants performed also a short intelligence test (KAI) to rule out significant differences in intelligence between groups. Descriptive statistics for the criterion measures completed by the participants are summarized in **Table 1**. All participants gave informed written consent, in accordance with procedures approved by the local ethics committee, and were paid for participation.

STIMULI

All participants heard two different classes of auditory stimuli in the context of a phonetic categorization task. One class formed the "speech condition" consisting of the same CV-syllables used in previous studies, such as Jäncke et al. (2002) and Meyer et al. (2007). These syllables (/ka/, /ta/, /pa/, /ga/, /da/, /ba/) were digitally recorded by a trained phonetician at a sampling rate of 44.1 kHz and a sampling depth of 16-bit. The onset, duration, intensity, and fundamental frequency of the stimuli were edited and synchronized by means of a speech editor. The criterion for temporal alignment of the syllables was the onset of articulatory release. The duration of the syllables ranged from 310 to 360 ms, depending on the VOT, with a vowel duration of 300 ms (VOTs in ms for the stops were approximately "k" = 59, "t" = 49, "p" = 39, "g" = 19, "d" = 04, "b" = 09). The second class, forming the "non-speech condition," consisted of noise-analogs, which were synthesized from the CV-syllables used in the speech condition. This was achieved using a variation of a formerly published procedure (Shannon et al., 1995). Thus, spectral information was removed from the syllables by replacing

the frequency specific information in a broad frequency region with band-limited white noise (Band 1: 500–1500 Hz, Band 2: 2500–3500 Hz) whereas normalized amplitudes and temporal cues were preserved in each spectral band, resulting in double-band-pass filtered noise with temporal CV-amplitude dynamics. Thus, these stimuli are devoid of any phonological, semantic, and syntactic cues but imitate the temporal envelope of proper speech signals (syllables, words, or sentences).

This non-speech condition was introduced in a non-task-relevant manner to reassess whether any found differences are due to different VOT processing and not to other varying characteristics between the speech stimuli. **Figure 1** shows both the speech and non-speech waveforms of the syllables “pa” and “ba” as an example of the stimuli used in the present study. All manipulations were carried out with the “Adobe Audition 1.5” software (Adobe Systems Incorporated, 1992–2004). All auditory stimuli were presented binaurally at a sound pressure level of about 70 dB using hi-fi headphones. Stimuli were presented and responses recorded using Presentation® software (Neurobehavioral Systems, USA).

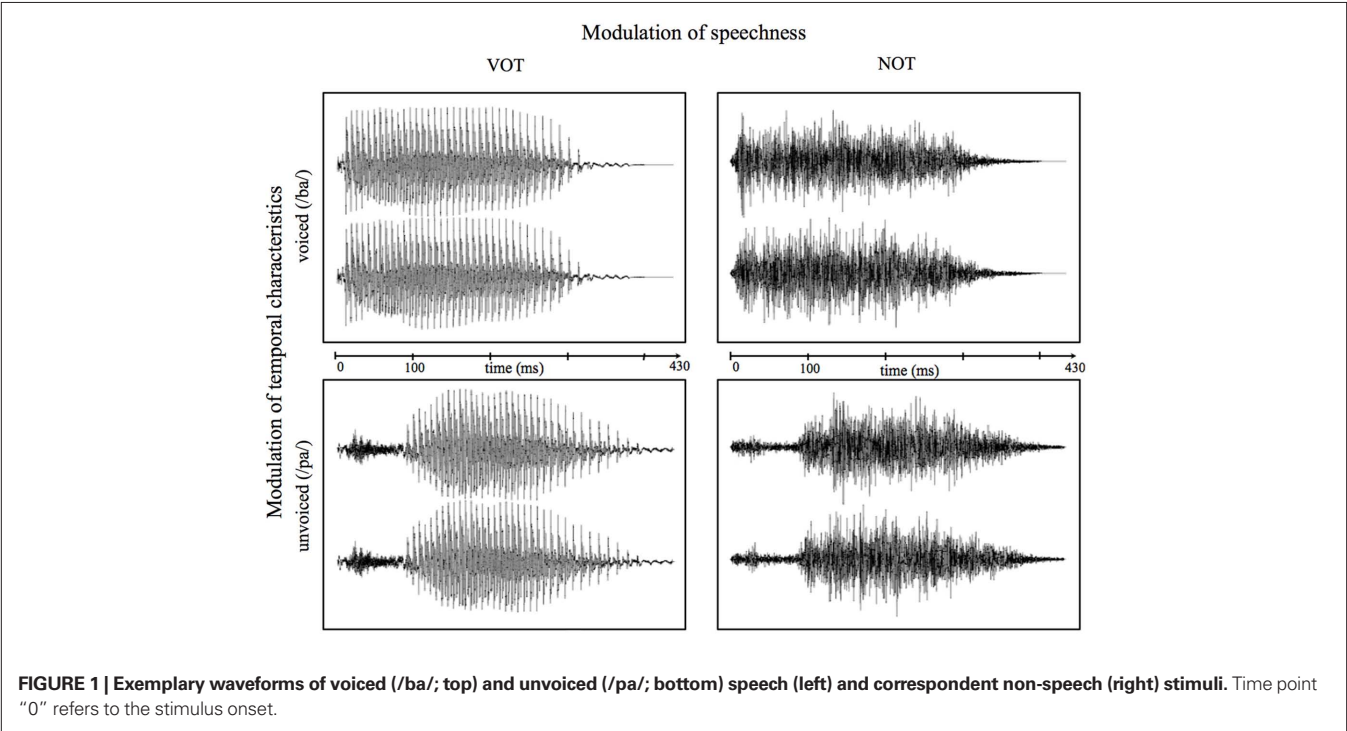
PROCEDURE

The EEG experiment comprised four blocks per condition (i.e., “speech” and “non-speech”), resulting in a total of eight blocks and an overall duration of the experiment of about 60 min. Each block contained 42 trials of each stimulus, thus presenting 252 stimuli per block and a total amount of 168 presentations per stimulus. In both conditions, stimuli were shuffled in a pseudo-randomized order. The condition order was randomized for every subject. Stimuli were of two different categories, assigned either “voiced” or “unvoiced,” depending on their particular VOT (“voiced” = /ga/,/da/,/ba/and corresponding noise-analogs; “unvoiced” = /ka/,/ta/,/pa/and corresponding noise-analogs). Participants were seated in a comfortable chair in an electrically shielded, dimly lit room, and were instructed to listen attentively to the auditory stimuli presented via headphones. They had to decide as quickly and accurately as possible whether the heard stimulus was of type “voiced” or “unvoiced,” and to indicate their response by pressing either the left or right mouse button. We utilized this task to ensure the participant’s vigilance throughout the

Table 1 | Descriptive statistics for the criterion measures completed by the subjects are listed group-wise (M = musicians; NM = non-musicians; GMA = Gordon musical aptitude).

	Age		Gender		IQ		GMA*		Handedness		
	Mean	SD	m	f	Mean	SD	Mean	SD	Left	Right	Ambi
M	26.7	10.1	5	11	131.4	6.5	68.93	14.28	1	13	2
NM	26.07	7.9	4	11	124.9	11.9	43.26	21.41	3	11	1

*Difference between experimental groups is significant at $p < 0.05$ [$T(29) = -3.951$; note that t-tests were only applied to variables “Age,” “IQ” and “GMA.” For “Gender” and “Handedness,” Chi-Square tests were used]. All statistical tests shown in this table were conducted comparing the two experimental groups.



experiment, even though it has previously been shown (Baumann et al., 2008) that an enhancement of N1 amplitudes in musicians rather reflects an influence of expertise than selective attention. In order to ensure that participants could solve the task as instructed, each participant performed practice trials before the experiment until complete comprehension of the task was reported. As each stimulus was presented 500 ms after the preceding mouse button press, the inter-stimulus interval depended mainly on the speed of the participant's response. When no response occurred, the next stimulus was presented automatically after 2 s.

AEP RECORDINGS

Electroencephalogram was recorded using a high-density Geodesic EEG system[®] (GSN300; Electrical Geodesic Inc., Oregon)¹ with 128-Channel HydroCel Geodesic Sensor Nets[®] (HCGSN120). Data was sampled at 250 Hz and band-pass filtered at 0.1–100 Hz. The vertex electrode (Cz) served as on-line recording reference. Impedance was kept below 30 k Ω . For various reasons, such as violations of attendance criteria regarding age or technical problems during EEG recording, behavioral and EEG data of five participants (three musicians and two non-musicians) were discarded from further analysis, leading to a total of 26 remaining datasets (13 per group).

DATA ANALYSIS – BEHAVIORAL DATA

The number of correct identifications, errors, and misses in each block was counted within the two groups (musicians vs. non-musicians) and averaged for each subject across blocks, categories (i.e., “voiced” vs. “unvoiced”), and classes (i.e., “speech” vs. “non-speech”). The same procedure was carried out with reaction times (RT), which were analyzed for correct identifications only. Thus, a $2 \times 2 \times 2$ repeated measure analysis of variance (ANOVA) with factors *group* (between-subject), *voiceness*, and *speechness* (within-subject) was computed for each of these variables in order to statistically evaluate any differences in identification accuracy and RT's.

DATA ANALYSIS – EEG, N1 COMPONENT ANALYSIS

Electroencephalogram recordings of each participant were imported and analyzed in the BrainVision Analyzer2 software (Version 2.0.1; Brain Products GmbH, D-82205 Gilching)². In a first step, data was band-pass filtered at 1.5–70 Hz and a 50-Hz notch filter was applied. An ICA (independent component analysis) was then performed to correct for ocular artifacts (e.g., Jung et al., 2000). Each EEG recording was visually inspected and trials with sweating and muscle artifacts, amplifier saturation, and remaining eye blinks or eye movements were excluded. Noisy channels were interpolated and the data was then re-referenced to the mastoid electrodes for ERP calculation. Each ERP waveform was an average of more than 100 repetitions of the EEG sweeps evoked by the same stimulus type. EEG-recordings were sectioned into 500 ms segments (100 ms pre-stimulus and 400 ms post-stimulus) and a baseline correction using the signal's pre-stimulus portion was carried out. Finally, ERP's for each stimulus were averaged for

each subject and grand-averaged across subjects within the two groups separately. Had peak values of grand-averages been used for statistical analysis, differences between experimental groups in the N1 peak values might have been blurred during averaging due to individual differences in their latencies. Therefore, grand-average waves were used only to reassess the N1 time window and for illustrative purposes.

In order to statistically confirm relevant differences between AEP's at Cz as a function of experimental group, stimulus category and condition, mean amplitude ERP's time-locked to the auditory stimulation were therefore measured in a specific latency window. This was individually pre-determined for each subject and stimulus by visual inspection of the event-related signal. These individual latency windows were centered at the peak of the prominent N1 component and covered a total signal length of 20 ms around the center. Individually chosen latency windows were used to ensure that the mean amplitude actually reflect the N1 peak values of every subject and stimulus. Mean amplitudes were then averaged separately within groups (i.e., “musicians” vs. “non-musicians”), depending on their category (i.e., “voiced” vs. “unvoiced”), and class (i.e., “speech” vs. “non-speech”). Subsequently, a $2 \times 2 \times 2$ repeated measure ANOVA with a between-subject factor (*group*) and two within-subjects factors (*speechness* and *voiceness*) was computed for the central electrode (Cz), and the reported *p*-values were adjusted for non-sphericity using Greenhouse–Geisser Epsilon when equal variances could not be assumed. Subsequently, Bonferroni–Holm adjusted *post hoc t*-tests were applied. The global level of significance was set at $p < 0.05$ for all statistical analyses.

DATA ANALYSIS – EEG, TOPOGRAPHIC PATTERN ANALYSIS

Topographic Pattern Analysis was performed using the Cartool software (Version 3.43/869; The Cartool Community group)³. Subject- and grand-averaged ERP's were imported into Cartool and recalculated against the average reference. To identify the dominant map topographies on the scalp, segments of stable voltage topography (or EP maps) were then defined by using a topographic atomize and agglomerate hierarchical cluster analysis (T-AAHC) in the grand-averaged ERP's across conditions and groups over the full post-stimulus EEG segment length of 400 ms. These template maps are the mean maps over the period where the stable voltage topography segments were found. The optimal number of templates was determined by the combination of a modified cross-validation and the Krzanowski–Lai criterion (e.g., Pascual-Marqui et al., 1995; Murray et al., 2008). In order to assess any differences regarding map duration between groups and conditions in and around the N1 time window (i.e., 50–200 ms), the presence of each map that was identified in the group-averaged data within that particular epoch was verified statistically over the same period in the ERP's of the individual participants (i.e., “single-subject fitting”; Murray et al., 2008). This step allowed us to determine the duration of any given template map for each condition within the musician's and control's group between 50 and 200 ms. This time window was accordingly chosen to encompass the N1 time window itself plus the periods immediately before and after it. These duration values were then statistically evaluated

¹<http://www.egi.com>

²<http://www.brainproducts.com>

³sites.google.com/site/cartoolcommunity

for each map of interest by means of a repeated measures ANOVA with the factors *group* (between-subjects), *speechness*, and *voiceness* (within-subjects), as in the classical N1 component analysis. Subsequently, any significant interactions were further examined by applying *post hoc t*-tests.

RESULTS

BEHAVIORAL DATA

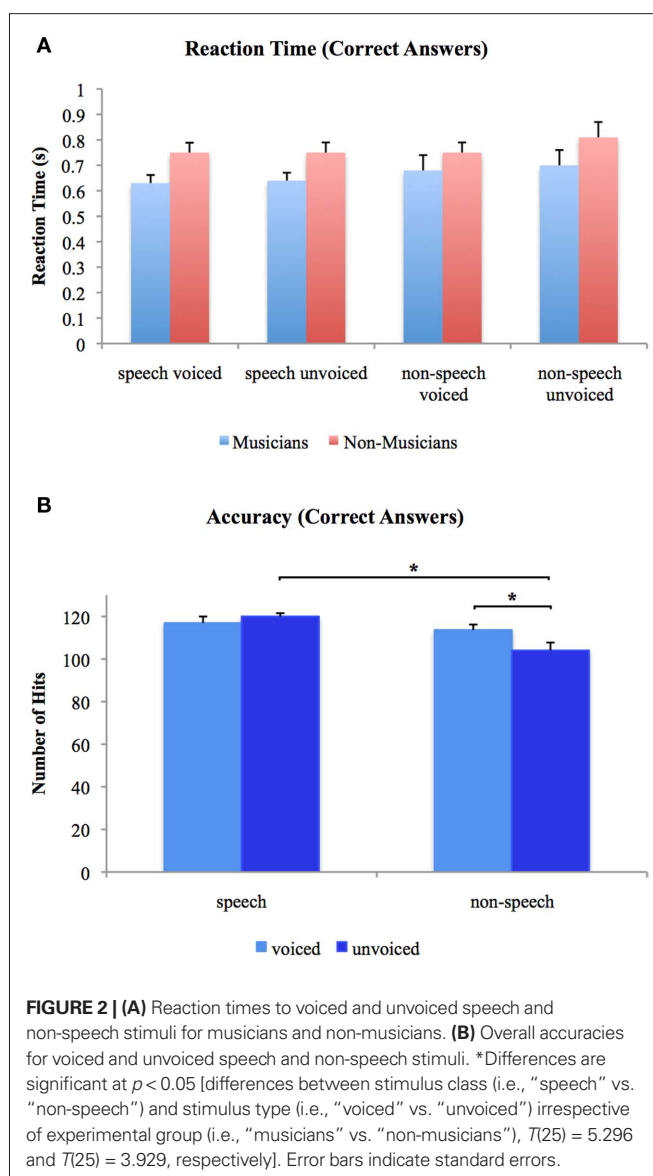
We analyzed accuracy and RT using $2 \times 2 \times 2$ ANOVAs with one between-subjects factor (*group*: musicians vs. non-musicians) and two repeated measurement factors (*voiceness*: voiced vs. unvoiced; *speechness*: speech vs. non-speech). With regard to accuracy, significant main effects for *voiceness* [$F(1, 24) = 6.096, p < 0.05$] and *speechness* [$F(1, 24) = 6.742, p < 0.05$] were found with higher accuracies for speech stimuli in general. Furthermore, there was a significant *voiceness* by *speechness* [$F(1, 24) = 1.217, p < 0.05$] interaction characterized by higher accuracies for voiced than unvoiced non-speech stimuli [$T(25) = 5.296, p < 0.05$] and for unvoiced speech than unvoiced non-speech stimuli [$T(25) = 3.929, p < 0.05$]. There were no significant differences between groups neither for the main effects nor the interactions. With respect to RT, neither significant main nor interaction effects were found in the $2 \times 2 \times 2$ ANOVA of speech and non-speech stimuli. Mean reaction times for the two experimental groups are shown in **Figure 2A**. **Figure 2B** shows overall accuracies irrespective of group. Note that only RT of correct answers were analyzed.

EEG DATA – N1 COMPONENT ANALYSIS

Grand-averaged waveforms evoked by each category and class of stimuli recorded at Cz are shown in **Figure 3** for the speech and in **Figure 4** for the non-speech condition, respectively. We observed that all stimuli elicited a prominent N1 component. In the non-musician group, the strongest N1 amplitudes were found for stimuli with short VOT's (i.e., "voiced" stimuli), thus replicating the results of Zaehle et al. (2007) and various others (e.g., Simos et al., 1998; Sharma et al., 2000). However, in the musician group, no difference between N1 amplitudes was revealed. In order to statistically examine the ERP effects, mean amplitudes of each stimulus category (i.e., "voiced" vs. "unvoiced") and class (i.e., "speech" vs. "non-speech") were analyzed for each subject in a 20-ms latency window around the N1 peaks.

Results of the $2 \times 2 \times 2$ repeated measures ANOVA with the factors *group* (musicians vs. non-musicians), *voiceness* (voiced vs. unvoiced), and *speechness* (speech vs. non-speech) for the N1 amplitudes showed significant main effects for the factors *voiceness* [$F(1, 24) = 4.733, p < 0.05$] and *speechness* [$F(1, 24) = 29.33, p < 0.05$]. In addition, a significant *group* by *voiceness* interaction was revealed [$F(1, 24) = 6.209, p < 0.05$].

Separate *post hoc* analyses were conducted for the two experimental groups, using Bonferroni–Holm adjusted *t*-tests and accepting therefore differences as significant only when the corresponding *p*-values were below the adjusted significance-level. **Figures 3 and 4** show group-wise plots of the mean amplitudes recorded at Cz for each stimulus category. **Figure 5** shows the corresponding results of *post hoc t*-tests collapsed for both the speech and non-speech condition (i.e., speech and non-speech stimuli are treated as one class). In the non-musician group, separate *t*-tests revealed significantly



stronger N1 potentials elicited by voiced than by unvoiced stimuli [$T(12) = -3.480, p < 0.05$]. In contrast, neither significant differences were found within the musician group nor between groups. Moreover, N1 amplitudes elicited by speech were stronger in general than those elicited by non-speech [$T(25) = 5.446, p < 0.05$].

EEG DATA – TOPOGRAPHIC PATTERN ANALYSIS

Figure 6 shows the results of the topographical EP mapping of the grand-averaged data for each group and condition. In order to provide consistency with the examination of the classical N1 component, we first inspected visually the resulting template maps in and around the N1 time window (50–200 ms). Subsequently, we restricted our further analyses to the one template map, which expressed a typical N1-like auditory topography (map 6) and those which appeared immediately before and after that map (maps 3 and 7). Durations of these maps were then extracted and compared by means of a repeated measure ANOVA.

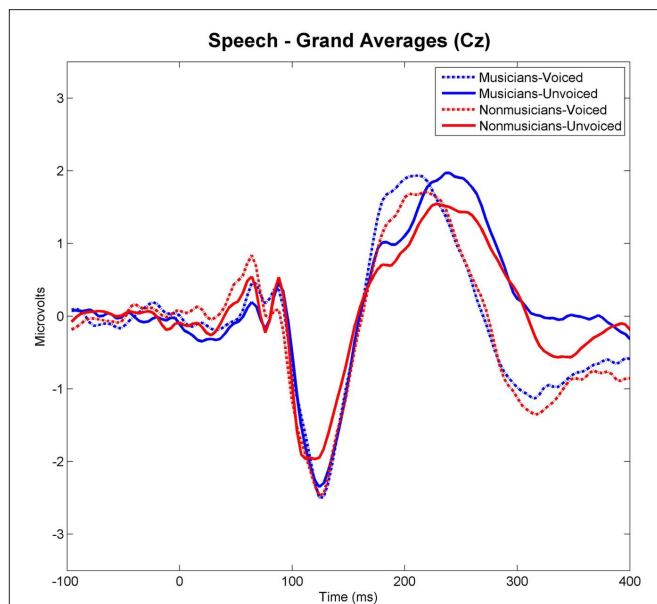


FIGURE 3 | Grand-averaged waveforms elicited by voiced and unvoiced speech stimuli. Time point “0” refers to the stimulus onset.

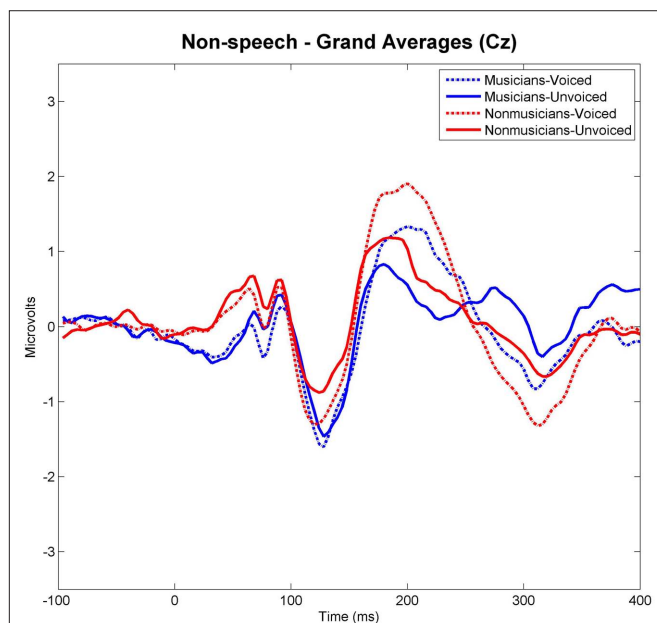


FIGURE 4 | Grand-averaged waveforms elicited by voiced and unvoiced non-speech stimuli. Time point “0” refers to the stimulus onset.

Results of the $2 \times 2 \times 2$ repeated measure ANOVA with between-subjects factor *group* and within-subjects factors *speechness* and *voiceness* showed significant main effects for the factor *speechness* regarding the durations of map 3 [$F(1, 24) = 10.046$, $p < 0.05$] and map 7 [$F(1, 24) = 5.139$, $p < 0.05$]. Furthermore, a significant *speechness by group* interaction was found for map 3 [$F(1, 24) = 6.684$, $p < 0.05$], whereas map 6 exhibited a significant *voiceness by group* interaction [$F(1, 24) = 5.933$, $p < 0.05$].

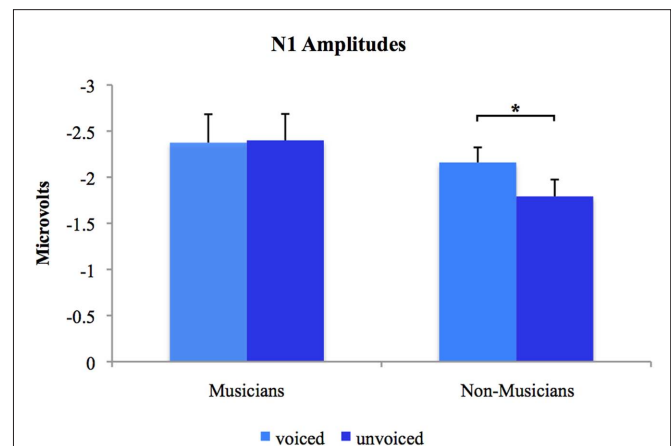


FIGURE 5 | Group-wise plots of N1 mean amplitudes for voiced and unvoiced stimuli. Speech and non-speech conditions are collapsed.

*Difference is significant at $p < 0.05$ [difference between stimulus type (i.e., “voiced” vs. “unvoiced”) irrespective of stimulus class (i.e., “speech” vs. “non-speech”) within the non-musician group, $T(12) = -3.480$]. Error bars indicate standard errors.

Subsequent *post hoc* Bonferroni–Holm adjusted *t*-tests revealed significant differences regarding the duration of map 3 between speech and non-speech stimuli [$T(12) = -3.582$, $p < 0.05$] within the musician group. In particular, map 3 was longer for musicians in the non-speech condition. Non-musicians on the other hand showed similar durations for map 3 in both conditions.

The topography of map 6 strongly resembles the typical auditory N1 topography. *Post hoc t*-tests for the durations of this microstate revealed significantly longer durations for voiced than for unvoiced stimuli within the non-musicians group [$T(12) = 2.533$, $p < 0.05$]. Within the musicians group, no significant difference in the duration of map 6 was found at all.

DISCUSSION

The purpose of this study was to examine the extent to which musicians and non-musicians process phonetic information differently. We were specifically interested in the early processing steps within the first 200 ms of stimulus onset during which the basic acoustic properties of acoustic stimuli are processed. Our hypothesis was that musical practice would have shaped the auditory system of musicians and that this use-dependent influence might also exert its influence on early phoneme perception. We anticipated therefore different neurophysiological activation patterns during early auditory processing stages.

For this study we used speech and non-speech stimuli with short (voiced) and long (voiceless) VOT. High-density EEG-recordings were obtained during the early acoustic processing stages and analyzed in two different ways: Firstly we calculated and analyzed conventional AEP components and focused on the N1 component, which is particularly important for auditory analysis (e.g., Sharma and Dorman, 1999; Steinschneider et al., 1999, 2005; Sharma et al., 2000; Zaehle et al., 2007). In addition to the conventional ERP analysis we also applied a topographic pattern analysis. With this method we exploit the entire spatial information of our high-density EEG

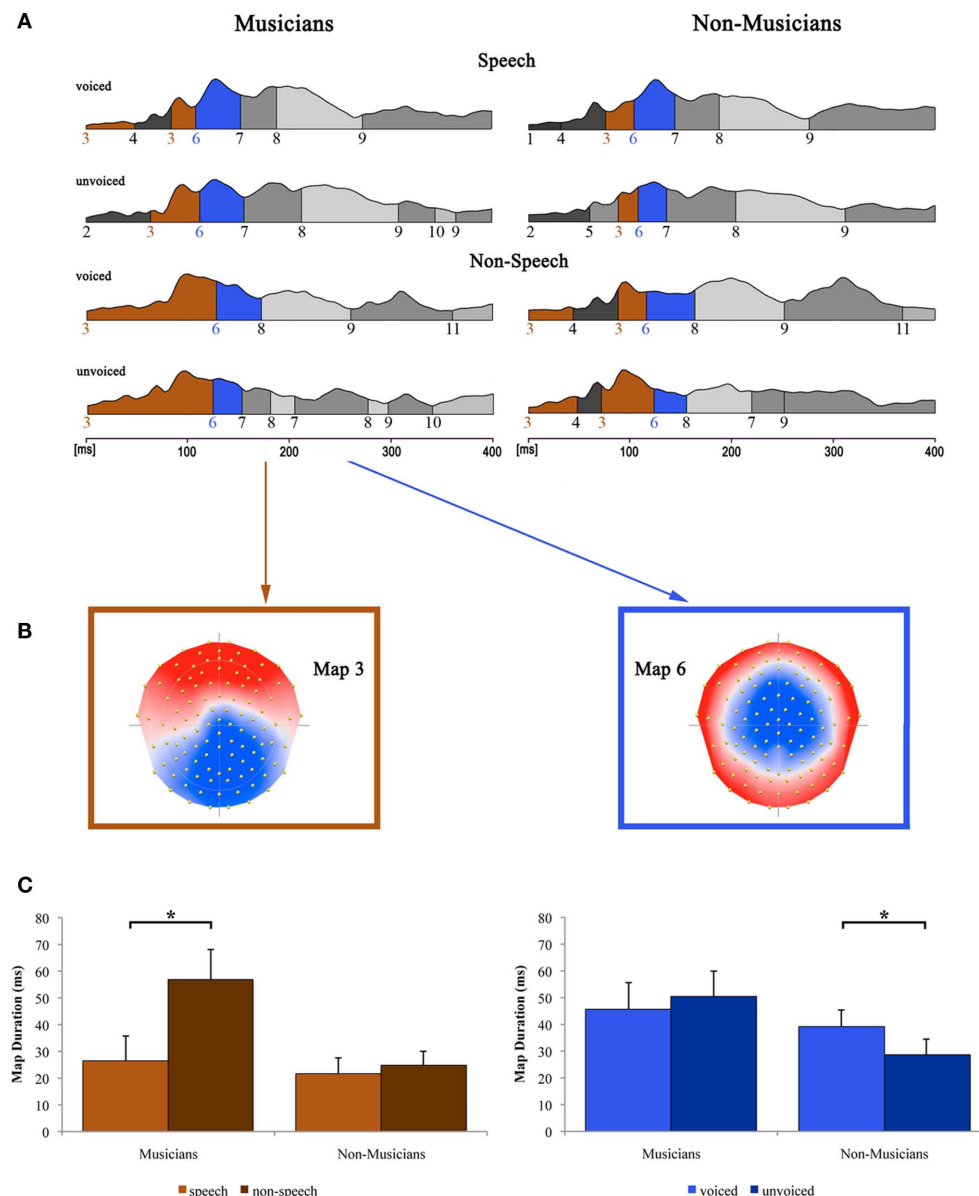


FIGURE 6 | Evoked potential data. (A) Segments of stable map topography are shown group-wise (left: musicians, right: non-musicians) for each condition under the global field power curve from 0 to 400 ms. Auditory N1-like template map 6 (blue) was found at ~100 ms and was significantly longer for voiced than for unvoiced stimuli within the non-musicians group whereas within the musicians group, durations were comparable. Template map 3 (brown) preceded map 6 and lasted significantly longer in the non-speech condition compared to the speech condition within the musicians group only. **(B)** Map topography and **(C)** duration of maps 3 (left) and 6 (right) for musicians and non-musicians in the respective conditions (voiced and unvoiced stimuli are

collapsed for map 3; speech and non-speech conditions are collapsed for map 6). *Difference is significant at $p < 0.05$ [left: Difference between stimulus class (i.e., “speech” vs. “non-speech”) irrespective of stimulus type (i.e., “voiced” vs. “unvoiced”) within the musician group, $T(12) = -3.582$; right: Difference between stimulus type irrespective of stimulus class within the non-musician group, $T(12) = 2.533$]. Note that the data shown in **(A,B)** are derived of grand-averaged waveforms, whereas map durations shown in **(C)** are obtained from individual ERPs (for a detailed description of the microstate analysis procedure see, e.g., Murray et al., 2008). Error bars indicate standard errors.

recording and reveal a more detailed activation pattern. Using this approach we focused on microstates around the N1 time window. Microstates are identified in an entirely data-driven approach, which substantially decreases the subjective influence on the data analysis. Thus, this method can be considered a more objective method to analyze evoked electrical responses (Michel et al., 2009).

As in various aforementioned studies (e.g., Sharma et al., 2000; Zaehle et al., 2007), the N1 amplitude is different to voiced and unvoiced acoustic stimuli in non-musicians. In particular, the N1 is larger for voiced than for unvoiced stimuli. Musicians on the other hand demonstrated large N1 amplitudes to all stimuli with no difference between voiced and unvoiced stimuli. Correspondingly,

for non-musicians the microstate durations around the N1 time window are significantly shorter for unvoiced stimuli irrespective of whether these stimuli are “speech” or “non-speech.” In addition, there is no difference between the microstate durations for musicians.

The findings with respect to the N1 amplitude and the corresponding microstate supports the idea that non-musicians process voiced and unvoiced stimuli differently and utilize different neurophysiological processes at least at processing stages around the N1 component. Simos et al. (1998) found the N1m peak equivalent current dipole (ECD) to be shifted toward more medial locations with increasing VOT. Due to the authors, this finding suggests that the peak of the N1m response elicited by stimuli with short VOT values may reflect activity from distinct subpopulations of auditory cortex neurons compared to the N1m response elicited by stimuli with long VOT values. That different neurophysiological processes are used to analyze voiced and unvoiced stimuli reflects the fact that there are qualitative differences between voiced and unvoiced phonemes. A pivotal feature in the perception of unvoiced phonemes is the relatively long VOT and the characteristic VOT differences between different unvoiced phonemes. For example, the VOT for/k/is much longer than the one for/p/. The VOT of voiced phonemes on the other hand is fairly short in general and much harder to resolve. Therefore, the processing of other acoustic features such as formant transitions and the specific composition of formants becomes more important in order to perceive and differentiate voiced phonemes, and spectral information is more relevant for this kind of analysis. Voiced phonemes are harder to detect in monitoring tasks (not in our study since the task used in this experiment was fairly easy) resulting in reduced hit rates and longer RT (Jäncke et al., 1992), thus it is possible that the processing of voiced stimuli requires more neurophysiological resources indicated by larger N1 amplitudes. Analysis of VOTs for perceiving unvoiced stimuli is a bit easier and thus does not need that much neurophysiological resource. A further possibility could be that the long VOT in unvoiced syllables provides the opportunity for some pre-processing after the first consonant, which then limits the quantity of processing needed after the occurrence of the following vowel. This preparation (or “warning effect”) elicited by the long VOT could explain the reduced N1 for unvoiced stimuli. If this argument is correct then musicians would show either less of this “preparation effect” or their auditory system would recover more quickly (or efficiently) after the “warning” initial consonant than in non-musicians. Finally, N1 amplitudes in response to unvoiced stimuli could also be reduced due to a second positive-going wave elicited by voicing onset that truncates the N1 evoked by consonant release (e.g., Sharma et al., 2000; Steinschneider et al., 2005).

Taken together, non-musicians utilize different processing modes at the time window of the N1 component for voiced and unvoiced stimuli.

The major finding of our study is that musicians process unvoiced stimuli differently than non-musicians. The N1 amplitude and the corresponding microstate are practically identical for voiced and unvoiced stimuli in musicians. Does that mean that musicians process voiced and unvoiced stimuli similarly? Although we are not in the position to answer this question unequivocally on the basis of our data, we believe that musicians indeed process

voiced and unvoiced stimuli at least partly similar. Due to their life-long musical practise to perceive, distinguish and memorize pitch and supra-segmental information, they are experts in this regard. In this context, the primary and secondary auditory cortex has been adapted anatomically and neurophysiologically to effectively analyze pitch and supra-segmental information (e.g., Pantev et al., 1998, 2003; Schneider et al., 2002; Baumann et al., 2008; Brancucci et al., 2009; Geiser et al., 2009). This expertise can even be identified when the musicians pre-attentively process pitch and supra-segmental information as has been shown in many studies using the mismatch negativity potential (e.g., Pantev et al., 2003; Herholz et al., 2008, 2009; Chandrasekaran et al., 2009; Nikjeh et al., 2009). Because of this preponderance, it might be that musicians are prone to use the neurophysiological processes applied for pitch-relevant information also to the analysis of unvoiced stimuli. Thus, they activate the pitch information analysis mode (*voiced analysis mode*) even during analyses of unvoiced stimuli. A further possibility could be that they activate the two analysis modes simultaneously when analyzing unvoiced stimuli, namely the *voiced* and the *unvoiced mode*. This would imply that musicians run a kind of multi-tasking analysis, which possibly provides more refined information about the spectro-temporal features of an auditory event. It might also be the case that highly trained musicians invest processing resources to a higher extent in the analysis of unvoiced stimuli, in order to ensure the analysis to be as precise as possible. Whether these explanations hold true has to be shown in future experiments.

Map 3, a microstate which immediately precedes the microstate within the N1 time window (here map 6) also revealed differences between musicians and non-musicians. This microstate lies in the P1 AEP time window (50–100 ms) and reflects early auditory neurophysiological processes. For this microstate musicians demonstrate longer durations during processing of non-speech stimuli irrespective of whether these stimuli are voiced or unvoiced. Thus, this neurophysiological process is longer active in musicians than in non-musicians. Even though our experiment was not primarily focused on this particular time window and we thus had no *a priori* hypothesis in this respect, we explain this finding as follows: We believe that in musicians, non-speech stimuli automatically evoke the *non-speech processing mode* at this very early processing stage. This processing mode is most likely specific for musicians because they have practised a lot to improve their capabilities to analyze speech-unspecific acoustic features (e.g., pitch, rhythm, intensity, metrum). Thus, they remain longer in the processing stage during which these basic acoustic features are analyzed than non-musicians.

Beside these neurophysiological differences there was also a slight advantage with respect to the RT for musicians. Though not significant, this was a consistent effect, especially for the speech stimuli. Whether this behavioral advantage is related to superior auditory processing or due to a more efficient auditory-motor coupling (see for example Bangert and Altenmüller, 2003) is difficult to infer. Future studies need to be conducted using different experimental designs in order to disentangle auditory processing from audio-motor coupling.

Most of the studies published so far which have examined differences between musicians and non-musicians with respect to basic auditory analyses have focused on supra-segmental acoustic features and pitch processing (e.g., Pantev et al., 1998, 2003; Schneider et al.,

2002; Magne et al., 2003; Moreno and Besson, 2005, 2006; Chartrand and Belin, 2006; Baumann et al., 2008; Brancucci et al., 2009; Geiser et al., 2009; Oechslin et al., 2009). Others examined short-term effects of musical training on brain plasticity and its behavioral outcome with respect to speech (e.g., Magne et al., 2006; Moreno et al., 2008; Draganova et al., 2009; Nikjeh et al., 2009). They clearly show that the auditory cortex is functionally altered in musicians compared to non-musicians, thus reflecting their exceptional expertise in encoding and analyzing auditory events. Moreover, structural and functional alterations can be seen already after a relatively short exposure to musical training. The present study complements these findings insofar as the results at hand show also different processing of voiced and unvoiced phonemes on a segmental level. This is reflected by different activation patterns between musicians and non-musicians with respect to N1 amplitudes and corresponding map durations.

It is however important to point out that it is not clear yet if and to what extent these differences might be driven by functionally altered processes at an even earlier level. Musacchia et al. (2007) clearly demonstrated that musical expertise affects the processing of speech as well as music at the level of the brainstem. In particular, they found stronger amplitudes in musicians to both CV-syllables and musical sounds in the transient and sustained portions of the brainstem response to auditory events. Several further studies examining the frequency-following response (FFR), which is interpreted as a marker of brainstem activity, have produced

results in line with these findings. For example, Bidelman et al. (2011) found stronger subcortical pitch encoding in musicians compared to non-musicians as reflected in an enhanced FFR in musicians for both speech and non-speech stimuli. It is therefore conceivable that the results of the present study regarding functional differences at the cortical level are influenced by or even solely based on an already altered afferent signal from the brainstem.

CONCLUSION

Our findings indicate that musical expertise transfers to basal speech processing on a very basic level of processing. Musicians utilize a similar neurophysiological processing mode to analyze voiced and unvoiced stimuli at the time window of the N1 component. Maybe they activate both modes (voiced and unvoiced analyses) simultaneously. At an earlier processing step preceding the voiced–unvoiced processing mode they strongly react to non-speech stimuli indicating a kind of automatic processing mode for non-speech stimuli. Although there are substantial neurophysiological differences between musicians and non-musicians, there are no strong behavioral differences. However, our results provide further support for the view that language and music share neural resources.

ACKNOWLEDGMENT

This study was supported by a grant of the Swiss National Science Foundation (SNF) to Lutz Jäncke (32003B_121927).

REFERENCES

- Annett, M. (1970). A classification of hand preference by association analysis. *Br. J. Psychol.* 61, 303–321.
- Bangert, M., and Altenmüller, E. O. (2003). Mapping perception to action in piano practice: a longitudinal DC-EEG study. *BMC Neurosci.* 4, 26. doi: 10.1186/1471-2202-4-26
- Baumann, S., Meyer, M., and Jäncke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *J. Cogn. Neurosci.* 20, 2238–2249.
- Besson, M., Schön, D., Moreno, S., Santos, A., and Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restor. Neurol. Neurosci.* 25, 399–410.
- Bidelman, G. M., 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.
- Brancucci, A., di Nuzzo, M., and Tommasi, L. (2009). Opposite hemispheric asymmetries for pitch identification in absolute pitch and non-absolute pitch musicians. *Neuropsychologia* 47, 2937–2941.
- Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain Lang.* 108, 1–9.
- Chartrand, J. P., and Belin, P. (2006). Superior voice timbre processing in musicians. *Neurosci. Lett.* 405, 164–167.
- Colombo, L., Deguchi, C., Boureux, M., Sarlo, M., and Besson, M. (2011). Detection of pitch violations depends upon the familiarity of intonational contour of sentences. *Cortex* 47, 557–568.
- Davis, M. H., and Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear. Res.* 229, 132–147.
- Draganova, R., Wollbrink, A., Schulz, M., Okamoto, H., and Pantev, C. (2009). Modulation of auditory evoked responses to spectral and temporal changes by behavioral discrimination training. *BMC Neurosci.* 10, 143. doi: 10.1186/1471-2202-10-143
- Geiser, E., Ziegler, E., Jäncke, L., and Meyer, M. (2009). Early electrophysiological correlates of meter and rhythm processing in music perception. *Cortex* 45, 93–102.
- Gordon, E. E. (1989). *Advanced Measures of Music Audiation*. Chicago, IL: GIA Publications, Inc.
- Gordon, R. L., Schön, D., Magne, C., Astesano, C., and Besson, M. (2010). Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PLoS ONE* 5e, 9889. doi: 10.1371/journal.pone.0009889
- Herholz, S. C., Lappe, C., Knief, A., and Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *Eur. J. Neurosci.* 28, 2352–2360.
- Herholz, S. C., Lappe, C., and Pantev, C. (2009). Looking for a pattern: an MEG study on the abstract mismatch negativity in musicians and non-musicians. *BMC Neurosci.* 10, 42. doi: 10.1186/1471-2202-10-42
- Hillenbrand, J., Canter, G. J., and Smith, B. L. (1990). Perception of intraphonemic differences by phoneticians, musicians, and inexperienced listeners. *J. Acoust. Soc. Am.* 88, 655–662.
- Jäncke, L. (2009). The plastic human brain. *Restor. Neurol. Neurosci.* 27, 521–538.
- Jäncke, L., Steinmetz, H., and Volkman, J. (1992). Dichotic listening: what does it measure? *Neuropsychologia* 30, 941–950.
- Jäncke, L., Wüstenberg, T., Scheich, H., and Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage* 15, 733–746.
- Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., and Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37, 163–178.
- Liégeois-Chauvel, C., de Graaf, J. B., Laguitton, V., and Chauvel, P. (1999). Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cereb. Cortex* 9, 484–496.
- Lisker, L., and Abramson, A. S. (1964). Across language study of voicing in initial stops: acoustical measurements. *Word* 20, 384–411.
- Magne, C., Schön, D., and Besson, M. (2003). Prosodic and melodic processing in adults and children: behavioral and electrophysiologic approaches. *Ann. N. Y. Acad. Sci.* 999, 461–476.
- 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.
- Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., and Besson, M. (2010a). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *J. Cogn. Neurosci.* doi: 10.1162/jocn.2010.21585. [Epub ahead of print].
- Marie, C., Kujala, T., and Besson, M. (2010b). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech

- sounds. *Cortex*. doi:10.1016/j.cortex.2010.11.006. [Epub ahead of print].
- Marques, C., Moreno, S., Castro, S. L., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463.
- Meyer, M., Baumann, S., and Jancke, L. (2006). Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. *Neuroimage* 32, 1510–1523.
- Meyer, M., Elmer, S., Baumann, S., and Jancke, L. (2007). Short-term plasticity in the auditory system: differential neural responses to perception and imagery of speech and music. *Restor. Neurol. Neurosci.* 25, 411–431.
- Michel, C. M., Th. Koenig, Brandeis, D., Gianotti, L. R. R., and Wackermann, J. (2009). “Electrical neuroimaging in the time domain - microstate analysis,” in *Electrical Neuroimaging*, Chap. 6, eds C. M., Michel, Th. Koenig, D. Brandeis, L. R. R. Gianotti, and J. Wackermann (Cambridge: Cambridge University Press), 123–125.
- Moreno, S., and Besson, M. (2005). Influence of musical training on pitch processing: event-related brain potential studies of adults and children. *Ann. N. Y. Acad. Sci.* 1060, 93–97.
- Moreno, S., and Besson, M. (2006). Musical training and language-related brain electrical activity in children. *Psychophysiology* 43, 287–291.
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S., and Besson, M. (2008). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex* 19, 712–723.
- Münte, T. F., Altenmüller, E., and Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nat. Rev. Neurosci.* 3, 473–478.
- Murray, M. M., Brunet, D., and Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr.* 20, 249–264.
- Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15894–15898.
- Näätänen, R., and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425.
- Nikjeh, D. A., Lister, J. J., and Frisch, S. A. (2009). Preattentive cortical-evoked responses to pure tones, harmonic tones, and speech: influence of music training. *Ear Hear.* 30, 432–446.
- Oechslin, M. S., Meyer, M., and Jäncke, L. (2009). Absolute pitch-functional evidence of speech-relevant auditory acuity. *Cereb. Cortex* 20, 447–455.
- Ostroff, J. M., Martin, B. A., and Boothroyd, A. (1998). Cortical evoked response to acoustic change within a syllable. *Ear Hear.* 19, 290–297.
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature* 392, 811–814.
- Pantev, C., Ross, B., Fujioka, T., Trainor, L. J., Schulte, M., and Schulz, M. (2003). Music and learning-induced cortical plasticity. *Ann. N. Y. Acad. Sci.* 999, 438–450.
- Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear Hear.* 30, 653–661.
- Pascual-Marqui, R. D., Michel, C. M., and Lehmann, D. (1995). Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Trans. Biomed. Eng.* 42, 658–665.
- Picton, T. W., Skinner, C. R., Champagne, S. C., Kellett, A. J., and Maiste, A. C. (1987). Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. *J. Acoust. Soc. Am.* 82, 165–178.
- Pratt, H., Starr, A., Michalewski, H. J., Bleich, N., and Mittelman, N. (2007). The N1 complex to gaps in noise: effects of preceding noise duration and intensity. *Clin. Neurophysiol.* 118, 1078–1087.
- Schlaug, G. (2001). The brain of musicians. A model for functional and structural adaptation. *Ann. N. Y. Acad. Sci.* 930, 281–299.
- Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* 5, 688–694.
- Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303–304.
- Sharma, A., and Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J. Acoust. Soc. Am.* 106, 1078–1083.
- Sharma, A., Marsh, C. M., and Dorman, M. F. (2000). Relationship between N1 evoked potential morphology and the perception of voicing. *J. Acoust. Soc. Am.* 108, 3030–3035.
- Simos, P. G., Diehl, R. L., Breier, J. I., Molis, M. R., Zouridakis, G., and Papanicolaou, A. C. (1998). MEG correlates of categorical perception of a voice onset time continuum in humans. *Cogn. Brain Res.* 7, 215–219.
- Steinschneider, M., Volkov, I. O., Fishman, Y. I., Oya, H., Arezzo, J. C., and Howard, M. A. (2005). Intracortical response in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cereb. Cortex* 15, 170–186.
- Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., and Howard, M. A. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J. Neurophysiol.* 82, 2346–2357.
- Thompson, W. F., Schellenberg, E. G., and Husain, G. (2003). Perceiving prosody in speech. effects of music lessons. *Ann. N. Y. Acad. Sci.* 999, 530–532.
- Trébuchon-Da Fonseca, A., Giraud, K., Badier, J., Chauvel, P., and Liégeois-Chauvel, C. (2005). Hemispheric lateralization of voice onset time (VOT) comparison between depth and scalp EEG recordings. *Neuroimage* 27, 1–14.
- Zaehle, T., Jancke, L., and Meyer, M. (2007). Electrical brain imaging evidences left auditory cortex involvement in speech and non-speech discrimination based on temporal features. *Behav. Brain Funct.* 3, 63.

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: 21 March 2011; accepted: 03 August 2011; published online: 06 September 2011.

Citation: Ott CGM, Langer N, Oechslin MS, Meyer M and Jäncke L (2011) Processing of voiced and unvoiced acoustic stimuli in musicians. *Front. Psychology* 2:195. doi: 10.3389/fpsyg.2011.00195
This article was submitted to *Frontiers in Auditory Cognitive Neuroscience*, a specialty of *Frontiers in Psychology*.
Copyright © 2011 Ott, Langer, Oechslin, Meyer and Jäncke. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



Tone language fluency impairs pitch discrimination

Isabelle Peretz*, Sébastien Nguyen and Stéphanie Cummings

International Laboratory for Brain, Music and Sound Research (BRAMS), Department of Psychology, University of Montreal, Montreal, QC, Canada

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Gottfried Schlaug, Beth Israel Deaconess Medical Center and Harvard Medical School, USA
Steven Brown, McMaster University, Canada

*Correspondence:

Isabelle Peretz, International Laboratory for Brain, Music and Sound Research (BRAMS), Université de Montréal, Pavillon 1420 Mont-Royal, Montreal, QC, Canada H2V 4P3.
e-mail: isabelle.peretz@umontreal.ca

Here we present evidence that native speakers of a tone language, in which pitch contributes to word meaning, are impaired in the discrimination of falling pitches in tone sequences, as compared to speakers of a non-tone language. Both groups were presented with monotonic and isochronous sequences of five tones (i.e., constant pitch and intertone interval). They were required to detect when the fourth tone was displaced in pitch or time. While speakers of a tone language performed more poorly in the detection of downward pitch changes, they did not differ from non-tone language speakers in their perception of upward pitch changes or in their perception of subtle time changes. Moreover, this impairment cannot be attributed to poor musical abilities since the impairment remains unchanged when individual differences in musical pitch-based processing is taken into account. Thus, the impairment appears highly specific and may reflect the influence of statistical regularities of tone languages.

Keywords: tone language, pitch perception, individual differences in musical abilities

INTRODUCTION

A fundamental question that is currently hotly debated is to what extent music processing shares perceptual mechanisms with language. One way to examine this question is to study transfer effects between musical and language abilities, as such a transfer may be mediated by shared mechanisms. For example, Slevc and Miyake (2006) showed that native Japanese speakers with high musical aptitude spoke English with better pronunciation than did their peers with less musical aptitude. Similarly, speakers of tone languages are better able to reproduce musical pitch patterns in singing than English speakers (Pfordresher and Brown, 2009). The latter domain-transfer effect from language to music suggests that tone language acquisition, that makes use of small pitch contrasts to convey meaning, fine-tunes pitch perception which in turn can be carried over to musical tasks. The objective of the present study was to re-examine this issue.

In the present study, we tested whether early exposure to a tone language may confer a sensory advantage in discriminating subtle pitch deviations in non-speech tone sequences. To this aim, we asked both speakers of a tone language, mostly Mandarin, and speakers of non-tone languages (mostly French), to detect pitch changes that were at or just above threshold. To assess the specificity of the transfer, we used a control condition in which the task was to detect equally subtle time changes. In both conditions, participants heard standard sequences of five piano tones presented monotonically (constant pitch) and isochronously [constant intertone interval (ITI)], as well as comparison sequences in which the fourth tone was displaced in pitch or time. Their task was to detect a change in the fourth tone of a sequence.

Musicians were excluded so as to control for possible transfer effects due to musical training rather than language experience. Indeed, it has been shown that musicians outperformed non-musicians in the processing of lexical tones (Wong et al., 2007; Lee and Hung, 2008). In such cases, the positive transfer effects observed between music and language abilities could be the consequence of

other factors than shared sensory processing. One likely mediating factor is that musicians have enhanced attentional or executive control capacities as compared to non-musicians (Bialystok and DePape, 2009). Yet, non-musicians might differ in musical capacity. In order to measure individual differences in musical abilities, all participants were tested with a standardized battery – the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003) – that was developed for Westerners but that showed similar sensitivity in speakers of a tone language (Nan et al., 2010).

MATERIALS AND METHODS

PARTICIPANTS

Twenty-four speakers of a tone language (referred to as tone speakers) and 25 speakers of a non-tone language (referred to as non-tone speakers) were tested in Montreal (see Table 1). The native language of tone speakers was Mandarin ($n = 18$), Vietnamese ($n = 4$), or Cantonese ($n = 2$); all had immigrated to Canada in adulthood and had spoken French and/or English as a second language for 6 years on average ($SD = 5$). Otherwise, tone and non-tone speakers were matched for age, education, duration of musical lessons, and handedness. None of the participants reported any hearing disorder and all were paid for their participation.

To screen for any musical deficit and to assess individual differences in musical abilities, all participants were tested on the MBEA, a validated and standardized collection of six tests developed to assess normal music perception and memory (Peretz et al., 2003). Each test contained 30 same – different classification trials in which a standard melody was followed by a comparison melody. For trials where the comparison melody differed from the standard, one note was altered in one of the following ways: it was out of the scale of the melody while preserving the contour (scale test), it changed the contour of the melody while preserving the scale (contour test), or it changed the intervallic distance between two successive notes while maintaining the contour and the scale (interval test). The rhythm test also used a same – different classification task with 30

trials. In half the trials, the duration of two adjacent notes was changed so that the rhythmic grouping of the melody was altered. For the metric test, half the 30 trials were written in duple meter and half in triple meter and the participant was asked to classify each trial as either a march or a waltz. Finally, recognition of melodies (memory test) was evaluated. In this last test, participants had to discriminate “old” melodies that they heard in the previous five tests from “new” melodies similar in structure to the old melodies but never heard before. The melodies were novel, written in the Western tonal musical system and presented on a piano timbre; they contained 10 notes on average.

The scores obtained on the MBEA revealed that one tone speaker performed 2 SD below the tone speakers’ average performance ($M = 84\%$; $SD = 7$). The data of this amusic participant were excluded. As can be seen in **Table 2**, the two groups obtained similar scores on the MBEA tests, except for the scale test on which tone speakers obtained lower scores than non-tone speakers. Nevertheless, there was no significant Group effect [$F(1, 47) = 1.66, p = 0.20$] nor interaction between Group and the MBEA tests [$F(4, 184) = 1.38, p = 0.23$].

MATERIAL AND PROCEDURE

The material and procedure were adapted from Hyde and Peretz (2004). Stimuli consisted of 21 different sequences, each containing five successive tones. In the standard sequence, all tones were

120 ms long, played at the pitch level of C6 (1047 Hz), and synthesized with a piano timbre; the ITI (onset to onset) was 350 ms. In the comparison sequences, the fourth tone was altered. In the 10 pitch-altered sequences, the fourth tone was displaced by one of five pitch distances upward or downward from C6; the five pitch distances used ranged from 1 Hz (1.56 cents) to 15 Hz (25 cents; 100 cents corresponds to 1 semitone; **Figure 1**). The 1 Hz distance was chosen so as to be at threshold at a reference frequency above 1000 Hz (e.g., Moore, 1973). In the 10 time-altered sequences, the fourth tone was displaced by one of five temporal increments earlier or later than its isochronous position; the five temporal increments ranged from 6 to 14% of the ITI.

Participants were tested individually, with separate sessions for the pitch and time tasks, presented in a counterbalanced order across subjects. In both tasks, they were asked to press a “yes” button when they detected a change and a “no” button when they were unable to detect a change. Trials were randomized, and half contained no change (i.e., the standard monotonic, isochronous sequence). Participants did not receive feedback but were informed about the nature of the possible change and the location where it could occur. They also received 20 practice trials with feedback after each trial. Each test session comprised 360 sequences (180 standard sequences, 18 of each of 10 altered sequences). The stimuli were presented bilaterally through Sennheiser HD450 headphones in a quiet room, at an intensity level of 70 dB SPL(A).

RESULTS

The percentage of hits (corresponding to a “yes” response when there was a change) minus false alarms (F.A., corresponding to a “yes” response when there was no change) was computed for each participant as a function of the type, direction, and size of change. It is worth noting that the groups did not differ in correct rejections (a “no” response to a trial containing no pitch change) with 68 and 73% for the tone and non-tone speakers, respectively, $t(47) = 1.28, p = 0.21$.

A 2 (group) \times 2 (direction) \times 5 (distance) mixed-design analysis of variance (ANOVA) was computed on the percentage of hits-F.A. obtained in the pitch task. As a three-way interaction

Table 1 | Participants’ characteristics.

	Tone speakers (<i>n</i> = 24)	Non-tone speakers (<i>n</i> = 25)	<i>t</i> -test ^a
Age (SD)	34 (7)	31 (11)	n.s.
Gender	13 F	15 F	–
Handedness	22 R	22 R	–
Education duration in years (SD)	16 (1)	17 (3)	n.s.
Musical training ^b (SD)	2.36 (1.26)	2.22 (1.45)	n.s.

F, female; R, right-handed.

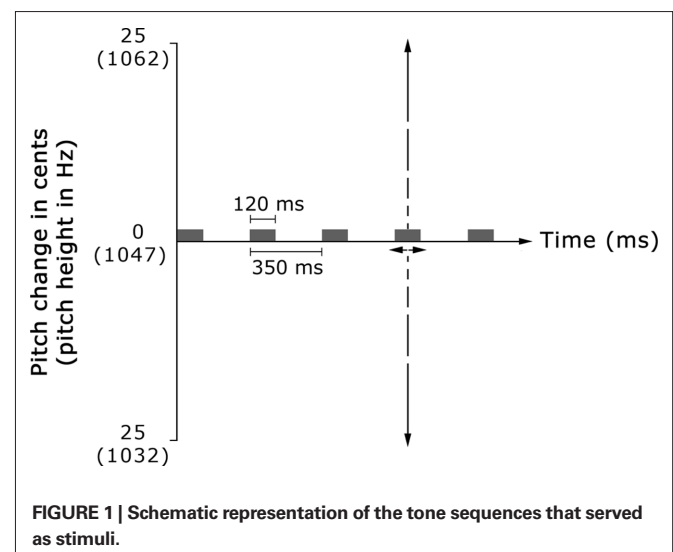
^aTwo-tailed.

^bBased on a scale of musical training with 1 = less than a year; 2 = 1–3 years; 3 = 4–6 years; 4 = 7–10 years; 5 = more than 10 years.

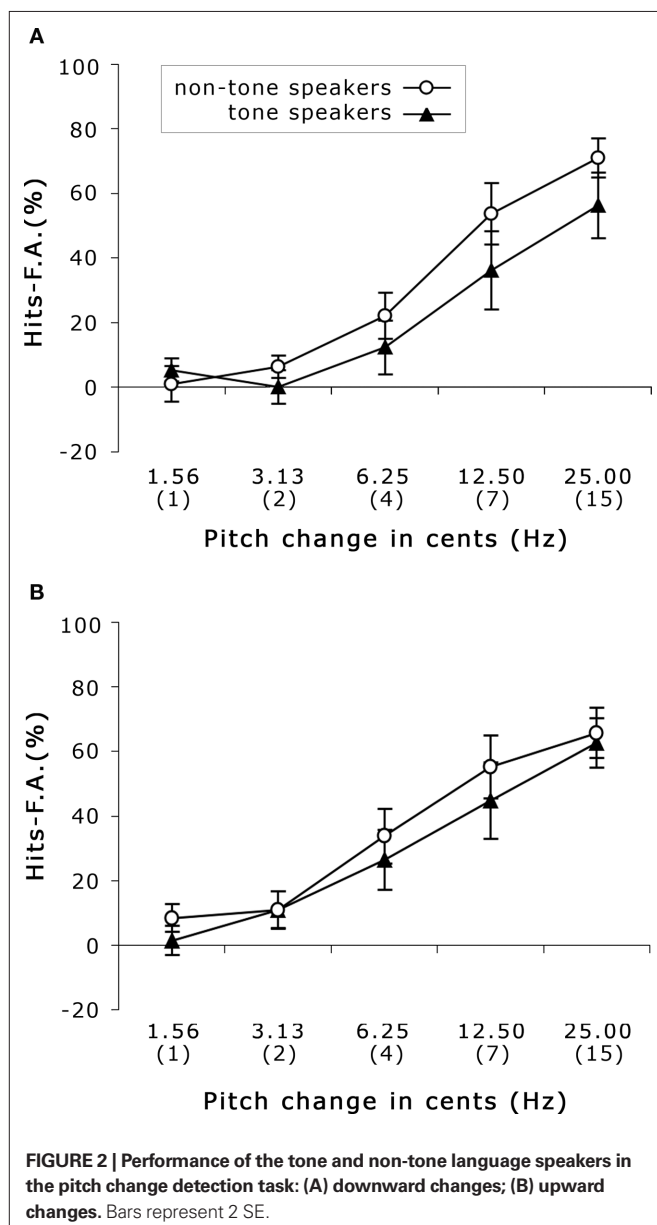
Table 2 | Percentage of correct responses (SD) obtained on the Montreal Battery of Evaluation of Amusia (MBEA) by speakers of a tone and non-tone language.

	Tone speakers (<i>n</i> = 24)	Non-tone speakers (<i>n</i> = 25)	<i>t</i> -test ^a
Scale	82 (8)	87 (10)	$p = 0.04$
Contour	82 (10)	87 (8)	n.s.
Interval	84 (8)	87 (9)	n.s.
Rhythm	90 (7)	88 (9)	n.s.
Meter	84 (17)	87 (11)	n.s.
Memory	89 (9)	89 (9)	n.s.
Global	85 (6)	87 (6)	n.s.

^aTwo-tailed.



between these factors was obtained, $F(4,188) = 4.65$, $p < 0.01$, separate ANOVAs for each direction were conducted, with Group as the between-subjects factor and Distance as the within-subjects factor. A significant two-way interaction emerged for downward pitch changes, $F(4,188) = 5.09$, $p < 0.01$. As can be seen in **Figure 2A**, tone speakers consistently performed below non-tone speakers in the detection of downward pitches, except at 1 Hz where groups did not differ, $t(47) = 1.32$, $p = 0.19$. At this smallest interval change distance, 17 tone speakers and 19 non-tone speakers performed at chance level. Statistically significant group differences emerged at 2 Hz, $t(47) = 2.05$, $p < 0.05$, and remained present at the largest pitch falls of 25 cents. In contrast, no interaction [$F(4,188) = 1.06$, $p = 0.38$] or main effect involving Group appeared for upward pitch changes [$F(1,48) = 1.67$, $p = 0.20$; **Figure 2B**].



In the time task, the ANOVA revealed no main effect or interaction involving Group ($F < 1.15$; p 's > 0.10), but a main effect of time distance, with $F(4,188) = 213.29$, $p < 0.001$ (**Figure 3**).

In order to assess whether the observed difference between the groups could be attributed to different experience of a tone language rather than to differences in musical aptitude, we assessed whether performance on each MBEA test could predict performance in the pitch change detection task (all conditions pooled together). As can be seen in **Table 3**, we found significant correlations between all pitch-related musical (Scale, Contour, and Interval) tests and performance in the pitch change detection task. It is worth pointing out that this correlation between fine-grained pitch perception and musical pitch discrimination is a general trend. It is not limited to the tone language speakers (**Figure 4**). Thus, group differences in musical aptitude are unlikely to account for the impairment observed for falling pitches in speakers of a tone language. To assess this statistically, the scores on the pitch-based musical tests were averaged and included as a covariate in a 2 (group) \times 2 (direction) \times 5 (distance) mixed-design ANCOVA. Essentially the same results as reported above were obtained, with a three-way interaction between all three

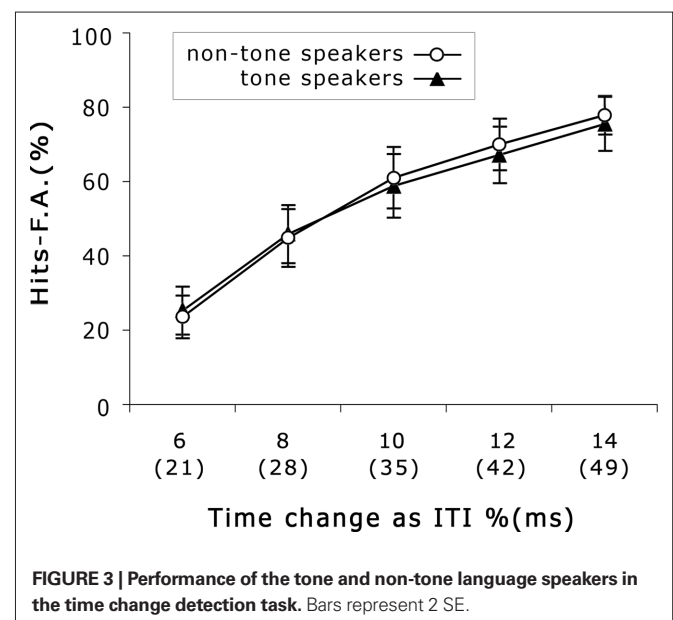
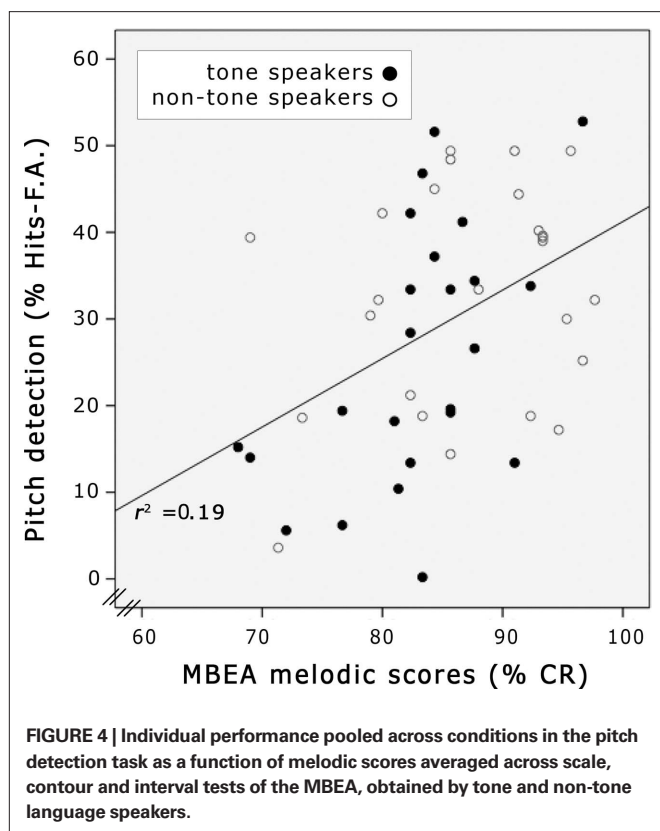


Table 3 | Bivariate correlations between performance on the pitch detection task (pooled across all conditions) and performance on each test of the MBEA.

MBEA	Pitch detection
Scale	0.299*
Contour	0.345*
Interval	0.490**
Rhythm	0.181
Meter	0.185
Memory	0.184

* $p < 0.05$; ** $p < 0.01$ (two-tailed).



factors, $F(4,184) = 5.94$, $p < 0.005$, and a significant two-way interaction for the falling pitch changes only, $F(4, 184) = 4.24$, $p < 0.005$.

DISCUSSION

Contrary to expectations, we find no evidence for enhanced perception of pitch in tone language speakers. On the contrary, we find an impairment that is limited to downward pitch changes. This impairment appears specific as tone language speakers do not differ from non-tone language speakers in their perception of upward pitch changes or in their perception of subtle time changes. Moreover, this impairment cannot be attributed to low musical aptitude since the impairment remains when individual differences in musical pitch-based processing are taken into account. Thus, fluency in a tone language interferes with the perception of falling pitches in a non-speech context.

This negative impact of tone language on pitch perception is not an isolated finding. Bent et al. (2006) found a similar pattern in the identification of pitch contours as rising, falling or flat. For the rising pitch contours, the Mandarin and English speakers performed similarly. In contrast, the tone language speakers misidentified the falling pitch contours more often than did the non-tone language

speakers, which is consistent with the direction-specific impairment reported here. This response pattern can be attributed to long-term exposure to Mandarin tones in which larger pitch ranges are typically used for falling tones than for rising tones (e.g., Xu, 1994). Although the pitch rises and falls are larger in naturally produced tones than the rises and falls of the stimuli used here and in Bent et al.'s (2006) study, the differences between rising and falling that characterizes lexical tones may have influenced the criteria used for non-speech stimuli presented at near-threshold levels. In sum, the statistical pitch regularities of tone languages may lead to the use of a language-specific perceptual strategy in a non-speech context.

Recourse to a language strategy in pitch discrimination is probably optional and promoted by the use of pitch deviations near threshold. As we have observed in two other studies (Tillmann et al., 2011; Tillmann et al., submitted), the more severe is a pitch perception deficit, the more beneficial is the speech context as compared to a non-speech context. More specifically, we found that fine-grained pitch discrimination was better in spoken syllables than in acoustically matched tones in amusic individuals while normal controls showed the reverse pattern (Tillmann et al., submitted). Similarly, in another study using lexical tones, we found that amusics with a severe pitch disorder performed better on syllables than on musical analogs (Tillmann et al., 2011). Thus, speech-specific processes can influence pitch perception when differences are difficult to hear. When pitch differences are large, there would be no need to use speech processes as a compensatory strategy. This may account for the fact we do not replicate the perception results obtained by Pfordresher and Brown's (2009). While we test pitch changes that are less than 25 cents apart, Pfordresher and Brown (2009) use distances from 25 to 800 cents and find both a tone language advantage and no pitch contour effect.

While we find no support for the notion that tone language experience fine-tunes pitch perception, we find evidence that individual differences in pitch perception carry over to music perception. In both tone and non-tone language speakers, performance in the pitch detection task predicts performance in the pitch-based melodic tests of the MBEA. Thus, individual differences in fine-grained pitch perception are linked to musical abilities independently of native language.

In sum, fine-tune pitch perception might be a prerequisite for achieving high levels of melodic abilities. In the advent of pitch discrimination difficulties, as in near-threshold discrimination or in the case of amusia, speech-specific processes may contribute to the pitch discrimination task albeit in a less effective manner than by a musical or acoustic mechanism.

ACKNOWLEDGMENTS

This research was supported by a summer fellowship from Natural Sciences and Engineering Research Council of Canada (NSERC) to Sébastien Nguyen, and by a grant from the same agency and a Canada Research Chair to Isabelle Peretz.

REFERENCES

- Bent, T., Bradlow, A. R., and Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 97–103.
- Bialystok, E., and DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 565–574.
- Hyde, K. L., and Peretz, I. (2004). Brains that are out of tune but in time. *Psychol. Sci.* 15, 356.
- Lee, C., and Hung, T. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *J. Acoust. Soc. Am.* 124, 3235–3248.
- Moore, B. C. J. (1973). Frequency difference limens for short-duration tones. *J. Acoust. Soc. Am.* 54, 610–619.
- Nan, Y., Sun, Y., and Peretz, I. (2010). Congenital amusia in speakers of a tone language: association with lexical tone agnosia. *Brain* 133, 2635–2642.
- 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.
- Pfordresher, P. Q., and Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attent. Percept. Psychophys.* 71, 1385–1398.
- Slevc, L. R., and Miyake, A. (2006). Individual differences in second-language proficiency. *Psychol. Sci.* 17, 675–681.
- Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., and Peretz, I. (2011). Amusia (or tone-deafness) interferes with tone language perception. *Front. Aud. Cogn. Neurosci.* (in press).
- 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.
- Xu, Y. (1994). Production and perception of coarticulated tones. *J. Acoust. Soc. Am.* 95, 2240–2253.
- 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: 13 February 2011; paper pending published: 28 April 2011; accepted: 16 June 2011; published online: 04 July 2011.*
- Citation: Peretz I, Nguyen S and Cummings S (2011) Tone language fluency impairs pitch discrimination. Front. Psychology 2:145. doi: 10.3389/fpsyg.2011.00145*
- This article was submitted to Frontiers in Auditory Cognitive Neuroscience, a specialty of Frontiers in Psychology.*
- Copyright © 2011 Peretz, Nguyen and Cummings. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.*



Music and language processing share behavioral and cerebral features

Clara E. James*

Faculty of Psychology and Educational Sciences, University of Geneva, Geneva Neuroscience Center, Geneva, Switzerland

*Correspondence: clara.james@unige.ch

A commentary on

Musical expertise and statistical learning of musical and linguistic structures

by Schön, D., and François, C. (2011). *Front. Psychology* 2:167. doi: 10.3389/fpsyg.2011.00167

Both language and music consist of discrete elements organized in embedded hierarchical structures. Schön and François nicely expose in this review that musical expertise facilitates learning of both linguistic and musical structures. At the behavioral level, the musicians did not outperform the non-musicians. However, ERP analyses showed that acquisition of boundary perception (segmentation) between units improved with musical training. The experimental strategy typically used to investigate segmentation relies in a learning phase on passive exposition to artificially constructed linguistic and musical material (cf. Figure 1). The authors plausibly argue, also based on a solid literature in this field, that such perceptual learning partially relies on statistics. The probability that a certain element is followed by another is different between and within units (words or tone sequences). In the test phase, participants should discriminate units from non-units. Statistical learning is by no means restricted to the auditory domain.

Why would musical expertise facilitate such learning in language? Musical and linguistic syntactical capacities seem correlated (Jentschke et al., 2008; Jentschke and Koelsch, 2009, also see the cited sources by Schön and François, p. 5). Moreover, brain substrates for language and music production and perception partially neighbor or overlap each other, although hemispheric dominances for music and language manifest (Zatorre, 2001; Koelsch et al., 2002; Brown et al., 2006). Schön and François observed similar ERP responses to linguistic and musical test-items (cf. Figure 4).

Shared cerebral networks and behavioral features involved in processing of complex sound suggest common roots. As already suggested by Darwin in his book “The Descent of Man and Selection in Relation to Sex,” a precursor or “proto language” may have preceded the emergence of separated musical and linguistic human capacities, explaining the observed brain and behavior commonalities. Vocal learning capacities possibly contributed to the “survival of the fittest.” We share akin vocal learning capacities with other higher order vertebrates (birds, whales, etc.), as shown in recent comparative research (Huron, 2001; Hauser and McDermott, 2003). More precisely, not vocal discrimination as such, but learning of vocal discrimination seems innate.

Now learning is synonymous with plasticity. We can become experts in very different domains, and behavior and brain adapt accordingly, comprising brain adaptations on the functional and the structural level (Maguire et al., 2000; Pascual-Leone, 2001; Brecht and Schmitz, 2008; James et al., 2008; Oechslin et al., 2009; Schlaug et al., 2009). In this context it is not surprising, as the authors also state, that experts in the musical domain show increased learning capacities for segmentation in both music and language. I would argue that trained musicians segment more efficiently not because their statistical learning is better, but because their discrimination and memory of complex sound is better, therefore allowing improved statistical learning as compared to non-musicians, also in a non-musical domain such as language.

In conclusion, joint examination of music and language constitutes a powerful means to gain further insight into the processing of highly structured complex sounds in language and music, and their shared behavioral and cerebral features. Schön and François provide us with compelling examples of such research.

REFERENCES

- Brecht, M., and Schmitz, D. (2008). Neuroscience. Rules of plasticity. *Science* 319, 39–40.
- 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.
- Hauser, M. D., and McDermott, J. (2003). The evolution of the music faculty: a comparative perspective. *Nat. Neurosci.* 6, 663–668.
- Huron, D. (2001). Is music an evolutionary adaptation? *Ann. N. Y. Acad. Sci.* 930, 43–61.
- James, C. E., Britz, J., Vuilleumier, P., Hauert, C. A., and Michel, C. M. (2008). Early neuronal responses in right limbic structures mediate harmony incongruity processing in musical experts. *Neuroimage* 42, 1597–1608.
- Jentschke, S., and Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *Neuroimage* 47, 735–744.
- Jentschke, S., Koelsch, S., Sallat, S., and Friederici, A. D. (2008). Children with specific language impairment also show impairment of music-syntactic processing. *J. Cogn. Neurosci.* 20, 1940–1951.
- Koelsch, S., Gunter, T. C., v Cramon, D. Y., Zysset, S., Lohmann, G., and Friederici, A. D. (2002). Bach speaks: a cortical “language-network” serves the processing of music. *Neuroimage* 17, 956–966.
- Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., and Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proc. Natl. Acad. Sci. U.S.A.* 97, 4398–4403.
- Oechslin, M. S., Imfeld, A., Loenneker, T., Meyer, M., and Jancke, L. (2009). The plasticity of the superior longitudinal fasciculus as a function of musical expertise: a diffusion tensor imaging study. *Front. Hum. Neurosci.* 3:76. doi: 10.3389/fpsyg.2012.00052
- Pascual-Leone, A. (2001). The brain that plays music and is changed by it. *Ann. N. Y. Acad. Sci.* 930, 315–329.
- Schlaug, G., Forgeard, M., Zhu, L., Norton, A., and Winner, E. (2009). Training-induced neuroplasticity in young children. *Ann. N. Y. Acad. Sci.* 1169, 205–208.
- Zatorre, R. J. (2001). Neural specializations for tonal processing. *Ann. N. Y. Acad. Sci.* 930, 193–210.

Received: 10 February 2012; accepted: 10 February 2012; published online: 28 February 2012.

Citation: James CE (2012) Music and language processing share behavioral and cerebral features. *Front. Psychology* 3:52. doi: 10.3389/fpsyg.2012.00052

This article was submitted to Frontiers in Auditory Cognitive Neuroscience, a specialty of Frontiers in Psychology.

Copyright © 2012 James. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.



Musical expertise and statistical learning of musical and linguistic structures

Daniele Schön* and Clément François

Institut de Neurosciences Cognitives de la Méditerranée, Marseille, France

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Christo Pantev, University of Muenster, Germany

Clara Eline James, University of Geneva, Switzerland

***Correspondence:**

Daniele Schön, Institut de Neurosciences Cognitives de la Méditerranée, 31 chemin J. Aiguier, 13402 Marseille, Cedex 20, France.
e-mail: schon@incm.cnrs-mrs.fr

Adults and infants can use the statistical properties of syllable sequences to extract words from continuous speech. Here we present a review of a series of electrophysiological studies investigating (1) Speech segmentation resulting from exposure to spoken and sung sequences (2) The extraction of linguistic versus musical information from a sung sequence (3) Differences between musicians and non-musicians in both linguistic and musical dimensions. The results show that segmentation is better after exposure to sung compared to spoken material and moreover, that linguistic structure is better learned than the musical structure when using sung material. In addition, musical expertise facilitates the learning of both linguistic and musical structures. Finally, an electrophysiological approach, which directly measures brain activity, appears to be more sensitive than a behavioral one.

Keywords: statistical learning, word segmentation, musical expertise, music and language processing

Language and music are both highly complex and articulated systems; both involve the combination of a small number of elements according to rules that allow the generation of unlimited numbers of utterances or musical phrases. Despite the complexity in language input, most infants acquire their mother tongue with astonishing ease and without formal instruction. This ability to learn new languages continues, though to a lesser extent, during adulthood. Musical skills, such as tapping along to a musical beat and singing, are also acquired without effort in early life. This similarity between speech and music acquisition, along with the extensive periods of musical training undergone by musicians allow us make several predictions. Firstly, that the learning of linguistic and musical structures may be similar. Secondly, that the learning of linguistic structures may be influenced by musical structures, and vice versa. And thirdly, that musical expertise may *transfer* to language learning.

STATISTICAL LEARNING AND SPEECH SEGMENTATION

In this paper we will focus on one of the first major difficulties in speech perception encountered by infants and second language learners: the ability to segment speech into separate units (words). Since the speech stream does not provide consistent acoustic cues to mark word boundaries, such as pauses or stresses, speech is more likely to be perceived as a continuous stream of sounds. The challenge for the learner, therefore, is to be able to segment this continuous speech stream into separate units, words, that will then be mapped onto conceptual representations.

In addition to prosodic cues the statistical structure of language is also an important cue for implicit speech segmentation (Saffran et al., 1996b; Kuhl, 2004; Rodríguez-Fornells et al., 2009). In general, “syllables that are part of the same word tend to follow one another predictably, whereas syllables that span word boundaries do not” (Saffran et al., 2001). The design typically used in this kind of investigation begins with a passive listening phase lasting few minutes. During this phase, participants listen to a synthetic and artificial language built in such a way that the only cues to segment words

are conditional probabilities between syllable pairs (high within words and lower across word boundaries). A two-alternative forced choice test follows in which participants must choose between two items based on which sounds the most familiar: one item is a word from the language while the other is not.

IMPLICIT LEARNING: SPECIFIC OR GENERAL?

Eight-month-old infants can learn to segment words in a simple artificial language in 2 min (Saffran et al., 1996a). Similar results have also been reported for newborns and adults (Aslin et al., 1998; Gervain et al., 2008; Teinonen et al., 2009), indicating a certain degree of independence from the developmental stage. *Statistical learning* has also been described in non-human primates and rats (Toro and Trobalón, 2005; Saffran et al., 2008), suggesting that it is not unique to humans. Statistical learning also seems to be « modality independent » insofar as several results have shown efficient learning using visual stimuli (Fiser and Aslin, 2005), but also visual movement or shape sequences (Fiser and Aslin, 2002) as well as sequences of tactile stimulations (Conway and Christiansen, 2005). Finally, learning takes also place when syllables are replaced with tones (Saffran et al., 1999). On a similar vein, Tillmann and McAdams (2004) replicated previous findings using complex non-verbal auditory material (sounds of different musical instruments). They showed that listeners are sensitive to statistical regularities despite acoustical characteristics (timbre) in the material that are supposed to affect grouping. Thus, although certain processing steps such as harmonic processing might be domain dependent (Peretz and Coltheart, 2003), these results reveal that this type of learning (i.e., statistical learning) is not specific to speech (and thus is not *modular*), but seems to be a more general-domain process.

PROSODY AND SEGMENTATION

When comparing music and language processing, it is difficult not to acknowledge the similarity between the *hierarchy of musical structures* and the one of prosodic structures in speech (Nespor

and Vogel, 1983). Prosody refers to the patterns of rhythm, stress, and intonation in speech. Prosodic variations are crucial cues for communicating emotions, disambiguating syntax, and also at a segmental level. Metrical cues play an important role in *word segmentation*: for instance, lexical stress typically implies increased duration, pitch, and loudness of specific syllables in a word, thus creating different perceptual rhythmic patterns (Abercrombie, 1967) which affect segmentation. These metrical cues constrain the ordering of stressed and unstressed syllables and are at the origin of the perception of strong and weak points in speech. A similar perception of strong and weak beats occurs in music. When prosodic cues are inserted, even subliminally, within the speech stream, performance on a speech segmentation task improves (Peña et al., 2002; Cunillera et al., 2006, 2008; De Diego Balaguer et al., 2007). An interesting example of how prosody can ease speech segmentation is the specific and exaggerated intonation that is used when talking to infants (infant-directed speech, ID speech). Compared to adult directed (AD) speech, ID speech is characterized by a slower rate, larger pitch and contour variations, and longer pauses (Fernald, 1992) and is easier to segment (Thiessen et al., 2005). Thus, if ID speech can be said to possess more exaggerated and more melodic prosodic features, a further step might be to consider how similar exaggerated and melodic features of singing might also ease speech segmentation.

BETTER LEARNING OF LINGUISTIC STRUCTURES IN SONG

Most of the studies investigating the relationship between music and language have focused on one domain or the other separately and this is also the case for statistical learning investigations. This is somewhat problematic insofar as a comparison of results issued from different tasks, subjects, and types of analyses is not straightforward. Therefore, singing is particularly well-suited to the study of the relation between language and music, the advantage being that both linguistic and musical information are merged into one acoustic signal with two salient dimensions, allowing for a direct comparison within the same experimental material. Moreover, this allows asking subjects to perform a task while manipulating the relation between linguistic and musical dimensions, thus studying the potential interferences from one dimension to the other. Of course, care must be taken because, in song, the phonological and metrical structures of language are strongly influenced by the type of melody that is sung. For instance, vowels sung in a very high register have the tendency to be less recognizable (Scotto di Carlo, 1994). We will review here three studies using a sung language.

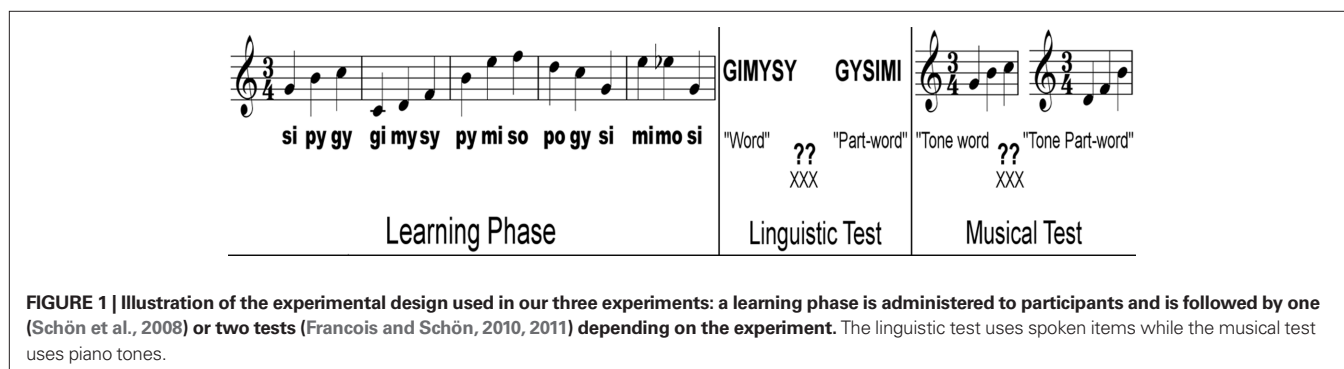
The design we used was very similar to that used by Saffran et al. (1999) Participants listened to a continuous stream of sung syllables lasting a few minutes. The stream was built by a pseudo-random concatenation of 6 (Experiments 1 and 2) or 5 (Experiment 3) trisyllabic words. Each syllable was sung at a constant pitch. This learning phase was followed by a two-alternative forced choice test wherein one item was a word from the language while the other was not (i.e., a part-word built by merging the end of a word with the beginning of another). Participants had to choose the most familiar of the two presented items.

In Experiment 1 (Schön et al., 2008), we compared segmentation of a spoken language versus segmentation of a sung language (see Figure 1). Two groups of participants listened to either 7 min of a flat contour artificial language or to 7 min of an artificial language with the same statistical structure wherein each syllable was sung on a specific pitch. The test, a two-alternative forced choice test, was identical for both groups and used spoken words only (flat contour).

While 7 min of listening were not enough to segment a language when it was spoken, they were sufficient to learn the same language when it was sung. These results suggest that redundancy in statistical musical and linguistic structures benefits the learning process, even if more general factors linked to motivation and arousal can not be excluded. Indeed, the sung version might well have been more « arousing » than the spoken version. To gain a finer understanding of the mechanisms responsible for these differences between spoken and sung languages we ran a further experiment using a sung language, wherein linguistic and musical statistical structures mismatched. More precisely, musical boundaries were shifted of one step to the right (i.e., one syllable later). Thus, while the statistical of linguistic and musical structures were preserved, word and pitch boundaries did not occur at the same time anymore. Performances laid in between the spoken version and the original sung version obtained in the two first experiments. These results point to a beneficial effect of both structural (“statistical”) and motivational properties of music in the very first steps of language acquisition (Thiessen et al., 2005; Schön et al., 2008).

COMPARING LEARNING OF LINGUISTIC AND MUSICAL STRUCTURES

The test used in Experiment 1 only contained spoken items. Thus, we could not know whether learning of the musical structure took place. At this aim and in order to better understand how musical information facilitated speech segmentation, we designed a similar experi-



ment that tested both learning of musical and linguistic structures contained in the sung language, and also recorded event-related potentials during the tests (ERPs). Thus, two tests followed the learning phase, one using spoken items at a fixed pitch, and another using piano tone sequences. First, we ran a study with a random sample (François and Schön, 2010, henceforth Experiment 2), then we compared a group of musicians and non-musicians (François and Schön, 2011, henceforth Experiment 3). At the behavioral level, results showed that linguistic structures were better learned than musical ones (though significance was reached only when pooling the data across the two studies). Comparison of performance with chance level (0.5) showed that while the participants' level of performance was above chance in the test using spoken items, it was not in the test using tones. Indeed, as can be seen on **Figure 2**, the distribution of the performances in the musical test were highly heterogeneous across participants and only half of the subjects exhibited performances higher than chance level. Several non-exclusive explanations can account for this difference.

First, linguistic storage and/or retrieval of “lexical” items might have been more efficient or less sensitive to interferences caused by part-words/tone-part-words. Schendel and Palmer (2007) recently showed that, overall, verbal items were better recalled than musical items. Moreover, they found a larger interference effect on musical recall than on verbal recall. Researches investigating memory representations of song have also shown an advantage for words over melodies (Serafine et al., 1984; Crowder et al., 1990; Morrongiello

and Roes, 1990; Hébert and Peretz, 2001; Peretz et al., 2004). Altogether, these results seem to suggest that linguistic information may be more resistant to interferences than musical information.

The second explanation is more stimulus related. Indeed, while the statistical structure of melodic and phonologic information was the same in our material, other important factors may come into play that make melodic segmentation more difficult than word segmentation. For instance, phoneme discrimination is mostly influenced by transient information in the specter (e.g., “py” versus “gy”), while pitch discrimination is influenced by more stationary information (e.g., fundamental frequency). Moreover, phoneme transitions in our language were probably sensitive to the implicit phonotactic knowledge humans have in their mother tongue, dictating which sound sequences are allowed to occur within the words of a language (Friederici and Wessels, 1993). This is hardly transposable to music, wherein the implicit knowledge affecting music perception concerns tonal relationships between notes (Tillmann et al., 2000). Overall, it is very difficult to control and balance the perceptual saliency of the linguistic and musical structures and care must be taken in comparing and interpreting results across dimensions.

Nonetheless, when pooling the data across the two studies (Experiments 2 and 3), we observe a significant positive correlation between performances in music and linguistic tests: the higher the performance in the linguistic test, the higher in the music test (**Figure 3**). This fits well with results showing a correlation between linguistic and musical abilities. For instance, several studies report a positive correlation between phonological awareness and music perception (Butzlaff, 2000; Anvari et al., 2002; Foxton et al., 2003), suggesting that these competences may share some of the same auditory mechanisms. Interestingly, phonological awareness requires the ability to segment speech into its component sounds and to extract the phonological invariants. Similarly, the perception of music also requires the listener to be able to segment the stream of tones into relevant units and to be able to recognize these units when played with different timbers tempos, keys, and styles.

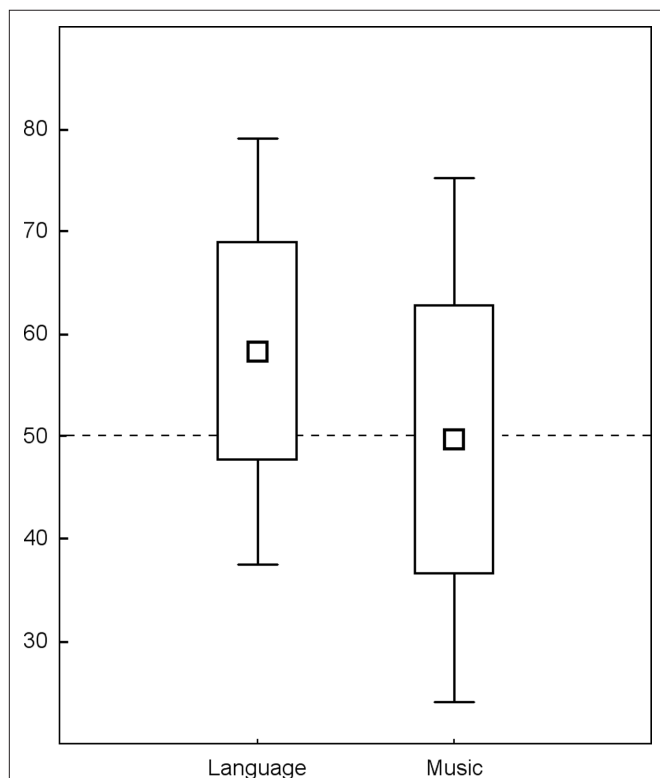


FIGURE 2 | Percentage of correct responses: box plot of performances in the linguistic (left) and musical tests (right). Data from Experiments 2 and 3 (François and Schön, 2010, 2011), are pooled together leading to $N = 50$ subjects (dashed line = 50%, chance level).

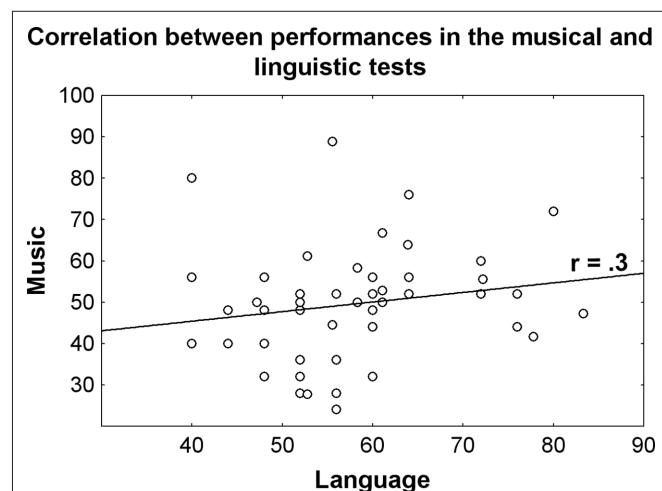


FIGURE 3 | Correlation between performances in the linguistic and musical tests for 50 participants (François and Schön, 2010, 2011). The correlation is significant ($p = 0.05$), that is the better participants performance in the musical test, the better are performances in the linguistic test.

STATISTICAL LEARNING: COMPARING ERPS AND BEHAVIOR

Overall, behavioral data alone could not reveal whether participants could or could not segment the musical structure (Experiments 2 and 3). Indeed, the lack of significance might simply be related to a lack in sensitivity of the testing procedure and of the dependent variable. ERPs seem to be more sensitive to the subtle mechanisms underlying implicit learning. Indeed, McLaughlin et al. (2004), in a study focusing on second language learning, showed that ERPs can be more sensitive in reflecting implicit learning than do explicit categorical judgments and that behavioral assessment might underestimate the extent to which learning has taken place. While behavior reflects the combined effects of several processing stages, ERPs, thanks to their continuous time resolution, can (sometimes) show differences in brain activity even when this is not backed by an overt behavioral response. Several examples of ERP-behavior dissociations have been reported pointing the greater ERP sensitivity, for instance in the field of subliminal perception (Sergent et al., 2005) and learning (Tremblay et al., 1998).

In our studies (Experiments 2 and 3), ERPs were also more sensitive than behavior in that they revealed a significant difference between familiar and unfamiliar items in both linguistic and musical test. Indeed, even subjects with a performance that could be hardly interpreted as higher than chance (e.g., subjects with 55% of correct responses), did show different ERPs to words and part-words. These results seem to point to a greater sensitivity of the ERP data compared to the behavioral data in implicit learning designs.

LEXICALITY, FAMILIARITY, AND THE N400

Electrophysiological data of the linguistic test showed a negative component at fronto-central sites in the 450- to 800-ms latency band which was larger for part-words than for words (see Figure 4). In order to interpret this ERP difference it is important to clarify the respective “linguistic” status of the two conditions. What

researchers in field of statistical learning typically call “words” are in fact nonsense words that are repeated in a random order for N minutes during the learning phase. Therefore, “words” do not have an entry in the lexicon. However, because they are repeated so many times (typically >100 each), they “pop out” of the continuous stream of syllables, due to statistical learning or associative mechanisms. Thus, to a certain extent, “words” become familiar, and as such they can be distinguished from “part-words” acquiring a kind of proto-lexical status (Fernandes et al., 2009; Rodríguez-Fornells et al., 2009). Indeed, in an experiment using X words, if, during the learning phase, “words” are heard N times each, “part-words” are heard $N/(X - 1)$ times (e.g., 216 and 44 times, respectively). This is because “part-words” are built by merging the end of a word with the beginning of another (and each word can be followed by $X - 1$ possible words). Therefore, the linguistic status of the items used in the test is that of pseudo-words which are more or less familiar (words and part-words, respectively).

The most common use of pseudo-words in the ERP literature is found in studies concerned with lexicality. These studies most often compare a pseudo-word to a word condition. For instance, O’Rourke and Holcomb (2002) used a lexical decision task and found a larger negativity for pseudo-words than for words, most prominent at anterior sites. Moreover, the negative component to pseudo-words peaked around 600 ms, 200 ms later than the classical N400 effect described for words. The authors claimed that pseudo-word processing may undergo an additional time consuming step, consisting of a top-down checking process. Interestingly, in our studies, the negative component to both familiar and unfamiliar items was also peaking around 600 ms.

Turning to familiarity effects, previous ERP studies have shown that the amplitude of the N400 is sensitive to familiarity. Indeed, low frequency words elicited a larger negativity peaking around 400 ms compared to high frequency words (Van Petten and Kutas,

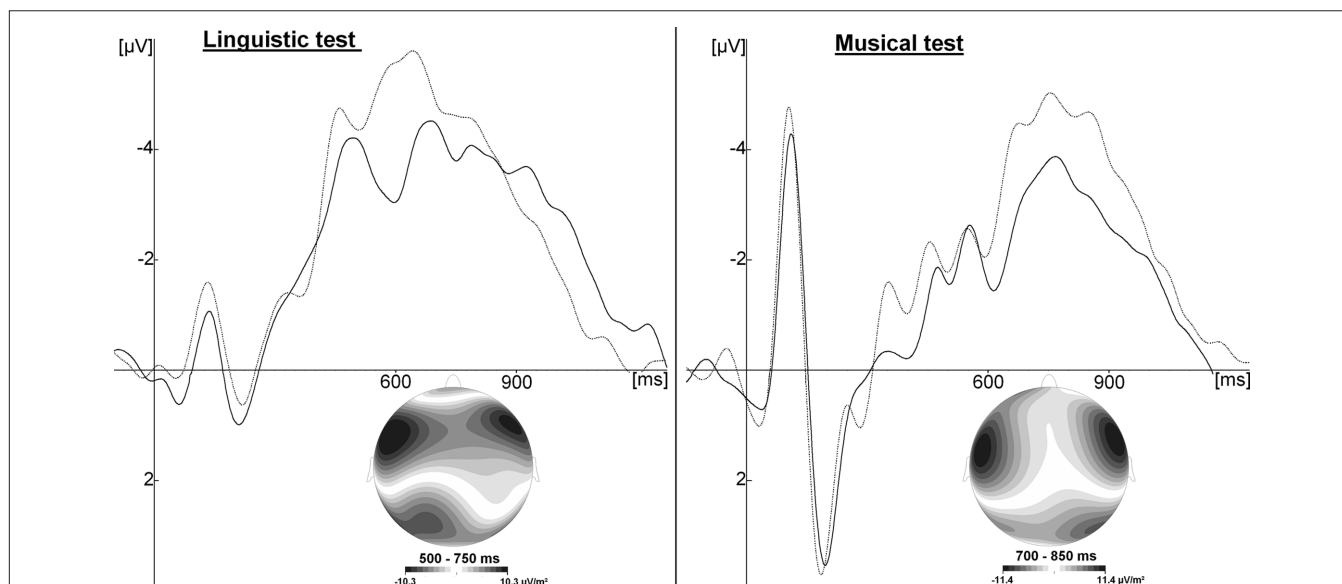


FIGURE 4 | Grand average across two experiments (François and Schön, 2010, 2011; $N = 50$), for the linguistic (left), and musical test (right) at Cz electrode. Solid lines illustrate ERPs to familiar words/tone-words; dashed lines illustrate ERPs to unfamiliar part-words/tone-part-words. Current Source Density Maps illustrate the topographic distribution of the familiarity effect in the significant time windows.

1990; Young and Rugg, 1992). Thus, the N400 amplitude seems to be sensitive to the ease of accessing the information from long-term memory (Federmeier and Kutas, 2000). In statistical learning experiments, during the passive listening phase, only sufficiently reinforced items may “survive” concurrent interference and become stored in a long lasting manner (Perruchet and Vinter, 2002). Therefore, the observed differences between words (familiar) and pseudo-words (unfamiliar) in the negative component, might reflect the difficulty of accessing and/or retrieving item representations.

Electrophysiological data of the musical test revealed a similar ERP component whose amplitude was larger for unfamiliar than for familiar items: a fronto-central negative component peaking around 900 ms after stimulus onset (Experiments 2 and 3).

In the studies described, insofar as the sung language was built by concatenating trisyllabic words, participants were (implicitly) familiarized with the three-pitch melodies corresponding to the melodic contour of the trisyllabic words. Therefore the same issues on lexical representation and familiarity described for linguistic items also apply when testing with musical sequences. Interestingly, in both Experiments 2 and 3, we found a fronto-central negative component larger for unfamiliar than familiar musical items, with a topographic distribution quite similar to the one observed in the linguistic test for the same contrast (unfamiliar – familiar, see **Figure 4**). However, although sensitive to the same factor (familiarity) and with a similar topography, this negative component peaked at around 600 ms to linguistic items and at 900 ms to musical items, possibly due to the fact that participants found the musical test harder than the linguistic test. One may interpret these differences in latencies in terms of functional differences, that is, different underlying processes. We favor the interpretation of these differences in terms of difficulty in item retrieval: participants probably needed more time to “search” whether an item was familiar or not in the musical test than in the linguistic test. Although less commonly described in the literature, late negative components have been observed in response to musical stimuli following for instance, unexpected harmonies (Steinbeis and Koelsch, 2008) but also following musical excerpts that do not match the concept of a previous verbal context (Daltrozzo and Schön, 2009) or at the familiarity emergence point of a melody (Daltrozzo et al., 2010).

MUSICAL EXPERTISE AFFECTS BEHAVIORAL AND BRAIN INDICES

Comparing experts to non-experts in a given domain is an elegant way to study the effect of extensive training on brain plasticity. The musicians’ brain has long been considered a model of plasticity (Münste et al., 2002). Thus, comparing musicians to non-musicians allows the effects of extensive audio-motor training on the functional and structural organization of the brain to be studied (even if causation is only showed using a longitudinal design). At a perceptual level, it has been shown that musicians outperform non-musicians on a variety of music related task: musicians have a lower frequency threshold (Spiegel and Watson, 1984; Kishon-Rabin et al., 2001; Micheyl et al., 2006) and exhibit finer rhythmic processing than non-musicians (Yee et al., 1994; Jones et al., 1995). Compared to non-musicians, musicians also show an enhanced cortical attentive and pre-attentive processing of linguistic and musical features as reflected by larger amplitude and/or shorter

latencies of many ERP components such as the N1, P2, MMN, P300 (Pantev et al., 1998; Koelsch et al., 1999; Shahin et al., 2003; Van Zuijen et al., 2005). In our study on statistical learning (Experiment 3) the N1 component was larger for musicians than non-musicians in both the linguistic and the musical tasks. Interestingly, recent results support the view that increased auditory evoked potentials in musicians (N1–P2) reflect an enlarged neuronal representation for specific sound features of these tones rather than selective attention biases (Baumann et al., 2008). Even at the sub-cortical level, musicians show more robust encoding of linguistic and musical features as reflected by earlier and larger brainstem responses compared to non-musicians (Kraus and Chandrasekaran, 2010). Overall, these differences can be interpreted as reflecting a greater efficiency of musicians’ auditory system in processing sound features and can be accompanied by morphological differences, showing that musicians have a larger gray matter concentration in the auditory cortex (Bermudez and Zatorre, 2005), an increased gray matter density and volume in the left inferior frontal gyrus, Broca’s area (Sluming et al., 2002) and a larger planum temporale (Schlaug et al., 1995; Keenan et al., 2001).

The fact that musicians perceive some sound features more accurately than non-musicians is not so surprising. After all, they spent hours and hours of their life focusing on sounds and the way they are generated, paying close attention to pitch, timber, duration, and timing. It would be rather surprising and even deceiving if this did not affect the way they hear, and as a consequence their brain functioning. What seems to us less evident is that such an intensive musical practice also seems to affect non-musical abilities. For example, both adult and child musicians perform better than matched controls when asked to detect fine contour modifications in the prosody of an utterance (Schön et al., 2004; Magne et al., 2006). Adult musicians also have better performances and larger ERP components to metric incongruities at the end of an utterance (Marie et al., 2011). There is also evidence for a possible correlation between musical and linguistic aptitudes in both children (Anvari et al., 2002; Milovanov et al., 2008, 2009) and adults (Foxton et al., 2003; Slevc and Miyake, 2006) as well as a benefit of music training on linguistic skills (Butzlaff, 2000; Overy, 2003; Gaab et al., 2005; Tallal and Gaab, 2006; Forgeard et al., 2008; Moreno et al., 2009; Parbery-Clark et al., 2009).

BETTER SEGMENTATION OF THE MELODIC STRUCTURE IN MUSICIANS

When using a sung language in artificial language learning, one needs to choose a series of pitches that are mapped onto syllables. The choice of pitches is an important one as it determines the tonal or atonal character of the stream. Although one may guess that a tonal structure will benefit language learning more than an atonal one, further work is needed to understand how and to what extent the tonality of the musical structure influences learning of the phonological one. In the experiments described here we used a musical structure with a rather strong tonal center (10 notes out of 11 were in C major). Thus, it is possible that participants keep in memory a representation of a tonal center during the post-learning test, which influences the processing of the musical items rather early on. In fact, once a tonal center is established, each individual pitch can be processed and categorized relative to that tonal center, even in the absence of absolute pitch abilities. Indeed, musicians, in the musical

test, showed a significant effect of familiarity 200 ms after the first tone onset and before the beginning of the second tone on the P2 component (larger to unfamiliar than familiar items), possibly due to a more difficult categorization of the unfamiliar than familiar items (Liebenthal et al., 2010). Thus, musicians were sensitive to the tonal structure of the language, which in turn influenced their perception of the first tone of the items in the test, despite poor behavioral performance.

This benefit of musical expertise in segmenting the musical structure was also evident in later ERP components. For instance, compared to familiar items, unfamiliar items yielded a larger mismatch negativity (MMN) in response to the second tone of the melody (Experiment 3). Overall, both familiarity effects on the P2 and MMN-like components suggest that musicians did indeed learn the musical structure better than non-musicians and this in turn affected the way they processed the musical items presented during the test. This is in line with previous findings showing that implicit learning of 12-tone music is influenced by expertise. Indeed, only participants with routine exposure to atonal music do (implicitly) perceive the distinction between different types of transforms (Dienes and Longuet-Higgins, 2004).

MUSICAL TRAINING FACILITATES IMPLICIT LEARNING OF BOTH LINGUISTIC AND MUSICAL STRUCTURES

The most striking finding of the series of experiments reviewed here is that compared to non-musicians, musicians seem to have more “robust” representations of both musical and linguistic structures that have been shaped during the listening phase. Indeed, musicians showed a larger familiarity N400-like effect than non-musicians, in both dimensions (Experiment 3).

We interpret these findings as evidence that musical expertise facilitates regularity extractions and sequence learning in general (Janata and Grafton, 2003). For instance, it is known that musicians can organize a sound sequence according to number regularity, implicitly distinguishing segments containing four tones from the segments containing five tones. By contrast, such a perceptual organization of sound in terms of number is less relevant for non-musicians (Van Zuijen et al., 2005). Moreover, recent findings showed that deaf children with cochlear implants are impaired in visual sequence learning, suggesting that a period of auditory deprivation may have a major impact on cognitive processes that are not specific to the auditory modality (Conway et al., 2010). Therefore, sound seems to provide a cognitive scaffolding for the development of serial-order behavior: whether sound processing is impaired or whether it is extensively practiced have opposite effects. Interestingly, Sluming et al. (2002) reported an increased gray matter density and volume in the left inferior frontal gyrus, Broca’s area. Moreover, neuroplastic development throughout a musicians’ life seems to promote the retention of cortical tissue (Sluming et al., 2002). Broca’s area is known to be involved in on-line speech and tone stream segmentation as well as in music harmonic perception (Tillmann et al., 2003, 2006; McNealy et al., 2006; Abia and Okanoya, 2008; Cunillera et al., 2009). Overall our results support an “auditory scaffolding hypothesis” (Conway et al., 2009) as we presently show that increased exposure to sounds leads to a benefit for implicit learning, putatively via anatomical and/or functional modifications going beyond the auditory regions.

CONCLUSION

Based on the similarity between music and speech acquisition we made several predictions. Firstly, we predicted that the learning of linguistic and musical structures may be similar. We found that ERP responses to linguistic and musical test-items present a similar (although not identical) morphology and topography (Experiments 2 and 3). While this cannot be interpreted in terms of a similar learning process, it seems that the linguistic and musical representations that are a consequence of the learning process are similar. This further supports previous findings showing that statistical learning can apply to non-linguistic stimuli. Secondly, we predicted that speech segmentation may be influenced by musical structure. Results of Experiment 1 directly address and confirm this hypothesis. By clarifying the structure of the language and by increasing the motivation/arousal of participants, music seems to facilitate speech segmentation. Finally, we predicted that musical expertise may transfer to language learning. Experiment 3 addressed and confirmed this prediction, by showing that musicians have more robust representations of both musical and linguistic structures, possibly via a more efficient brain network involving both auditory and more integrative processing.

FUTURE QUESTIONS

LEARNING PROCESS VERSUS LEARNING OUTCOME

Most studies on word segmentation and language learning rely on behavioral measures. These are typically obtained using the head-turning procedure in infants and a familiarity two-alternative forced choice procedure in adults. These procedures are used after the listening phase, in which learning takes place. Thus, these procedures test the result of learning rather than the learning process itself. The data presented here (Experiments 2 and 3) suffer from the same limitation. Researchers in the field of implicit learning have more recently realized the importance of better defining the status of the representations derived from the learning processes and promote the combined use of both subjective behavioral measures (guessing criterion and zero correlation criterion) and EEG or fMRI measures (Seth et al., 2008). In this sense, studies measuring EEG during learning are promising, pointing to several possible electrophysiological indices, ERPs, or oscillations that may reflect how learning takes place over time (Rodríguez-Fornells et al., 2009; De Diego Balaguer et al., 2011). The challenge for the coming years is to access to a temporally detailed learning curve. Ideally, one would want to know how each word presented during the learning has been processed. As some studies seem to point out, learning curves may differ across individuals (Abia et al., 2008). This might be crucial in studying pathology, since similar behavioral results (null effects) may rely upon different learning curves.

EXTENDING MUSIC AND LANGUAGE COMPARISONS

The literature comparing music and language processing (and also in studies on language learning) is mostly focused on the perceptual side, with production having received little attention. While the statistical learning framework is a brilliant way of studying language learning in a laboratory, it lacks by construction the rich context and complex behaviors that are typically found in language acquisition. For instance the natural context typically implies a word to world mapping: once a unit/word is identified, it is mapped onto one or several concepts. Moreover, learning

acquisition outside the laboratory relies on coupling both listening and imitation. Thus, further efforts need to be done to elaborate new experimental paradigms comparing language and music using a richer learning context as well as linguistic and musical unit production. Also, while the direct comparison between music and language is certainly a privileged one, it should also be extended to other domains and modalities (e.g., visual sequence learning, number sequences) in order to better understand to what extent underlying mechanisms are domain-general and to provide support to the auditory scaffolding hypothesis (Conway et al., 2009).

MUSIC TRAINING AND BRAIN DYNAMICS

The study of the effects of music training on language learning also poses a challenge in terms of brain plasticity. Language learning probably relies on a distributed network involving the

superior temporal gyrus (phonological representation), the dorsal stream and premotor cortex for mapping sounds to movement (phonological–articulatory trace) and the medial temporal lobe (lexical trace, Hickok and Poeppel, 2007; Rodríguez-Fornells et al., 2009). Thus, changes due to music training may take place at each of these different processing steps. Moreover, it is highly likely that music training influences the connectivity between these regions (Oechslin et al., 2010). Thus, the future challenge is to track to what extent music training influences the temporal evolution of the brain dynamics of this complicated network as learning takes place.

ACKNOWLEDGMENT

This study has been supported by the French National Research Agency (ANR BLAN: Does Music Boost the Brain) to D. Schön.

REFERENCES

- Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- Abla, D., Katahira, K., and Abla, O. K. (2008). On-line assessment of statistical learning by event-related potentials. *J. Cogn. Neurosci.* 20, 952–964.
- Abla, D., and Okanoya, K. (2008). Statistical segmentation of tone sequences activates the left inferior frontal cortex: a near-infrared spectroscopy study. *Neuropsychologia* 46, 2787–2795.
- 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.
- Aslin, R. N., Saffran, J. R., and Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychol. Sci.* 9, 321–324.
- Baumann, S., Meyer, M., and Jäncke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *J. Cogn. Neurosci.* 20, 2238–2249.
- Bermudez, P., and Zatorre, R. J. (2005). Differences in gray matter between musicians and nonmusicians. *Ann. N. Y. Acad. Sci.* 1060, 395–399.
- Butzlaff, R. (2000). Can music be used to teach reading? *J. Aesthet. Educ.* 34, 167–178.
- Conway, C. M., and Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 24–39.
- Conway, C. M., Pisoni, D. B., Anaya, E. M., Karpicke, J., and Henning, S. C. (2010). Implicit sequence learning in deaf children with cochlear implants. *Dev. Sci.* 11, 69–82.
- Conway, C. M., Pisoni, D. B., and Kronenberg, W. G. (2009). The importance of sound for cognitive sequencing abilities: the auditory scaffolding hypothesis. *Curr. Dir. Psychol. Sci.* 18, 275–279.
- Crowder, R. G., Serafine, M. L., and Repp, B. (1990). Physical interaction and association by contiguity in memory for the words and melodies of songs. *Mem. Cognit.* 18, 469–476.
- Cunillera, T., Càmarà, E., Toro, J. M., Marco-Pallares, J., Sebastián-Gallés, N., Ortiz, H., Pujol, J., and Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage* 48, 541–553.
- Cunillera, T., Gomila, A., and Rodríguez-Fornells, A. (2008). Beneficial effects of word final stress in segmenting a new language: evidence from ERPs. *BMC Neurosci.* 18, 9. doi: 10.1186/1471-2202-9-23
- Cunillera, T., Toro, J. M., Sebastián-Gallés, N., and Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. *Brain Res.* 1123, 168–178.
- Daltrozzo, J., and Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *J. Cogn. Neurosci.* 21, 1882–1892.
- Daltrozzo, J., Tillmann, B., Platel, H., and Schön, D. (2010). Temporal aspects of the feeling of familiarity for music and the emergence of conceptual processing. *J. Cogn. Neurosci.* 22, 1754–1769.
- De Diego Balaguer, R., Fuentesmilla, L., and Rodríguez-Fornells, A. (2011). Brain dynamics sustaining rapid rule extraction from speech. *J. Cogn. Neurosci.* doi: 10.1162/jocn.2011.21636. [Epub ahead of print].
- De Diego Balaguer, R., Toro, J. M., Rodríguez-Fornells, A., and Bachoud-Lévi, A. C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS ONE* 2, e1175. doi: 10.1371/journal.pone.0001175
- Dienes, Z., and Longuet-Higgins, C. (2004). Can musical transformations be implicitly learned? *Cogn. Sci.* 28, 531–558.
- Federmeier, K. D., and Kutas, M. (2000). It's about time. *Brain Lang.* 71, 62–64.
- Fernald, A. (1992). "Human maternal vocalizations to infants as biologically relevant signals," in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, eds J. Barkow, L. Cosmides, and J. Tooby (Oxford: Oxford University Press), 391–428.
- Fernandes, T., Kolinsky, R., and Ventura, P. (2009). The metamorphosis of the statistical segmentation output: Lexicalization during artificial language learning. *Cognition* 112, 349–366.
- Fiser, J., and Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape-sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 458–467.
- Fiser, J., and Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *J. Exp. Psychol. Gen.* 134, 521–537.
- 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
- Foxton, J. M., Talcott, J. B., Witton, C., Brace, H., McIntyre, F., and Griffiths, T. D. (2003). Reading skills are related to global, but not local, acoustic pattern perception. *Nat. Neurosci.* 6, 343–344.
- Francois, C., and Schön, D. (2010). Learning of musical and linguistic structures: comparing event-related potentials and behavior. *Neuroreport* 21, 928–932.
- Francois, C., and Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cereb. Cortex*. doi: 10.1093/cercor/bhr022. [Epub ahead of print].
- Friederici, A. D., and Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Percept. Psychophys.* 54, 287–295.
- Gaab, N., Tallal, P., and Kim, H. (2005). Neural correlates of rapid spectrotemporal processing in musicians and nonmusicians. *Ann. N. Y. Acad. Sci.* 1060, 82–88.
- Gervain, J., Macagno, E., Cogoi, S., Peña, M., and Mehler, J. (2008). The neonate brain detects speech structure. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14222–14227.
- Hébert, S., and Peretz, I. (2001). Are text and tune of familiar songs separable by brain damage? *Brain Cogn.* 46, 169–175.
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Janata, P., and Grafton, S. T. (2003). Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nat. Neurosci.* 6, 682–687.
- Jones, M. R., Jagacinski, R. J., Yee, W., Floyd, R. L., and Klapp, S. T. (1995). Tests of attentional flexibility in listening to polyrhythmic patterns. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 293–307.
- Keenan, J. P., Thangaraj, V., Halpern, A. R., and Schlaug, G. (2001). Absolute pitch and planum temporale. *Neuroimage* 14, 1402–1408.
- Kishon-Rabin, L., Amir, O., Vexler, Y., and Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *J. Basic Clin. Physiol. Pharmacol.* 12(2 Suppl.), 125–143.
- Koelsch, S., Schröger, E., and Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport* 10, 1309–1313.

- Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605.
- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843.
- Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., and Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. *Cereb. Cortex* 20, 2958–2970.
- Magne, C., Schön, D., and Besson, M. (2006). Musician children detect pitch violations in both music and language better than non-musician children. *J. Cogn. Neurosci.* 18, 199–211.
- Marie, C., Magne, C., and Besson, M. (2011). Musicians and the metric structure of words. *J. Cogn. Neurosci.* 23, 294–305.
- McLaughlin, J., Osterhout, L., and Kim, A. (2004). Neural correlates of second – language word learning: minimal instruction produces rapid change. *Nat. Neurosci.* 7, 703–704.
- McNealy, K., Mazziota, J. C., and Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *J. Neurosci.* 26, 7629–7639.
- Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hear. Res.* 219, 36–47.
- Milovanov, R., Huottilainen, 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.
- Milovanov, R., Huottilainen, M., Välimäki, V., Esquef, P. A., Alku, P., and Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: neural and behavioral evidence. *Brain Res.* 1194, 81–89.
- 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.
- Morrongio, B. A., and Roes, C. L. (1990). Children's memory for new songs: integration or independent storage of words and tunes? *J. Exp. Child Psychol.* 50, 25–38.
- Münste, T. F., Altenmüller, E., and Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nat. Rev. Neurosci.* 3, 473–478.
- Nespor, M., and Vogel, I. (1983). "Prosodic structure above the word," in *Prosody: Models and Measurements*, eds A. Cutler and D. R. Ladd (Berlin: Springer-Verlag), 123–140.
- Oechslin, M. S., Imfeld, A., Loenneker, T., Meyer, M., and Jäncke, L. (2010). The plasticity of the superior longitudinal fasciculus as a function of musical expertise: a diffusion tensor imaging study. *Front. Hum. Neurosci.* 8, 3–76.
- O'Rourke, T. B., and Holcomb, P. J. (2002). Electrophysiological evidence for the efficiency of spoken word processing. *Biol. Psychol.* 60, 121–150.
- Overly, K. (2003). Dyslexia and music: From timing deficits to musical intervention. *Ann. N. Y. Acad. Sci.* 999, 497–505.
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature* 392, 811–814.
- Parbery-Clark, A., Skoe, E., and Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* 29, 14100–14107.
- Peña, M., Bonatti, L., Nespor, M., and Mehler, J. (2002). Signal-driven computations in speech processing. *Science* 298, 604–607.
- Peretz, I., and Coltheart, M. (2003). Modularity of music processing. *Nat. Neurosci.* 6, 688–691.
- Peretz, I., Radeau, M., and Arguin, M. (2004). Two-way interactions between music and language: evidence from priming recognition of tune and lyrics in familiar songs. *Mem. Cognit.* 32, 142–152.
- Perruchet, P., and Vinter, A. (2002). The self-organizing consciousness. *Behav. Brain Sci.* 25, 297–300.
- Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., and De Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philos. Trans. R. Soc. B. Biol. Sci.* 364, 3711–3735.
- Saffran, J., Johnson, E., Aslin, R., and Newport, E. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.
- Saffran, J. R., Hauser, M., Seibel, R. L., Kapfhamer, J., Tsao, F., and Cushman, F. (2008). Grammatical pattern learning by infants and cotton-top tamarin monkeys. *Cognition* 107, 479–500.
- Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996a). Word segmentation: the role of distributional cues. *J. Mem. Lang.* 35, 606–621.
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996b). Statistical learning by 8-month old infants. *Science* 274, 1926–1928.
- Saffran, J. R., Senghas, A., and Trueswell, J. C. (2001). The acquisition of language by children. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12874–12875.
- Schendel, Z. A., and Palmer, C. (2007). Suppression effects on musical and verbal memory. *Mem. Cognit.* 35, 640–650.
- Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F., and Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia* 33, 1047–1055.
- Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., and Kolinsky, R. (2008). Song as an aid for language acquisition. *Cognition* 106, 975–983.
- Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349.
- Scotto di Carlo, N. (1994). L'intelligibilità de la voix chantée. *Méd. Arts* 10, 2–15.
- Serafine, M. L., Crowder, R. G., and Repp, B. H. (1984). Integration of melody and text in memory for songs. *Cognition* 16, 285–303.
- Sergent, C., Baillet, S., and Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nat. Neurosci.* 8, 1391–1400.
- Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., and Pessoa, L. (2008). Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends Cogn. Sci.* 12, 314–321.
- 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.
- Slevc, L. R., and Miyake, A. (2006). Individual differences in second-language proficiency: does musical ability matter? *Psychol. Sci.* 17, 675–681.
- 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.
- Spiegel, M. F., and Watson, C. S. (1984). Performance on frequency discrimination tasks by musicians and non-musicians. *J. Acoust. Soc. Am.* 76, 1690–1695.
- Steinbeis, N., and Koelsch, S. (2008). 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
- Tallal, P., and Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends Neurosci.* 29, 382–390.
- Teinonen, T., Fellman, V., Näätänen, R., Alku, P., and Huottilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* 13, 10–21.
- Thiessen, E. D., Hill, E., and Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy* 7, 49–67.
- Tillmann, B., Bharucha, J. J., and Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychol. Rev.* 107, 885–913.
- Tillmann, B., Janata, P., and Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cogn. Brain Res.* 16, 145–161.
- Tillmann, B., Koelsch, S., Escoffier, N., Bigand, E., Lalitte, P., Friederici, A. D., and Von Cramon, D. Y. (2006). Cognitive priming in sung and instrumental music: activation of inferior frontal cortex. *Neuroimage* 31, 1771–1782.
- Tillmann, B., and McAdams, S. (2004). Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis)similarities. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1131–1142.
- Toro, J. M., and Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Percept. Psychophys.* 67, 867–875.
- Tremblay, K., Kraus, N., and McGee, T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *Neuroreport* 9, 3557–3560.
- Van Petten, C., and Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Mem. Cognit.* 18, 380–393.
- Van Zuijlen, T. L., Sussman, E., Winkler, I., Näätänen, R., and Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity: a mismatch negativity study comparing musicians and non-musicians. *Brain Res. Cogn. Brain Res.* 23, 270–276.
- Yee, W., Holleran, S., and Jones, M. R. (1994). Sensitivity to event timing in regular and irregular

sequences – influences of musical skill. *Percept. Psychophys.* 56, 461–471.

Young, M. P., and Rugg, M. D. (1992). Word frequency and multiple repetition as determinants of the modulation of event-related potentials in a semantic classification task. *Psychophysiology* 29, 664–676.

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: 02 March 2011; accepted: 04 July 2011; published online: 18 July 2011.

Citation: Schön D and François C (2011) Musical expertise and statistical learning of musical and linguistic structures. *Front. Psychology* 2:167. doi: 10.3389/fpsyg.2011.00167

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

Copyright © 2011 Schön and François. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise

Dana L. Strait^{1,2} and Nina Kraus^{1,2,3,4,5*}

¹ Auditory Neuroscience Laboratory, Northwestern University, Evanston, IL, USA

² Institute for Neuroscience, Northwestern University, Evanston, IL, USA

³ Department of Communication Sciences, Northwestern University, Evanston, IL, USA

⁴ Department of Neurobiology and Physiology, Northwestern University, Evanston, IL, USA

⁵ Department of Otolaryngology, Northwestern University, Evanston, IL, USA

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Mireille Besson, Centre National de la

Recherche Scientifique, France

Barbara Tillmann, Centre National de la

Recherche Scientifique, France

*Correspondence:

Nina Kraus, Auditory Neuroscience Laboratory, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA.

e-mail: nkraus@northwestern.edu
<http://www.brainvolts.northwestern.edu>

Even in the quietest of rooms, our senses are perpetually inundated by a barrage of sounds, requiring the auditory system to adapt to a variety of listening conditions in order to extract signals of interest (e.g., one speaker's voice amidst others). Brain networks that promote selective attention are thought to sharpen the neural encoding of a target signal, suppressing competing sounds and enhancing perceptual performance. Here, we ask: does musical training benefit cortical mechanisms that underlie selective attention to speech? To answer this question, we assessed the impact of selective auditory attention on cortical auditory-evoked response variability in musicians and non-musicians. Outcomes indicate strengthened brain networks for selective auditory attention in musicians in that musicians but not non-musicians demonstrate decreased prefrontal response variability with auditory attention. Results are interpreted in the context of previous work documenting perceptual and subcortical advantages in musicians for the hearing and neural encoding of speech in background noise. Musicians' neural proficiency for selectively engaging and sustaining auditory attention to language indicates a potential benefit of music for auditory training. Given the importance of auditory attention for the development and maintenance of language-related skills, musical training may aid in the prevention, habilitation, and remediation of individuals with a wide range of attention-based language, listening and learning impairments.

Keywords: attention, cortical variability, auditory-evoked potentials, music, musicians, speech in noise, prefrontal cortex, language

"Attention is the holy grail. Everything that you're conscious of, everything you let in, everything you remember and you forget, depends on it."

D. Strayer¹

INTRODUCTION

The human nervous system is constantly faced with an astounding amount of sensory input. Despite the fact that our brains house over 10 billion neurons with more than 10 trillion synapses, accurate encoding of a complete environmental scene is a functional impossibility. Fortunately, the brain has evolved in ways that permit the modulation of neural activity according to environmental and systemic demands, permitting the selection, efficient encoding, and appropriate behavioral response to the stimuli of greatest biological interest. Selective attention makes this modulation possible, directing the allocation of neural resources to selectively encode one aspect of the environment while excluding competing aspects. Selective attention resolves the competition imposed by a mass of incoming signals through the activation of executive control regions

(e.g., prefrontal cortex) to promote increased spiking in neurons that represent the attended object (Desimone and Duncan, 1995). Variability in the activation of brain networks that underlie selective attention can bring about significant behavioral disadvantages, such as attention lapses (Weissman et al., 2006) and symptoms of an attention impairment (Depue et al., 2010).

Recently, scientific pursuits concerning attention have invoked a hotbed of discussion. With the dawn of attention-fracturing devices such as portable music players, texting, and the internet, the act of sustaining attention on a single task may be rapidly fading into oblivion. More than ever, scientists need to determine the neural mechanisms that underlie attention, their behavioral outcomes, and how we might strengthen them with training and life experience. Here, we emphasize the discussion of auditory attention given its importance for language processing and the development and maintenance of language-related skills, such as hearing speech in background noise.

Event-related potentials (ERPs) have provided striking insights into the neuronal underpinnings of selective auditory attention (Hillyard et al., 1973; Woldorff et al., 1993; Coch et al., 2005), especially as it relates to everyday auditory function in noise. Standard experimental procedures imitate listening requirements in noisy

¹Quoted from Matt Richtel, "Your Brain on Computers: Outdoors and Out of Reach, Studying the Brain," *New York Times* Aug 15, 2010, p. A1.

environments by presenting separate sound streams to the right and left ears, asking participants to pay attention to one side while ignoring the other. When ERPs to the attended relative to the ignored sounds are compared, the negative deflection occurring at ~100 ms post-stimulus (i.e., the N100) is deeper in amplitude (i.e., more negative). Although more information is needed to precisely define the neuronal mechanisms that drive such outcomes and their malleability with training, development, and life experiences, these findings reveal that attention has the power to modulate early sensory processing.

The brains of musicians may provide insight into neural attention mechanisms and their potential for experience-dependent plasticity (Kraus and Chandrasekaran, 2010). Musical practice and performance require sustained attentional control for the delicate online manipulation of sound and, for ensemble players, to permit coordination with other instrumentalists. Given that musicians traditionally initiate training during early developmental years, attention to sound is regularly practiced during pivotal periods of brain development. Recent evidence from our laboratory indicates enhanced auditory but not visual attention ability in musicians relative to non-musicians, with musicians demonstrating faster reaction times to a target sound than non-musicians, but not to the task's visual analog (**Figure 1A**; Strait et al., 2010; see Materials and Methods for task details). This finding may reflect decreased variability in musicians' sustained auditory attention task performance. Surprisingly little is known about the impact of musical training on the neural correlates of attention. We do know, however, that cortical networks that promote attention to music share considerable overlap with those that underlie general attention in other auditory domains, such as language. In addition to the primary auditory cortex, these sites include the fronto-parietal attention and working memory networks that comprise the prefrontal cortex, the intraparietal sulcus, the supplementary and presupplementary motor areas, and the precentral gyrus (Janata et al., 2002; Kane and Engle, 2002). This functional overlap between attention to language and music corroborates previous results suggesting that a combination of modality-specific and general attention and working memory mechanisms (e.g., the fronto-parietal attention network) contribute to sustained auditory attention (Zatorre et al., 1999; Petkov et al., 2004). The prefrontal cortex has been particularly emphasized for its role in sustaining attention by providing access to recently presented stimuli and directs sensory processing according to behavioral goals – especially in challenging perceptual environments (Kane and Engle, 2002). Although we lack direct evidence for how musical training shapes brain mechanisms that underlie auditory attention performance, that musical training tunes the brain's executive control network for auditory processing beyond the music domain – particularly for sustaining attention with minimal variability – would not be surprising.

While little is known about the neural correlates of attention ability in musicians, it is well established that musical training strengthens cortical and subcortical mechanisms for auditory processing. Despite the fact that neural specializations for music and speech have been established (Zatorre et al., 2002; Brown et al., 2006; Abrams et al., 2010; Rogalsky et al., 2011), there is no doubt that the human brain also recruits shared mechanisms for processing sound in both domains (Koelsch et al., 2002; Patel, 2003; Zatorre

and Gandour, 2008; Fedorenko et al., 2009; Slevc et al., 2009). Such shared mechanisms may account, at least in part, for musicians' structural (Schmithorst and Wilke, 2002; Schneider et al., 2002; Gaser and Schlaug, 2003; Hutchinson et al., 2003; Schlaug et al., 2009) and functional enhancements for general auditory processing that are not constrained to the domain of music but that extend to language (Schon et al., 2004; Marques et al., 2007; Musacchia et al., 2007; Wong et al., 2007; Moreno et al., 2009) and emotional communication sounds (Strait et al., 2009). Neural enhancements are particularly evident in musicians in the context of challenging listening environments, such as in the presence of background noise (Parbery-Clark et al., 2009a) and reverberation (Bidelman and Krishnan, 2010), with musically trained adults demonstrating less noise-induced degradation in the subcortical encoding of speech than non-musicians (**Figure 1D**). The degree of noise-induced subcortical response degradation is functionally correlated with speech-in-noise perceptual ability in that individuals with increased subcortical resilience to background noise demonstrating better speech-in-noise perception (Parbery-Clark et al., 2009a). These findings imply that musicians' nervous systems are fine-tuned for the extraction of meaning from a complex soundscape, being shaped through their extensive and consistent interactions with organized sound to better exclude competing noise and more accurately encode signals of interest.

The precise neurobiological mechanisms that bring about musical training-induced neuronal enhancements remain undetermined, although strengthened cognitive control over auditory processing, as would be directed by attention, provides a plausible agency (Strait and Kraus, in press; Strait et al., 2010). Increasing evidence has accrued to indicate that musicians more heavily recruit extra-sensory cortical areas associated with attention and working memory, such as the prefrontal, superior parietal, and inferior frontal cortices, during challenging auditory tasks that demand discriminatory alertness (e.g., when subjects are instructed to listen for certain auditory targets) compared to non-musicians (Stewart et al., 2003; Haslinger et al., 2005; Baumann et al., 2008). The prefrontal cortex has been attributed particular importance, being associated with goal-directed behavior and the top-down guiding of sensory processing according to internal states or intentions (Miller and Cohen, 2001). Whereas musically trained and non-trained adults demonstrate equivalent auditory cortex activation for the completion of pitch discrimination and sound recall tasks, musicians more extensively activate parietal and prefrontal extra-sensory networks – indicating more extensive involvement of attention and working memory networks that could facilitate heightened control over sensory processing (Gaab and Schlaug, 2003; Pallesen et al., 2010). Musicians' recruitment of extra-sensory networks involved in attention and working memory may account for their enhanced performance on auditory tasks such as pitch discrimination (Kishon-Rabin et al., 2001; Parbery-Clark et al., 2009a; Strait et al., 2010), sound recall and hearing speech in noise (**Figure 1**; Parbery-Clark et al., 2009a, 2011).

That strengthened cognitive control mechanisms guide general auditory processing enhancements in musicians in a top-down manner is particularly viable given recent observations to this effect in animal models, in which auditory training leads to modifications in spectrotemporal tuning curves in the primary auditory cortex

that appear to be facilitated by functional connections with the prefrontal cortex (cf. Bajo and King, 2010; Fritz et al., 2010). Even with regard to subcortical auditory plasticity, a primary role has been established for the reciprocal corticocollicular pathway, with training-induced changes in inferior collicular response properties being ablated with the targeted cooling of the cortex (Bajo et al., 2010). This is not surprising given the noted strength of cortical descending pathways in modulating subcortical (i.e., collicular) neuronal response properties (Suga et al., 2002). The resiliency of musicians' nervous systems for encoding signals of interest in the presence of background noise (**Figure 1D**) may indicate increased executive control over auditory function, or, in other words, strengthened top-down attentional mechanisms within the primary auditory cortex that guide the resolution of competition imposed by a mass of incoming signals.

As noted, when multiple auditory streams are present in a scene, they compete for cortical representation. Selective auditory attention provides a mechanism for determining which sounds will be most thoroughly processed and brought to awareness, to the exclusion of others. It is unlikely, however, that the human brain is able to invariably maintain attention on a specific sound stream of interest over a sustained period of time. Consistently sustaining attention – with minimal attention lapses – is particularly difficult in input-rich sensory environments, such as when tracking a single individual's voice amidst other conversations. Accordingly, Weissman and colleagues have demonstrated that brain regions associated with attention routinely demonstrate performance variability during the execution of sustained attention-demanding tasks, decreasing in activity while other brain regions increase in activity. Specifically, variability in the activation of the attention network during task performance has been linked to momentary lapses in attention, with continued activation of prefrontal and parietal regions underlying successful sustained attention performance (Weissman et al., 2006). Variability in the activation of prefrontal control regions is interpreted as the failure to accomplish attention's goal, being to maximally and consistently enhance the sensory processing of behaviorally relevant stimuli. Decreases in the fronto-parietal network's activation reduce its suppression of a default – or “day-dreaming” – network, which corresponds with poorer attention task performance.

Here, we aimed to define the impact of musical training on neural networks underlying selective auditory attention performance in a natural language-listening environment. In light of the functional importance of sustained fronto-parietal attention network activation combined with musicians' enhanced reliance on this network compared to non-musicians for the execution of auditory tasks, we asked two questions. First, we asked whether the act of sustaining auditory attention on a target speech stream leads to decreases in auditory-evoked response variability across all participants, especially within the primary auditory cortex and fronto-parietal attention areas. Second, we asked whether this decrease is larger in musicians. We hypothesized that musicians demonstrate less variability in neural responses to speech with auditory attention compared to non-musicians, particularly in prefrontal and parietal cortices, and that this decrease in variability correlates with musicians' training backgrounds.

MATERIALS AND METHODS

PARTICIPANTS

All experimental procedures were approved by the Northwestern University Institutional Review Board. Twenty-three normal hearing adults (≤ 20 dB pure tone thresholds at octave frequencies from 125 to 8000 Hz) between the ages of 18–35 participated in this study, for which they provided informed consent. All participants completed an extensive questionnaire addressing family history, musical experience and educational history. Musicians (Mus, $N = 11$) were self-categorized, began instrumental musical training at $< \text{age } 7$ and had consistently practiced for ≥ 11 years (consistently defined as practicing at least 3 days weekly; $M = 16.5$, $SD = 5.8$). Non-musicians (NonMus, $N = 12$) were self-categorized, had < 5 years of formal musical experience throughout their lifespans ($M = 1.2$, $SD = 1.8$), began musical training after age 11 and had not played a musical instrument in the 5 years leading up to the experiment. Nine of the 12 NonMus participants had no degree of musical experience. Mus and NonMus groups did not differ according to age ($F_{(1,22)} = 0.20$, $p = 0.66$), sex ($\chi^2 = 0.35$, $p = 0.68$), non-verbal I.Q. as measured by the Wechsler abbreviated matrix reasoning subtest ($F_{(1,22)} = 0.37$, $p = 0.55$; Harcourt Assessment, San Antonio, TX, USA), or performance on the attention task (as measured by quiz scores, described below) ($F_{(1,22)} = 0.32$, $p = 0.58$).

SPEECH-IN-NOISE AND AUDITORY ATTENTION PERFORMANCE

In order to clarify the relationship between speech-in-noise perception and auditory attention, we assessed these skills in 22 18- to 35-year-old musician and non-musician participants ($N = 14$; Mus = 8) using data collected for two separate experiments, the isolated results of which have since been published and are discussed above (Parbery-Clark et al., 2009a,b; Strait et al., 2010). Five of these participants also participated in the electrophysiological paradigm, described below. Speech-in-noise (SIN) perception was measured using the Hearing in Noise Test (Nilsson et al., 1994), in which participants are asked to repeat short sentences presented in speech-shaped background noise using a speaker placed one meter directly ahead. The noise presentation level was fixed at 65 dB SPL and the program adjusted perceptual difficulty by increasing or decreasing the intensity level of the target sentences until the threshold signal-to-noise ratio was determined. Perceptual SIN thresholds were defined as the level difference (in dB) between the speech and the noise presentation levels at which 50% of sentences are correctly repeated.

Auditory attention was assessed using the IHR Multicentre Battery of Auditory Processing's Auditory Attention subtest (Barry et al., 2010), which measures phasic alertness via reaction times induced by the presence or absence of a cue that occurred with a variable delay (0.5–1.0 s) before a target stimulus. We have previously reported between-group differences using this measure in musicians and non-musicians, with musicians demonstrating enhanced performance compared to non-musicians (**Figure 1A**; Strait et al., 2010). Participants were instructed to listen for a “beep” (presented at 80 dB SPL) and to press a button on a response box as soon as they heard it. Participants were cued by a second sound (a “siren,” presented at 70 dB SPL) on some trials and were asked *not* to respond to that cue. Reaction time was measured in milliseconds. Results reported here (**Figure 1C**) reflect subjects' average reaction time to the cued stimulus.

ELECTROPHYSIOLOGY

We employed a paradigm designed by Helen Neville and colleagues that has proven enormously successful for studying neural mechanisms of selective auditory attention in children and adults (Coch et al., 2005).

Stimulus

The evoking stimulus was a six-formant, 170 ms speech syllable synthesized in Klatt (1980) with a 5 ms voice onset time and a level fundamental frequency (100 Hz). The first, second and third formants were dynamic over the first 50 ms (F_1 , 400–720; F_2 , 1700–1240; F_3 , 2580–2500 Hz) and then maintained frequency for the rest of the duration. The fourth, fifth and sixth formants were constant throughout the entire duration of the stimulus (F_4 , 3300; F_5 , 3750; F_6 , 4900 Hz). The stimulus was presented using NeuroScan Stim2 (Compumedics, Charlotte, NC, USA).

Electrophysiologic recording parameters and procedure

Auditory-evoked potentials were recorded to the speech sound /da/ using a 31-channel tin-electrode cap (Electrocap International, Eaton, OH, USA) in NeuroScan Aquire 4.3 (Compumedics) while participants were seated in a sound-attenuated booth. Single electrodes were placed on the earlobes and on the superior and outer canthi of the left eye, thereby acting as reference and eye-blink monitors, respectively. Contact impedance for all electrodes was under <5 k Ω with less than 3 k Ω difference across channels. Neural recordings were on-line filtered from 0.05 to 100 Hz and digitally sampled at a rate of 500 Hz.

The evoking stimulus was presented in the context of short story recitations through two wall-mounted speakers located 1 m to the left and right of the participant. Participants were asked to attend to one of the two simultaneously presented stories, which differed in direction (left/right speaker), presentation voice (male/female), and story content. Instructions described both the direction of the attended story and its speaker's sex (*listen to the story on your right/left, which will be told by a male/female, and ignore the story presented from the other side by a speaker of the opposite sex*). The initial direction of the attended voice was randomized across participants to control for potential advantages or disadvantages of attending to one voice over the other. The evoking stimulus was presented randomly to the left or right (i.e., attended or ignored) sides of the head with a randomized inter-stimulus interval (ISI) that was either 600, 900, or 1200 ms. The stories and the evoking stimulus were presented with a 10 dB difference between the stories (65 dB SPL) and the stimulus (75 dB SPL). The recording was paused every 8 min, during which participants were given one minute to complete a five-question multiple choice quiz regarding the attended story content and one minute to stretch. An average score of $\geq 4/5$ correct answers was required for study inclusion. After each break, the attended story changed directions and participants were asked to change their attended side (left/right) in order to continue with the same voice. The entire recording session lasted 40 min and yielded 600 simultaneously recorded responses in both attended and ignored conditions.

Data processing and analysis

Continuous neural data for attended and ignored conditions were baseline corrected and the removal of eye-blink artifacts was accomplished using the spatial filtering algorithm in NeuroScan Edit 4.3

(Compumedics). Response variability was computed through generation of a variability index (VI) for each subject in each condition, following a procedure described in Smith and Goffman (1998), who applied it to assess variability in speech movements. Continuous files were epoched from –100 to 500 ms, referenced to the presentation of the stimulus (0 ms); epochs demonstrating amplitudes beyond ± 100 μ V were rejected as muscular artifact and the first 500 artifact-free responses from each participant were subjected to analysis. Epochs were grouped into twenty subsets of 25 individual responses; these 25 individual responses in each subset were then averaged, resulting in 20 averaged waveforms (i.e., subaverages). The VI was determined through calculation of amplitude variances across these subaverages. Specifically, amplitudes were determined for each of the 300 points that made up the evoked response subaverages. Rather than comparing amplitudes across subaverages on a point-by-point basis, we averaged point-by-point amplitudes across 50 equally spaced increments (comprised of six points each), computed the variances in these increments across the subaverages and summed them. This generated a single VI for each subject in both attended and ignored conditions. Although evoked response variability has been previously assessed in humans (Anderson et al., 1991), our method is unique in that it enables the assessment of variance over the entire evoked response, including early evoked potentials that are not observable in individual evoked responses (P1/N1). Because these early components are small, we performed our analysis on small subaverages. All data processing was executed with scripts generated in Matlab 7.5.0 (The Mathworks, Natick, MA, USA).

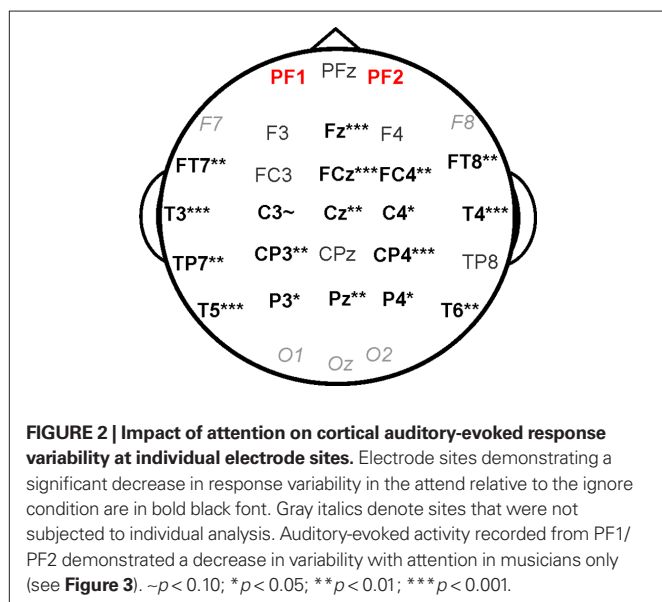
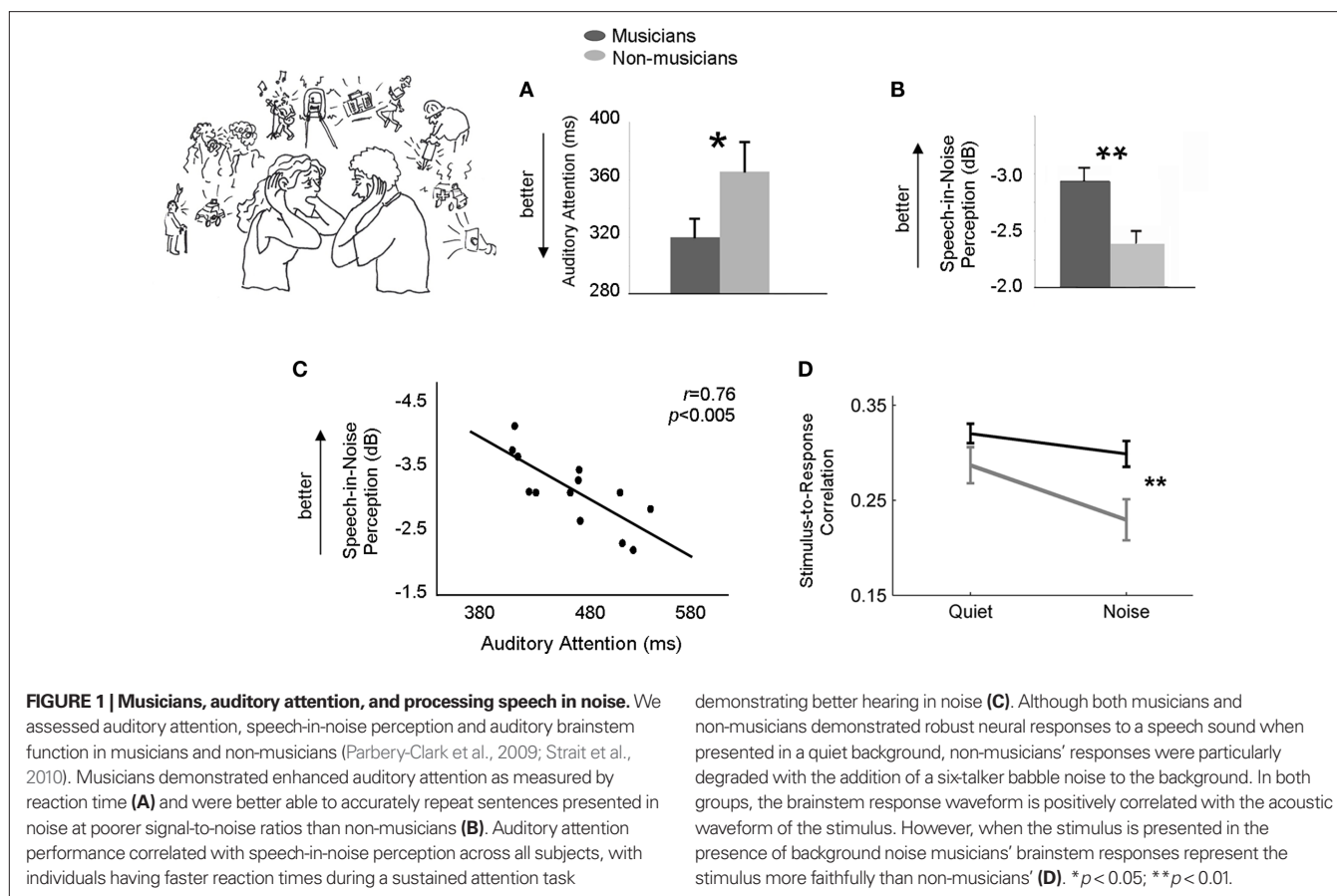
Differences in response variability between attend and ignore conditions were compared for all of the 31 electrode sites using a Repeated Measures ANOVA. Effects at individual electrode sites (Figure 2) were subsequently explored using post-hoc paired and independent samples *t*-tests for all electrode sites except for F7, F8, O1, O2, and OZ, which did not demonstrate clear responses characteristic of cortical auditory-evoked activity (i.e., the P1–N1–P2–N2 complex; all other electrode sites demonstrated clear responses characteristic of cortical auditory-evoked activity). Relationships among musical practice histories (i.e., age of onset of musical practice, years of musical practice) and cortical variability were examined with Pearson's correlations (SPSS Inc., Chicago, IL, USA). All results reported herein reflect two-tailed values and normality for all data was confirmed using the Kolmogorov–Smirnov test for equality.

RESULTS

SUMMARY OF RESULTS

Within the subset of participants who had both measures (as described in Materials and Methods), auditory attention performance correlated with speech-in-noise perceptual ability, with better auditory attention relating to the ability to accurately perceive speech in higher levels of background noise (i.e., at lower signal-to-noise ratios; Figure 1C).

All participants demonstrated less cortical auditory-evoked response variability over a majority of electrode sites in responses to the stimulus when it was presented in the attended compared to the ignored story (Figure 2). Only musicians, however, demonstrated decreased cortical response variability with auditory attention over



the prefrontal cortex (Figure 3), a region of particular importance for sustaining attention in challenging perceptual environments (Kane and Engle, 2002). The degree to which attention decreased prefrontal response variability correlated with musical practice histories and is interpreted in the context of musical training's impact on cortical mechanisms of selective auditory attention.

CORTICAL RESPONSE VARIABILITY DECREASES AS A FUNCTION OF SELECTIVE AUDITORY ATTENTION

A $2 \times 31 \times 2$ RMANOVA with condition (attend/ignore) and electrode site as within subject variables and group (Mus/NonMus) as between subject variable revealed a main effect of condition on response variability, indicating that cortical response variability varied as a function of attention ($F_{(1,22)} = 10.49$, $p < 0.005$). We also observed an interaction between condition electrode site, indicating that attention impacted response variability differentially across the scalp ($F_{(1,22)} = 17.08$, $p < 0.0001$), and an anticipated main effect of electrode site ($F_{(1,22)} = 17.49$, $p < 0.0001$). *Post hoc* paired *t*-tests demonstrated that, across all participants, response variability decreased to stimuli in the attended compared to the ignored story at all analyzed electrode sites except for seven (for these seven sites, all $t_{(22)} < 1.4$, all $p \geq 0.2$; Figure 2). An overall effect of attention across all participants was not observed for the three prefrontal sites ($F_{(1,22)} = 0.31$, $p = 0.40$).

EFFECT OF MUSICAL TRAINING ON CORTICAL RESPONSE VARIABILITY

Musicians and non-musicians did not differ based on overall response variability at any electrode site. That is, neither group was more or less variable in auditory-evoked responses to the ignored or to the attended stories individually, indicating no group difference in general auditory-evoked variability. Rather, differences between musicians and non-musicians were observed with regard to the extent to which attention decreased auditory-evoked response variability at prefrontal electrode sites. Specifically, a 2×2 RMANOVA

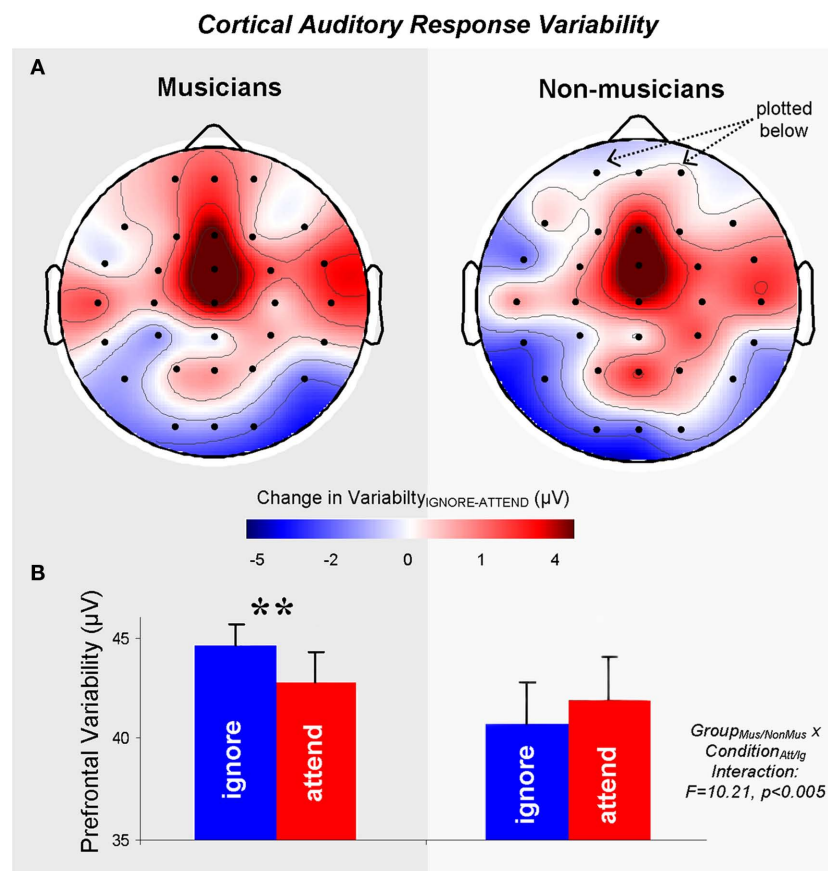


FIGURE 3 | Cortical auditory-evoked response variability in musicians and non-musicians. (A) 31-Channel headplots for musicians (left) and non-musicians (right) demonstrate the difference in cortical auditory-evoked response variability between ignore and attention conditions, plotting variability across the scalp as a function of attention. Because the difference was calculated by subtracting attend from ignore variability, positive values (red) indicate a decrease in response

variability in the attend relative to the ignore condition. Negative values (blue) indicate an increase in response variability in the attend relative to the ignore condition. **(B)** Musicians demonstrate an increased impact of attention on prefrontal response variability compared to non-musicians. Whereas musicians demonstrate a decrease in prefrontal response variability in the attend relative to the ignore condition, non-musicians do not. $**p < 0.01$.

with condition (attend/ignore) as within subject variable and group (Mus/NonMus) as between subject variable revealed a significant condition–group interaction at prefrontal electrode sites FP1 and FP2 ($F_{(1,22)} = 10.21, p < 0.005$). *Post hoc* within-group paired *t*-tests demonstrated that whereas musicians demonstrated decreased response variability over the prefrontal cortex with auditory attention ($t_{(10)} = 3.0, p < 0.01$), non-musicians did not ($t_{(11)} = 1.6, p = 0.2$; **Figure 3B**). Differences in response variability with attention between musicians and non-musicians were not observed for any of the other electrode sites (all $t < 1.4$, all $p > 0.12$).

Across all participants with some degree of musical training ($N = 14$; Mus = 11, NonMus = 3), the age of onset of musical training correlated with the extent to which response variability decreased in responses to the attended relative to the ignored story ($r = -0.54, p < 0.05$).

DISCUSSION

Here, we substantiate a relationship between auditory attention performance and speech-in-noise perception (**Figure 1C**) and reveal a novel neural index for selective auditory attention in musician

and non-musician adults, consisting of decreased auditory-evoked response variability in attended relative to ignored speech. Across all participants, attention decreased cortical auditory-evoked response variability at central, temporal, and parietal sites (**Figure 2**), and this effect was equivalent in musicians and non-musicians. Only musicians, however, demonstrated an impact of selective auditory attention on prefrontal evoked activity (**Figure 3**). These results provide evidence for the power of musical training to shape prefrontal neural activity involved in sustaining auditory attention and may contribute to the definition of a biological mechanism that would facilitate musicians' advantages in auditory tasks (Kraus and Chandrasekaran, 2010; Strait et al., 2010).

CORTICAL AUDITORY-EVOKED RESPONSE VARIABILITY UNDERLIES SELECTIVE AUDITORY ATTENTION

The goal of sustaining attention on a specific task is to reduce moment-to-moment variability in one's performance. Sustaining attention becomes particularly difficult in the presence of competing stimuli, such as when tracking a single voice amidst a noisy background. In this situation it is the listener's goal to absorb the

entirety of the attended speaker's content in order to adequately respond, and lapses in attention result in comprehension gaps that can lead to conversational confusion. Variability in attention performance (i.e., lapses in attention) can also have more drastic consequences, being responsible for accidents while operating mechanical equipment (e.g., cars) and, in educational scenarios, has the potential to diminish the quality of learning that takes place in young brains (Vaurio et al., 2009).

Moment-to-moment behavioral variability has been directly linked with variability in the brain's extra-sensory evoked activity during task performance (i.e., prefrontal, frontal, and parietal cortices; Carmena et al., 2005; Fox et al., 2005; Weissman et al., 2006). As tasks become more difficult, cortical response variability increases, concurrent with poorer task performance (Vogels et al., 1989). Given that the most frequent analysis technique for electrophysiologic data involves averaging, scientists might regularly overlook a crucially informative neural index for attention and human behavior. This is because evoked potentials traditionally necessitate the averaging of many individual responses to a repeated stimulus in order to maximize that which is consistent across trials (i.e., the average evoked response), effectively minimizing that which is inconsistent and discarding it as noise. This disregarded "noise," or response variability, is often as large or even larger than the average response itself (Figure 4; Vogels et al., 1989; Softky and Koch, 1993; Arieli et al., 1996). Arieli et al. (1996) encouraged the revision of what we regard as noise in the nervous system, proposing that in doing so we may discover that "response variability... provide[s] the neuronal substrate for the dependence of sensory information processing on behavioral and conscious states". The data we present here corroborate Arieli's suggestion by demonstrating a functional relevance for variability in cortical evoked potentials in humans, serving as an index for selective auditory attention. Further work comparing cortical response variability with more commonly employed

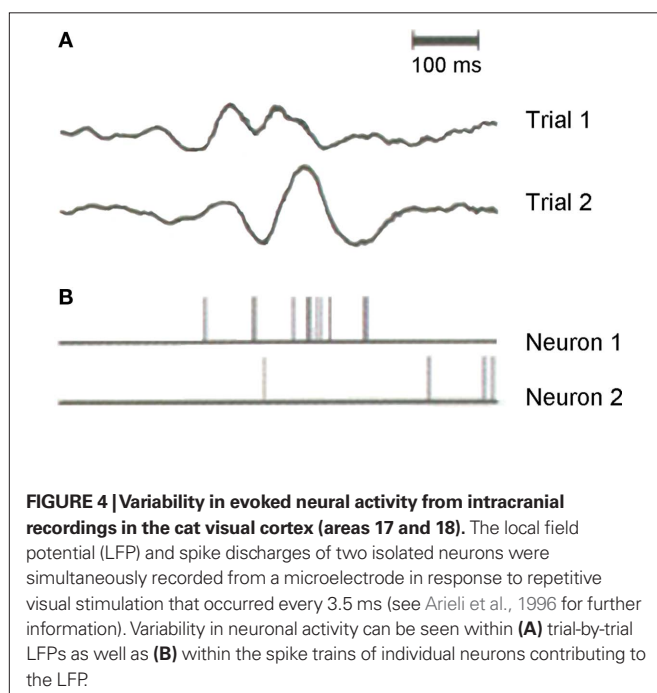
techniques for analyzing electrophysiologic recordings is likely to reveal relationships between response variability and both spontaneous and averaged evoked activity, such as average peak amplitudes/latencies and oscillatory activity within different frequency bands.

Although it is possible that cortical evoked response variability stems, at least in part, from stochastic noise (Faisal et al., 2008), evoked response variability can be predicted by deterministic interactions of sensory responses with ongoing spontaneous activity (Arieli et al., 1996; Curto et al., 2009) that can be modulated by an individual's brain state (Steriade et al., 2001) and cognitive capacity (Benasich et al., 2008). The decreased evoked response variability with selective auditory attention demonstrated here may indicate general changes in ongoing spontaneous activity between attended and ignored states, revealing a novel neural metric for selective auditory attention in behaving humans. Furthermore, group differences as a function of musical training reflect more consistent prefrontal activity in musicians with auditory attention, which may translate into increased control over the sensory competition imposed by competing auditory signals. This implication is particularly relevant given the role of the prefrontal cortex in directing goal-oriented behavior and the top-down shaping of sensory processing according to internal states or intentions (Miller and Cohen, 2001). Further studies coupling behavioral and neural indices of selective auditory attention are needed in order to better define the functional advantage of decreased prefrontal response variability in musicians. Furthermore, simultaneous recording of cortical and subcortical evoked activity may shed light on relationships between prefrontal response variability and subcortical response properties, such as to speech in background noise (Figure 1D).

MUSICAL TRAINING HONES CORTICAL MECHANISMS OF EXECUTIVE CONTROL THAT ARE IMPLICATED IN SELECTIVE AUDITORY ATTENTION

Our results demonstrate a selective impact of musical training on response variability with attention at prefrontal electrode sites. This outcome contributes to a growing literature suggesting that musical training shapes auditory function by training the brain to more extensively recruit extra-sensory mechanisms affiliated with cognitive control, such as working memory and attention, for the completion of general auditory tasks (Gaab and Schlaug, 2003; Stewart et al., 2003; Haslinger et al., 2005; Baumann et al., 2008; Pallesen et al., 2010). Previous experiments, however, have not explicitly investigated neural mechanisms of auditory attention in musicians but, rather, studied musicians' brain function during the execution of psychophysical auditory discrimination and memory tasks. Our data provide the first direct evidence for differential brain activation in musicians and non-musicians during selective auditory attention to speech. That these data are observed in an ecologically valid language-listening environment strengthens arguments for musical training's impact on functional brain networks that underlie language processing.

Our findings may indicate that musicians demonstrate more consistent ongoing (i.e., spontaneous) prefrontal activity during selective auditory attention, compared to non-musicians. As described above, the dynamics of ongoing neural activity convincingly predict variability in cortical evoked responses. Specifically, evoked activity is low when spontaneous activity is low and evoked activity is high when spontaneous activity is high, with spontaneous and evoked



activity positively correlating at an impressive $r = 0.9$ and $p < 10^{-12}$ (Arieli et al., 1996). Decreased variability in musicians' responses would imply increased consistency in ongoing prefrontal activity and, given the importance of consistency for sustaining attention (Weissman et al., 2006), provides a biological mechanism that could account for our previously reported advantage for sustained attention task performance in musicians (Figure 1A; Strait et al., 2010).

Distinctive neural activity during selective auditory attention in musicians and non-musicians may be attributed to the musicians' rehearsal of auditory cognitive mechanisms required for focused musical practice and performance, strengthening top-down contributors to auditory processing (Tervaniemi et al., 2009; Kraus and Chandrasekaran, 2010; Strait et al., 2010). Although the argument can be made for a genetic contributor to structural and functional neural differences between musicians and non-musicians, repeated evidence substantiates that these differences can be modulated, at least in part, by one's method of musical practice (Seppanen et al., 2007) or instrument of specialization (Pantev et al., 2001; Shahin et al., 2008; Margulis et al., 2009; Strait et al., 2011). Furthermore, data consistently reveal correlations between the extent of neural enhancement observed in musicians and their years of musical practice or age of practice onset (Gaser and Schlaug, 2003; Hutchinson et al., 2003; Wong et al., 2007; Parbery-Clark et al., 2009a,b; Strait et al., 2009). These data, together with the correlation reported here between prefrontal response variability and age of onset of musical practice, suggest a contribution of experience-induced neuroplasticity to musicians' auditory processing characteristics.

CLINICAL AND EDUCATIONAL IMPLICATIONS

That musical training has the power to shape neural mechanisms underlying selective attention to speech carries substantial implications for educators and clinicians involved in the remediation of attention-based listening and learning impairments. The ability to attend to a target signal and suppress competing noise is a primary concern for child educators and clinicians given its primacy in everyday learning and communication. It is also of concern to those involved in the treatment of aging-induced listening impairment, which may be prevented through the strengthening of auditory cognitive abilities, such as attention (Parbery-Clark et al., 2011). Accordingly, interest in learning to attend has increased in recent years (Tang and Posner, 2009); within the visual domain, outcomes reveal that task-specific training can improve the temporal allocation of attention (Makovski et al., 2008) and, as required by our paradigm here, increases the neural capacity to filter out competing irrelevant input (Dixon et al., 2009; Kelley and Yantis, 2009). Musical training may provide a naturalistic and entertaining means for strengthening auditory cognitive processing through increasing the consistency of prefrontal control over auditory function.

Although improving attention and the ability to tune in to a signal of interest would benefit the general population, the topic of behavioral and neural variability during selective attention has particular relevance for attention deficit/hyperactivity disorder (ADHD). This is because ADHD is characterized by moment-to-moment variability in behavioral performance (Mullins et al., 2005; Vaurio et al., 2009) and neuronal activity (Depue et al., 2010). Furthermore, structural and functional prefrontal anomalies have been associated with the disorder (Hynd et al., 1990; Casey et al., 1997; Filipek et al., 1997; Gilliam et al., 2011) and are reflected in decreased prefrontal activity during attention task performance (Bush et al., 1999; Rubia et al., 1999). Children with ADHD are particularly noted for an inability to suppress the neural processing of competing sensory input (Suskauer et al., 2008), contributing to frequent distraction. Our association between musical training and decreased prefrontal variability in a neuronal mechanism that underlies selective attention to language during the simultaneous suppression of a competing sound stream may suggest musical training as a viable remediation strategy in children with attention impairment. Still, more work should be done to test the efficacy of music as a remedial approach for ADHD. Population studies investigating the prevalence of ADHD in children and adults with musical training, particularly those with a family history of ADHD, could yield interesting insights.

CONCLUSION

Increasing effort is being expended to define activities that strengthen what might be considered the cornerstone of human perception: attention. While musical training is known to bolster auditory-specific cognitive skills, such as auditory short-term memory, and the ability to pull out speech signals from competing background noise, little is known about how musical training strengthens attention; even less is known about how music shapes the neural mechanisms that underlie it. Here, we present the first biological evidence for musical training's impact on neural mechanisms of selective auditory attention within a language context. Given the high prevalence of developmental attention disorders and their detrimental impacts on educational performance, musical training's power to shape neural mechanisms that underlie selective attention to speech may be of interest to individuals involved in the habilitation and remediation of attention and attention-based learning impairment.

ACKNOWLEDGMENTS

The authors wish to thank V. Abecassis for his assistance with data collection, T. Nicol for his contributions to the variability analyses and T. Nicol, A. Parbery-Clark, and K. Chan for their constructive comments on a previous version of this manuscript. This work was supported by NIH F31DC011457-01, NSF 0921275, and a Grammy Foundation Research Grant.

REFERENCES

- Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., and Menon, V. (2010). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cereb. Cortex*. doi: 10.1093/cercor/bhq198. [Epub ahead of print].
- Anderson, J., Rennie, C., Gordon, E., Howson, A., and Meares, R. (1991). Measurement of maximum variability within event related potentials in schizophrenia. *Psychiatry Res.* 39, 33–44.
- Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* 273, 1868–1871.
- Bajo, V. M., and King, A. J. (2010). Focusing attention on sound. *Nat. Neurosci.* 13, 913–914.
- Bajo, V. M., Nodal, F. R., Moore, D. R., and King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* 13, 253–260.
- Barry, J. G., Ferguson, M. A., and Moore, D. R. (2010). Making sense of listening: the IMAP test battery. *J. Vis. Exp.* 44.
- Baumann, S., Meyer, M., and Jancke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not

- selective attention. *J. Cogn. Neurosci.* 20, 2238–2249.
- Benasich, A. A., Gou, Z., Choudhury, N., and Harris, K. D. (2008). Early cognitive and language skills are linked to resting frontal gamma power across the first 3 years. *Behav. Brain Res.* 195, 215–222.
- Bidelman, G. M., and Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res.* 1355, 112–125.
- 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(10), 2791–2803.
- Bush, G., Frazier, J. A., Rauch, S. L., Seidman, L. J., Whalen, P. J., Jenike, M. A., Rosen, B. R., and Biederman, J. (1999). Anterior cingulate cortex dysfunction in attention-deficit/hyperactivity disorder revealed by fMRI and the counting Stroop. *Biol. Psychiatry.* 45, 1542–1552.
- Carmena, J. M., Lebedev, M. A., Henriquez, C. S., and Nicolelis, M. A. (2005). Stable ensemble performance with single-neuron variability during reaching movements in primates. *J. Neurosci.* 25, 10712–10716.
- Casey, B. J., Castellanos, F. X., Giedd, J. N., Marsh, W. L., Hamburger, S. D., Schubert, A. B., Vaituzis, A. C., Dickstein, D. P., Sarfatti, S. E., and Rapoport, J. L. (1997). Implication of right frontostriatal circuitry in response inhibition and attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry.* 36, 374–383.
- Coch, D., Sanders, L. D., and Neville, H. J. (2005). An event-related potential study of selective auditory attention in children and adults. *J. Cogn. Neurosci.* 17, 605–622.
- Curto, C., Sakata, S., Marguet, S., Itskov, V., and Harris, K. D. (2009). A simple model of cortical dynamics explains variability and state dependence of sensory responses in urethane-anesthetized auditory cortex. *J. Neurosci.* 29, 10600–10612.
- Depue, B. E., Burgess, G. C., Willcutt, E. G., Bidwell, L. C., Ruzic, L., and Banich, M. T. (2010). Symptom-correlated brain regions in young adults with combined-type ADHD: their organization, variability, and relation to behavioral performance. *Psychiatry Res.* 182, 96–102.
- Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.
- Dixon, M. L., Ruppel, J., Pratt, J., and De Rosa, E. (2009). Learning to ignore: acquisition of sustained attentional suppression. *Psychon. Bull. Rev.* 16, 418–423.
- Faisal, A. A., Selen, L. P., and Wolpert, D. M. (2008). Noise in the nervous system. *Nat. Rev. Neurosci.* 9, 292–303.
- 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.
- Filipek, P. A., Semrud-Clikeman, M., Steingard, R. J., Renshaw, P. F., Kennedy, D. N., and Biederman, J. (1997). Volumetric MRI analysis comparing subjects having attention-deficit hyperactivity disorder with normal controls. *Neurology* 48, 589–601.
- Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678.
- Fritz, J. B., David, S. V., Radtke-Schuller, S., Yin, P., and Shamma, S. A. (2010). Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat. Neurosci.* 13, 1011–1019.
- Gaab, N., and Schlaug, G. (2003). The effect of musicianship on pitch memory in performance matched groups. *Neuroreport* 14, 2291–2295.
- Gaser, C., and Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *J. Neurosci.* 23, 9240–9245.
- Gilliam, M., Stockman, M., Malek, M., Sharp, W., Greenstein, D., Lalonde, E., Clasen, L., Giedd, J., Rapoport, J., and Shaw, P. (2011). Developmental trajectories of the corpus callosum in attention-deficit/hyperactivity disorder. *Biol. Psychiatry.* 69, 839–846.
- Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., and Ceballos-Baumann, A. O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *J. Cogn. Neurosci.* 17, 282–293.
- Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science* 182, 177–180.
- Hutchinson, S., Lee, L. H., Gaab, N., and Schlaug, G. (2003). Cerebellar volume of musicians. *Cereb. Cortex* 13, 943–949.
- Hynd, G. W., Semrud-Clikeman, M., Lorys, A. R., Novey, E. S., and Eliopoulos, D. (1990). Brain morphology in developmental dyslexia and attention deficit disorder/hyperactivity. *Arch. Neurol.* 47, 919–926.
- Janata, P., Tillmann, B., and Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cogn. Affect. Behav. Neurosci.* 2, 121–140.
- Kane, M. J., and Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychon. Bull. Rev.* 9, 637–671.
- Kelley, T. A., and Yantis, S. (2009). Learning to attend: effects of practice on information selection. *J. Vis.* 9, 16.
- Kishon-Rabin, L., Amir, O., Vexler, Y., and Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *J. Basic Clin. Physiol. Pharmacol.* 12(2 Suppl.), 125–143.
- Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Amer.* 67, 13–33.
- Koelsch, S., Gunter, T. C., Cramon, D. Y., Zysset, S., Lohmann, G., and Friederici, A. D. (2002). Bach speaks: a cortical “language-network” serves the processing of music. *Neuroimage* 17, 956–966.
- Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605.
- Makovski, T., Vázquez, G. A., and Jiang, Y. V. (2008). Visual learning in multiple-object tracking. *PLoS ONE* 3, e2228. doi:10.1371/journal.pone.0002228
- Margulis, E. M., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., and Wong, P. C. M. (2009). Selective neurophysiologic responses to music in instrumentalists with different listening biographies. *Hum Brain Mapp* 30, 267–275.
- Marques, C., Moreno, S., Castro, S. L., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463.
- Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202.
- 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.
- Mullins, C., Bellgrove, M. A., Gill, M., and Robertson, I. H. (2005). Variability in time reproduction: difference in ADHD combined and inattentive subtypes. *J. Am. Acad. Child Adolesc. Psychiatry* 44, 169–176.
- Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15894–15898.
- Nilsson, M., Soli, S., and Sullivan, J. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.* 95, 1085–1099.
- Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., and Carlson, S. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PLoS ONE* 5, e11120. doi: 10.1371/journal.pone.0011120
- Pantev, C., Roberts, L. E., Schulz, M., Engelen, A., and Ross, B. (2007). Timbre-specific enhancement of auditory cortical representations in musicians. *Neurorep* 12, 169–174.
- Parbery-Clark, A., Skoe, E., and Kraus, N. (2009a). Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* 29, 14100–14107.
- Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009b). Musician enhancement for speech-in-noise. *Ear Hear.* 30, 653–661.
- 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. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681.
- Petkov, C. I., Kang, X., Alho, K., Bertrand, O., Yund, E. W., and Woods, D. L. (2004). Attentional modulation of human auditory cortex. *Nat. Neurosci.* 7, 658–663.
- 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.
- Rubia, K., Overmeyer, S., Taylor, E., Brammer, M., Williams, S. C., Simmons, A., and Bullmore, E. T. (1999). Hypofrontality in attention deficit hyperactivity disorder during higher-order motor control: a study with functional MRI. *Am. J. Psychiatry* 156, 891–896.
- Schlaug, G., Forgeard, M., Zhu, L., Norton, A., and Winner, E. (2009). Training-induced neuroplasticity in young children. *Ann. N. Y. Acad. Sci.* 1169, 205–208.
- Schmithorst, V. J., and Wilke, M. (2002). Differences in white matter architecture between musicians and non-musicians: a diffusion tensor imaging study. *Neurosci. Lett.* 321, 57–60.
- Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in

- the auditory cortex of musicians. *Nat. Neurosci.* 5, 688–694.
- Schon, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349.
- Seppanen, M., Brattico, E., and Tervaniemi, M. (2007). Practice strategies of musicians modulate neural processing and the learning of sound-patterns. *Neurobiol. Learn. Mem.* 87, 236–247.
- Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., and Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *NeuroImage* 41, 113–132.
- 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.
- Smith, A., and Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *J. Speech Lang. Hear. Res.* 41, 18–30.
- Softky, W. R., and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.* 13, 334–350.
- Steriade, M., Timofeev, I., and Grenier, F. (2001). Natural waking and sleep states: a view from inside neocortical neurons. *J. Neurophysiol.* 85, 1969–1985.
- Stewart, L., Henson, R., Kampe, K., Walsh, V., Turner, R., and Frith, U. (2003). Brain changes after learning to read and play music. *Neuroimage* 20, 71–83.
- Strait, D. L., Chan, K., Ashley, R., and Kraus, N. (2011). Specialization among the specialized: auditory brainstem function is tuned in to timbre. *Cortex*. doi: 10.1016/j.cortex.2011.03.015. [Epub ahead of print].
- Strait, D. L., and Kraus, N. (in press). Playing music for a smarter ear: cognitive, perceptual and neurobiological evidence. *Music Percept.*
- Strait, D. L., Kraus, N., Parbery-Clark, A., and Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear Res.* 261, 22–29.
- Strait, D. L., Kraus, N., Skoe, E., and Ashley, R. (2009). Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *Eur. J. Neurosci.* 29, 661–668.
- Suga, N., Xiao, Z., Ma, X., and Ji, W. (2002). Plasticity and corticofugal modulation for hearing in adult animals. *Neuron* 36, 9–18.
- Suskauer, S. J., Simmonds, D. J., Fotedar, S., Blankner, J. G., Pekar, J. J., Denckla, M. B., and Mostofsky, S. H. (2008). Functional magnetic resonance imaging evidence for abnormalities in response selection in attention deficit hyperactivity disorder: differences in activation associated with response inhibition but not habitual motor response. *J. Cogn. Neurosci.* 20, 478–493.
- Tang, Y. Y., and Posner, M. I. (2009). Attention training and attention state training. *Trends Cogn. Sci.* 13, 222–227.
- Tervaniemi, M., Kruck, S., De Baene, W., Schroger, E., Alter, K., and Friederici, A. D. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *Eur. J. Neurosci.* 30, 1636–1642.
- Vaurio, R. G., Simmonds, D. J., and Mostofsky, S. H. (2009). Increased intra-individual reaction time variability in attention-deficit/hyperactivity disorder across response inhibition tasks with different cognitive demands. *Neuropsychologia* 47, 2389–2396.
- Vogels, R., Spileers, W., and Orban, G. A. (1989). The response variability of striate cortical neurons in the behaving monkey. *Exp. Brain Res.* 77, 432–436.
- Weissman, D. H., Roberts, K. C., Visscher, K. M., and Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nat. Neurosci.* 9, 971–978.
- Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., and Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc. Natl. Acad. Sci. U.S.A.* 90, 8722–8726.
- 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.
- 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.
- 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.
- Zatorre, R. J., Mondor, T. A., and Evans, A. C. (1999). Auditory attention to space and frequency activates similar cerebral systems. *Neuroimage* 10, 544–554.

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: 15 February 2011; accepted: 14 May 2011; published online: 13 June 2011.

Citation: Strait DL and Kraus N (2011) Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Front. Psychology* 2:113. doi: 10.3389/fpsyg.2011.00113

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

Copyright © 2011 Strait and Kraus. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages

Barbara Tillmann^{1,2,3,4*}, Denis Burnham⁴, Sebastien Nguyen^{5,6}, Nicolas Grimault^{1,2,3}, Nathalie Gosselin^{5,6} and Isabelle Peretz^{5,6}

¹ CNRS, UMR5292; INSERM, U1028; Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics Team, Lyon, France

² University Lyon, France

³ University Lyon 1, Villeurbanne, France

⁴ MARCS Auditory Laboratories, University of Western Sydney, Sydney, NSW, Australia

⁵ International Laboratory for Brain, Music and Sound Research (BRAMS), Université de Montréal, Montréal, QC, Canada

⁶ Department of Psychology, Université de Montréal, Montréal, QC, Canada

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Stefan Koelsch, Freie Universität Berlin, Germany

Teppo Särkämö, University of Helsinki, Finland

*Correspondence:

Barbara Tillmann, Lyon Neuroscience Research Center, CNRS, UMR5292; INSERM, U1028, Université de Lyon, Team Auditory Cognition and Psychoacoustics, 50 Avenue Tony Garnier, F-69366 Lyon Cedex 07, France.

e-mail: barbara.tillmann@olfac.univ-lyon1.fr

Congenital amusia is a neurogenetic disorder that affects music processing and that is ascribed to a deficit in pitch processing. We investigated whether this deficit extended to pitch processing in speech, notably the pitch changes used to contrast lexical tones in tonal languages. Congenital amusics and matched controls, all non-tonal language speakers, were tested for lexical tone discrimination in Mandarin Chinese (Experiment 1) and in Thai (Experiment 2). Tones were presented in pairs and participants were required to make same/different judgments. Experiment 2 additionally included musical analogs of Thai tones for comparison. Performance of congenital amusics was inferior to that of controls for all materials, suggesting a domain-general pitch-processing deficit. The pitch deficit of amusia is thus not limited to music, but may compromise the ability to process and learn tonal languages. Combined with acoustic analyses of the tone material, the present findings provide new insights into the nature of the pitch-processing deficit exhibited by amusics.

Keywords: congenital amusia, pitch perception, music processing, tone-language processing

INTRODUCTION

A highly debated question is to what extent music and language share processing components (e.g., Patel, 2008). Our study contributes to this debate by investigating pitch-processing across domains in congenital amusia (or tone-deafness). Congenital amusia refers to a lifelong disorder of music processing that occurs despite normal hearing and other cognitive functions as well as normal exposure to music. We investigated here whether the impaired musical pitch perception typically found in congenital amusia might reflect a domain-general deficit that also affects pitch processing in speech.

Pitch processing is crucial in music, but also in speech processing, notably for discriminating questions and statements, as well as for emotional expressions in non-tone intonation languages (e.g., English, French); while in tone languages (e.g., Mandarin, Thai, Vietnamese), it is used for all these as well as for understanding word meaning. Tone-languages comprise 70% of the world's languages (Yip, 2002) and are spoken by more than 50% of the world's population (Fromkin, 1978). In these languages, tone variation changes (comprising predominantly F₀ height and contour parameters) at the syllabic level have the same effect on word meaning as do vowel and consonant variations in non-tone languages. For examples, see Figure 1. In the present study, we used natural speech samples of tone languages to investigate whether pitch variations in speech might be affected by the previously described musical pitch deficit in congenital amusia.

Expertise or training in a tonal language can facilitate pitch perception and production with musical material: Mandarin, Vietnamese, and Cantonese speakers have been found to be more

accurate at imitating musical pitch and discriminating intervals than English speakers (Pfordresher and Brown, 2009; see also Hove et al., 2010), as can be also reflected in subcortical pitch tracking (e.g., Krishnan et al., 2005). The influence of tone-language background has been mostly observed for relative pitch processing (e.g., intervals), and it might even lead to difficulties in pitch contour processing when non-speech target sounds resemble features of linguistic tones (Bent et al., 2006). However, it has been found that listeners with tone-language background did not differ from listeners with non-tone-language background for absolute pitch discrimination of non-speech sounds (e.g., Bent et al., 2006; Pfordresher and Brown, 2009). Interestingly, in musically trained participants, there is a link between tone-language background and single pitch processing: absolute pitch (i.e., the ability to label a tone without a reference pitch) is more prevalent among tone-language speakers than among non-tone-language speakers (Deutsch et al., 2006, 2009).

Conversely, it has been shown that musical training or expertise can improve pitch perception not only in musical contexts, but also in speech contexts. For example, musicians show improved pitch processing for the prosody of non-tonal language material (Schön et al., 2004; Magne et al., 2006) and for tone-language material, such as Thai tones (Burnham and Brooker, 2002; Schwahnhaüßer and Burnham, 2005) and Mandarin tones (Alexander et al., 2005; Wong et al., 2007; Lee and Hung, 2008; Delogu et al., 2010; Bidelman et al., 2011).

Previous research has thus shown some positive influences between music and speech due to expertise in music or in tone languages, and these effects suggest common pitch-processing mechanisms in

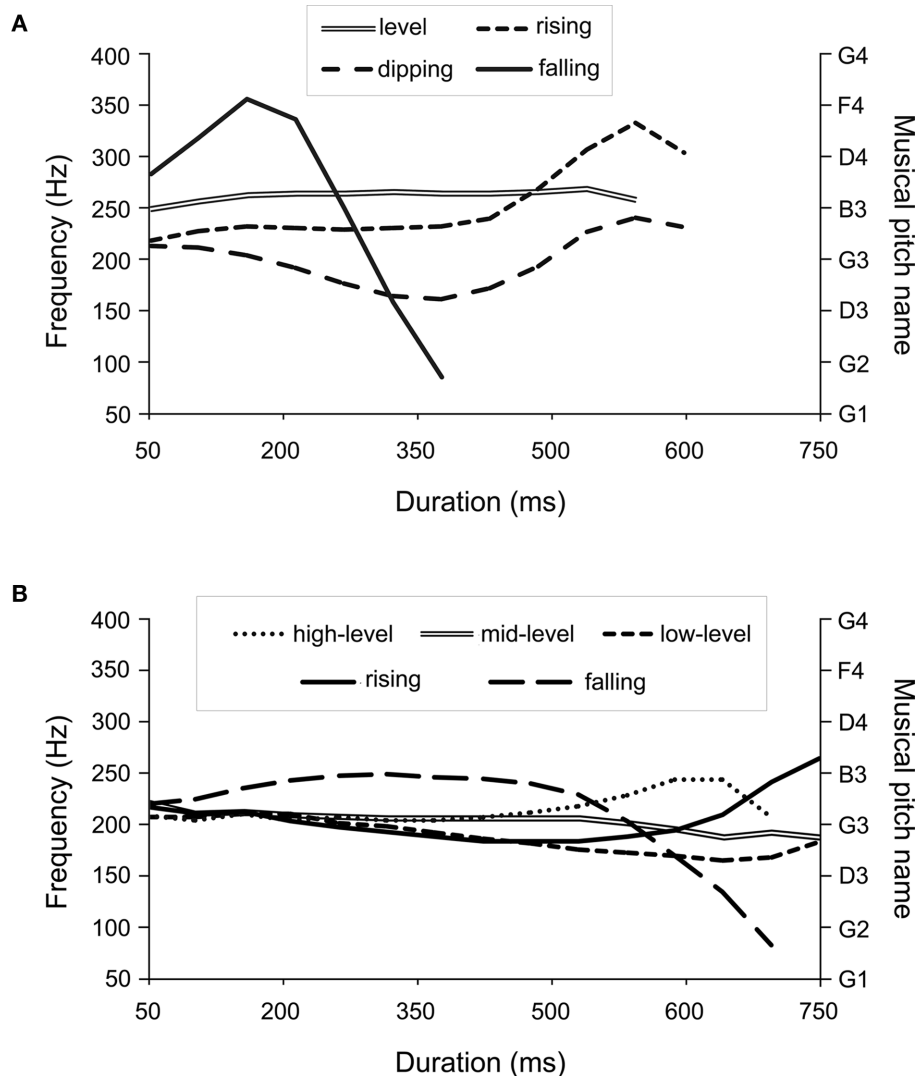


FIGURE 1 | (A) Fundamental frequency contours of the four Mandarin tones (spoken by a female speaker of Mandarin). Each tone on the syllable/ma/ represents a different lexical item. “ma1” is the level tone, “ma2” the rising tone, “ma3” the dipping tone, and “ma4” the falling tone. **(B)** Fundamental

frequency contours of the five Thai tones (spoken by a female speaker of Thai). Each tone on the syllable/ma/ represents a different lexical item. “ma0” is the mid level tone, “ma1” the low level tone, “ma2” the falling tone, “ma3” the high level tone, and “ma4” the rising tone.

music and speech. For example, musical training might shape basic sensory circuitry as well as corticofugal tuning of the afferent system, which is context-general and thus also has positive side-effects on linguistic pitch processing (e.g., Wong et al., 2007). Similar findings suggesting experience-dependent corticofugal tuning have been recently reported for the effects of tone-language expertise on musical pitch processing (Bidelman et al., 2011). In parallel to this previously observed training-related improvement of pitch processing from one domain to the other, the experiments reported here investigate the influence of a pitch perception deficit for music, as observed in congenital amusia, on pitch perception in speech.

Up to recently, congenital amusia has been thought to result from a musical pitch-processing disorder. Individuals with congenital amusia have difficulties recognizing familiar tunes without lyrics and detecting an out-of-key or out-of-tune note. They have

impaired perception of pitch directions for pure tones (Foxton et al., 2004) and for detecting pitch deviations that are smaller than two semitones in sequences of piano notes (Hyde and Peretz, 2004) as well as in note pairs (Peretz et al., 2002). Initial reports have suggested that the deficit was restricted to pitch processing in music, and did not extend to pitch processing in speech material. Individuals with congenital amusia have been reported to be unimpaired in language and prosody tasks, such as learning and recognizing lyrics, classifying a spoken sentence as statement or question based on final falling or rising pitch information (e.g., Ayotte et al., 2002; Peretz et al., 2002).

Peretz and Hyde (2003) suggested that the difference between pitch perception in speech and music is related to the relative size of relevant pitch variations. In speech (of non-tonal languages), pitch variations are typically coarse (e.g., more than 12 semitones

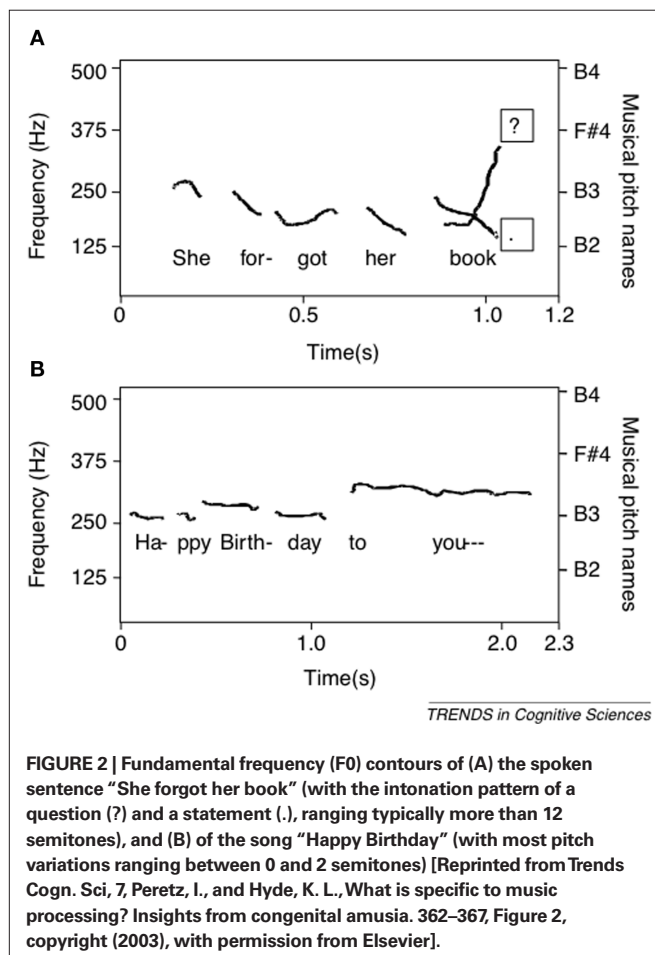
in the final pitch rise indicative of a question; see Fitzsimons et al., 2001), whereas in music, these are more fine-grained (1 or 2 semitones; Vos and Troost, 1989; see **Figure 2**). Accordingly, amusics' pitch deficit would affect music more than speech not because their deficit is music-specific, but because music is more demanding in pitch resolution. Thus, congenital amusia would represent a music-relevant deficit, not necessarily a music-specific deficit. However, when pitch changes of spoken sentences were embedded in a non-speech context (i.e., musical analogs preserving gliding-pitch changes or transforming these into discrete steps), amusics failed to discriminate these pitch changes – in contrast with their high performance level for the same pitch changes in the sentences (Patel et al., 2005). Conversely, recent data have shown that, for some amusic cases, the pitch-processing deficit can also affect the processing of speech intonation in the amusics' mother tongue (Patel et al., 2008; Jiang et al., 2010; Liu et al., 2010). In particular, a slow rate of gliding-pitch change might have deleterious effects on pitch perception in English, but not in French speech (Patel et al., 2008), although the influence of glide rate has not been replicated in a subsequent study for English in British amusics (Liu et al., 2010).

In addition to differences in the size of pitch changes, musical, and linguistic materials differ in their use of discrete, segmented events versus continuous pitch changes (i.e., glides), respectively. Foxton et al. (2004) have shown that congenital amusics have higher thresholds for segmented tones (exceeding one semitone) than for continuous

tone glides (below one semitone). These higher thresholds would affect music perception more markedly, because pitch in music is based upon discrete notes, while the better thresholds for glides might lead to less impairment for pitch processing in speech signals, with its gliding, continuous pitch changes. However, this is unlikely since amusics perform equally poorly on tone analogs of sentences made of pitch glides and discrete events (Patel et al., 2005). In a recent study using discrete segmented events (a tone for the musical material, the syllable/ka/for the verbal material), we have observed that fine-grained pitch discrimination (i.e., 25 cents) can be impaired in amusics not only for musical sounds, but also for verbal sounds. Interestingly, pitch discrimination is better when the pitch is carried by verbal material than by musical material (Tillmann et al., submitted).

Pitch perception in congenital amusia might thus be affected by the size of pitch changes and the nature of the material (verbal, musical). The present study investigates amusics' pitch processing in tonal language material in order to address the question: do congenital amusics show deficits for lexical tone perception, thus for speech material with continuous (rather than discrete) pitch changes, and with pitch changes larger than those that are relevant in music, but smaller than those used in statements/questions in their mother tongue (see **Figures 1 and 2**; Fitzsimons et al., 2001)? We tested French-speaking amusics for their perception of Mandarin tones (Experiment 1) and Thai tones (Experiment 2). We here used monosyllabic words (in contrast to sentences or phrases in previous studies) to keep memory load relatively low, in particular as amusic individuals show impaired short-term memory for pitch information (Gosselin et al., 2009; Tillmann et al., 2009; Williamson et al., 2010). Experiment 2 additionally tested the perception of the same pitch changes in non-verbal, musical analogs. Furthermore, for both Experiments 1 and 2, we present acoustic analyses of the tone-language stimulus materials, and compare the acoustic features of the stimulus materials with participants' behavioral performance, in order to locate the critical acoustic information used by amusic and control participants.

The overall objective of our study is to further understand the nature of the pitch-processing deficit experienced by individuals with congenital amusia, particularly because congenital amusia is now known to have neurogenetic correlates (Drayna et al., 2001; Peretz et al., 2007), and these may not be music-specific, but apply also to speech, as suggested by recent work on amusics' perception of pitch in speech of their mother tongue (e.g., Liu et al., 2010; Nan et al., 2010). Testing the perception of tone-language materials allowed us to use natural speech that contained smaller pitch differences than those occurring in native non-tonal speech (English or French), as in the sentences used in Patel et al. (2005) and Liu et al. (2010). Another advantage of using non-native pitch variations in speech is that, as the words have no meaning for the participants, they are free to respond to the acoustic parameters of the speech without any added complication of semantic significance. This also has the advantage that the non-native tones can be presented as speech that has full speech-shaped spectral information, albeit devoid of semantic significance; and as non-speech – in which the same tones are converted to musical stimuli. In this way, the same acoustic aspects of speech, notably here the pitch information, can be presented in speech and non-speech contexts in which the main difference between speech and more musical non-speech stimuli is the differences in spectral



make-up. In addition, testing amusics who were non-native speakers with speech signals that were tone-language materials also allowed us to aim for converging evidence with findings on tone-language processing recently reported for amusics who were native speakers (Jiang et al., 2010; Nan et al., 2010). Note that previous research has shown that non-native (non-amusic) listeners still engage a speech listening mode (as reflected by the linguistic constraints of their mother tongue) when processing non-native tone-language materials (e.g., Burnham et al., 1996; Burnham et al., submitted).

EXPERIMENT 1

Mandarin Chinese uses four tones characterized by their pitch trajectories, traditionally numbered as tones 1–4: tone 1 is high level, tone 2 is mid-rising, tone 3 mid-dipping (or mid-falling–rising), and tone 4 is high-falling (see **Figure 1A**). Tone 1 has little fundamental frequency (F₀) movement (and so is often referred to as a level tone), whereas tones 2–4 have more F₀ movement (and so often referred to as contour or dynamic tones, Abramson, 1978): tone 2 has a rising F₀ pattern, tone 3 a falling–rising F₀ pattern, and tone 4 a falling F₀ pattern.

Experiment 1 tested native French-speaking congenital amusics¹ for their pitch discrimination with this unfamiliar language material. Even though normal French listeners do not perceive tone contrasts categorically, they are sensitive to tone contour variations (Hallé et al., 2004). A same-different paradigm using monosyllabic Mandarin Chinese words was employed; it was taken from Klein et al. (2001) who showed that normal English or Mandarin speaking participants reached high levels of performance (although English-speakers performed at a level slightly below that of Mandarin speakers, 93 versus 98% accuracy). If the hypothesis of domain-general pitch-processing mechanisms is true, then it can be predicted that amusics' musical pitch deficit should lead to impaired performance in this discrimination task.

METHOD

Participants

The amusic group and the control group each comprised 20 adults who were native French speakers (from Canada and France). The groups were matched for gender, age, education, and musical training (see **Table 1**). All participants completed the Montreal battery of evaluation of amusia (MBEA; Peretz et al., 2003), which is currently widely used in research investigating congenital amusia. The MBEA involves six tests that aim to assess the various components that are known to contribute to melody processing. The stimuli are novel melodic sequences, played one note at a time on a piano; they are written in accordance to the rules of the tonal structure of the Western idiom. These melodies are arranged in various tests so as to assess abilities to discriminate pitch and rhythmic variations, and to recognize musical sequences heard in prior tests of the battery. Peretz et al. (2003) tested a large population and defined a cut-off score (78%) under which participants can be defined as amusics or above which participants were normal. Participants' individual scores for the full battery were below cut-off for the amusic group, but not for the control group (**Table 1**). One amusic participant reported living in China from age 6 to 10; he indicated that he took lessons in Mandarin Chinese but with difficulty and that he spoke either French (he attended a French school) or English, not Mandarin, during that period. Note that 16 out of the 20 amusics have also participated in the experiment testing pitch change detection in verbal and non-verbal materials (Tillmann et al., submitted).

Material

The 98 recordings of monosyllabic Mandarin words (produced by a native female Mandarin speaker) from Klein et al. (2001) were used (see **Table 3** for acoustic descriptors). In all there were 51 different words (13 words with level tone, 10 words with rising

Table 1 | Number of participants per group (amusic/controls) for Experiments 1 and 2, followed by mean age (years), gender distribution, mean education (years), level of musical instruction as well as mean scores obtained on the Montreal battery of evaluation of amusia (MBEA), for the entire test (global score), and the subtests.

	Experiment 1 (Mandarin)			Experiment 2 (Thai)		
	Amusics (<i>n</i> = 20)	Controls (<i>n</i> = 20)	<i>t</i> -Test	Amusics (<i>n</i> = 19)	Controls (<i>n</i> = 19)	<i>t</i> -Test
Age	48.5 (17.2); 19–71	47.9 (17.1); 19–73	<i>p</i> = 0.91	50.5 (15.9); 21–71	49.5 (15.9); 22–72	<i>p</i> = 0.85
Gender	12 F; 8 M	13 F; 7 M	–	11 F; 8 M	12 F; 7 M	–
Education (years)	16.1 (2.5)	15.2 (2.6)	<i>p</i> = 0.28	16.3 (2.3)	15.3 (2.5)	<i>p</i> = 0.19
Musical instruction ¹	1.6 (0.9)	1.4 (0.7)	<i>p</i> = 0.36	1.6 (0.9)	1.5 (0.8)	<i>p</i> = 0.68
MBEA ²	–	–	–	–	–	–
Global	66.4 (6.9)	89.9 (4.2)	<i>p</i> < 0.001	65.8 (6.9)	90.1 (4.1)	<i>p</i> < 0.001
Scale	62.1 (8.3)	90.7 (7.0)	<i>p</i> < 0.001	62.5 (8.0)	91.2 (7.1)	<i>p</i> < 0.001
Contour	61.1 (8.0)	90.5 (7.1)	<i>p</i> < 0.001	60.8 (8.3)	91.1 (6.9)	<i>p</i> < 0.001
Interval	61.4 (9.5)	88.0 (7.0)	<i>p</i> < 0.001	60.8 (9.8)	88.1 (7.4)	<i>p</i> < 0.001
Rhythm	76.2 (14.4)	92.2 (5.8)	<i>p</i> < 0.001	74.6 (14.4)	91.2 (6.9)	<i>p</i> < 0.001
Meter	66.5 (13.3)	88.3 (6.7)	<i>p</i> < 0.001	64.6 (14.0)	88.6 (6.0)	<i>p</i> < 0.001
Memory	71.8 (10.3)	91.2 (6.2)	<i>p</i> < 0.001	71.9 (9.9)	92.1 (6.1)	<i>p</i> < 0.001

Standard deviations are indicated in brackets. For age, the range is also indicated with minimum and maximum age (years).

F, female; M, male; ¹based on an interval scale of musical training (years): 1 = less than a year, 2 = 1–3 years, 3 = 4–6 years, 4 = 7–10 years, 5 = more than 10 years;

²percentage of correct responses.

tone, 15 words with dipping tone, and 13 words with falling tone), with multiple recordings of 25 words for use in same-word pairs (thus leading to acoustic variability between words used in the same-word condition). These words consisted of various consonant–vowel (CV) combinations (e.g., /nju//kual//tʃuən/). Words were presented in 49 pairs: 24 composed of word pairs with the same CV combination but differing in the tone, and 25 composed of different renditions of the same words, and so having the same tone¹. For all participants, word combinations presented in a pair (and word order within each pair) were the same. Each of the 98 recordings was used once in the task. The experiment was run with E-Prime software (Schneider et al., 2002).

Procedure

Within each pair, the first word was followed by a silent period of 350 ms, followed by the second word. Following each pair, listeners were asked to judge whether the two words of the pair were the same or different, by pressing one of the indicated keys on a computer keyboard. Participants were not explicitly told that the relevant dimension for discrimination was pitch. Listeners were first familiarized with the task by means of three practice pairs followed by error feedback, and then moved to the 49 experimental pairs without feedback. After participants' responses, the next pair was presented after a delay of 2 s. The order of presentation of word pairs was randomized for each participant. The experimental session lasted for about 10 min.

RESULTS

Performance was analyzed by calculating proportions of Hits (number of correct responses for different trials/number of different trials) minus False Alarms (FAs; number of incorrect responses for same trials/number of same trials). The amusic group performance was significantly below that of the control group, $F(1,38) = 11.63$, $p = 0.002$. Nevertheless, as can be seen in **Figure 3A**, there was substantial overlap between the groups. Only three amusic individuals performed 2 SD below average control performance. The amusic participant who had spent some time in China reached a performance level of 0.46, i.e., in the lower performance range of the amusic group and within the 2 SD of average control performance.

Correlations² between performance and the six subtests of the MBEA reached significance only for the interval subtest in amusics, $r(18) = 0.47$, $p = 0.038$.

To specify whether the observed group difference was associated with a reduced sensitivity to lexical tone pitch changes, or rather to a propensity to judge a “same” pair to entail a change, we ran two two-sided, independent t -tests for Hits and FAs, respectively. Amusics made fewer hits (0.66) and more FAs (0.09) than controls (0.79 versus 0.05), $t(38) = 2.70$, $p = 0.010$, and $t(38) = 2.23$, $p = 0.031$, respectively. Amusics performed more poorly than controls by both failing to discriminate pairs that were different, and erroneously judging same-word pairs to be different. In general then, it appears that amusics had a less clear grasp of this speech-based pitch discrimination task than did controls.

An additional analysis separated tone pairs as a function of the different tone comparisons (Level–Rising, Level–Dipping, Level–Falling, Rising–Dipping, Rising–Falling, Dipping–Falling). An 6×2 ANOVA on Hit–FA rates with tone comparisons as within-participants factor and Group (amusics/controls) as between-participants factor revealed a main effect of tone comparison [$F(5,190) = 10.81$, $p < 0.0001$, $MSE = 0.05$] and a main effect of group [$F(1,38) = 11.31$, $p = 0.002$, $MSE = 0.17$], but no interaction ($p = 0.39$). Control participants performed generally better than the amusic participants, but both groups showed higher performance level for the pairs comparing level and dipping tones (0.68 for amusics and 0.87 for controls) and for pairs comparing dipping and falling tones (0.71 for amusics and 0.86 for controls) than for the other tone pair comparisons (0.48 for amusics and 0.66 for controls).

DISCUSSION

Experiment 1 revealed that amusics, who were speakers of a non-tonal language (French), encountered difficulties in Mandarin lexical tone discrimination (in comparison to their matched controls). In addition, amusics' performance correlated with their performance in the interval test of the MBEA, which requires the discrimination of tone sequences differing by interval sizes: the lower their performance on the interval test with melodies, the lower their performance in the lexical tone discrimination for Mandarin. These findings support the conclusion that amusics' pitch deficit in melodies extends to the perception of pitch in speech material.

While amusics' average performance was below the group performance of controls, there was considerable overlap in performance ranges between the two groups. The relatively comparable performance of the amusics might be due to some pitch variations in the Mandarin tones (or the comparisons in some of the pairs) that might exceed amusics' thresholds (e.g., larger than two semitones). To increase the difficulty of the task and to assess the generality of the findings, Experiment 2 tested amusics and controls with a same-different paradigm using Thai tones.

EXPERIMENT 2

Standard Thai uses five tones: three level tones (low, mid, high) and two contour tones (rising and falling), referred to as static and dynamic tones respectively (see **Figure 1B**; Abramson, 1962). The tone systems of Thai and Mandarin are different in relation to the number of tones as well as their pitch height, durations and start/end points. They also show some similarities. For example, both Thai and Mandarin have one rising and one falling contour tone and at least one level tone. Thai, however, contains five (not four) different tones, and these tones are based on smaller pitch changes together with weaker contribution of durational differences than in Mandarin. Previous studies have shown that English-speaking children and adults can discriminate Thai tones (Burnham et al., 1996; Burnham and Francis, 1997; Burnham and Brooker, 2002), and Experiment 2 tested for the first time congenital amusics on this material.

In addition, to further our understanding of domain-generality versus -specificity of pitch processing, Experiment 2 compared amusics' performance for the lexical tones to their performance on musical analogs thereof (pitch variations of the Thai tones applied

¹One “different” pair had to be discarded due to distortion in the acoustic signal.

²All reported correlations were Pearson correlation coefficients that were tested with two-tailed tests.

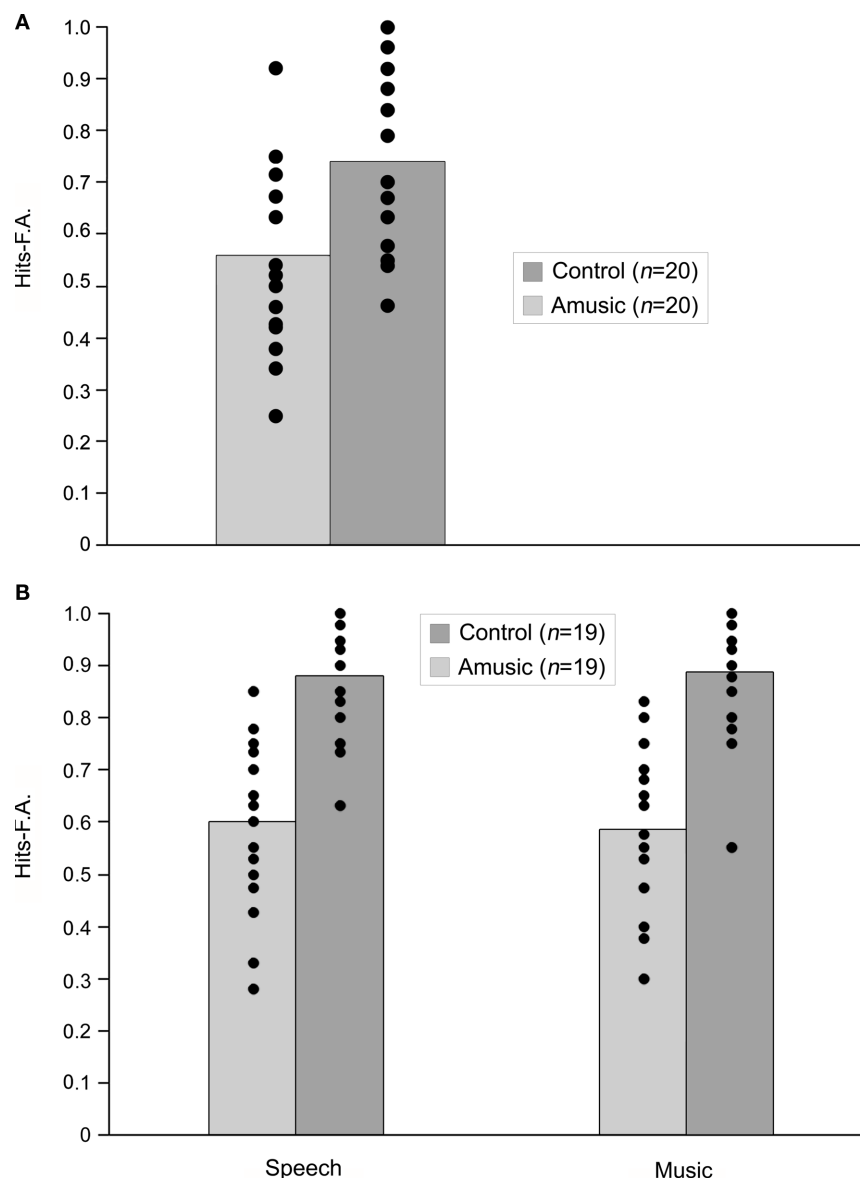


FIGURE 3 | Performance [expressed as Hits – False Alarms (FA)] for amusic and control participants in Experiment 1 with the Mandarin material (A) and in Experiment 2, with the Thai material (speech) and its musical analogs [music; (B)]. Dots represent individual participants.

to a violin sound). Previous studies have shown that (1) normal, English-speaking listeners performed worse for the speech signals than for musical analogs or low-pass filtered versions of the speech signals (Burnham et al., 1996; Burnham and Brooker, 2002), and (2) musical training boosted overall performance levels, with musicians without absolute pitch performing better than non-musicians, and musicians with absolute pitch performing the best (Burnham and Brooker, 2002). Based on these positive transfer effects of musical training, Burnham and Brooker (2002) concluded that speech and music perception are not independent, and that musical training and absolute pitch ability can affect speech processing. Thus, amusic's musical deficit would predict impaired processing for pitch in the speech material here, in line with data of Experiment 1 as well

as some previous work on amusic's perception of their mother tongue, whether non-tonal (Patel et al., 2008; Liu et al., 2010) or tonal (Nan et al., 2010).

Impaired pitch processing for speech material has been also reported for a task requiring fine-grained pitch change detection in sequences of a repeated syllable (/ka/) (Tillmann et al., submitted). Even though impaired, amusic's performance was less impaired for these syllable sequences than for sequences with repeated tones (carefully matched to the syllables for their acoustic features). This performance difference between speech and musical sounds might be linked to differences in the energy distribution of the sounds' spectrum, notably by the presence versus absence of formants, and/or to higher-level processing related to strategic influences (see

below for further discussion). In the Tillmann et al. (submitted) study, the to-be-detected pitch changes were instantiated between syllables (or tones), thus between segmented events, and not within a given event with continuous pitch changes as in the materials of Experiments 1 and 2 here. In Experiment 2, we thus hypothesized that for the processing of pitch in Thai tones, amusics' performance should benefit from the speech signal, leading to some boost in pitch processing (compared to musical analogs), at least for the most severely impaired amusics (Tillmann et al., submitted). Controls, however, should perform better for the musical material (Burnham and Brooker, 2002; Tillmann et al., submitted).

METHOD

Participants

The amusic group and the control group each comprised 19 French-speaking adults (from Canada and France); 18 of the amusics and 17 of the controls also participated in Experiment 1 with the Mandarin tones. As in Experiment 1, the groups were matched for gender, age, education, and musical training, with MBEA scores below cut-off for the amusic group only (Table 1). Note that 14 out of the 19 amusics tested have also participated in the experiment testing pitch change detection in verbal and non-verbal material (Tillmann et al., submitted).

Materials

Five tokens of /ba/ for each of the five Thai tones, recorded by a female speaker, were taken from Burnham et al. (1996) (see Table 3 for acoustic descriptors). Instead of using musical stimuli that were created by a professional musician imitating the lexical tones on the violin as in Burnham et al. (1996), the musical analogs were created by applying the pitch contour, temporal envelope, duration, and intensity of the Thai tones to a violin sound (a steady-state violin sound of an original duration of 1 s, which was then shortened to that of the tones). First, overall duration, pitch contour, and temporal envelope (computed as the half-wave rectified signal low-pass filtered at 80 Hz) were extracted from each of the verbal tokens using STRAIGHT (Kawahara et al., 1999). Then, duration and pitch contour of the verbal sound were applied to the violin sound with STRAIGHT. Finally, the temporal envelope and the RMS value of the verbal sound were applied to the transformed violin sound. Overall, 25 musical sounds, corresponding to the 25 verbal sounds (five tokens for each of the five Thai tones) were generated. For each type of material (verbal, musical), 40 same pairs and 40 different pairs were created. For the 40 same pairs, 8 pairs were created for each of the five tones. For the 40 different pairs, each tone was presented with one of the other four tones four times. Same pairs consisted of different tokens of the same tone and over all pairs, different tokens were used across participants. The experiment was run with PsyScope software (Cohen et al., 1993).

Procedure

Within each pair, the first item was followed by a silent period of 500 ms, followed by the second item (as in Burnham et al., 1996; Burnham and Brooker, 2002). For each pair, listeners were asked to judge whether the speaker (or the musician) pronounced (played) the two syllables (notes) in the same way or in a different way. As in Experiment 1, no explicit references were made to the pitch

dimension, the changes were referred to as "different ways of pronouncing the syllable" or "different ways of producing the violin sound." Participants indicated their answers by pressing one of two keys on a computer keyboard. Listeners were first familiarized with the materials by two example pairs (one different pair, one same pair), which could be repeated for clarification if required. The 80 verbal and 80 musical pairs were each separated into two blocks. The four resulting blocks were presented in either the order verbal-musical-musical-verbal or the order musical-verbal-verbal-musical, counterbalanced across participants. Within each block, trials were presented in randomized orders for each participant. No feedback was given for the experimental trials, and the next trial started when participants pressed a third key. The experimental session lasted for about 20 min.

Pretest

A pretest was run to confirm that the musical material newly constructed for Experiment 2 replicated the result pattern of Burnham et al. (1996). Twenty English-speaking students from the University of Western Sydney participated in the pretest (mean age: 23.2 ± 8.73), with average instruction on a musical instrument of $0.33 \text{ years} \pm 0.73$ and a median of 0 years. Performance was analyzed by calculating proportions of Hits (number of correct responses for different trials/number of different trials) minus FAs (number of incorrect responses for same trials/number of same trials). Results replicated better performance for musical material (0.85 ± 0.03) than for verbal material (0.78 ± 0.04), $F(1,19) = 4.37$, $p = 0.05$, as previously observed by Burnham et al. (1996), Burnham and Brooker (2002), Burnham and Francis (1997) in non-musicians. In addition, as in previous data, performance reflected the degree of acoustic changes between the sounds of a pair: performance was better for pairs combining two contour tones (rising, falling) than for pairs with two level tones (low, mid, high) or mixed pairs, $F(2,38) = 6.68$, $p = 0.003$.

RESULTS

As in Experiment 1, proportions of Hits (number of correct responses for different trials/number of different trials) and FAs (number of incorrect responses for same trials/number of same trials) were calculated. These proportions were analyzed by a 2×2 ANOVA with Material (verbal, musical) as a within-participant factor and Group (amusics, controls) as a between-participants factor. Only the main effect of group was significant, $F(1,36) = 46.08$, $p < 0.0001$: the amusic group performed below the level of the control group, although there was substantial overlap between the groups (see Figure 3B). Twelve amusics for the verbal material and 11 amusics for the musical material performed 2 SD below average control performance. The amusic who had lived in China (see Experiment 1) had performance levels of 0.48 and 0.40 for verbal and musical materials, respectively. No other effects were significant, $ps > 0.36$. Thus, performance did not differ significantly for verbal and musical materials, either for amusics or controls.

For the verbal material, the only correlation between performance and scores in the subtests of the MBEA was for the meter subtest in controls, $r(17) = 0.47$, $p < 0.04$. In amusics, the correlation with the interval test was marginally significant, $r(17) = 0.41$, $p < 0.08$, in agreement with the findings for the Mandarin material

in Experiment 1. For the musical material, no correlations were significant, although the correlation with the meter subtest for controls was marginally significant, $r(17) = 0.45$, $p < 0.053$. Correlation for performance between verbal and musical tasks were significant both for amusics, $r(17) = 0.64$, $p < 0.003$, and for controls, $r(17) = 0.91$, $p < 0.0001$.

As for Experiment 1, we also analyzed group differences with two-tailed, independent t -tests for Hits and FAs, respectively (Table 2). As in Experiment 1, amusics made fewer Hits and more FAs than controls for both the verbal material, $t(36) = 6.13$, $p < 0.0001$, and $t(36) = 3.39$, $p = 0.001$, respectively, and the non-verbal material, $t(36) = 5.86$, $p < 0.0001$, and $t(36) = 2.75$, $p = 0.009$, respectively.

As in Burnham et al. (1996), we separated performance (Hits–FAs) for pairs combining the two contour tones (i.e., rising, falling), or two of the level tones (i.e., low, mid, or high) or mixed pairs with one contour tone and one level tone (Table 2). Performance was analyzed in a $3 \times 2 \times 2$ ANOVA with Type (contour–contour, level–level, contour–level) and Material (verbal, musical) as within-participants factors and Group (amusics, controls) as a between-participants factor. The main effects of group and of type were significant, $F(1,36) = 40.85$, $p < 0.0001$, and $F(2,72) = 62.05$, $p < 0.0001$, respectively, as was their interaction, $F(2,72) = 23.38$, $p < 0.0001$. Overall amusics performed worse than controls; this difference was smaller for pairs with two contour tones, while still being significantly below controls, $F(1, 36) = 5.11$, $p = 0.03$. Even though the three-way interaction between Group, Type, and Material was not significant ($p = 0.38$), we ran two additional 3×2 ANOVAs with Type and Material as within-participants factors in each group separately, with the goal to further investigate amusics' and controls' sensitivity to the type of comparison pairs. Only the main effect of type was significant in both amusics, $F(2,36) = 57.54$, $p < 0.0001$, $MSE = 0.02$, and controls, $F(2,36) = 7.76$, $p = 0.002$, $MSE = 0.01$. For controls, performance was better for contour–contour pairs than for contour–level ($p = 0.003$) and level–level pairs ($p = 0.01$), while these latter two did not differ ($p = 0.23$). For amusics, performance was also better for contour–contour pairs than for contour–level pairs ($p < 0.0001$) and level–level pairs ($p < 0.0001$), but in addition, amusics were sensitive to the difference between these two latter pairs, with lower performance

for the level–level pairs ($p = 0.004$). Note that for the Level–Level pairs, the advantage of the verbal material over the musical material observed in the overall mean performance was not significant in amusics $p = 0.24$. No other effects were significant, $ps > 0.13$. The relative difficulty of the pair types was also reflected in the number of amusics performing 2 SD below average control performance: only 2 amusics (out of 19) performed below average control performance for contour–contour pairs, but 12 for contour–level and level–level, respectively (i.e., for verbal materials).

The analyses of the entire amusic group, reported above, did not show a significant advantage of the verbal material over the musical material, either for the overall material set or for pairs comparing two level tones. Following Tillmann et al. (submitted), in which the advantage of verbal over musical material was mostly observed for amusics who had severe pitch deficits, we divided the amusics based on their pitch thresholds. Fifteen of the 19 amusics had previously participated in a pitch perception threshold test³; their overall average threshold was 1.49 semitones (± 1.13), but with thresholds ranging from 0.13 to 4. We separated amusics into two groups: eight amusics with thresholds below one semitone (mean of 0.68, ranging from 0.13 to 0.97) and seven amusics with thresholds above one semitone (mean of 2.41, ranging from 1.3 to 4).

For amusics with thresholds above one semitone, performance was significantly better for the verbal material than the musical material for the entire material set (0.55 versus 0.45, $p = 0.03$) and for pairs comparing two level tones (0.52 versus 0.29, $p = 0.04$). For both comparisons, six of the seven amusic showed higher mean performance for the verbal material. This advantage was not observed for amusics with thresholds below one semitone, either for the entire material set (0.60 versus 0.65, $p = 0.38$) or for the level–tone pairs (0.51 versus 0.52, $p = 0.90$). To further investigate this group comparison, we run a 2×2 ANOVA with the two subgroups of amusics as between-participants factor and material (verbal, non-verbal) as within-participants factor. Amusics with thresholds below 1 ST tended to perform better than amusics with thresholds above one semitone, $F(1,13) = 3.49$, $p = 0.08$, $MSE = 0.03$. The main effect of material was not significant ($p = 0.38$), but material interacted with group, $F(1,13) = 5.77$, $p = 0.03$, $MSE = 0.007$: for amusics with thresholds below one semitone, performance did not differ between the two sets of materials ($p = 0.30$), while for amusics with thresholds above one semitone, performance was better for the verbal material than for the musical material ($p = 0.04$). When the amusics with threshold below one semitone were directly compared to the control group in an additional ANOVA, the interaction between group and material was not significant ($p = 0.41$). Control participants performed better than amusic participants, as shown by the main effect of Group, $F(1,25) = 33.28$, $p < 0.0001$, $MSE = 0.02$. In addition, overall performance tended to be better for the musical material than the verbal material, even though the effect of material failed to reach significance, $F(1,25) = 3.21$, $p = 0.09$, $MSE = 0.004$.

³To determine pitch perception thresholds, a two-alternative forced-choice task was employed with adaptive tracking using a two-down, one-up staircase procedure. Participants were presented with two pairs of pure tones: in one pair the tones had the same pitch, and in the other pair the tones differed in pitch. They were asked to decide whether the first or the second pair contained a pitch difference (see Tillmann et al., 2009, for details).

Table 2 | Discrimination performance for amusic and control participants for the Thai tones (speech material) and their musical analogs (music) in Experiment 2. The two top lines present Hits and False Alarms. The three bottom lines present Hits minus False Alarms separated for pairs comparing two types of contour tones (rising, falling), two types of level tones (low, mid, high) or one contour tone with a level tone.

	Amusic group		Control group	
	Speech	Music	Speech	Music
Hits	0.71 (0.03)	0.70 (0.04)	0.92 (0.02)	0.94 (0.02)
False alarms	0.11 (0.02)	0.12 (0.02)	0.04 (0.01)	0.05 (0.02)
Contour–contour	0.86 (0.04)	0.86 (0.04)	0.95 (0.02)	0.94 (0.02)
Level–level	0.54 (0.04)	0.47 (0.05)	0.86 (0.03)	0.86 (0.04)
Contour–level	0.58 (0.04)	0.60 (0.04)	0.87 (0.02)	0.90 (0.02)

SE are indicated in brackets.

Note that in the 15 amusics considered, thresholds correlated with performance level for musical material, $r(13) = -0.62, p < 0.02$, but not for the verbal material, $r(13) = -0.29$.

DISCUSSION

Experiment 2 tested amusics' perception of Thai tones and their musical analogs in a same-different paradigm. Findings supported those of Experiment 1 on Mandarin tones and suggest that amusics' pitch deficit extends to the perception of pitch in speech material. As in Experiment 1, the French-speaking amusics encountered difficulties in lexical tone discrimination, and their performance tended to correlate with their score on the interval test of the MBEA. While we observed again an overlap in performance ranges in amusic and control groups, considerably more amusics performed below the controls' distribution (i.e., 12 out of the 19 amusics for the verbal material). This observation suggests that for amusics the task with the Thai material was more difficult than the task with the Mandarin material. This might be due to the larger set of tones used (five instead of four), the smaller pitch range covered by Thai tones (see **Figure 1B**), the more standardized material solely using the syllable/ba/(instead of using a range of CV syllables), or the fact that Thai tones do not vary in duration as much as do Mandarin tones. In addition, the delay between the to-be-compared syllables was slightly longer in Experiment 2 than in Experiment 1 (500 versus 350 ms). In this regard, the recently reported memory deficit of amusics for pitch material might thus have contributed to make the task with the Thai materials more difficult (e.g., Gosselin et al., 2009; Tillmann et al., 2009; Williamson et al., 2010). However, even if we cannot exclude the contribution of any pitch memory deficit in amusics, its contribution should be rather minor as the delays in our tasks were considerably shorter than the delays tested in the previous studies (e.g., between 1 and 15 s in Williamson et al., 2010).

The analyses separating pairs as a function of tone categories (i.e., pairs comparing contour tones only, level tones only or mixed pairs) revealed amusics' sensitivity to the acoustic features in the presented material. As for controls in the present experiment as well as participants in previous studies (see Burnham and Brooker, 2002), amusics performed better with pairs that required the comparison of two contour tones, which involve larger acoustic differences than the comparison of two level tones, for example.

Experiment 2 tested amusics' perception not only with lexical tones, but also with musical analogs of these tones. For amusics, the findings support previous conclusions for discrete pitches (Tillmann et al., submitted): even though amusics appeared impaired overall for speech and musical materials, the amusics with the largest pitch deficits benefited from the speech material, leading to improved performance. In controls, the reverse pattern (worse performance for the speech materials) was only observed in the Australian-English language student group in the pretest, but not for the French-language matched control group (even though the mean performance difference pointed in the expected direction, but $p = 0.14$). This difference between the control group and participants in our pretest as well as participants in Burnham and Brooker (2002) might be attributed to the fact that our controls were French-speaking, while participants in the pretest and the study by Burnham and Brooker were English-speaking (see Patel

et al., 2008, for pitch-processing differences in French- and English-speaking participants). The comparison of performance patterns (0.88 and 0.89 Hits-FAs for matched controls for verbal and musical material, versus 0.60 and 0.80 in Burnham and Brooker, 2002) suggest that both language groups perform similarly for the music material, while English-speaking participants show lower performance level for the speech materials than did the French-speaking participants. This might be linked to the observation that native English-speakers disregard supra-segmental cues to stress in word recognition (e.g., Cutler 2009), which might thus attenuate their performance for the speech material here⁴. Despite these differences between English and French controls, the important point here is that the amusics with greater pitch deficits *benefitted* from speech material, whereas for controls, be they French or English speakers, the reverse is the case – there is a benefit for the non-speech, musical materials.

ACOUSTIC CORRELATES OF PERFORMANCE

In order to investigate the acoustic information used by amusic and control participants in Experiments 1 and 2, we analyzed Mandarin and Thai tones for the information contained in the pitch dimension (F_0 mean, slope and movement, and duration of the voiced pitch component of the syllables) as well as the overall sound duration of each word and mean intensity. We added these latter non-pitch features into the analyses as we reasoned that amusic participants might use these alternative cues to aid their discrimination. The purpose of these analyses was to estimate the difficulty of the same-different task for the experimental pairs on the basis of the acoustic differences for the stimuli. Accordingly, we calculated for each acoustic feature the distance between the two items presented in each pair. If listeners use a given acoustic feature in their judgments, then larger differences on this acoustic dimension between the items should lead to higher accuracy, and smaller differences to lower accuracy.

ACOUSTIC ANALYSES

All acoustic analyses were conducted in a Matlab computing environment. STRAIGHT (Kawahara et al., 1999) was used to compute pitch contours for all stimuli. All pitch contours were subsequently fitted in a least squares sense with a 4th-degree polynomial to avoid fast pitch variation and to capture the overall shape of the pitch contour. Prior visual inspection of the pitch contours suggested the fitting of this polynomial degree: it allowed for four inflexion points or three reversals in the pitch contour. These pitch contours are displayed both individually and as average pitch contours in **Figure 4**. Note that the degree of similarity between tokens in the Thai material (all based on the syllable /ba/) was understandably higher than that between the items of the Mandarin material (which were based on various CV syllables).

For each sound file, the following steps were performed. First, the duration of the whole spoken sound was extracted. Second, several parameters were computed from the fitted pitch contour

⁴Note that for the more artificial speech material (synthesized syllable /ka/ and its matched musical analog in Tillmann et al., submitted), the French speakers showed a disadvantage for pitch change detection solely for the smallest pitch changes used (i.e., 25 cents), but not for the larger changes for which performance was overall high.

for each wave file: the *duration* of the pitch contour, the mean *pitch* (average F₀ over the entire syllable), the mean *slope* (in semitone/s), which provided an estimate of the direction of variation of the pitch contour, and the mean of the *absolute value of the slope* of the pitch contour (in semitone/s), which provided an estimate of the overall pitch movement. Third, the mean *RMS* value was calculated using an arbitrary dB scale. These values were then averaged for each of the linguistic categories for Mandarin tones (level tone, rising, dipping, falling) and for Thai tones (low, mid, high, rising, falling) and are presented in **Table 3**.

“ACOUSTIC DISTANCES” BETWEEN THE PAIR MEMBERS AND LINK TO BEHAVIORAL DATA

Calculation of acoustic distances

For each acoustic feature, we calculated the “acoustic distance” between the two items of a pair using Euclidean distance computation. For the Mandarin tones, for which each participant was presented with the exact same set of tone pairs with only order of trials differing between participants, we calculated the distances between the items of each different pair as well as each same pair (i.e., consisting of different recordings of the same word, see

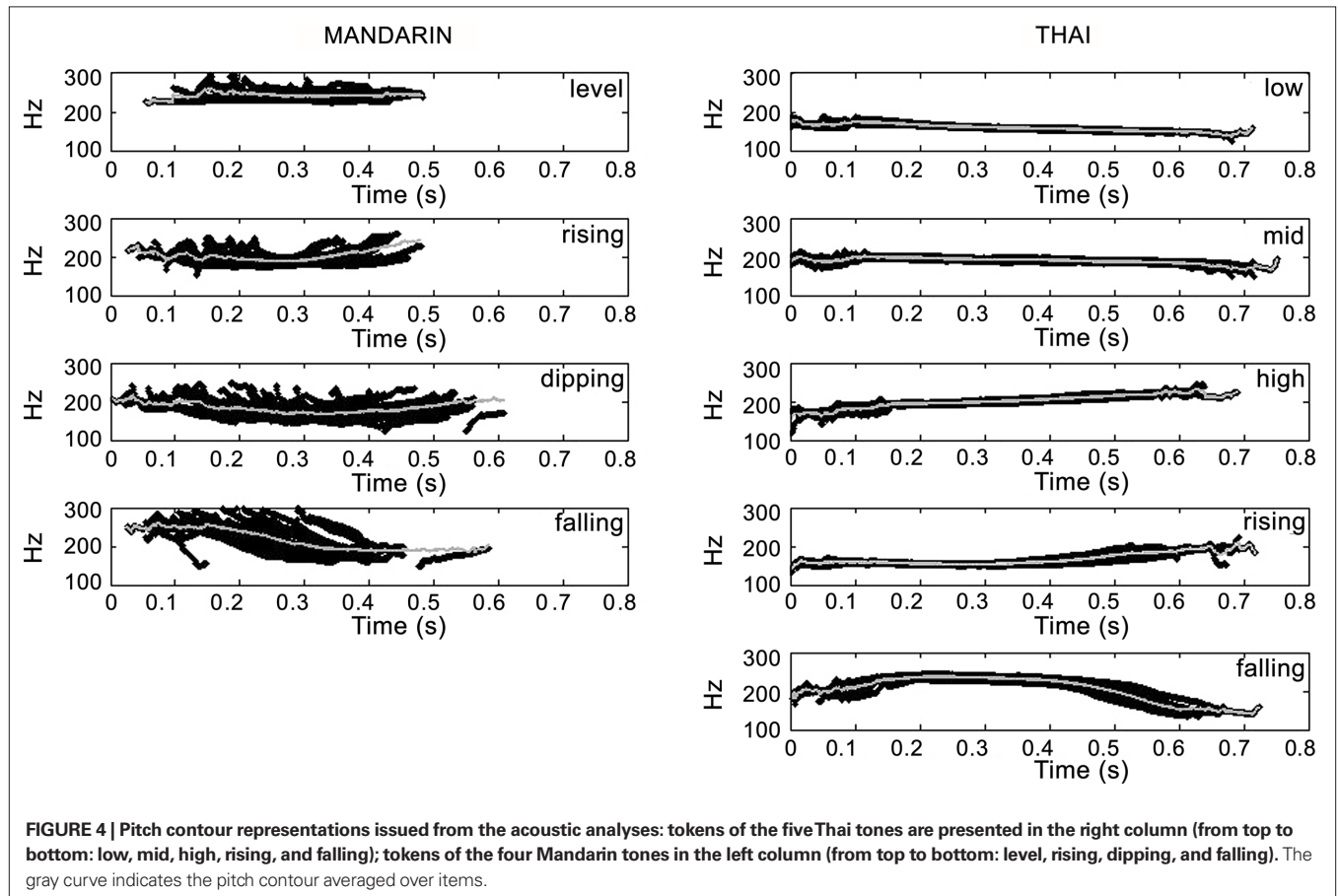


FIGURE 4 | Pitch contour representations issued from the acoustic analyses: tokens of the five Thai tones are presented in the right column (from top to bottom: low, mid, high, rising, and falling); tokens of the four Mandarin tones in the left column (from top to bottom: level, rising, dipping, and falling). The gray curve indicates the pitch contour averaged over items.

Table 3 | Mean values of the acoustic descriptors for the four Mandarin tones and the five Thai tones.

Language	Tone	Pitch, F ₀ (Hz)	Slope (ST/s)	Absolute slope (ST/s)	Pitch duration (s)	Sound duration (s)	RMS (arb dB)
Mandarin	Level	243.41 (8.03)	−4.90 (3.95)	13.18 (5.26)	0.28 (0.03)	0.41 (0.04)	47.05 (2.80)
	Rising	204.64 (5.31)	8.35 (3.74)	28.80 (7.20)	0.32 (0.03)	0.38 (0.05)	47.78 (3.34)
	Dipping	184.03 (5.94)	−0.85 (4.44)	28.75 (6.15)	0.40 (0.07)	0.51 (0.05)	45.61 (2.29)
	Falling	212.22 (5.56)	−14.42 (10.53)	32.21 (7.70)	0.30 (0.07)	0.37 (0.05)	47.33 (3.24)
Thai	Low	158.99 (1.42)	−4.38 (1.01)	5.20 (0.81)	0.66 (0.05)	0.67 (0.06)	66.24 (2.84)
	Mid	188.75 (3.94)	−3.07 (1.35)	5.99 (1.88)	0.71 (0.03)	0.72 (0.02)	69.01 (1.53)
	High	199.64 (2.31)	9.91 (2.58)	10.39 (2.48)	0.61 (0.02)	0.66 (0.06)	65.53 (2.73)
	Rising	164.53 (1.71)	6.58 (0.74)	10.64 (1.56)	0.64 (0.06)	0.75 (0.05)	62.56 (1.13)
	Falling	209.04 (3.76)	−9.30 (1.78)	23.75 (1.79)	0.63 (0.02)	0.69 (0.05)	68.47 (1.64)

Mean values were calculated over all items in one linguistic category. For Mandarin, there were 23 items for level tone and rising tone, 25 for the dipping tone, and 27 for the falling tone. For Thai, there were five items in each category. SD are indicated in brackets. ST, semitones; arb dB, arbitrary dB units.

Method of Experiment 1). For the Thai tones, we calculated acoustic distances using the average values over the five tokens because the specific associations of tokens for tones varied across participants. Consequently, distances were calculated for different pairs only, as the average distance in the same pairs would be 0.

Link with behavioral performance

To investigate whether acoustic distances correlated with participants' performance, we calculated mean performance (percentage correct) for the 24 same pairs and the 25 different pairs with Mandarin tones and for the 10 different pairs with Thai tones in each set of material (verbal or musical). Mean performance over item pairs was then correlated with mean distances for each acoustic descriptor.

For Mandarin "same pairs," negative correlations were observed in amusics for mean pitch [$r(23) = -0.44, p < 0.05$] and pitch duration [$r(23) = -0.44, p < 0.05$]: the larger the changes in pitch and in pitch duration, the lower their performance. This suggests that amusics detected some changes in mean pitch and pitch duration, and that these changes led them to respond "different" and thus to err. In contrast to amusics, controls showed such a "distraction" effect only for the absolute slope [$r(23) = -0.49, p < 0.05$]. For Mandarin "different pairs," no correlations were significant.

For Thai "different pairs," positive correlations ($ps < 0.05$) were observed in both amusics and controls for distances in mean pitch [amusics: $r(8) = 0.79$ (verbal), $r(8) = 0.74$ (musical); controls: $r(8) = 0.80$ (verbal only)] and in absolute slope [amusics: $r(8) = 0.64$; controls: $r(8) = 0.75$, both for musical materials]: the larger the changes in pitch and in absolute slope, the higher participants' performance for the different pairs. Finally, it is interesting to point out that the slope seems to have been used more strongly by controls than amusics in the musical material [$r(8) = 0.54$ for

controls; $r(8) = 0.09$ for amusics], although this correlation in controls did not reach significance, which might be related to controls' overall high performance level.

In sum, these correlational analyses suggest that amusics' performance was influenced by changes in pitch parameters, notably mean pitch or absolute slope (i.e., pitch movement). While this information correlated positively and thus facilitated performance for the Thai material, which was based on the same CV syllable, it was misleading for the Mandarin material. In addition, the amusic participants also erroneously used the duration of the pitch information in Mandarin, which resulted in attenuated performance. As Mandarin tones do differ considerably in duration, it appears that the amusics focused on this cue (given their impaired sensitivity to pitch), but that this cue was ultimately misleading.

To further investigate the link between acoustic distances and behavioral performance, we performed an additional analysis for the Thai material (both verbal and musical). For this, we related for each participant the acoustic distances to performance differences as follows: we selected among the 10 different tone pairs, the three pairs with the largest acoustic distance and the three pairs with the smallest acoustic distance, for each of the descriptors. For these subgroups of test items, we computed mean performance (in Hits–FAs) for each participant. **Figure 5** represents the performance difference between stimulus pairs with large and with small acoustic distances. Positive differences represent better performance when the items of a pair differing markedly on the acoustic descriptor. Negative differences represent better performance when the items of a pair differed only weakly on the acoustic descriptor. Two-tailed t -tests on amusics' data showed that the observed differences were significantly different from 0 for all predictors and both the verbal and the musical materials ($ps < 0.05$), except for pitch duration in both speech and musical materials ($ps > 0.42$) and for RMS in the musical material ($p = 0.11$).

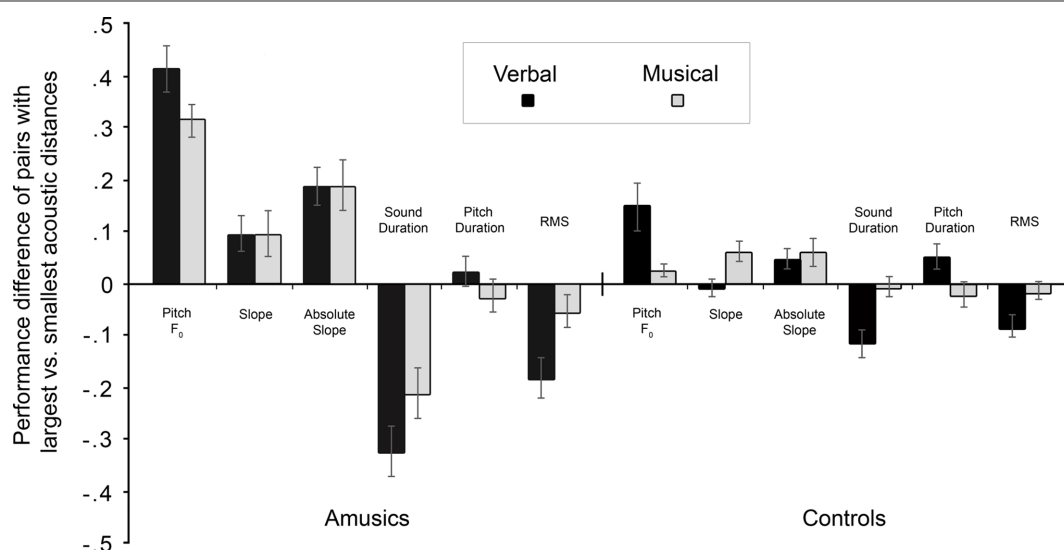


FIGURE 5 | Performance differences of amusic participants (left) and the control participants (right) for the Thai material (speech and musical analogs) between trials selected for large acoustic differences and trials selected for small acoustic differences. Positive values represent better performance when the items of a pair differed strongly on a given acoustic

descriptor. Negative values represent better performance when the items of a pair differed only weakly on a given descriptor. Note that controls' data were presented here only for the sake of completeness as control participants reached relatively high performance levels, thus leaving little room for measuring changes as a function of acoustic differences.

The difference graph (**Figure 5**) shows that amusics benefited from large differences in mean pitch and in absolute slope, but less so from differences in slope. This might be due to amusics' previously described difficulty in using pitch direction information (see Foxton et al., 2004). For total sound duration and for intensity, however, smaller differences led to better performance (or larger differences led to worse performance). When the two items in a pair differed markedly on these features (i.e., these were misleading cues), amusics seemed to use this rather irrelevant information and thereby perform less accurately. When, however, the two items differed only minimally on these features, the strategy to base judgments on these cues (duration, intensity) was less disruptive. In such cases, amusics might then also consider other features that might be more difficult for them to perceive (but that are more relevant, such as pitch), leading to better performance.

GENERAL DISCUSSION

The present study investigated pitch perception in congenital amusia for tone-language materials (Experiments 1 and 2) and musical analogs (Experiment 2). Our goal was to investigate whether the previously described musical pitch-processing deficit of congenital amusics might also impair pitch processing in tone-language speech signals. Overall, our findings suggest that this is indeed the case: amusics showed impaired performance for lexical tones. This finding is consistent with recent studies showing mild deficits of amusics in speech intonation discrimination, identification, and imitation in their native language, notably British English (Liu et al., 2010) and Mandarin Chinese (Jiang et al., 2010). In Mandarin Chinese, Nan et al. (2010) further showed that a subgroup of native speakers who are amusic shows impaired identification of lexical tones.

Our results show that amusics' pitch deficit can extend to non-native, meaningless speech material, that is when no semantic content might distract listeners' attentional focus away from the task-relevant pitch information. The use of non-native speech material (Mandarin, Thai) also allowed us to investigate French native speakers (who are amusic) with natural speech material that required the processing of pitch changes smaller than those used in previous intonation tasks (Patel et al., 2008; Liu et al., 2010). The comparison of the observed deficits for Mandarin and Thai materials, respectively, further shows the impact of the size of the to-be-processed pitch changes on amusics' performance (see below). In addition, it is worth noting that the deficit was here observed with an experimental task that required the comparison of two syllables rather than two sentences (as in Patel et al., 2005; Liu et al., 2010), thereby decreasing memory load. Our results further reveal that amusics' deficit is less important for speech than for music, at least in the most severe cases of amusia. Combined with the acoustic analyses, the present study provides new insights into the nature of the pitch-processing deficit exhibited by amusics as set out below.

Amusics' pitch deficit was first documented with musical material (thus leading to the term "amusia"). Subsequent tests of the consequences of this pitch deficit on speech processing have implications for the understanding of the overall phenomenon of what has been labeled "amusia": the pitch deficit does not seem to be domain-specific and restricted to musical material, but is rather a domain-general deficit that was first discovered in a musical setting. Indeed, the findings show that a pitch deficit can be observed in

speech material, even though it is not systematic (not all amusics show the deficit), relatively mild (both quantitatively and qualitatively small), and is not as pronounced as in musical analogs.

Experiments 1 and 2 revealed that congenital amusics, as a group, showed deficits in tasks requiring discrimination of Mandarin and Thai tones, respectively. This deficit also correlated with amusics' deficit in a musical task that required interval processing between tones (i.e., a subtest of the MBEA). The additional analyses of the Thai tone pairs further showed that amusics performed only 9% below controls for the contour–contour pair (the rising versus falling tones), but dropped to 33% below controls for other pairs requiring finer pitch contour discriminations (averaged over verbal and musical materials; **Table 2**).

In Experiment 2 with the Thai tones, the comparison between speech material and musical analogs suggests that in the presence of a severe musical pitch deficit, pitch information might be slightly better perceived in speech materials. Speech might thus enhance pitch processing in amusics, even if it does not restore normal processing. As discussed in Tillmann et al. (submitted) for discrete pitch changes, it remains to be investigated whether this boost for verbal material might be due to acoustic features of the speech sound that facilitate pitch extraction, or whether top-down influences come into play to modulate pitch-processing depending on material type (speech, music). According to the latter view, which suggests influences related to strategies, attention or memory, pitch extraction of tones in congenital amusia might not be the sole impairment, but rather the deficit may also incorporate later processing stages and be related to material-specific top-down processes.

Comparisons between Experiments 1 and 2 suggested that the amusics' deficit affected more strongly the processing of Thai tones than of Mandarin tones: for Thai, 63% of the amusics performed 2 SD below the controls' mean, whereas this was only the case of 15% of the amusics for Mandarin. The greater difficulty level of the Thai material might be linked to the set of five tones and the acoustic features of the to-be-detected changes, which were less varied in the Thai than Mandarin stimuli. We further acknowledge that the difference in performance for the two experiments might be exacerbated by the longer delay between the two items for the Thai tone pairs, which may have selectively influenced amusics' performance due to their pitch memory deficit (Gosselin et al., 2009; Tillmann et al., 2009; Williamson et al., 2010). Furthermore, the discrimination of Mandarin tones (but less so of Thai tones) might also be partly based on cues other than pitch, such as length and intensity contours (Whalen and Xu, 1992; Hallé et al., 2004), even if those cues are less reliable, and as was found here, misleading.

Even though the group of amusics performed below the level of the group of controls, our findings showed some overlap in performance between the groups (see **Figure 3**). This overlap might reflect the influence of large individual differences reported in the normal population, notably for the perception of pitch change and pitch direction (Semal and Demany, 2006; Foxton et al., 2009) as well as for the learning of pitch contours in syllables (Golestani and Zatorre, 2004; Wong and Perrachione, 2007). Some of these previous findings also suggest that pitch processing/learning is not independent between music and speech: for example, Wong and Perrachione (2007) reported an association between participants' ability to learn pitch patterns in syllables and their ability to perceive pitch patterns in

non-lexical contexts as well as with their previous musical experience. Finally, in light of the observed variability in learning, Golestani and Zatorre (2004) discussed a “connectivity hypothesis,” notably that fast and slow learners might differ in white matter connectivity, with greater myelination leading to more rapid neural transmission. This suggestion, along with our present findings, even if based on behavior only, might be related to recent data in congenital amusics having neural anomalies in white matter concentration, cortical thickness and fiber tracts in the right hemisphere (Hyde et al., 2006, 2007; Loui et al., 2009), and in functional connectivity between the auditory and inferior frontal gyrus (Hyde et al., 2011). Previous brain imaging research has suggested that pitch processing in tone-languages involves a right-hemisphere network rather than a left-hemisphere network for non-native listeners (e.g., Hsieh et al., 2001). This might explain the overall deficit for the amusic group observed here for the tone-language material (all non-native listeners). However, other brain imaging data have suggested the implication of left-hemisphere networks in speech processing independently of language background, but rather depending on the acoustic features of verbal sounds (see Zatorre and Gandour, 2007 for a review). In our study, the observed advantage of Thai tones over non-verbal, musical analogs in the amusics who exhibited higher pitch thresholds might suggest some implication of the left-hemispheric network, which seems to be unimpaired in congenital amusia.

Aiming to further understand amusics' performance and deficits, we calculated a series of acoustic measures and conducted analyses to investigate their relation to the behavioral data. These analyses provide some interesting insights into features related to amusics' discrimination performance. It was found that that amusics' tone discrimination is related not only to some pitch characteristics of the sounds, but also to features that are unrelated to pitch changes or pitch movement. Regarding first the pitch-related features, the analyses of the Thai material (**Figure 5**) suggested that small differences in F_0 mean and absolute slope are associated with more difficult item pairs: amusics benefited from large differences in F_0 mean and absolute slope (i.e., mean of the absolute value of the slope of the pitch contour), but this was less the case for slope (i.e., providing information about the direction of the pitch movement). This finding is in agreement with previous observations suggesting that amusics can detect pitch movement, but that they have difficulties in perceiving the direction of these movements (e.g., Foxton et al., 2004; Patel et al., 2008). Another pitch-related feature that seemed to be used by the amusics for the Mandarin materials is the duration over which pitch information was present in the tones but here, larger differences in pitch duration was a misleading cue; it led to more rather than less errors.

The acoustic analyses further showed that, in contrast to control participants, amusics seemed to use non-pitch-related information, notably sound duration and intensity, which are less relevant for lexical tone discrimination. The result pattern suggests that amusics use these acoustic features as some kind of replacement strategy: as amusics are less efficient in using pitch cues, they focus on these cues, whose variations they *can* discriminate well and this erroneously results in lower performance levels. When, however, items of a pair differ less strongly on these non-pitch features, amusics seem able to use other cues (probably also pitch-related cues), leading to higher performance levels.

In sum, the various acoustic analyses suggest that same-different task performance might not be based on single features, but that participants (in particular controls who have better performance on the task) used a weighted combination of acoustic features. Relative weighting of acoustic cues in lexical tone perception has been previously shown for native and non-native speakers (for example, for Yoruba and Thai, see Gandour and Harshman, 1978). Amusic individuals might (a) use the various pitch features less efficiently (because of their pitch deficit), (b) not weight/combine them in the adequate way (or also under-use some components), and (c) get waylaid by other, irrelevant but, for them, more easily discriminable cues (duration, intensity). Based on our findings, future experiments can now directly (i.e., parametrically) manipulate the various pitch and non-pitch cues in extent as well as in combination to further investigate the importance of these parameters, their relative weighting, as well as the deficits of amusics in processing these parameters.

This study provides evidence that the pitch deficit of congenital amusics previously observed for musical material extends to speech material, notably lexical tones of Mandarin and Thai. Our findings provide further motivation to investigate processing of pitch in amusics' native language (e.g., non-tonal languages). Specifically, future studies should focus not only on statements/questions, as has been done up to now, but investigate also prosody-based perception of emotion and humor, as well as the use of subtle pitch cues. Regarding this latter aspect, pitch variations can make a syllable more salient, create focus stress in syllables or allow for differentiation between, for example, “a hot dog” and “a hotdog.” For example, Spinelli et al. (2010) tested the influence of F_0 changes in the first vowel for word segmentation in French (e.g., “la fiche” versus “l'affiche”). Future research should investigate the use of these subtle pitch cues in amusics' native language perception (e.g., English, French), and even extend this to language production. Finally, our results further suggest that congenital amusics may experience difficulties in acquiring a tonal language. Actually, this possible relation between amusia and tone-language processing has been recently confirmed by Nan et al. (2010) who reported cases of congenital amusia among native speakers of Mandarin. Importantly, the musical deficit was found to be associated to impairments in lexical tone discrimination and identification, but not production – a finding that mirrors the previously observed mismatch between perception and production of musical intervals in congenital amusia (Loui et al., 2008).

ACKNOWLEDGMENTS

This research was supported by grants from the Cluster 11 of the Region Rhône-Alpes and the Eminent Visiting Research Scheme of the University of Western Sydney to Barbara Tillmann, by the Australian Research Council Discovery grant (DP0666981) to Denis Burnham and by grants from the Natural Sciences and Engineering Research Council of Canada, the Canada Institute of Health Research and a Canada Research Chair to Isabelle Peretz. Experiment 1 has been partly presented at the Neurosciences and Music Conference in Montreal in 2008 (see Nguyen et al., 2009, for a summary). We wish to thank Dr. Karen Mattock for advice and discussion about the Thai tone stimuli used in Experiment 2; Dr. Michael Tyler for insightful comments; Dr. Nan Xu for phonetic analyses of the stimuli; and Ms Benjawan Kasisopa for assistance with the figures.

REFERENCES

- Abramson, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Lang. Speech* 21, 319–325.
- Abramson, A. S. (1962). *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments*. Bloomington, IN: Indiana University Research Center in Anthropology, Folklore and Linguistics, 20.
- Alexander, J. A., Wong, P. C. M., and Bradlow, A. R. (2005). “Lexical tone perception in musicians and non-musicians,” in *Proceedings of Interspeech 2005*, Lisbon.
- Ayotte, J., Peretz, I., and Hyde, K. L. (2002). Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain* 125, 238–251.
- Bent, T., Bradlow, A. R., and Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 97–103.
- 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.
- Burnham, D., and Brooker, R. (2002). “Absolute pitch and lexical tones: tone perception by non-musician, musician, and absolute pitch nontonal language speakers,” in *7th International Conference on Spoken Language Processing*, Denver, 257–260.
- Burnham, D., and Francis, E. (1997). “The role of linguistic experience in the perception of Thai tones,” in *Southeast Asian Linguistic Studies in Honour of Vichin Panupong* (Science of Language, Vol. 8.), ed. A. S. Abramson (Bangkok: Chulalongkorn University Press), 29–47.
- Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiaboon, C., Lacerda, F., and Keller, P. (1996). “Perception of lexical tone across languages: evidence for a linguistic mode of processing,” in *Fourth International Conference on Spoken Language Processing (ICSLP'96)*, Philadelphia.
- Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behav. Res. Methods Instrum. Comput.* 25, 257–271.
- Cutler, A. (2009). Greater sensitivity to prosodic goodness in non-native than in native listeners. *J. Acoust. Soc. Am.* 125, 3522–3525.
- Delogu, F., Lampis, G., and Belardinelli, M. O. (2010). From melody to lexical tone: musical ability enhances specific aspects of foreign language perception. *Eur. J. Cogn. Psychol.* 22, 46–61.
- Deutsch, D., Dooley, K., Henthorn, T., and Head, B. (2009). Absolute pitch among students in an American music conservatory: association with tone language fluency. *J. Acoust. Soc. Am.* 125, 2398–2403.
- Deutsch, D., Henthorn, T., Marvin, E., and Xu, H.-S. (2006). Absolute pitch among American and Chinese conservatory students: prevalence differences, and evidence for a speech-related critical period. *J. Acoust. Soc. Am.* 119, 719–722.
- Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., and Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science* 291, 1969–1972.
- Fitzsimons, M., Sheahan, N., and Staunton, H. (2001). Gender and the integration of acoustic dimensions of prosody: implications for clinical studies. *Brain Lang.* 78, 94–108.
- Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., and Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying “tone deafness”. *Brain* 127, 801–810.
- Foxton, J. M., Weisz, N., Bauchet-Lecaigard, F., Delpuech, C., and Bertrand, O. (2009). The neural bases underlying pitch processing difficulties. *Neuroimage* 45, 1305–1313.
- Fromkin, V. (1978). *Tone: A Linguistic Survey*. New York: Academic.
- Gandour, J., and Harshman, R. (1978). Cross-language differences in tone perception: a multidimensional scaling investigation. *Lang. Speech* 21, 1–33.
- Golestani, N., and Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *Neuroimage* 21, 494–506.
- Gosselin, N., Jolicoeur, P., and Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Ann. N. Y. Acad. Sci.* 1169, 270–272.
- Hallé, P., Chang, Y.-C., and Best, C. (2004). Categorical perception of Taiwan Mandarin Chinese tones by Chinese versus French native speakers. *J. Phon.* 32, 395–421.
- Hove, M. J., Sutherland, M. E., and Krumhansl, C. L. (2010). Ethnicity effects in relative pitch. *Psychon. Bull. Rev.* 17, 310–316.
- Hsieh, L., Gandour, J., Wong, D., and Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain Lang.* 76, 227–252.
- Hyde, K., Zatorre, R., and Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cereb. Cortex* 21, 292–299.
- Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., and Peretz, I. (2007). Cortical thickness in congenital amusia: when less is better than more. *J. Neurosci.* 27, 13028–13032.
- Hyde, K. L., and Peretz, I. (2004). Brains that are out of tune but in time. *Psychol. Sci.* 15, 356–360.
- Hyde, K. L., Zatorre, R. J., Griffiths, T. D., and Peretz, I. (2006). Morphometry of the amusic brain: a two-site study. *Brain* 129, 2562–2570.
- Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., and Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia* 48, 2630–2639.
- Kawahara, H., Masuda-Katsuse, I., and Cheveigne, A. D. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous frequency-based f0 extraction. *Speech Commun.* 27, 187–207.
- Klein, D., Zatorre, R. J., Milner, B., and Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage* 13, 646–653.
- Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res. Cogn. Brain Res.* 25, 161–168.
- Lee, C.-Y., and Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *J. Acoust. Soc. Am.* 124, 325–3248.
- Liu, F., Patel, A. D., Fourcin, A., and Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification, and imitation. *Brain* 133, 1682–1693.
- Loui, P., Alsop, D., and Schlaug, G. (2009). Tone deafness: a new disconnection syndrome? *J. Neurosci.* 29, 10215–10220.
- Loui, P., Guenther, F. H., Mathys, C., and Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Curr. Biol.* 18, R331–R332.
- 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.
- Nan, Y., Sun, Y., and Peretz, I. (2010). Congenital amusia in speakers of a tonal language: association with lexical tone agnosia. *Brain* 133, 2635–2642.
- Nguyen, S., Tillmann, B., Gosselin, N., and Peretz, I. (2009). Tonal language processing in congenital amusia. *Ann. N. Y. Acad. Sci.* 1169, 490–493.
- Patel, A. D. (2008). *Music, Language, and the Brain*. New York, NY: Oxford University Press.
- Patel, A. D., Foxton, J. M., and Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain Cogn.* 59, 310–313.
- Patel, A. D., Wong, M., Foxton, J., Lochy, A., and Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Percept.* 25, 357–368.
- Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., and Jutras, B. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron* 33, 185–191.
- 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.
- Peretz, I., Cummings, S., and Dubé, M. P. (2007). The genetics of congenital amusia (tone deafness): a family-aggregation study. *Am. J. Hum. Genet.* 81, 582–588.
- Peretz, I., and Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends Cogn. Sci. (Regul. Ed.)* 7, 362–367.
- Pfordresher, P. Q., and Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Atten. Percept. Psychophys.* 71, 1385–1398.
- Schneider, W., Eschmann, A., and Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh, PA: Psychology Software Tools, Inc.
- Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349.
- Schwanhäuser, B., and Burnham, D. (2005). “Lexical tone and pitch perception in tone and non-tone language speakers,” in *Proceedings of the 9th European Conference on Speech Communication and Technology* (Bonn: ISCA), 1701–1704.
- Semal, C., and Demany, L. (2006). Individual differences in the sensitivity to pitch direction. *J. Acoust. Soc. Am.* 120, 3907–3915.
- Spinelli, E., Grimault, N., Meunier, F., and Welby, P. (2010). An intonational cue to word segmentation in phonemically identical sequences. *Atten. Percept. Psychophys.* 2, 775–787.

- Tillmann, B., Schulze, K., and Foxton, J. (2009). Congenital amusia: a short-term memory deficit for nonverbal, but not verbal sounds. *Brain Cogn.* 71, 259–264.
- Vos, P. G., and Troost, J. M. (1989). Ascending and descending melodic intervals: statistical findings and their perceptual relevance. *Music Percept.* 6, 383–396.
- Whalen, D. H., and Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica* 49, 25–47.
- Williamson, V., McDonald, C., Deutsch, D., Griffiths, T., and Stewart, L. (2010). Faster decline of pitch memory over time in congenital amusia. *Adv. Cogn. Psychol.* 6, 15–22.
- Wong, P. C. M., and Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Appl. Psycholinguist.* 28, 565–585.
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
- Zatorre, R. J., and Gandour, J. T. (2007). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1087–1104.
- 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: 24 February 2011; accepted: 27 May 2011; published online: 17 June 2011.
- Citation: Tillmann B, Burnham D, Nguyen S, Grimault N, Gosselin N and Peretz I (2011) Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. *Front. Psychology* 2:120. doi: 10.3389/fpsyg.2011.00120
- This article was submitted to *Frontiers in Auditory Cognitive Neuroscience*, a specialty of *Frontiers in Psychology*. Copyright © 2011 Tillmann, Burnham, Nguyen, Grimault, Gosselin and Peretz. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



Non-invasive brain stimulation enhances the effects of melodic intonation therapy

Bradley W. Vines^{1,2}, Andrea C. Norton¹ and Gottfried Schlaug^{1*}

¹ Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

² Department of Psychiatry, Institute of Mental Health, University of British Columbia, Vancouver, BC, Canada

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Lutz Jäncke, University of Zurich, Switzerland

Eckart Altenmüller, University of Music and Drama Hannover, Germany

*Correspondence:

Gottfried Schlaug, Music, Neuroimaging, and Stroke Recovery Laboratories, Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Avenue, Palmer 127, Boston, MA 02215, USA.
e-mail: gschlaug@bidmc.harvard.edu

Research has suggested that a fronto-temporal network in the right hemisphere may be responsible for mediating melodic intonation therapy's (MIT) positive effects on speech recovery. We investigated the potential for a non-invasive brain stimulation technique, transcranial direct current stimulation (tDCS), to augment the benefits of MIT in patients with non-fluent aphasia by modulating neural activity in the brain during treatment with MIT. The polarity of the current applied to the scalp determines the effects of tDCS on the underlying tissue: *anodal*-tDCS increases excitability, whereas *cathodal* tDCS decreases excitability. We applied *anodal*-tDCS to the posterior inferior frontal gyrus of the right hemisphere, an area that has been shown both to contribute to singing through the mapping of sounds to articulatory actions and to serve as a key region in the process of recovery from aphasia, particularly in patients with large left hemisphere lesions. The stimulation was applied while patients were treated with MIT by a trained therapist. Six patients with moderate to severe non-fluent aphasia underwent three consecutive days of *anodal*-tDCS + MIT, and an equivalent series of *sham*-tDCS + MIT. The two treatment series were separated by 1 week, and the order in which the treatments were administered was randomized. Compared to the effects of *sham*-tDCS + MIT, *anodal*-tDCS + MIT led to significant improvements in fluency of speech. These results support the hypothesis that, as the brain seeks to reorganize and compensate for damage to left hemisphere language centers, combining *anodal*-tDCS with MIT may further recovery from post-stroke aphasia by enhancing activity in a right hemisphere sensorimotor network for articulation.

Keywords: melodic intonation therapy, transcranial direct current stimulation, tDCS, Broca's aphasia, stroke, neurorehabilitation, singing

INTRODUCTION

Of the more than 750,000 strokes that occur every year in the US, approximately 20% of the survivors suffer from some form of aphasia, an impairment in the ability to express and/or understand speech and language (Kertesz and McCabe, 1977; Mohr et al., 1978; Pedersen et al., 2004; Schlaug et al., 2008a). Though traditional therapies for post-stroke language recovery have been shown to have a beneficial effect (Robey, 1994; Holland et al., 1996), in most cases, patients' recovery is incomplete, particularly when they have large lesions in the left hemisphere. To date, relatively few speech therapy techniques have succeeded in helping these severely impaired patients. However, an intonation-based intervention, melodic intonation therapy (MIT), may offer hope for patients who suffer from severe non-fluent aphasia (Albert et al., 1973; Sparks et al., 1974; Schlaug et al., 2008a,b, 2010). In addition, the synthesis of effective behavioral therapy with complementary brain stimulation methods to further engage neural centers that are important for recovery and to facilitate neuroplastic changes (Stefan et al., 2000; Adkins-Muir and Jones, 2003; Nair et al., 2008; Celnik et al., 2009; Baker et al., 2010; Lindenberg et al., 2010; Floel et al., 2011; Fridriksson et al., 2011) is a promising frontier for post-stroke neurorehabilitation. Indeed, a combined

peripheral and central stimulation approach has recently been shown to enhance synaptic plasticity more than central stimulation alone in an experimental animal study (Fritsch et al., 2010). Thus, combining a speech or intonation-based therapy with non-invasive brain stimulation may be particularly effective as a means for treating non-fluent aphasia.

A stroke affecting the left frontal lobe can cause a form of aphasia known as "non-fluent" or "Broca's" aphasia (Goodglass and Geschwind, 1976; Kertesz et al., 1977; Mohr et al., 1978). Broca's aphasia is characterized by either a deficit in speech output without significant impairment to comprehension, or an impairment in the ability to organize the elements of speech (e.g., phonemes) into streams of sound that form meaningful utterances. Prior research has suggested that there are two neural pathways with the potential to facilitate recovery from Broca's aphasia. One pathway primarily recruits peri-lesional cortex in the left hemisphere, and may also elicit varying degrees of engagement from the right hemisphere for recovery of language function. This route for recovery is only possible when patients have smaller left hemisphere lesions that have not completely destroyed Broca's region or the surrounding cortex, and thus, may have enough peri-lesional cortex remaining to support language and speech output. Typically, these

patients have a milder form of aphasia. The other route to recovery relies almost exclusively on the right hemisphere, and may be the only recovery option for patients with large lesions that have destroyed all language-capable structures in the left hemisphere (Mimura et al., 1998; Pizzamiglio et al., 2001; Thiel et al., 2001; Blasi et al., 2002; Winhuisen et al., 2005; Schlaug et al., 2008b). Natural outcome studies in patients with lesions of this kind provide evidence that neural centers in the right hemisphere are capable of supporting vocal-motor output and compensating for damaged “eloquent” areas in the left hemisphere to facilitate language recovery (Mimura et al., 1998; Rosen et al., 2000; Winhuisen et al., 2005; Heiss and Thiel, 2006; Saur et al., 2006). What remains undetermined is how to specifically engage these right-hemispheric vocal-motor centers and enhance their contribution to the recovery process. Based upon the existing literature and our own data, it appears that these right fronto-temporal centers respond well to interventions that emphasize melodic contour and re-map sounds to articulatory actions with the help of a right-hemispheric sensorimotor network that is engaged by rhythmic activities involving the left hand (Schlaug et al., 2008a,b, 2009, 2010; Wan et al., 2010). Although impairments in melodic and temporal music information processing have been associated with acquired lesions in both right and left hemispheres, right hemisphere-lesioned patients have revealed more of an impairment in melodic contour and meter (Schuppert et al., 2000) suggesting that interventions emphasizing these aspects of an intonation-based therapy might particularly engage right hemisphere structures.

Neuroimaging studies have revealed both unique and overlapping brain networks for speaking and singing (Brown et al., 2004; Ozdemir et al., 2006); although speaking tends to be lateralized to the left hemisphere and singing to the right (Riecker et al., 2000; Jeffries et al., 2003; Sparing et al., 2007), these two behaviors involve some of the same brain areas. In an fMRI study from our group (Ozdemir et al., 2006) we found that singing and speaking shared some bilateral fronto-temporal neural correlates, but singing, or intoned speaking, led to additional activation of the right more than left superior temporal and right more than left central operculum compared to the speaking condition. Thus, singing could be a way to access language-capable regions in the right hemisphere for the purpose of facilitating language recovery (Racette et al., 2006; Schlaug et al., 2008a,b, 2009, 2010; Wan et al., 2010).

One speech therapy that capitalizes on the possible rehabilitative effects of singing in patients with motor aphasia is MIT. This technique was inspired by the common clinical observation that severely aphasic patients can sing the lyrics of songs better than they can speak them (Goldstein, 1942; Gerstman, 1964; Geschwind, 1971; Keith and Aronson, 1975; Kinsella et al., 1988; Hebert et al., 2003). MIT uses a simplified, exaggerated prosody to train patients to intone and tap out high probability words and phrases syllable by syllable. Research has found MIT to be effective in facilitating significant improvements in language production (Albert et al., 1973; Sparks et al., 1974; Laughlin et al., 1979; Bonakdarpour et al., 2000; Wilson et al., 2006; Schlaug et al., 2008a, 2009). MIT's efficacy may be due to its unique ability to engage language-capable brain regions in both hemispheres, but most particularly, in the right hemisphere (Albert et al., 1973; Sparks et al., 1974; Schlaug et al., 2008a). The posterior inferior frontal

gyrus (IFG) very likely plays a critical role in the network underlying the rehabilitative effect of MIT (Schlaug et al., 2008a), and it is important to note that this brain region, which includes the right hemisphere homolog of Broca's area, is easily accessible for application of non-invasive brain stimulation.

Transcranial direct current stimulation (tDCS) is a technique that influences excitability in the brain by modulating the spontaneous firing rate of neurons (Priori et al., 1998; Nitsche and Paulus, 2000). Research suggests that the polarity of the current determines the effects of tDCS: *anodal*-tDCS *increases* cortical excitability, whereas *cathodal* tDCS *decreases* excitability. Activity in sodium and calcium ion channels as well as the efficacy of NMDA receptors may mediate the effects of tDCS (Liebetanz et al., 2002; Nitsche et al., 2003a). Applying tDCS has been shown to improve cognitive and behavioral performance on tasks involving the stimulated brain area (Nitsche et al., 2003b; Antal et al., 2004; Kincses et al., 2004; Fregni et al., 2005; Boggio et al., 2006a; Vines et al., 2006a). For example, studies have found that applying *anodal*-tDCS to regions in the left frontal cortex significantly improved fluency in healthy participants (Iyer et al., 2005; Cattaneo et al., 2011). We have also recently shown (Zheng et al., 2011) that tDCS can lead to modality dependent regional blood flow increases in targeted brain regions, and that resting regional blood flow after tDCS is turned off shows longer lasting effects that differ between *anodal* and *cathodal* stimulation. These findings are in agreement with the longer-term electrophysiological effects of either *increased* excitability (*anodal*) or *decreased* excitability (*cathodal*). Researchers have already employed tDCS as a tool to facilitate recovery from dysfunctions caused by a stroke (Schlaug and Renga, 2008; Schlaug et al., 2008c; Baker et al., 2010; Fiori et al., 2010; Lindenberg et al., 2010; Floel et al., 2011; Fridriksson et al., 2011; Schlaug et al., 2011). In the context of speech recovery, at least one study has investigated the potential for tDCS alone to facilitate recovery from non-fluent aphasia (Monti et al., 2008). And there is a precedent for combining tDCS with a behavioral speech therapy with positive results (Baker et al., 2010; Floel et al., 2011; Fridriksson et al., 2011). We sought to combine tDCS with the intonation-based speech therapy, MIT, for two reasons: (1) preliminary imaging results indicate a leading role for the right posterior IFG in the MIT-induced recovery process, and (2) no study has yet tested whether applying *anodal stimulation vs sham stimulation in combination with MIT* to undamaged right hemisphere structures will lead to an improvement in speech output/fluency for patients with non-fluent aphasia (see also Hamilton et al., 2011 and Schlaug et al., 2011 for more details).

The present study contrasted the effects of two tDCS conditions (*anodal* and *sham*) when applied over the right IFG during MIT sessions. We hypothesized that, compared to *sham*, applying *anodal*-tDCS concurrently with MIT would enhance neural activity, and synaptic plasticity in the right-hemispheric brain regions that drive the positive effects of MIT on speech production and fluency, and thereby promote a greater degree of recovery.

MATERIALS AND METHODS

PARTICIPANTS

Six patients (all males) with non-fluent aphasia participated in this study. The protocol was approved by the Beth Israel Deaconess

Medical Center Institutional Review Board and all participants gave written informed consent. All patients were at least 1 year post-onset of their first (and only) ischemic stroke. All patients' strokes affected their left frontal lobe (see combined lesion map of all six patients shown in **Figure 1**). Lesion volumes ranged from 60 to 218 cc (see **Table 1**; for more details on lesion calculations see Marchina et al., 2011). Based on our baseline assessments which included the Boston Diagnostic Aphasia Examination (BDAE), our patients were classified as having moderate to severe non-fluent Broca's aphasia with relatively unimpaired comprehension. All six patients had previously been part of a "proof-of-concept" study in our laboratory assessing the behavioral and neural effects of MIT (for more details on preliminary results of this study see Schlaug et al., 2008a,b, 2009, 2010), which they had completed at least 6 months prior to participating in the current protocol of tDCS + MIT. Thus, our patients were familiar with MIT, but had not participated in any study involving MIT for at least 6 months prior to being enrolled into the current study. All patients had a stable baseline prior to enrollment in this study. The ages of the patient participants ranged from 30 to 81 years. Five of the patients were right handed and one was mixed-handed. One of the six participants was bi-lingual (Russian and English); all other participants were native speakers of English.

PROCEDURE

Participants underwent two series (one series per tDCS condition) of three therapy sessions each. Treatment sessions were

administered one per day for three consecutive days. The two series were separated by 1 week. During the treatment sessions, we applied tDCS to the right posterior IFG with the electrode angled toward the temporal lobe. The stimulation period lasted for 20 min and overlapped with a 20-min session of MIT administered by a trained therapist. The therapist tailored each MIT session to the skill level of the participant. For one 3-day therapy series, we applied *anodal*-tDCS, and for the other, *sham*-tDCS. The ordering of the two stimulation conditions was counter-balanced across participants such that half of them received the *sham*-tDCS series first, and whether participants started with the *sham* or *anodal* stimulation was determined in a randomized fashion. tDCS was applied for 20 min, with the active electrode positioned over the participants' right IFG, and centered approximately 2.5 cm posterior to F8 of the 10–20 International EEG system for electrode placement. Neuroimaging studies have confirmed the correspondence between F8 and the right midportion of the IFG (Homan et al., 1987; Okamoto et al., 2004), including our own pilot study using high resolution (1 mm three voxel size) structural MRI ($N = 5$). We chose to position the active electrode slightly posterior to F8 (about 2.5 cm) based upon our pilot study investigating the location of the right hemisphere's homolog of Broca's area. A number of TMS and tDCS studies have used the 10–20 EEG system to identify the location of brain structures for stimulation (Kincses et al., 2004; Rogalewski et al., 2004; Fregni et al., 2005; Iyer et al., 2005; Vines et al., 2006a,b, 2008a,b). Due to the size of the active electrode (area = 16.3 cm²),

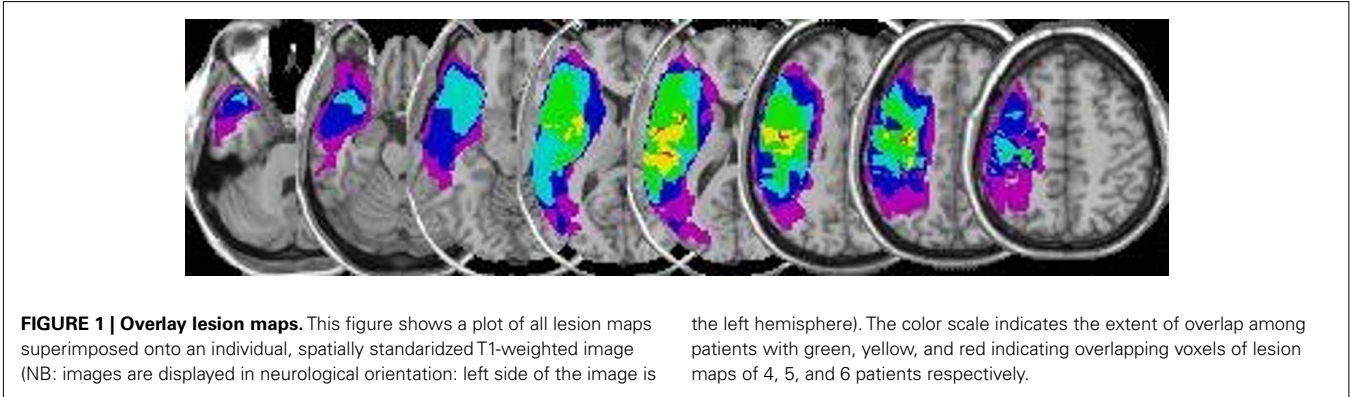


Table 1 | Age at treatment (in years), time post-stroke (in years), total lesion volume (in cubic centimeter), lesion volume overlying a canonical arcuate fasciculus derived from a group of healthy elderly control subjects (AF lesion-load; for more details see Marchina et al., 2011), and fluency changes (in seconds) either after three sessions of anodal-tDCS + MIT or after three sessions of sham-tDCS + MIT.

Age @ Tx	Times post-stroke (years)	Lesion volume (cc)	AF-lesion-load (cc)	Fluency, percentage	Fluency, percentage
				Δ in duration (s)	Δ in duration (s)
				After anodal-tDCS + MIT	After sham-tDCS + MIT
31.3	10.0	191.7	10.6	−8.3	12.9
48.8	2.5	154.8	7.6	−19.8	1.1
80.9	1.3	60.0	3.8	−12.8	5.3
55.3	5.8	218.4	14.8	−3.0	−6.9
58.2	5.1	93.2	4.0	−9.7	9.0
62.7	2.8	86.1	7.3	−11.9	−6.3

the stimulation may have extended into anterior temporal cortex and ventral premotor cortex, which make up the network of fronto-temporal regions that may underlie MIT's therapeutic effect. The reference electrode (area = 30 cm²) was positioned over the left supraorbital region. This location for the reference electrode was functionally ineffective in the experimental design (Nitsche et al., 2003b).

A battery-driven, constant-current stimulator (Phoresor, Iomed Inc., Salt Lake City, UT, USA) delivered 1.2 mA of electrical current to a participant's scalp by means of saline-dampened electrodes. For the *anodal* stimulation, the tDCS current ramped up over the first few seconds, and then remained on during the 20-min stimulation period. The *sham-(control)* tDCS was identical to the *anodal* stimulation, except that the experimenter reduced the current to zero after it ramped up for 30 s; the current then stayed at zero for the remaining time period. Participants reported a tingling or itchy sensation at the start of the stimulation, which typically faded away after a few seconds. This sensation was present for both *anodal*- and *sham*-tDCS. Gandiga et al. (2006) found that naive participants were not able to distinguish between actual tDCS and *sham*-tDCS as we applied them in the present study. The application of tDCS began 5 min after the start of MIT, and continued for 5 min after the end of the MIT session. During the 5-min break after the end of MIT and before the end of the stimulation, the patients rested before completing the verbal fluency tasks.

TASK

Participants completed a battery of verbal fluency tests before and after each stimulation session. The tasks included the automatic production of verbal sequences (e.g., counting from 1 to 21, reciting the days of the week, months of the year, and the United States' Pledge of Allegiance), describing flash card-like scenes, and picture naming. We chose the flash card-like scenes from drawings associated with the MIT intervention (Helm-Estabrooks and Albert, 1991). For the naming task we selected a set of colored pictures from the Snodgrass-Vanderwart inventory for naming, and a black and white line drawing of a scenario from the (BDAE 2nd Edn; Goodglass and Kaplan, 1983). Participants completed the automatic response items in the same order for each testing session. Pictures for the naming task were shown in a new random order for each session. We determined the number of items used in the automatic speech and picture naming tasks based on the ability of the participant, so that no testing session exceeded 30 min. We limited the duration of the testing sessions in order to avoid excessive fatigue for the patients. Patients were instructed to simply try their best during each testing session, both pre- and post-therapy.

This was a randomized, cross-over, sham-controlled, double-blind study in which the participants, the therapist, and the experimenter conducting and scoring the testing were all blind to stimulation condition.

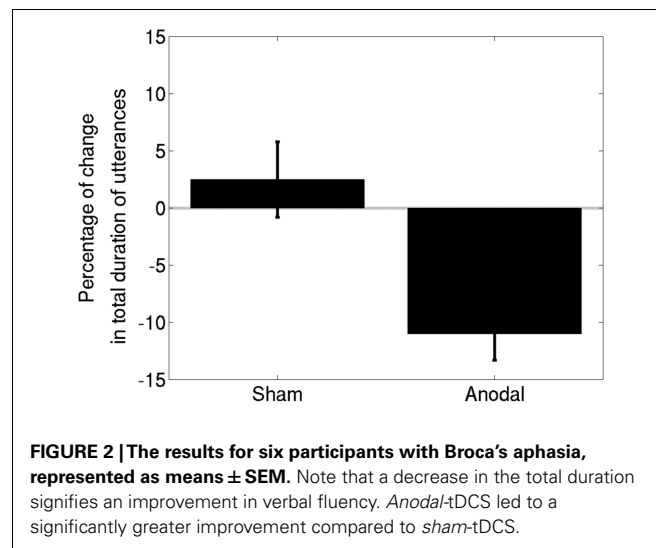
DATA ANALYSIS

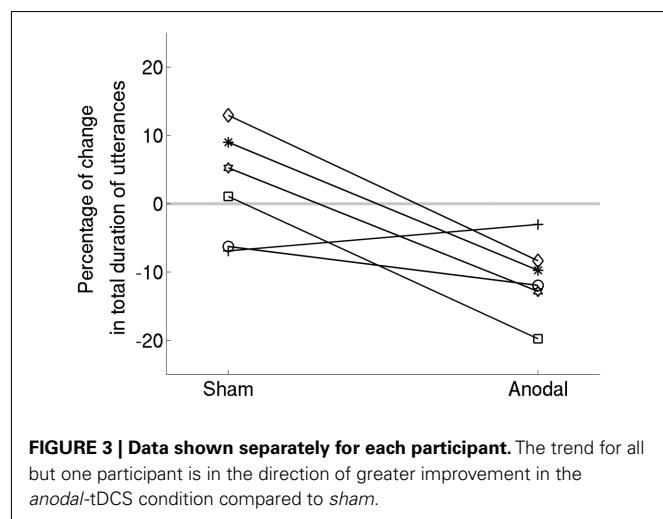
We calculated the dependent variable as the percentage of change ("proportional change") in the sum duration of fluency measures from before the first of three stimulation sessions to after the last of three stimulation sessions. This calculation produced

two values for each participant: (1) (post-*anodal* series – pre-*anodal* series)/(pre-*anodal* series), (2) (post-*sham* series – pre-*sham* series)/(pre-*sham* series). To ensure equality for all four time-points of interest (pre-*anodal*, post-*anodal*, pre-*sham*, post-*sham*), the calculation of the dependent variable only included durations for fluency items, or portions thereof, that were intact at all of these time-points. For example, on the task of counting from 1 to 21, if a participant was able to count to 21 at all time-points except one, during which he only made it to 18, the duration for counting from 1 to 18 was used at all time-points. Similarly, we only used pictures that a participant was able to name at all four time-points of interest. Doing this ensured that the material at each time-point was identical in terms of content and what was actually spoken clearly. (Please see **Table A1** in Appendix for details on individual participants' verbal output that met these criteria for inclusion in the analysis.) A rater who was blind to experimental condition measured the duration of the participants' fluency-test utterances using a waveform editor. The dependent variable was not sensitive to changes in the *amount* of verbal production, but to changes in the *rate* of verbal production, that is, how quickly the participant was able to utter appropriate, intelligible responses. The measure reflects fluency, with regard to ease of speech production. We compared the percentage of change for *anodal*- and *sham*-tDCS combined with MIT using a planned, two-tailed paired-samples *t*-test.

RESULTS

All six participants completed the experimental procedure. All patients tolerated the tDCS well and no adverse effects related to the application of tDCS were experienced. Averaged data for the effects of combining *anodal* and *sham* stimulation with MIT appear in **Figure 2**; and individual data for each participant are shown in **Figure 3**. We tested normality using the Kolmogorov-Smirnov test and found that the data were normal for both the *sham* [$Z(6) = 0.48$, $p = 0.98$] and *anodal* [$Z(6) = 0.49$, $p = 0.97$] conditions. The *t*-test comparing the effects of *anodal*- and *sham*-tDCS yielded a significant result [$t(5) = 3.22$, $p = 0.02$]. We measured the effect size as Cohen's $d = 1.98$. There was no difference





between baseline performance in the *anodal*-tDCS condition and baseline performance in the *sham* condition. A paired-samples *t*-test comparing pre-*anodal* performance to pre-*sham* performance did not yield a significant result [$t(5) = -0.31$, $p = 0.77$]. Thus, compared to *sham*-tDCS, applying *anodal*-tDCS to the right IFG during MIT produced a significantly greater improvement in verbal fluency while the verbal fluency in the *sham*-tDCS-MIT condition did not change significantly.

DISCUSSION

The results of this study provide evidence that applying “real” *anodal*-tDCS to the right IFG during MIT can augment the beneficial effects of the intonation-based speech therapy. By up-regulating excitability in the right IFG, the *anodal*-tDCS may have increased synaptic plasticity in brain areas that are engaged by MIT and likely, drive the MIT-induced recovery process. We posit that increasing excitability in the right IFG with tDCS further engaged that area of the brain and thus, facilitated verbal output and fluency.

Future research will investigate whether the positive effects of tDCS were due to the particular placement of the anodal electrode over the right posterior IFG, or if anodal stimulation over other brain areas, such as the right anterior temporal cortex, could also improve upon the beneficial effects of MIT. It also remains unknown whether tDCS, as applied in this study, exerts a positive influence on language recovery only when combined with a behavioral speech therapy, or if tDCS could be effectively used on its own to improve verbal fluency for stroke patients such as those who participated in this study. The evidence from several recent cross-over studies suggests that beneficial effects are achieved only when real tDCS is simultaneously combined with behavioral therapy – neither behavioral nor speech therapy alone led to beneficial effects in these studies (Baker et al., 2010; Floel et al., 2011; Fridriksson et al., 2011), although one must consider the relatively few behavioral/speech therapy sessions that were done in these studies, as well as the total number of MIT sessions (3) in our current study. Nevertheless, the effects that were achieved by adding brain stimulation to behavioral interventions

suggest that the combined approach might lead to much greater effects than a behavioral intervention alone. A previous review examining the intensity of the behavioral intervention in experimental aphasia studies (Bhogal et al., 2003) concluded that positive studies testing behavioral interventions had applied at least 60 h of treatment. This is many more hours than has been tested in the sham-control arm of experimental studies (Baker et al., 2010; Floel et al., 2011; Fridriksson et al., 2011; current study). Nevertheless, effects in the real-stimulation arm of recent experimental studies suggest that the addition of simultaneous brain stimulation to behavioral interventions might lead to stronger effects with fewer therapy sessions. Because the brain’s plasticity that facilitates post-stroke language recovery may involve the development of latent neural connections in the undamaged brain, it is possible that modulating cortical excitability with non-invasive brain stimulation will have its greatest impact when a behavioral therapy takes advantage of the resulting neuro-plastic changes, as has been suggested in experimental animal studies (Fritsch et al., 2010).

While some aspects of other studies that also applied tDCS to facilitate recovery from non-fluent aphasia may have been similar to ours, a striking difference was that we stimulated a different area of the brain. Monti et al. (2008), for example, stimulated left fronto-temporal cortex, centered on Broca’s area, and found an effect when applying 2 mA of *cathodal* stimulation over the left fronto-temporal electrode. Of particular interest here is that the *anodal* electrode in the Monti et al. (2008) design was placed over the right supraorbital region. Baker et al. (2010) and Fridriksson et al. (2011) both targeted peri-lesional areas in the left hemisphere with 1 mA and found increased accuracy on trained items in non-fluent patients (Baker et al., 2010; as well as an improvement in reaction time on trained items and a trend on untrained items in fluent patients; Fridriksson et al., 2011). We did not include a left hemisphere stimulation condition in the current study for three primary reasons. First, the patients in our study had little or no surviving brain tissue in the posterior IFG and surrounding brain structures of the left hemisphere (Figure 1; shows that at least four patients had lesions that involved the posterior IFG directly). For these patients, the best, and possibly *only* chance of recovery involved recruitment of brain regions in the right hemisphere to compensate for damaged language centers in the left hemisphere. Second, we chose to stimulate the right hemisphere because preliminary brain-imaging research provided evidence that MIT-induced improvements in speech output correlated with increased activity in the right IFG (Schlaug et al., 2008a). Our choice of location for stimulation was aimed at complementing the effects of MIT on brain activity. And third, stimulating over the damaged hemisphere of patients with relatively large lesions could generate an unpredictable current distribution and lead to correspondingly unpredictable effects on brain activity. This may explain some of the variability in results obtained from stimulating the left hemisphere in the treatment of non-fluent aphasia – Fridriksson et al. (2011) and Baker et al. (2010) found that *anodal* stimulation led to speech improvements, whereas Monti et al. (2008) found that *cathodal* but not *anodal* stimulation produced improvements. Furthermore, a new study was just published in which *anodal* stimulation was applied to the right hemisphere

and coupled with picture naming training (similar to Baker et al., 2010; Fridriksson et al., 2011), which led to a significant increase in accuracy on trained items (Floel et al., 2011).

The question of whether it would be best to apply tDCS before, during, or after a behavioral therapy also remains unanswered. If the effect of tDCS is greatest on performance, then applying the stimulation before or during therapy would be ideal for priming the brain in a specific area. However, if the beneficial effect of tDCS is on consolidation, then applying the stimulation either during or after therapy would have a greater effect. We chose to apply tDCS during MIT for the following reasons: (1) a number of tDCS studies have found effects on performance during stimulation (Antal et al., 2001; Nitsche et al., 2003b; Kincses et al., 2004; Rogalewski et al., 2004; Fregni et al., 2005; Hummel et al., 2005; Iyer et al., 2005; Boggio et al., 2006b; Hesse et al., 2007); (2) in the context of neurorehabilitation, applying stimulation during behavioral therapy is more efficient as it saves time for both the patient and the therapist/physician; (3) stimulation during the therapy has the potential to influence both the performance and the consolidation phases of learning; and (4) experiments in animals have shown that the combination of peripheral and central stimulation enhances synaptic plasticity more than central stimulation alone (Fritsch et al., 2010).

One primary limitation of this study was that we did not collect data to measure the longevity of any positive effect. As a result, we cannot comment on how long the effects of combined tDCS and MIT might last beyond the treatment period. For example, will the effects on performance continue for several days or weeks after the end of the combined therapy? It will be important to answer this question in order to determine whether and how often to use maintenance sessions to prolong potential benefits of the treatment. It also remains unknown how many tDCS + MIT sessions would be most effective. The current study used three, however, the benefits are likely to grow with additional sessions. Future studies will investigate the dose effects of combining tDCS with MIT to identify the optimal number of sessions.

Another limitation of this study concerns determining the exact site of stimulation. Based on previous research and our own pilot

study, we located the right posterior IFG as the right hemisphere homolog of Broca's area, using the 10–20 International EEG system for electrode placement. Considering the anatomical variability from one participant to the next, we may have stimulated slightly different brain areas in each of the six patients. The use of frameless stereotaxy in combination with MRI structural images for each participant would have enabled us to center the stimulation electrode with greater accuracy. However, due to the size of the electrode we used, it is highly likely that we did stimulate the area of interest even if there was some variability in its exact location across participants.

Given the positive nature of the results, this study shows that tDCS has the potential to augment the benefits of an intonation-based speech therapy. We found that three sessions of real tDCS + MIT led to a significant improvement in verbal fluency compared to an equivalent series of sham-tDCS + MIT. Although this is a preliminary finding with only six patients, the general implications extend beyond the domain of speech recovery, and into other areas of rehabilitation, such as motor recovery, cognitive training, and psychiatric treatment. Potentially, adding tDCS to any behavioral therapy may augment the positive effects. In the domain of speech therapy for severe non-fluent aphasia, this study provides evidence that anodal-tDCS + MIT applied to the right posterior IFG in non-fluent aphasic patients can facilitate improvements in linguistic fluency, and, presumably, in quality of life for stroke survivors.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institute of Neurological Disorders and Stroke to Bradley W. Vines (NS053326) and to Gottfried Schlaug (RO1 DC008796). Gottfried Schlaug also acknowledges support from the Mary Crown and William Ellis Family Fund and the Rosalyn and Richard Slifka Family Fund. Bradley W. Vines also acknowledges support from the Grammy Foundation and the Michael Smith Foundation for Health Research. We thank Sarah Marchina for her help in generating Figure 1.

REFERENCES

- Adkins-Muir, D. L., and Jones, T. A. (2003). Cortical electrical stimulation combined with rehabilitative training: enhanced functional recovery and dendritic plasticity following focal cortical ischemia in rats. *Neurol. Res.* 25, 780–788.
- Albert, M. L., Sparks, R. W., and Helm, N. A. (1973). Melodic intonation therapy for aphasia. *Arch. Neurol.* 29, 130–131.
- Antal, A., Nitsche, M. A., Kruse, W., Kincses, T. Z., Hoffmann, K.-P., and Paulus, W. (2004). Direct current stimulation over V5 enhances visuo-motor coordination by improving motion perception in humans. *J. Cogn. Neurosci.* 16, 521–527.
- Antal, A., Nitsche, M. A., and Paulus, W. (2001). External modulation of visual perception in humans. *Neuroreport* 12, 3553–3555.
- Baker, J. M., Rorden, C., and Fridriksson, J. (2010). Using transcranial direct-current stimulation to treat stroke patients with aphasia. *Stroke* 41, 1229–1236.
- Bhagal, S. K., Teasell, R., and Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke* 34, 987–993.
- Blasi, V., Young, A. C., Tansy, A. P., Petersen, S. E., Snyder, A. Z., and Corbetta, M. (2002). Word retrieval learning modulates right frontal cortex in patients with left frontal damage. *Neuron* 36, 159–170.
- Boggio, P. S., Castro, L. O., Savagim, E. A., Braitte, R., Cruz, V. C., Rocha, R. R., Rigonatti, S. P., Silva, M. T., and Fregni, F. (2006a). Enhancement of non-dominant hand motor function by anodal transcranial direct current stimulation. *Neurosci. Lett.* 404, 232–236.
- Boggio, P. S., Ferrucci, R., Rigonatti, S. P., Covre, P., Nitsche, M., Pascual-Leone, A., and Fregni, F. (2006b). Effects of transcranial direct current stimulation on working memory in patients with Parkinson's disease. *J. Neurol. Sci.* 249, 31–38.
- Bonakdarpour, B., Eftekhazadeh, A., and Ashayeri, H. (2000). Preliminary report on the effects of melodic intonation therapy in the rehabilitation of Persian aphasic patients. *Iran. J. Med. Sci.* 25, 156–160.
- Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., and Parsons, L. M. (2004). The song system of the human brain. *Brain Res. Cogn. Brain Res.* 20, 363–375.
- Cattaneo, Z., Pisoni, A., and Papagno, C. (2011). Transcranial direct current stimulation over Broca's region improves phonemic and semantic fluency in healthy individuals. *Neuroscience* 183, 64–70.
- Celnik, P., Paik, N. J., Vandermeeren, Y., Dimyan, M., and Cohen, L. G. (2009). Effects of combined peripheral nerve stimulation and brain polarization on performance of a motor sequence task after chronic stroke. *Stroke* 40, 1764–1771.
- Fiori, V., Coccia, M., Marinelli, C. V., Vecchi, V., Bonifazi, S., Ceravolo, M. G., Provinciali, L., Tomaiuolo, F., and Marangolo, P. (2010). Transcranial direct current stimulation improves word retrieval in healthy and non-fluent aphasic subjects. *J. Cogn. Neurosci.* 23, 2309–2323.

- Floel, A., Meinzer, M., Kirstein, R., Nijhof, S., Deppe, M., Knecht, S., and Breitenstein, C. (2011). Short-term anomia training and electrical brain stimulation. *Stroke* 42, 2065–2067.
- Fregni, F., Boggio, P. S., Nitsche, M., Bormpohl, F., Antal, A., Feredoes, E., Marcolin, M. A., Rigonatti, S. P., Silva, M. T., Paulus, W., and Pascual-Leone, A. (2005). Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. *Exp. Brain Res.* 166, 23–30.
- Fridriksson, J., Richardson, J. D., Baker, J. M., and Rorden, C. (2011). Transcranial direct current stimulation improves naming reaction time in fluent aphasia: a double-blind, sham-controlled study. *Stroke* 42, 819–821.
- Fritsch, B., Reis, J., Martinowich, K., Schambra, H. M., Yuan, Y., Cohen, L. G., and Lu, B. (2010). Direct current stimulation promotes BDNF-dependent synaptic plasticity: potential implications for motor learning. *Neuron* 66, 198–204.
- Gandiga, P. C., Hummel, F. C., and Cohen, L. G. (2006). Transcranial DC stimulation (tDCS): a tool for double-blind sham-controlled clinical studies in brain stimulation. *Clin. Neurophysiol.* 117, 845–850.
- Gerstman, H. L. (1964). A case of aphasia. *J. Speech Hear. Disord.* 29, 89–91.
- Geschwind, N. (1971). Current concepts: aphasia. *N. Engl. J. Med.* 284, 654–656.
- Goldstein, K. (1942). *After Effects of Brain-Injuries in War: Their Evaluation and Treatment*. New York: Grune and Stratton.
- Goodglass, H., and Geschwind, N. (1976). "Language disorders," in *Handbook of Perception: Language and Speech*, Vol. 7, eds E. Carterette and M. P. Friedman (New York: Academic Press), 390–428.
- Goodglass, H., and Kaplan, E. (1983). *Boston Diagnostic Aphasia Examination*, 2nd Edn. Philadelphia: Lea and Febiger.
- Hamilton, R. H., Chrysikou, E. G., and Coslett, B. (2011). Mechanisms of aphasia recovery after stroke and the role of noninvasive brain stimulation. *Brain Lang.* 118, 40–50.
- Hebert, S., Racette, A., Gagnon, L., and Peretz, I. (2003). Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain* 126, 1–13.
- Heiss, W. D., and Thiel, A. (2006). A proposed regional hierarchy in recovery of post-stroke aphasia. *Brain Lang.* 98, 118–123.
- Helm-Estabrooks, N., and Albert, M. L. (1991). *Manual of Aphasia Therapy*. Austin: Pro-Ed.
- Hesse, S., Werner, C., Schonhardt, E. M., Bardeleben, A., Jenrich, W., and Kirker, S. G. (2007). Combined transcranial direct current stimulation and robot-assisted arm training in subacute stroke patients: a pilot study. *Restor. Neurol. Neurosci.* 25, 9–15.
- Holland, A. L., Fromm, V., DeRuyter, F., and Stein, M. (1996). Treatment efficacy: aphasia. *J. Speech Hear. Res.* 39, S27–S36.
- Homan, R. W., Herman, J., and Purdy, P. (1987). Cerebral location of international 10–20 system electrode placement. *Electroencephalogr. Clin. Neurophysiol.* 66, 376–382.
- Hummel, F., Celnik, P., Giraux, P., Floel, A., Wu, W.-H., Gerloff, C., and Cohen, L. G. (2005). Effects of non-invasive cortical stimulation on skilled motor function in chronic stroke. *Brain* 128, 490–499.
- Iyer, M. B., Mattu, U., Grafman, J., Lomarev, M., Sato, S., and Wassermann, E. M. (2005). Safety and cognitive effect of frontal DC brain polarization in healthy individuals. *Neurology* 64, 872–875.
- Jeffries, K. J., Fritz, J. B., and Braun, A. R. (2003). Words in melody: an H-2 O-15 PET study of brain activation during singing and speaking. *Neuroreport* 14, 749–754.
- Keith, R. L., and Aronson, A. E. (1975). Singing as therapy for apraxia of speech and aphasia: report of a case. *Brain Lang.* 2, 483–488.
- Kertesz, A., and McCabe, P. (1977). Recovery patterns and prognosis in aphasia. *Brain* 100, 1–18.
- Kertesz, A., Lesk, D., and McCabe, P. (1977). Isotope localization of infarcts in aphasia. *Arch. Neurol.* 34, 590–601.
- Kincses, T. Z., Antal, A., Nitsche, M. A., Bartfai, O., and Paulus, W. (2004). Facilitation of probabilistic classification learning by transcranial direct current stimulation of the prefrontal cortex in the human. *Neuropsychologia* 42, 113–117.
- Kinsella, G., Prior, M. R., and Murray, G. (1988). Singing ability after right and left sided brain damage. A research note. *Cortex* 24, 165–169.
- Laughlin, S. A., Naeser, M. A., and Gordon, W. P. (1979). Effects of three syllable durations using the melodic intonation therapy technique. *J. Speech Hear. Res.* 22, 311–320.
- Liebetanz, D., Nitsche, M. A., Tergau, F., and Paulus, W. (2002). Pharmacological approach to the mechanisms of transcranial DC-stimulation-induced after-effects of human motor cortex excitability. *Brain* 125, 2238–2247.
- Lindenberg, R., Renga, V., Zhu, L. L., Nair, D., and Schlaug, G. (2010). Bihemispheric brain stimulation facilitates motor recovery in chronic stroke patients. *Neurology* 75, 2176–2184.
- Marchina, S., Zhu, L. L., Norton, A., Zipse, L., Wan, C. Y., and Schlaug, G. (2011). Impairment of speech production predicted by lesion load of the arcuate fasciculus. *Stroke* 42, 2251–2256.
- Mimura, M., Kato, M., Sano, Y., Kojima, T., Naeser, M., and Kashima, H. (1998). Prospective and retrospective studies of recovery in aphasia. Changes in cerebral blood flow and language functions. *Brain* 121, 2083–2094.
- Mohr, J. P., Pessin, M. S., Finkelstein, S., Funkenstein, H. H., Duncan, G. W., and Davis, K. R. (1978). Broca aphasia: pathologic and clinical. *Neurology* 28, 311–324.
- Monti, A., Cogiamanian, F., Marceglia, S., Ferrucci, R., and Mameli, F. (2008). Improved naming after transcranial direct current stimulation in aphasia. *J. Neurol. Neurosurg. Psychiatry* 79, 451–453.
- Nair, D. N., Renga, V., Hamelin, S., Pascual-Leone, A., and Schlaug, G. (2008). Improving motor function in chronic stroke patients using simultaneous occupational therapy and tDCS. *Stroke* 39, 542.
- Nitsche, M. A., Fricke, K., Henschke, U., Schlitterlau, A., Liebetanz, D., Lang, N., Henning, S., Tergau, F., and Paulus, W. (2003a). Pharmacological modulation of cortical excitability shifts induced by transcranial DC stimulation. *J. Physiol.* 553, 293–301.
- Nitsche, M. A., Schauenburg, A., Lang, N., Liebetanz, D., and Exner, C. (2003b). Facilitation of implicit motor learning by weak transcranial direct current stimulation of the primary motor cortex in the human. *J. Cogn. Neurosci.* 15, 619–626.
- Nitsche, M. A., and Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *J. Physiol.* 527, 633–639.
- Okamoto, M., Dan, H., Sakamoto, K., Takeo, K., Shimizu, K., Kohno, S., Oda, I., Isobe, S., Suzuki, T., Kohyama, K., and Dan, I. (2004). Three-dimensional probabilistic anatomical cranio-cerebral correlation via the international 10–20 system oriented for transcranial functional brain mapping. *Neuroimage* 21, 99–111.
- Ozdemir, E., Norton, A., and Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage* 33, 628–635.
- Pedersen, P. M., Vinter, K., and Olsen, T. S. (2004). Aphasia after stroke: type, severity and prognosis. The Copenhagen aphasia study. *Cerebrovasc. Dis.* 17, 35–43.
- Pizzamiglio, L., Galati, G., and Committeri, G. (2001). The contribution of functional neuroimaging to recovery after brain damage: a review. *Cortex* 37, 11–31.
- Priori, A., Berardelli, A., Rona, S., Accornero, N., and Manfredi, M. (1998). Polarization of the human motor cortex through the scalp. *Neuroreport* 9, 2257–2260.
- Racette, A., Bard, C., and Peretz, I. (2006). Making non-fluent aphasics speak: sing along! *Brain* 129, 2571–2584.
- Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., and Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport* 11, 1997–2000.
- Robey, R. R. (1994). The efficacy of treatment for aphasic persons: a meta-analysis. *Brain Lang.* 47, 582–608.
- Rogalewski, A., Breitenstein, C., Nitsche, M. A., Paulus, W., and Knecht, S. (2004). Transcranial direct current stimulation disrupts tactile perception. *Eur. J. Neurosci.* 20, 313–316.
- Rosen, H. J., Petersen, S. E., Linenweber, M. R., Snyder, A. Z., White, D. A., Chapman, L., Dromerick, A. W., Fiez, J. A., and Corbetta, M. D. (2000). Neural correlates of recovery from aphasia after damage to left inferior frontal cortex. *Neurology* 55, 1883–1894.
- Saur, D., Lange, R., Baumgaertner, A., Schraknepper, V., Willmes, K., Rijntjes, M., and Weiller, C. (2006). Dynamics of language reorganization after stroke. *Brain* 129, 1371–1384.
- Schlaug, G., Marchina, S., and Norton, A. (2008a). From singing to speaking: why singing may lead to recovery of expressive language function in patients with Broca's aphasia. *Music Percept.* 25, 315–323.
- Schlaug, G., Norton, A., and Marchina, S. (2008b). The role of the right hemisphere in post-stroke language recovery. *Stroke* 39, 542–543.
- Schlaug, G., Renga, V., and Nair, D. (2008c). Transcranial direct current stimulation in stroke recovery. *Arch. Neurol.* 65, 1571–1576.

- Schlaug, G., Marchina, S., and Norton, A. (2009). Evidence for plasticity in white-matter tracts of patients with chronic Broca's aphasia undergoing intense intonation-based speech therapy. *Ann. N. Y. Acad. Sci.* 1169, 385–394.
- Schlaug, G., Marchina, S., and Wan, C. Y. (2011). The use of non-invasive brain stimulation techniques to facilitate recovery from post-stroke aphasia. *Neuropsychol. Rev.* 21, 288–301.
- Schlaug, G., Norton, A., Marchina, S., Zipse, L., and Wan, C. Y. (2010). From singing to speaking: facilitating recovery from non-fluent aphasia. *Future Neurol.* 5, 657–665.
- Schlaug, G., and Renga, V. (2008). Transcranial direct current stimulation – a non-invasive tool to facilitate stroke recovery. *Expert Rev. Med. Devices* 5, 759–768.
- Schuppert, M., Münte, T., Wieringa, B., and Altenmüller, E. (2000). Receptive amusia: evidence for cross-hemispheric neural networks underlying music processing strategies. *Brain* 123, 546–559.
- Sparing, R., Meister, I. G., Wiennemann, M., Buelte, D., Staedtgen, M., and Boroojerdi, B. (2007). Task-dependent modulation of functional connectivity between hand motor cortices and neuronal networks underlying language and music: a transcranial magnetic stimulation study in humans. *Eur. J. Neurosci.* 25, 319–323.
- Sparks, R., Helm, N., and Albert, M. (1974). Aphasia rehabilitation resulting from melodic intonation therapy. *Cortex* 10, 303–316.
- Stefan, K., Kunesch, E., Cohen, L. G., Benecke, R., and Classen, J. (2000). Induction of plasticity in the human motor cortex by paired associative stimulation. *Brain* 123, 572–584.
- Thiel, A., Herholz, K., Koyuncu, A., Ghaemi, M., Kracht, L. W., Habedank, B., and Heiss, W. D. (2001). Plasticity of language networks in patients with brain tumors: a positron emission tomography activation study. *Ann. Neurol.* 50, 620–629.
- Vines, B. W., Cerruti, C., and Schlaug, G. (2008a). Dual-hemisphere tDCS facilitates greater improvements in motor performance compared to uni-hemisphere stimulation. *BMC Neurosci.* 9, 103. doi: 10.1186/1471-2202-9-103
- Vines, B. W., Nair, D. G., and Schlaug, G. (2008b). Modulating activity in the motor cortex affects performance for the two hands differently depending upon which hemisphere is stimulated. *Eur. J. Neurosci.* 28, 1667–1673.
- Vines, B. W., Nair, D. G., and Schlaug, G. (2006a). Contralateral and ipsilateral motor effects after transcranial direct current stimulation. *Neuroreport* 17, 671–674.
- Vines, B. W., Schnider, N., and Schlaug, G. (2006b). Testing for causality with tDCS: pitch memory and the left supramarginal gyrus. *Neuroreport* 17, 1047–1050.
- Wan, C., Rüber, T., Hohmann, A., and Schlaug, G. (2010). The therapeutic effects of singing in neurological disorders. *Music Percept.* 27, 287–295.
- Wilson, S. J., Parsons, K., and Reutens, D. C. (2006). Preserved singing in aphasia: a case study of the efficacy of melodic intonation therapy. *Music Percept.* 24, 23–35.
- Winhuisen, L., Thiel, A., Schumacher, B., Kessler, J., Rudolf, J., Haupt, W. F., and Heiss, W. D. (2005). Role of the contralateral inferior frontal gyrus in recovery of language function in post stroke aphasia: a combined repetitive transcranial magnetic stimulation and positron emission tomography study. *Stroke* 36, 1759–1763.
- Zheng, X., Alsop, D. C., and Schlaug, G. (2011). Effects of transcranial direct current stimulation (tDCS) on human regional cerebral blood flow. *Neuroimage* 58, 26–33.

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: 02 April 2011; accepted: 26 August 2011; published online: 26 September 2011.

Citation: Vines BW, Norton AC and Schlaug G (2011) Non-invasive brain stimulation enhances the effects of melodic intonation therapy. *Front. Psychology* 2:230. doi: 10.3389/fpsyg.2011.00230

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

Copyright © 2011 Vines, Norton and Schlaug. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.

APPENDIX

Table A1 | For each participant, this table includes spoken material that was intact at all four time-points, and thus, was used in the analysis.

Task category	Specific task	Intact portion(s)
PARTICIPANT 1		
Automatic speech	Pledge of allegiance	"I pledge allegiance to the flag of the United States of America"
	Counting	1–15
	Days of the week	Thursday through Sunday
	Months of the year	March through May
Picture description	Describe these scenes	"Cup of tea," "cream and sugar," "call the doctor," "...are my shoes," "brush your teeth," "clean it up" "my name is _____, what's your name?" "turn off the light,"
Picture naming	Name the object pictured	Abacus, acorn, apple, bathroom, beaver, cactus, giraffe, helicopter, money, pelican, pencil, pyramid, scissors, unicorn, water, whistle, window, zebra
PARTICIPANT 2		
Automatic speech	Pledge of allegiance	"I pledge allegiance to the flag of the United States of America, and to the Republic for which it stands, one nation under God, indivisible, with liberty and justice for all."
	Counting	1–10
	Days of the week	Monday through Sunday
	Months of the year	January through October
Picture description	Describe these scenes	"Open up," "go to bed," "call the doctor," "close the window," "where are my shoes?" "brush your teeth," "wait for me," "hi, my name is _____," "cup of coffee," "cream and sugar," "turn off the light," "I don't understand," "pass the butter"
Picture naming	Name the object pictured	Abacus, accordion, acorn, asparagus, beaver, cactus, camel celery compass, dominoes, hammock, hanger, harmonica, igloo, mushroom, octopus, overalls, palette, pelican, piano, pretzel, racket, rhinoceros, seahorse, strawberry, toothbrush, tripod, unicorn, volcano, wheelchair
PARTICIPANT 3		
Automatic speech	Pledge of allegiance	–
	Counting	1–8
	Days of the week	Monday through Tuesday
	Months of the year	March through June
Picture description	Describe these scenes	"Butter," "call the doctor, please," "shut the window, please," "brush your teeth," "where are my shoes?" "come over here," "my name is _____," "...you doing?" "coffee," "...and cream," "turn off the light," "I don't understand"
Picture naming	Name the object pictured	Bathroom, cat, celery, dominoes, flower, giraffe, hammer, money, overalls, pencil, piano, scissors, strawberry, toaster, whistle, zebra
PARTICIPANT 4		
Automatic speech	Pledge of allegiance	–
	Counting	1–11
	Days of the week	Monday through Wednesday
	Months of the year	January through August
Picture description	Describe these scenes	"Money"
Picture naming	Name the object pictured	–
PARTICIPANT 5		
Automatic speech	Pledge of allegiance	–
	Counting	1–5
	Days of the week	Monday through Tuesday
	Months of the year	–
Picture description	Describe these scenes	"Open up," "money," "ice cream," "I am hot," "fine," "bathroom," "I am sick," "I am hungry," "Where are my keys?" "watch TV," "what's your name?"
Picture naming	Name the object pictured	Duck, nose, pizza, sock, watch

(Continued)

Table A1 | Continued

Task category	Specific task	Intact portion(s)
PARTICIPANT 6		
Automatic speech	Pledge of allegiance	"...pledge allegiance to the flag of the United States of America, and to the Republic for which it stands, one nation under God, indivisible ..."
	Counting	1–20
	Days of the week	Monday through Sunday
	Months of the year	February through December
Picture description	Describe these scenes	–
Picture naming	Name the object pictured	–



How singing works

Katrin Schulze*

Developmental Cognitive Neuroscience Unit, UCL Institute of Child Health, London, UK

*Correspondence: kschulze@ich.ucl.ac.uk

A commentary on

Disorders of pitch production in tone deafness

by Dalla Bella, S., Berkowska, M., and Sowiński, J. (2011). *Front. Psychol.* 2:164. doi: 10.3389/fpsyg.2011.00164

Researchers have recently intensified their efforts to investigate the underlying neural correlates of music perception and processing (for an overview see Koelsch, 2011), music production, in comparison, has been studied rather sparsely.

The review article by Dalla Bella et al. (2011) on the other hand focuses on the cognitive and neural underpinnings of the human music production system that enables us to sing. Humans, as vocal learners, are not only capable of singing, but the authors point out that “singing is as natural as speaking for the majority of people.” Indeed, the ability of vocal learning, i.e., to imitate our auditory environment can be observed extremely early in human development: Newborns’ cry melody is influenced by the speech prosody of the surrounding spoken language (Mampe et al., 2009).

The review by Dalla Bella et al. (2011) defines the term “singing” and importantly categorizes and introduces us to several methods used to investigate singing. First, singing is an umbrella term for many different processes involving pitch production that differ in complexity: Pitch matching or repetition of tone sequences involve working memory to a different degree, whereas singing a well-known song requires the support of long-term memory. Second, there are different techniques to analyze singing, for example (i) the subjective rating of the singing output or (ii) the computing of an objective measurement of accuracy (deviation of produced pitch/interval from the target pitch/interval) and precision (consistency in producing pitch/interval). It is emphasized that different criteria – the type of

singing as well as the analysis used – can lead to very different estimates of the ability to sing.

Mainly based on functional neuroimaging studies, the authors developed the vocal sensorimotor loop (VSL) model to explain the cognitive and neural processes underlying singing. The model depicts the interplay between memory components, motor and auditory sensory areas, and emphasizes the role of sensorimotor integration for human singing. The process of sensorimotor integration (Hickok et al., 2011) has also been observed during verbal working memory and speech production in humans (Hickok et al., 2003, 2011; Koelsch et al., 2009) as well as for singing in songbirds (Prather et al., 2008; Mooney, 2009).

Furthermore, research that investigated the influence of perception on singing is reviewed. Studies do not provide a consistent picture so far, but a dissociation between perception and production is suggested: Whereas poor-pitch singing and perceptual deficits are in general associated in congenital amusia, cases of spared vocal performance and a deficient pitch perception and vice versa cases of intact perception and poor-singing have been reported. In line with the latter results, there are reports indicating that some musicians with absolute (perfect) pitch, a rare ability to name tones, are similarly not able to sing perfectly in tune (for an overview see Takeuchi and Hulse, 1993).

An important aspect of the review is the question of how much the neural networks supporting music and language production overlap, especially because research has mainly focused on comparing the perception and processing between language and music (for an overview see Koelsch, 2011). The comparison of the production of language and music indicates so far, that speech production seems to involve a more left-lateralized network, whereas singing seems to rely on a more bilateral network. The authors present preliminary

data suggesting that inaccuracy in pitch production does not extend to speech production in tone deafness, indicating that independent mechanisms are subserving imitation in music and language. An intriguing corroborating finding comes from a study investigating members of the KE family. The affected members of this family have an inherited speech–language disorder (verbal and orofacial dyspraxia) caused by a mutation of the FOXP2 gene, but they are not deficient in either the perception or production of pitch (Alcock et al., 2000).

To summarize, considering the importance of music for humans (Jancke, 2008, 2009; Koelsch, 2011), the review article by Dalla Bella et al. (2011) contributes to our understanding of human cognition by furthering our knowledge of normal and poor-pitch singing.

REFERENCES

- Alcock, K. J., Passingham, R. E., Watkins, K., and Vargha-Khadem, F. (2000). Pitch and timing abilities in inherited speech and language impairment. *Brain Lang.* 75, 34–46.
- Dalla Bella, S., Berkowska, M., and Sowiński, J. (2011). Disorders of pitch production in tone deafness. *Front. Psychol.* 2:164. doi: 10.3389/fpsyg.2011.00164
- 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.
- Hickok, G., Houde, J., and Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69, 407–422.
- Jancke, L. (2008). Music, memory and emotion. *J. Biol.* 7, 21.
- Jancke, L. (2009). The plastic human brain. *Restor. Neurol. Neurosci.* 27, 521–538.
- Koelsch, S. (2011). Toward a neural basis of music perception – a review and updated model. *Front. Psychol.* 2:110. doi: 10.3389/fpsyg.2011.00110
- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., and Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an fMRI study. *Hum. Brain Mapp.* 30, 859–873.
- Mampe, B., Friederici, A. D., Christophe, A., and Wermke, K. (2009). Newborns’ cry melody is shaped by their native language. *Curr. Biol.* 19, 1994–1997.
- Mooney, R. (2009). Neural mechanisms for learned bird-song. *Learn. Mem.* 16, 655–669.

Prather, J. F., Peters, S., Nowicki, S., and Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* 451, 305–310.

Takeuchi, A. H., and Hulse, S. H. (1993). Absolute pitch. *Psychol. Bull.* 113, 345–361.

Received: 16 December 2011; accepted: 10 February 2012; published online: 28 February 2012.

Citation: Schulze K (2012) How singing works. Front. Psychology 3:51. doi: 10.3389/fpsyg.2012.00051

This article was submitted to Frontiers in Auditory Cognitive Neuroscience, a specialty of Frontiers in Psychology.

Copyright © 2012 Schulze. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.



Disorders of pitch production in tone deafness

Simone Dalla Bella^{1,2*}, Magdalena Berkowska¹ and Jakub Sowiński¹

¹ Department of Cognitive Psychology, University of Finance and Management, Warsaw, Poland

² International Laboratory for Brain, Music and Sound Research, Montreal, QC, Canada

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Mari Tervaniemi, University of Helsinki, Finland

Katrin Schulze, UCL Institute of Child Health, UK

*Correspondence:

Simone Dalla Bella, Department of Cognitive Psychology, University of Finance and Management, Ul. Pawia 55, 01-030 Warsaw, Poland.
e-mail: sdallabella@vizja.pl

Singing is as natural as speaking for the majority of people. Yet some individuals (i.e., 10–15%) are poor singers, typically performing or imitating pitches and melodies inaccurately. This condition, commonly referred to as “tone deafness,” has been observed both in the presence and absence of deficient pitch perception. In this article we review the existing literature concerning normal singing, poor-pitch singing, and, briefly, the sources of this condition. Considering that pitch plays a prominent role in the structure of both music and speech we also focus on the possibility that speech production (or imitation) is similarly impaired in poor-pitch singers. Preliminary evidence from our laboratory suggests that pitch imitation may be selectively inaccurate in the music domain without being affected in speech. This finding points to separability of mechanisms subserving pitch production in music and language.

Keywords: music cognition, pitch production, tone deafness, congenital amusia, speech production, vocal performance, poor-pitch singing, cognitive neuroscience

INTRODUCTION

Making music (e.g., singing and dancing) is a universal form of expression, which is widespread across societies and cultures (Mithen, 2006). In particular, singing is as natural as speaking for the majority of people (Dalla Bella et al., 2007; Pfordresher and Brown, 2007; Dalla Bella and Berkowska, 2009). Adult singing is accurate (although not necessarily precise, see Pfordresher et al., 2010) on both pitch and time dimensions (Dalla Bella et al., 2007; Dalla Bella and Berkowska, 2009) and remarkably consistent both within and across individuals (e.g., Levitin, 1994; Levitin and Cook, 1996; Bergeson and Trehub, 2002). Extensive vocal training is not a *sine qua non* for singing in tune. The majority of individuals do not require formal vocal training or musical tutoring to sing proficiently. Singing emerges spontaneously during development, as observable vocalizations during the first months of life (Papoušek, 1996). This behavior, likely facilitated by the universality of maternal singing (e.g., Trehub and Trainor, 1999), is promptly imitated by infants. Eighteen-month-old children can produce recognizable songs, by repeating short musical phrases (e.g., see Ostwald, 1973; Welch, 2006, for a review). Finally, far from being merely a cultural frill, singing (and more generally vocalizations) is likely to have played a role during evolution. It is a common observation that people particularly enjoy singing when in group contexts (e.g., during religious ceremonies, in the military). This participatory aspect of singing is thought to foster group bonding, one of the reasons, together with sexual selection and mood regulation, why music may have some adaptive value (Wallin et al., 2000; Huron, 2001; Mithen, 2006).

In spite of the fact that singing is widespread, there are noticeable exceptions. In the general population, a few individuals, referred to as “tone deaf,” have notorious difficulties in carrying a tune. Recent studies estimate that approximately 10–15% of the general population may be particularly inaccurate in producing pitch, by singing quite far from the target pitches (in experiments requiring production of familiar melodies or imitation of single pitches, intervals, or simple

melodies; Dalla Bella et al., 2007; Pfordresher and Brown, 2007; Dalla Bella and Berkowska, 2009). This estimate rises to about 55%, however, when the consistency of repeated attempts to produce a pitch (i.e., precision) is taken into account (Pfordresher et al., 2010). A few studies have been carried out recently on poor-pitch singing showing that different sources of impairment may underline this disorder, such as perceptual deficits or auditory–motor integration (see Pfordresher and Brown, 2007; Berkowska and Dalla Bella, 2009b).

In the present article we review recent findings on normal singing in adults, and on poor-pitch singing with particular attention to those studies making use of quantitative estimates of singing proficiency (i.e., based on acoustic analyses). We mostly focus on singing proficiency in the majority (i.e., non-musicians). Note that there is a rich literature regarding acoustical features of professional singing (for reviews, see Sundberg, 1987, 1999). However, in only a few isolated studies has pitch production in professional singers been examined (Vurma and Ross, 2006; Zurbruggen et al., 2006). These studies revealed that when professional singers are asked to produce pitch intervals, they can be out of tune by 20–25 cents, with respect to the equally tempered scale. This error is typically not noticed by expert listeners (Vurma and Ross, 2006). In addition, features like the accuracy of the first note of the melody and melodic contour play a role in motor planning, as shown by asking singers to prepare for producing a melody (Zurbruggen et al., 2006).

In spite of the rich literature on vocal performance during development in Music Education (e.g., see Welch, 1979, 2006, for early studies on poor-pitch singing), this review is devoted to adult singing. We review behavioral and neuroimaging evidence, in order to examine the mechanisms which are likely malfunctioning in poor-pitch singers. Finally, since accurate production of pitch variations is a key process in both music and speech, we examine whether pitch production deficits in tone-deaf individuals extend to the language domain. To this aim, we present some preliminary and intriguing findings coming from our laboratory, suggesting that pitch production mechanisms may show domain-specificity.

NORMAL SINGING

HOW SINGING WORKS

Singing, like speaking, involves three independent physical components: respiratory, laryngeal (i.e., the vocal folds), and articulatory mechanisms (i.e., the vocal tract). Lungs provide the air supply needed for vocalization. The vocal folds modulate the airstream coming from the lungs (i.e., by chopping it into air pulses), a process referred to as “phonation.” Finally the vocal tract conveys to each sound the spectral and temporal properties characteristic of sung voice (e.g., Titze, 1994; Sundberg, 1999). The quality of the vocal output, telling apart professional from amateur singers, depends on the fine coordination of these mechanisms. The acoustical properties of the singing voice in professional singers have been the subject of a substantial body of research (for reviews, see Sundberg, 1987, 1999). For example, particular attention has been devoted to the so-called *singer’s formant* (i.e., partials falling in the frequency range of 2.5–3.0 KHz; Sundberg, 1987), which in professional singers is much stronger in sung vowels than in spoken vowels. The intensity of the singer’s formant, the presence of vibrato, the maximum phonational frequency range, and loudness all increase with musical experience (e.g., Brown et al., 2000; Mendes et al., 2003; Hunter et al., 2006). Until recently, however, evidence was scant on the mechanisms underlying singing proficiency (i.e., on the pitch and time dimensions).

Singing proficiency can be assessed with a variety of tasks. A quite natural task, which does not require vocal training, is to have participants sing a well-known song (i.e., “Happy birthday”) with lyrics from memory (e.g., Dalla Bella et al., 2007, 2009). A variant of the task is to ask participants to sing the same melody on a syllable (e.g., /la/); this typically results in enhanced performance, likely due to reduced memory load (Berkowska and Dalla Bella, 2009). Another possibility consists in presenting a model stimulus (e.g., a single pitch, an interval, or a short novel melody) which participants have to imitate, using a vowel or a syllable (pitch-matching tasks). Single-pitch-matching is a quite common task (Goetze et al., 1990; Pfordresher and Brown, 2007; Hutchins et al., 2010a) and is considered an important factor in assessing musical talent (Watts et al., 2003). In both singing from memory and imitation tasks, singing proficiency can be assessed with natural feedback, but also while the auditory feedback is augmented, for example by providing the correct pitch or melody as one sings (e.g., Pfordresher and Brown, 2007; Tremblay-Champoux et al., in press) or an altered pitch or melody (e.g., Hafke, 2008; Zarate and Zatorre, 2008). In sum, there are different ways to obtain a measure of singing proficiency. Interestingly, different tasks are likely to reflect the activity of partly different functional components of the song system. For example, singing from memory will particularly tap the retrieval of musical information from long-term memory. In contrast, working memory and auditory–motor mapping mechanisms are mainly targeted by imitation tasks (e.g., by augmenting or altering auditory feedback). Thus, different tasks may serve to assess the correct functioning of different components of the song system.

To shed light on the functional components of the song system underlying normal pitch production in the execution of the aforementioned tasks, and eventually to account for poor-pitch singing, we present in **Figure 1** a schema of the vocal sensorimotor loop (VSL; see also Berkowska and Dalla Bella, 2009b). This schema

is inspired, among others, by the Perceptual loop theory (Levelt, 1989), an account of performance self-monitoring and correction in speech, processes which are similarly relevant in music performance. The VSL includes perceptual, motor, auditory–motor mapping, and memory components. According to the VSL, singing of well-known melodies from memory is based on the retrieval from long-term memory of pitch and temporal information, followed by fine motor planning/implementation. The ongoing vocal production is fed back to the system (i.e., perception), matched with the planned melody, in some cases leading to error correction, thereby affecting planning of upcoming events. Vocal imitation of novel pitch sequences relies on short-term memory, and auditory–vocal mapping, without tapping retrieval from long-term memory. The target pitches to be imitated are perceptually analyzed, stored in short-term memory, and eventually mapped into motor gestures. As with singing from memory, self-monitoring of vocal performance is made possible by mechanisms allowing feedback analysis, auditory–motor mapping, and in some cases error correction. It is worth noting that overt and covert pathways for pitch perception are possible. The covert pathway is involved in tasks requiring explicit judgments of pitch differences (e.g., pitch discrimination). In some cases, participants are very inaccurate in judging pitch differences, still exhibiting proficient singing. In this condition it is hypothesized that pitch differences are analyzed via covert mechanisms, thus affording proficient singing (Griffiths, 2008; Loui et al., 2008; Dalla Bella et al., 2009).

MEASURING SINGING ACCURACY AND PRECISION

Singing proficiency has been mostly assessed by asking expert musicians to subjectively rate recordings (e.g., Hébert et al., 2003; Schön et al., 2004; Racette et al., 2006; Wise and Sloboda, 2008). This method provides a general and fast assessment of vocal performance. However, raters are not always consistent in providing their judgments (Kinsella et al., 1988; Prior et al., 1990). In addition, peers can hardly provide fine estimates of proficiency on the pitch dimension, independent of the time dimension, such as the exact deviation from the model pitch, variability over repetitions, and so forth. This is mostly due to the fact that musicians often categorize pitch and duration information with respect to the closest musical value. An alternative which has proven successful is to compute objective measurement of accuracy with acoustic methods (e.g., Murayama et al., 2004; Terao et al., 2006; Dalla Bella et al., 2007, 2009; Pfordresher et al., 2010). This method consists in computing note pitch onsets and pitch height after acoustic segmentation of the auditory signal. This information can be used to compute measures of accuracy and precision in vocal performance (Pfordresher et al., 2010).

Accuracy and precision can be computed separately for absolute pitch (i.e., the absolute pitch height of musical notes) and for relative pitch (i.e., the discrepancy between two subsequent pitches, or interval, expressed in semitones). For absolute pitch, accuracy indicates the average difference between sung and target pitches. Typically, such difference does not take into account the direction of the change (i.e., whether the sung pitch is on average higher or lower than the target pitch; Pfordresher et al., 2010; but see Pfordresher and Brown, 2007, for a measure of signed error). Another measure of accuracy in terms of absolute pitch, referred

to as “initial pitch deviation” (i.e., difference between the first produced pitch and the first note of a target melody), has been used in our laboratory (Berkowska and Dalla Bella, 2009a; Dalla Bella and Berkowska, 2009). Precision in terms of absolute pitch is the consistency in repeating the target pitch (i.e., whether a repeated note similarly deviates from the target across repetitions; Pfordresher et al., 2010). Another measure of variability related to precision, referred to as “pitch stability” (Dalla Bella et al., 2007, 2009), consists in computing the deviation of two reproductions of a single phrase in a melody. Similar measures of accuracy and precision of relative pitch can be computed for tasks where participants sang from memory or imitated pitch sequences. In this case, accuracy refers to the average difference between sung pitch intervals and target intervals based on the notation. This measure, sometimes referred to as “pitch interval deviation,” has been adopted in a few studies as a measure of singing proficiency (e.g., Dalla Bella et al., 2007, 2009; Berkowska and Dalla Bella, 2009a; Dalla Bella and Berkowska, 2009; Pfordresher et al., 2010). To our knowledge, precision of relative pitch (i.e., consistency in repeating the same target interval) has been examined in only one study (Pfordresher et al., 2010).

Objective acoustically based measures of singing accuracy and precision have the advantage of making explicit the criteria for teasing apart good from poor-pitch singers. To this aim, the choice of the measure(s) of singing proficiency has to be carefully made. Indeed, different criteria can lead to very different estimates of the prevalence of poor-pitch singing (much higher when pitch precision is considered instead of pitch accuracy; Pfordresher et al., 2010). In the majority of studies, accuracy in producing or imitating pitches is considered instead of precision. Individuals can be qualified as poor-pitch singers based on an absolute criterion, namely when in a pitch-matching task their produced pitches depart from a target pitch by more than a semitone (e.g., Pfordresher and Brown, 2007; Pfordresher et al., 2010). In other cases, individuals can be classified as poor-pitch singers relative to a control/comparison group, as often observed in single-case studies of patients with brain damage (e.g., Schön et al., 2004; Satoh et al., 2007), thus adopting a variable criterion. An alternative is to treat as poor-pitch singers those individuals who are outliers in a given group, for example deviating from the average of the group by more than 2 SD (e.g., Dalla Bella and Berkowska, 2009). A final distinction which is worth mentioning is between measures of accuracy based on absolute pitch differences (i.e., the deviation of produced pitch from the target pitch, in imitation tasks; e.g., Pfordresher and Brown, 2007) and measures based on relative pitch differences (i.e., the deviation of produced intervals from the target interval in singing from memory or imitation tasks; e.g., Dalla Bella et al., 2007). Due to the various criteria for defining poor-pitch singers, comparison of different studies is meaningful provided that poor-pitch singers have been selected using similar criteria.

SINGING IN THE GENERAL POPULATION

Until recently, relatively little was known about singing abilities in the general population. People generally tend to underestimate their ability to carry a tune. For example, almost 60% of 1000 university students reported that they cannot accurately imitate melodies (Pfordresher and Brown, 2007). Moreover, self-declared tone-deaf individuals, that is about 17% of the student population, believe

that they cannot sing proficiently (Cuddy et al., 2005). Yet, systematic assessments of singing proficiency in the general population indicate that around 85–90% can sing in tune (Dalla Bella et al., 2007; Pfordresher and Brown, 2007; Dalla Bella and Berkowska, 2009; but see Pfordresher et al., 2010, for a lower estimate when considering precision instead of accuracy).

We examined singing proficiency in the majority by testing a group of 62 occasional singers in Montreal (20 university students in the lab, and 42 participants recruited in a public park) as compared to 4 professional singers (Dalla Bella et al., 2007). Participants sang the refrain of a well-known song with lyrics. Renditions were submitted to acoustical analyses. Occasional singers were less accurate in producing pitch intervals (with a deviation of 0.6 semitones from the correct intervals, on average) than professional singers (deviation of 0.3 semitones). At the same time, occasional singers sang faster than professionals, a phenomenon tied to lower pitch accuracy. Further tests on 15 participants indicated that slowing down tempo typically enhances accuracy in producing pitch intervals. Yet, two participants (i.e., poor-pitch singers) did not exhibit any improvement as a result of the slower tempo. That the majority can carry a tune was confirmed more recently on a larger sample of occasional singers with familiar musical material, and comparing production (i.e., singing from memory) with imitation (Berkowska and Dalla Bella, 2009a; Dalla Bella and Berkowska, 2009). Occasional singers were less accurate when they sang from memory than in the imitation task. Moreover, their performance was more accurate when they sang on a syllable (i.e., with lower memory load) than with lyrics.

Other studies focused on vocal imitation abilities (e.g., pitch matching). The first studies examining single-pitch-matching (i.e., imitation of single pitches) revealed that adults perform poorly on this task (Ternström et al., 1988; Murbe et al., 2002; Amir et al., 2003). For example, non-musicians typically deviate by 1.3 semitones on average as compared to 0.5 semitones for musicians (Ternström et al., 1988; Murry, 1990; Murry and Zwiner, 1991; Amir et al., 2003). This estimate of accuracy in pitch imitation in non-musicians may be too defeatist, though. Low accuracy in imitating pitch does not characterize all individuals without musical training (Estis et al., 2009). Moreover, poorer performance in non-musicians may partly result from using pure tones as models for imitation. When imitating synthesized voices or sung performances, non-musicians achieved higher accuracy, exhibiting pitch deviations around 0.5 semitones or less (Pfordresher and Brown, 2007; Wise and Sloboda, 2008; see also Watts and Hall, 2008). Hence, accuracy in pitch-matching depends on the acoustic features of the stimulus to be imitated (for similar results with children, see also Small and McCachern, 1983; Green, 1990). Target stimuli sharing acoustical properties (i.e., spectral and temporal features) with the vocal production are likely to facilitate mapping onto sensorimotor representations, thus entailing enhanced accuracy.

Imitation of single pitches, intervals, and short novel melodies has been systematically assessed by Pfordresher and Brown (2007). A large sample of university students without musical training imitated various pitch sequences (i.e., a single repeated note, a sequence including a single change of pitch, and short four-note melodies). Most participants were able to perform the task accurately (i.e., with renditions within 1 semitone from the

target pitches). In addition, occasional singers were less accurate in terms of both absolute and relative pitch in imitating short melodies as compared to single pitches (as in Wise and Sloboda, 2008). That accuracy in pitch-matching decreases with the number of elements in a sequence is likely reflecting working memory constraints. Note that impaired working memory is a relevant factor in defining the profile of individuals with congenital music disorders (Tillmann et al., 2009; Williamson et al., 2010b), which is likely to affect their vocal production (Dalla Bella et al., 2009; Tremblay-Champoux et al., in press). The finding that occasional singers are typically quite accurate in imitating short unfamiliar melodies was replicated recently by Pfordresher et al. (2010). Nevertheless, these authors found that the majority was imprecise (i.e., the SD of the fundamental frequency for renditions of the same pitch class or interval exceeded 1 semitone). This intriguing finding suggests that precision, instead of accuracy, may be taken into consideration when subjectively judging our own performance. This may account for the very high percentage of individuals with singing difficulties by self-report. To sum up, even though early studies suggest that occasional singers are quite inaccurate in imitating single pitches, recent studies yielded more optimistic results. Nevertheless, accuracy in imitating pitch rapidly decreases with increasing sequence length and complexity. Moreover, even though occasional singers are quite accurate in imitating pitches, they may still be not very consistent over repetitions. The most recent studies devoted to singing proficiency in the general population are summarized in **Table 1**.

It is worth mentioning that some studies focused on the relation between accuracy in pitch-matching tasks and pitch discrimination skills. Indeed, it is possible that some occasional singers are particularly accurate in imitating pitch sequences due to fine monitoring of their own performance, allowing for efficient error correction.

For example, pitch matching in untrained singers covaries positively with the ability to discriminate pitches (i.e., good singers are more accurate in discriminating pitches than poor singers; e.g., Watts et al., 2003, 2005). Yet other studies failed to replicate this finding (Bradshaw and McHenry, 2005; Moore et al., 2008). This situation is reminiscent of studies comparing perception and performance skills in accurate singing during development, similarly yielding conflicting results (for studies showing a link between pitch perception and production, see Phillips and Aitchinson, 1997; Demorest, 2001; Demorest and Clements, 2007; for lack of replication, see Roberts and Davis, 1975; Geringer, 1983; Apfelstadt, 1984). In sum, whether (and to what extent) pitch perception and production are linked in adult occasional singers is still a matter of debate. The possibility of a dissociation of perception and action in vocal performance is addressed when we will discuss poor-pitch singing in tone deafness.

NEURONAL UNDERPINNINGS OF THE SONG SYSTEM

Singing is supported by a complex neural network involving motor and sensory areas, as well as auditory-motor integration regions. Several neuroimaging studies have been conducted with the goal of uncovering the neuronal underpinnings of the human song system. In this review we focus on the neuronal mechanisms underlying the main components of the VSL (see **Figure 1**). Motor areas (e.g., primary motor cortex), and in particular the mouth region (e.g., Brown et al., 2004) and the larynx/phonation area, are recruited during singing (by adduction/abduction and tension/relaxation of the vocal folds; see Brown et al., 2008). Sensory areas, such as the superior temporal gyrus, are also engaged by vocal performance, for example when repeating a single note (Perry et al., 1999), or singing more complex melodies (Brown et al., 2004; Kleber et al., 2007). Other cortical areas which systematically are recruited by vocal

Table 1 | Summary of the main recent studies using acoustical measures of pitch accuracy and precision in normal individuals without musical training.

Reports	Tasks	Main results
Amir et al. (2003)	Single pitch-matching	Non-musicians are more inaccurate (1.3 semitones) than musicians (0.5 semitones). High accuracy is related to superior pitch discrimination.
Watts et al. (2005)	Single pitch-matching	Accuracy in the general population spans from 0.9 semitones (good singers) to 2.2 semitones (poor singers).
Dalla Bella et al. (2007)	Singing from memory of familiar melodies	Most occasional singers (90%) are accurate on the pitch (<0.5 semitones) and time dimensions, when performing at a slow tempo.
Pfordresher and Brown (2007)	Imitation of single pitches, intervals, and short novel melodies	85–90% of the non-musicians can imitate the target stimuli (error <1 semitone). Accuracy in pitch-matching is not related to pitch discrimination abilities.
Wise and Sloboda (2008)	Imitation of single pitches, and of short pitch patterns (2, 3, 5 notes)	Non-musicians exhibit good accuracy (<1 semitone), in particular for short stimuli (around 0.2/0.3 semitones for single pitches).
Dalla Bella and Berkowska (2009); Berkowska and Dalla Bella (2009a)	Singing from memory and imitation of familiar melodies	Around 90% of occasional singers can imitate or sing from memory a familiar melody (accuracy = 0.5–0.6 semitones). Imitation (in particular on a syllable) affords higher accuracy.
Pfordresher et al. (2010)	Imitation of 5-note unfamiliar pitch sequences, and singing from memory of familiar songs	54–60% of non-musicians are imprecise singers in terms of absolute and/or relative pitch. The relation between precision and accuracy is asymmetric. Whereas low accuracy is almost always associated with low precision, the reverse is not true.

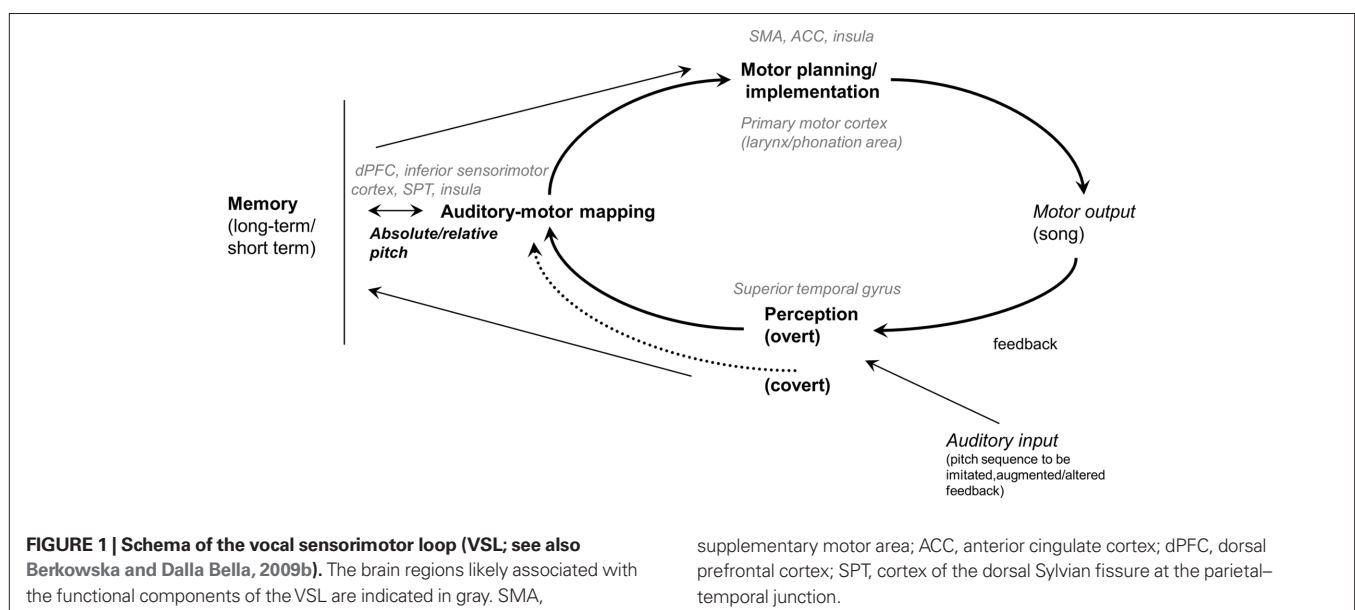
performance are the supplementary motor area (SMA), the anterior cingulate cortex (ACC), and the insula. For example, the SMA is notoriously engaged in high-level motor control, and needed for efficient motor planning in sequence production, such as in overt speech production (e.g., Turkeltaub et al., 2002). The ACC is a region associated with initiation of vocalization (see Jurgens, 2002, for a review), and activated during overt speech and singing (Perry et al., 1999; Paus, 2001). Finally, singing recruits the insula (Perry et al., 1999; Brown et al., 2004; Kleber et al., 2007; Zarate and Zatorre, 2008). This region, in particular the anterior insula, is mostly associated with articulation processes during vocalization (e.g., Dronkers, 1996), and given its connections with both the ACC and with auditory areas, it may play a role in integrating auditory feedback with motor output (Riecker et al., 2000; Ackermann and Riecker, 2004).

Other studies focused on the neuronal mechanisms acting as an interface between the sensory and the motor systems, thus affording sensorimotor mapping/integration (see **Figure 1**). Such areas are for example the dorsal prefrontal cortex, inferior sensorimotor cortex, and the superior temporal gyrus and sulcus, which are active both when speaking and singing (Özdemir et al., 2006; Gunji et al., 2007; Zarate and Zatorre, 2008). Another region involved in auditory–motor integration in vocal performance is area SPT (i.e., cortex of the dorsal Sylvian fissure at the parietal–temporal junction), which is recruited by both covert speech and covert humming (Hickok et al., 2003; Pa and Hickok, 2008). Area SPT has gained particular attention as a key region functioning as a sensorimotor interface between speech perception, working memory, and speech production (Hickok and Poeppel, 2007; Hickok et al., 2009). To sum up, neuroimaging evidence points to a complex neuronal network supporting vocal performance in singing. Further studies are needed to clarify the involvement of this network in various vocal tasks (e.g., pitch matching and singing from memory) and its relations with observed individual differences in terms of accuracy and precision.

Since both singing and speech involve vocalization and analysis of auditory feedback, it is reasonable to ask to what extent they rely on dedicated processes or rather share the same neuronal network (for a review, see Gordon et al., 2006). Brain areas underlying speaking and singing significantly overlap in non-musicians (e.g., Brown et al., 2006; Wilson et al., 2010). Nevertheless, singing appears to predominantly recruit right-hemisphere regions whereas speech production recruits primarily areas in the left hemisphere. Covert singing of familiar tunes without lyrics is correlated with greater activation in right sensorimotor cortex; in contrast, speaking an overlearned word string involves left sensorimotor cortex (Wildgruber et al., 1996; Riecker et al., 2000; Ackermann and Riecker, 2004). A similar lateralization pattern was found when speaking and singing with lyrics were contrasted (Callan et al., 2006, with covert performance; Jeffries et al., 2003, with overt performance). These findings are partly supported by brain stimulation studies (with transcranial magnetic stimulation, TMS). Applying TMS over left-hemisphere regions associated with speech production (e.g., near Broca's area) disrupts speech production. Comparable stimulation over homologous regions of the right hemisphere revealed inconsistent results across participants. Stimulation disrupted singing in only a minority of participants (Epstein et al., 1999; Lo and Fook-Chong, 2004), a finding recently replicated with subdural cortical stimulation (Suarez et al., 2010). Thus, song production shows, in general, more bilateral involvement than speech production.

POOR-PITCH SINGING

Both a brain insult and neurogenetic (i.e., congenital) disorders can disrupt the functioning of the song system, thereby leading to poor-pitch singing. In the present review we selectively focus on poor-pitch singing in otherwise healthy participants without musical training (i.e., tone-deaf individuals). Studies on vocal amusia or oral-expressive amusia consequent to brain damage have been reviewed elsewhere (Marin and Perry, 1999; Gordon et al.,



2006; Ackermann et al., 2006; Berkowska and Dalla Bella, 2009b; Stewart et al., 2009). About 10–15% of the population is inaccurate when asked to sing a melody from memory or when imitating a pitch sequence (Dalla Bella et al., 2007; Pfordresher and Brown, 2007; Dalla Bella and Berkowska, 2009). Poor-pitch singing is often treated as a sign of more general lack of musicality, or tone deafness, a term which became widespread (see Sloboda et al., 2005, for a discussion). The term “tone deafness” literally suggests that poor-pitch singing may stem from a deficient perceptual system. That impoverished perception may lead to poor-pitch singing is consistent with the VSL schema, where impaired perception would hinder self-monitoring during performance, affect auditory–motor mapping, and thereby limit error correction (but see below for the possibility of covert perception).

Poor-pitch perception characterizes a condition referred to as “congenital amusia” (Peretz, 2001; Ayotte et al., 2002; Peretz et al., 2002; Peretz and Hyde, 2003). Amusics are typically unable to tell the difference between melodies differing by a single note, exhibit difficulties in pitch discrimination, and thereby cannot recognize familiar tunes (Ayotte et al., 2002; Peretz et al., 2002; Hyde and Peretz, 2004). Congenital amusia is a neurogenetic disorder observed in about 4% of the general population (Kalmus and Fry, 1980; Peretz and Hyde, 2003; Peretz et al., 2007) and is associated with brain anomalies in the auditory cortex, inferior frontal cortex, and reduced connectivity between these areas (Hyde et al., 2006, 2007, 2011; Mandell et al., 2007; Loui et al., 2009). In a recent study we showed that poor-pitch singing and perceptual deficits are in general associated in congenital amusia (Dalla Bella et al., 2009). Eleven individuals with congenital amusia (determined with the Montreal battery of evaluation of amusia, MBEA; Peretz et al., 2003) sang a familiar melody from memory. Nine of them were inaccurate in producing pitch intervals when singing with lyrics. However, more than half of them could not sing more than a few notes when asked to perform the same tune without lyrics (i.e., on a syllable), a condition which was expected to improve accuracy (Berkowska and Dalla Bella, 2009a). The pattern of results may have arisen because of weak memory traces of the musical components of songs (e.g., Dalla Bella et al., 2009). In general, amusics who were the least accurate in producing pitch intervals also exhibited the highest pitch discrimination thresholds (i.e., low sensitivity to pitch difference) in a perceptual task (Hyde and Peretz, 2004), a finding consistent with the hypothesis that perception and action are tightly coupled in vocal performance (but see below for exceptions). Similar impairments in pitch production in amusics were observed with pitch-matching tasks (Hutchins et al., 2010a), showing in addition that, because of their perceptual disorders, amusics do not benefit from perceptual information (e.g., additional feedback) to improve or correct their performance. In sum, in general congenital amusics are inaccurate in singing from memory and pitch-matching tasks, a deficit associated with their impoverished pitch perception.

DISSOCIATIONS BETWEEN PERCEPTION AND ACTION IN TONE DEAFNESS

That poor-pitch singing is typically associated with perceptual disorders in congenital amusia does not entail that inaccurate singing cannot occur in isolation. Indeed, deficient motor planning or inaccurate auditory–motor mapping, regardless of spared

perception, are sufficient conditions leading to poor-pitch singing (see the VSL schema). Accordingly, some individuals exhibit poor-pitch singing without deficient pitch perception (Bradshaw and McHenry, 2005; Dalla Bella et al., 2007; Pfordresher and Brown, 2007; Wise and Sloboda, 2008), a condition referred to as “purely vocal tone deafness” (Dalla Bella et al., 2007). For example, in a previous study we found that among 15 occasional singers asked to sing a well-known melody at a slow tempo, a condition which should have increased pitch accuracy, 13 sang proficiently at a slow tempo; yet, 2 participants were still inaccurate singers (Dalla Bella et al., 2007), with sung intervals departing by at least 1 semitone from the notated intervals (vs. 0.3 semitones on average for other participants). Moreover, participants were able to detect pitch and time incongruities in unfamiliar melodies, thus showing normal perception. A similar dissociation between perception and action is found when considering pitch-matching tasks and imitation (Pfordresher and Brown, 2007). In a study by Pfordresher and Brown (2007), 13% of the tested sample of 79 non-musicians were poor-pitch singers, because the produced pitches departed by at least 1 semitone from the target. In spite of inaccurate pitch imitation, poor-pitch singers performed as accurately as proficient singers in a pitch discrimination task, confirming that poor-pitch singing is not merely the outcome of impoverished perception (see also Wise and Sloboda, 2008, for additional evidence in favor of a dissociation between perception and action with pitch-matching tasks).

Surprisingly, cases of spared vocal performance with deficient perception are also described. In a study by Loui et al. (2008) congenital amusics imitated tone intervals and in a second task judged whether the second tone of a pair was higher or lower than the first. Both congenital amusics and controls could imitate pitch direction. Yet amusics were unable to detect pitch direction, thus suggesting that there may be two separate streams for auditory perception and action (Griffiths, 2008). The two streams are indicated in the VSL schema as overt and covert perceptual pathways. We replicated this finding in a group of five congenital amusics who had difficulties in discriminating melodies differing in terms of melodic contour. In spite of their perceptual deficit, they produced the correct pitch direction when they sang a melody from memory (Dalla Bella et al., 2009). Interestingly, the found dissociation between perception and action is not confined to pitch direction. In a study where we assessed singing proficiency in congenital amusia, we found two individuals who, in spite of severely deficient pitch perception as revealed by the MBEA, could sing with lyrics as proficiently as controls (Dalla Bella et al., 2009).

Dissociations in the pitch domain between perception and performance suggest that poor-pitch singing may stem from different sources of malfunctioning within the song system. The next step would be to try to clarify which mechanisms do not work properly within the VSL in different poor-singing “phenotypes” and whether (or to what extent) they are music-specific or rather general-purpose mechanisms which underlie vocal performance also in other domains such as language. A thorough description of the possible causes of poor-pitch singing goes beyond the scope of this article (for reviews, see Pfordresher and Brown, 2007; Berkowska and Dalla Bella, 2009b). Here we briefly summarize these accounts, which are most relevant to the question of domain-specificity, referring to the VSL (see **Figure 1**). Poor-pitch singing resulting from perceptual

deficits (i.e., in congenital amusics) can be accounted for by the malfunctioning of (covert and overt) extraction of pitch information from the auditory input (herein, the “perceptual account”). The inability to extract pitch information hinders appropriate monitoring of the ongoing performance thereby leading to inadequate error correction and to diminished accuracy. Due to impaired perceptual monitoring congenital amusics are typically not aware of their deficit. However, the observation that production deficits can co-occur with spared perception (e.g., Dalla Bella et al., 2007; Pfordresher and Brown, 2007; Wise and Sloboda, 2008) and that perception is spared in vocal amusia consequent to brain damage (Schön et al., 2004) suggest that in those cases post-perceptual mechanisms are faulty. Pfordresher and Brown (2007; but see also Mandell et al., 2007) have proposed that these processes rather involve sensorimotor integration (herein, the “sensorimotor account”), or memory retrieval/motor planning (e.g., Pfordresher and Brown, 2007; Wise and Sloboda, 2008).

The sensorimotor account deserves particular attention. According to this account, a correct auditory representation of the vocal performance (and, for example, of feedback) would be inaccurately mapped to motor representations for phonation, which is spared in itself. Inaccurate mapping may concern the reproduction of local musical features (absolute pitch, and secondarily pitch intervals) without affecting global features (e.g., melodic contour; see Pfordresher and Brown, 2007). In addition, the relative independence of measures of absolute and relative pitch accuracy reported in poor-pitch singers (Dalla Bella and Berkowska, 2009) suggest that mapping relative and absolute musical features to motor movement may be selectively disrupted. That production of absolute and relative pitch features may engage at least partly independent mechanisms is supported by differential effects of feedback on pitch accuracy (i.e., choral singing enhances pitch accuracy in producing intervals and contour, but is detrimental for producing absolute pitch; Pfordresher and Brown, 2007). The sensorimotor account is plausible from a neurobiological point of view. The idea that poor-pitch singing results from disrupted or underdeveloped pathways bridging perception and action is supported by recent evidence of abnormally reduced connectivity of the fasciculus arcuatus (i.e., a pathway connecting temporal and frontal brain areas) in tone deafness (Loui et al., 2009).

In the following section we will focus on the question of domain-specificity of the pitch deficits reported in poor-pitch singers; particular attention will be paid to perceptual and sensorimotor mechanisms.

DOES INACCURACY IN PITCH PRODUCTION EXTEND TO SPEECH IN TONE DEAFNESS?

The findings reviewed so far indicate that there is a variety of deficits under the label “poor-pitch singing” and which refer to the malfunctioning of different components within the VSL. Are these deficits the result of disrupted mechanisms which are specifically engaged in vocal production of music, or rather equally subserving other functions, such as speech production? Indeed, pitch plays a prominent role in the structure of both speech and music (for a thorough review, see Patel, 2008). Whether music and language are subserved by independent or shared neuronal networks has been the object of a number of studies, mostly in the area of perception.

For example, the recent modular account of melody perception and recognition proposed by Peretz and Coltheart (2003) assumes that music and language are mostly independent systems, including a network of processes triggered selectively by music or speech material (but see increasing evidence of shared mechanisms across the two domains, e.g., Hickok et al., 2003; Koelsch, et al., 2009; Williamson et al., 2010a). The separate domain view is based on long-standing neuropsychological evidence of double dissociations between music and speech processing in patients with brain damage and, more recently, in individuals with congenital amusia (for reviews, see Dalla Bella and Peretz, 1999; Peretz, 2001; Peretz and Hyde, 2003; Peretz and Zatorre, 2005). For example there are cases of patients who are unable to recognize familiar tunes, while they can normally recognize lyrics and speech in general. The opposite condition is also observed. A brain insult can disrupt the ability to recognize spoken words while leaving intact the ability to recognize music (see Peretz, 1993; Peretz and Zatorre, 2005; Stewart et al., 2009).

Nevertheless, far from reflecting complete integration or total independence of music and language mechanisms, current theories rather propose a more mitigated (and probably realistic) account. Even scholars strongly inclined toward a modular account of music and speech perception (e.g., Peretz and Coltheart, 2003) acknowledge that the representation of pitch direction is likely to be common to music and speech (i.e., melodic contour and intonation, respectively). This conclusion is supported by evidence from patients with brain damage, who display deficits in perceiving both melodic contour and speech intonation (e.g., Patel et al., 1998). Moreover, in spite of early claims that individuals with congenital amusia have spared pitch processing in speech perception (Ayotte et al., 2002; Peretz et al., 2002), further studies revealed impaired discrimination of speech intonation in this condition (e.g., Patel, 2008). Congenital amusics exhibit a deficit in processing fine-grained pitch differences, irrespective of the domain. Because pitch differences underlying prosodic differences in speech (e.g., between questions and statements) are coarser than in melodies, impaired pitch perception is more visible within a musical context than in speech. Deficits in the two domains are detectable in congenital musics provided that pitch variation is comparable across the domains (Hutchins et al., 2010b).

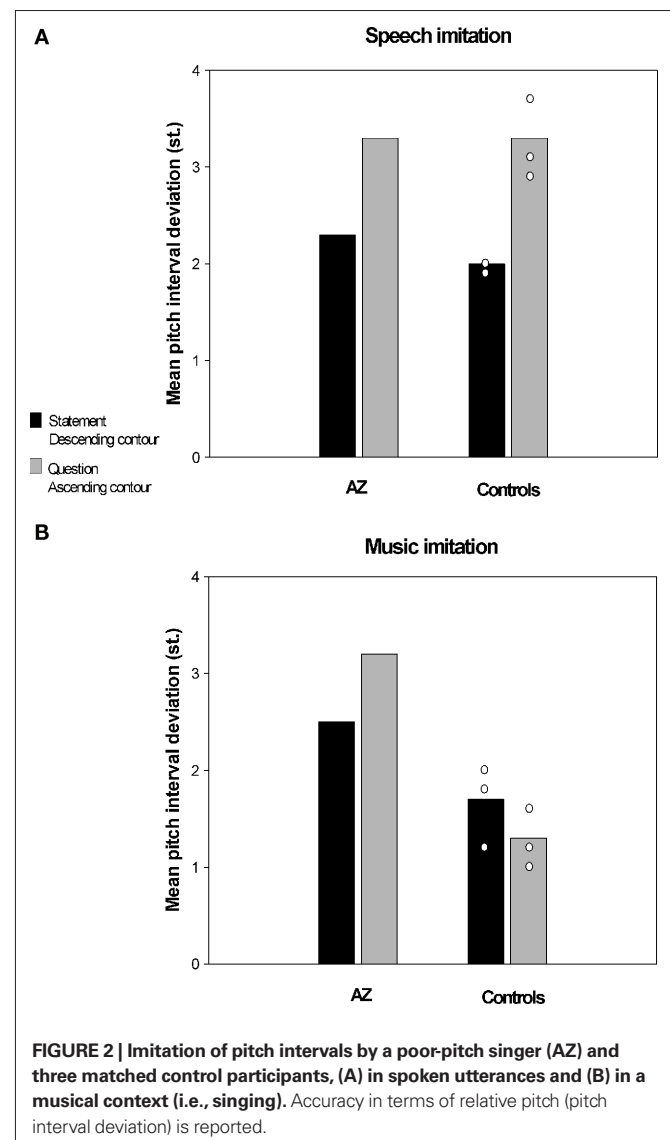
Dissociations between music and speech in pitch production are also reported in brain-damaged patients. A common observation in clinical neurology is that non-fluent aphasics exhibit major difficulty in speaking intelligibly whereas they can produce recognizable songs (e.g., Assal et al., 1977; Yamadori et al., 1977; but see Hébert et al., 2003; Racette et al., 2006; see also Gordon et al., 2006, for a review). In contrast there are cases of amusic patients who cannot sing while they can speak normally (e.g., Peretz et al., 1994). For example, Schön et al. (2004) reported the case of a tenor singer with lesions in right frontotemporoparietal regions, a pure case of vocal amusia exhibiting a specific deficit of the production of musical intervals. Interestingly, rhythm and contour were spared, as was musical perception and language abilities. In particular, perception and production of pitch variations in speech (i.e., intonation) were not impaired. For example, the patient was able to read texts using the appropriate accentuated expression. However, notice that poor singing often co-occurs with linguistic deficits resulting from

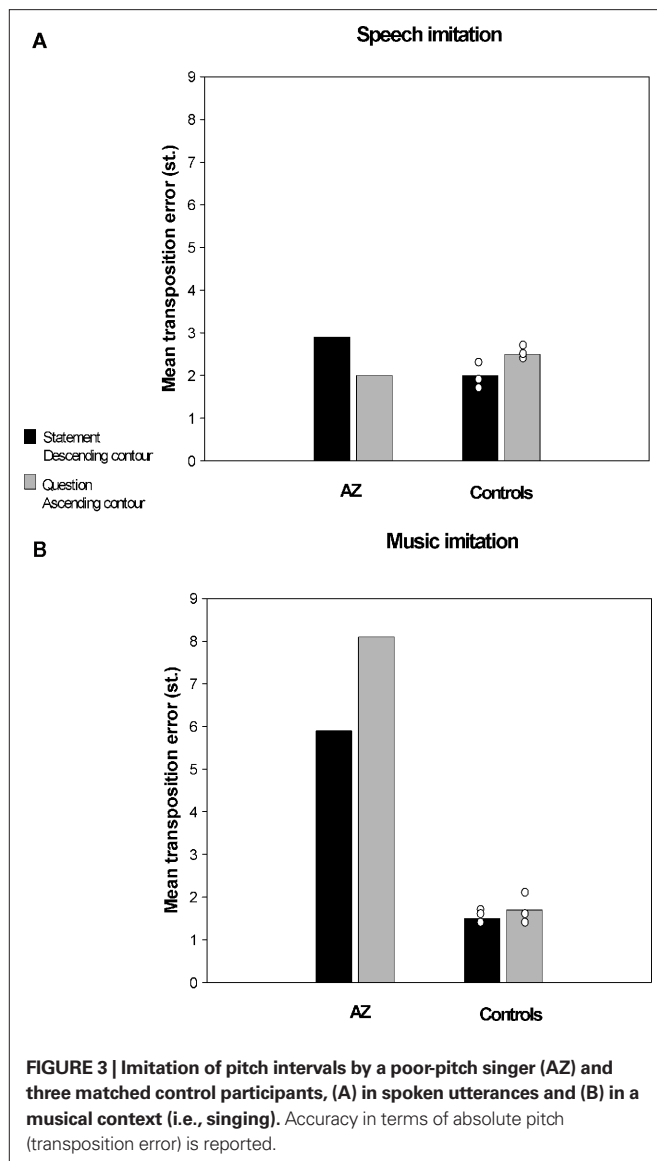
left-hemisphere damage (e.g., Benton, 1977). Furthermore, bilateral hemispheric involvement in sung performance is substantiated by evidence that lesions in either of the two hemispheres impair sung performance (Kinsella et al., 1988; Prior et al., 1990; for a review of further evidence, see Berkowska and Dalla Bella, 2009b). In sum, data coming from the study of brain-damaged patients indicate, as observed in perception, that pitch production in music and language can be independently disrupted by a brain injury, pointing toward different underlying mechanisms. However, the co-occurrence of singing and speech deficits, and the fact that singing involves both hemispheres suggests that some mechanisms may be shared (e.g., production of prosody, as observed in perception).

Studies on poor-pitch singing in tone deafness are theoretically an important source of evidence to test the domain-specificity of the mechanisms underlying pitch production. An intriguing question would be whether poor imitation in poor-pitch singers is similarly observed when pitch variations occur in a linguistic context. Since poor-pitch singing is mostly the outcome of perceptual deficits (see the perceptual account, above) or from inaccurate auditory–motor mapping (see the sensorimotor account), it is possible that inaccurate pitch processing is observed in both speech and singing. Indeed, poor-pitch singers who do not perceive pitch accurately (i.e., congenital amusics, Dalla Bella et al., 2009) also show difficulties in treating pitch in a linguistic context (e.g., Hutchins et al., 2010b). Moreover, auditory–motor integration is likely to be underpinned by the same neuronal substrate during speech and singing (Hickok et al., 2003; Pa and Hickok, 2008). In sum, impaired processes within the VSL underlying poor-pitch singing may also affect pitch production in the context of speech. This possibility is consistent with recent evidence showing that linguistic background (e.g., comparing tone vs. non-tone languages) shapes both perception and imitation of musical pitch (Pfordresher and Brown, 2009).

In a recent study conducted in our laboratory (Dalla Bella and Berkowska, in preparation) we examined imitation of pitch in music and speech contexts in AZ, a tone-deaf individual. AZ is a university student with 14 years of general education and without musical training. Her performance on singing from memory tasks and familiar melody imitation tasks shows that AZ is highly inaccurate in pitch production. For example, she deviates from the correct pitch intervals on average by 1.4 semitones (vs. 0.4 semitones for matched controls), and makes on average 6.4 contour errors (vs. 1.5 for controls). Interestingly, poor-pitch singing is not accompanied by major perceptual deficits. AZ normally perceives interval differences; yet, she reveals slightly impaired perception of melodic contour as shown by the MBEA (Peretz et al., 2003). That AZ's perception is mostly intact is confirmed by the fact that she enjoys music, and also paradoxically her own singing, which indicate that AZ, as is often the case in tone deafness, is not aware of her disorder. In order to assess whether AZ's poor-pitch singing extends to imitation of pitch in speech context she was asked to perform an interval imitation task. AZ and control participants imitated a short spoken or sung fragment with words (e.g., “klub gra mecz,” eng. “the team is playing the game”). Sentences to be imitated were questions (i.e., with ascending intonation) and statements (i.e., with descending intonation). Similarly, sung stimuli had an ascending or descending contour. The material to be imitated was recorded by a professional singer and manipulated so that the spoken and sung

fragments had the same pitch content. Accuracy of pitch imitation was computed using acoustic analyses (as in Dalla Bella et al., 2007, 2009). Accuracy in imitating relative pitch (i.e., pitch interval deviation) and absolute pitch (i.e., transposition error) for AZ and three control participants is reported in **Figures 2 and 3** respectively. Pitch interval deviation is the absolute deviation of produced intervals from the intervals to be imitated (in semitones). Transposition error is the absolute deviation of the first produced pitch from the first pitch of the stimulus to be imitated (in semitones). As can be seen, AZ was very inaccurate both in terms of absolute and relative pitch when imitating pitch in a musical context. Yet, she was comparable to controls when imitating the same pitch intervals while repeating sentences. Note that in this case the observed differences cannot be accounted for by differences in interval size between speech and music material, since pitch variations were the same in both cases. To our knowledge, this dissociation between pitch production depending on the context (i.e., musical or linguistic) is reported





REFERENCES

- Ackermann, H., and Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: a review and a hypothesis. *Brain Lang.* 89, 320–328.
- Ackermann, H., Wildgruber, D., and Riecker, A. (2006). “Singing in the (b) rain: cerebral correlates,” in *Music, Motor Control and the Brain*, eds E. Altenmüller, M. Wiesendanger, and J. Kesselring (Oxford: Oxford University Press), 205–222.
- Amir, O., Amir, N., and Kishon-Rabin, L. (2003). The effect of superior auditory skills on vocal accuracy. *J. Acoust. Soc. Am.* 113, 1102–1108.
- Apfelstadt, H. (1984). Effects of melodic perception instruction on pitch discrimination and vocal accuracy of kindergarten children. *J. Res. Music Educ.* 32, 15–24.
- Assal, G., Buttet, J., and Javet, R. C. (1977). Musical aptitude in aphasics. *Rev. Med. Suisse Romande* 97, 5–12.
- 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.
- Benton, A. L. (1977). “The amusias,” in *Music and the Brain*, eds M. Critchley and R. A. Henson (London: William Heinemann), 378–397.
- Bergeson, T. R., and Trehub, S. E. (2002). Absolute pitch and tempo in mothers’ songs to infants. *Psychol. Sci.* 13, 72–75.
- Berkowska, M., and Dalla Bella, S. (2009a). Reducing linguistic information enhances singing proficiency in occasional singers. *Ann. N. Y. Acad. Sci.* 1169, 108–111.
- Berkowska, M., and Dalla Bella, S. (2009b). Acquired and congenital disorders of sung performance: a review. *Adv. Cogn. Psychol.* 5, 69–83.
- Bradshaw, E., and McHenry, M. A. (2005). Pitch discrimination and pitch-matching abilities of adults who sing inaccurately. *J. Voice* 14, 431–439.
- Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. R., and Parsons, L. M. (2004). The song system of the human brain. *Cogn. Brain Res.* 20, 363–375.
- 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.
- Brown, S., Ngan, E., and Liotti, M. (2008). A larynx area in the human motor cortex. *Cereb. Cortex* 18, 837–845.
- Brown, W. S. Jr., Rothman, H. B., and Sapienza, C. (2000). Perceptual and acoustic study of professionally trained versus untrained voices. *J. Voice* 14, 301–309.
- Callan, D. E., Tsytarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., and Turner, R. (2006). Song and speech: brain regions involved with perception and covert production. *Neuroimage* 31, 1327–1342.
- Cuddy, L. L., Balkwill, L.-L., Peretz, I., and Holden, R. R. (2005). Musical difficulties are rare. A study of “tone deafness” among university students. *Ann. N. Y. Acad. Sci.* 1060, 311–324.
- Dalla Bella, S., and Berkowska, M. (2009). Singing proficiency in the majority: normality and “phenotypes” of poor singing. *Ann. N. Y. Acad. Sci.* 1169, 99–107.
- Dalla Bella, S., Giguère, J.-F., and Peretz, I. (2007). Singing proficiency in the

for the first time in poor-pitch singers. This finding suggests that the mechanisms underlying imitation of pitch differences in music and language may enjoy functional separability.

CONCLUSION

In the present article we sought to review evidence regarding poor-pitch singing, as compared to normal singing, and briefly overview the causes responsible for this condition. Poor-pitch singing may result from different sources, as indicated in the VSL schema. Particular attention was paid to perceptual mechanisms and sensorimotor mapping. Malfunctioning of these mechanisms can result in poor performance in tasks, such as singing familiar melodies from memory or vocal imitation, leading to a variety of disorders. Interestingly, these processes are similarly crucial for pitch production in a language context. Therefore, co-occurrence of pitch production deficits in music and language would be expected. In this case we focused on poor-pitch singing in a population of otherwise normal individuals (i.e., with tone deafness). Previous studies have demonstrated that pitch perception in music and language are not completely independent in tone deafness (e.g., the case of melodic contour and speech intonation). However, we provided here preliminary evidence in a tone-deaf individual showing that imitation of pitch intervals is very inaccurate in terms of both absolute and relative pitch while singing whereas no such deficits are observed while speaking. This intriguing finding points to an independence of mechanisms subserving imitation in music and language in the production domain. Further studies on a larger sample of tone-deaf individuals are required to examine whether these findings more generally characterize this condition and eventually to clarify which mechanisms within the VSL are domain-specific.

ACKNOWLEDGMENTS

This research was supported by a research grant from the European Commission (ITN-EBRAMUS, n. 238157), and by internal grants from the University of Finance and Management in Warsaw. We thank two anonymous reviewers for helpful comments on an earlier version of this manuscript.

- general population. *J. Acoust. Soc. Am.* 121, 1182–1189.
- Dalla Bella, S., Giguère, J.-F., and Peretz, I. (2009). Singing in congenital amusia. *J. Acoust. Soc. Am.* 126, 414–424.
- Dalla Bella, S., and Peretz, I. (1999). Music agnosias: selective impairments of music recognition after brain damage. *J. New Music Res.* 3, 209–216.
- Demorest, S. M. (2001). Pitch-matching performance of junior high boys: a comparison of perception and production. *Bull. Counc. Res. Music Educ.* 151, 63–70.
- Demorest, S. M., and Clements, A. (2007). Factors influencing the pitch-matching of junior high boys. *J. Res. Music Educ.* 55, 190–203.
- Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature* 384, 159–161.
- Epstein, C. M., Meador, K. J., Loring, D. W., Wright, R. J., Weissmann, J. D., Sheppard, S., Lah, J. J., Puhallovich, F., Gaitan, L., and Davey, K. R. (1999). Localization and characterization of speech arrest during transcranial magnetic stimulation. *Clin. Neurophysiol.* 110, 1073–1079.
- Estis, J. M., Coblenz, J. K., and Moore, R. E. (2009). Effects of increasing time delays on pitch-matching accuracy in trained singers and untrained individuals. *J. Voice* 23, 439–445.
- Geringer, J. M. (1983). The relationship of pitch-matching and pitch discrimination abilities of preschool and fourth-grade students. *J. Res. Music Educ.* 31, 93–99.
- Goetze, M., Cooper, N., and Brown, C. J. (1990). Recent research on singing in the general music classroom. *Bull. Counc. Res. Music Educ.* 104, 16–37.
- Gordon, R. L., Racette, A., and Schön, D. (2006). “Sensory-motor networks in singing and speaking: a comparative approach,” in *Music, Motor Control and the Brain*, eds E. Altenmüller, M. Wiesendanger, and J. Kesselring (Oxford: Oxford University Press), 205–222.
- Green, G. A. (1990). The effect of vocal modeling on pitch-matching accuracy of elementary schoolchildren. *J. Res. Music Educ.* 38, 225–231.
- Griffiths, T. D. (2008). Sensory systems: auditory action streams? *Curr. Biol.* 18, R387–R388.
- Gunji, A., Ishii, R., Chau, W., Kakigi, R., and Pantev, C. (2007). Rhythmic brain activities related to singing in humans. *Neuroimage* 34, 426–434.
- Hafke, H. Z. (2008). Nonconscious control of fundamental voice frequency. *J. Acoust. Soc. Am.* 123, 273–278.
- Hébert, S., Racette, A., Gagnon, L., and Peretz, I. (2003). Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain* 126, 1838–1850.
- Hickok, G., Buchsbaum, B., Humpries, 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.
- Hickok, G., Okada, K., and Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J. Neurophysiol.* 101, 2725–2732.
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Hunter, E. J., Svec, J. G., and Titze, I. R. (2006). Comparison of the produced and perceived voice range profiles in untrained and trained classical singers. *J. Voice* 20, 513–526.
- Huron, D. (2001). Is music an evolutionary adaptation? *Ann. N. Y. Acad. Sci.* 930, 43–61.
- Hutchins, S., Zarate, J. M., Zatorre, R. J., and Peretz, I. (2010a). An acoustical study of vocal pitch matching in congenital amusia. *J. Acoust. Soc. Am.* 127, 504–512.
- Hutchins, S., Gosselin, N., and Peretz, I. (2010b). 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., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A., and Peretz, I. (2007). Cortical thickness in congenital amusia: when less is better than more. *J. Neurosci.* 27, 13028–13032.
- Hyde, K. L., and Peretz, I. (2004). Brains that are out of tune but in time. *Psychol. Sci.* 15, 356–360.
- 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.
- Hyde, K. L., Zatorre, R. J., and Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cereb. Cortex* 21, 292–299.
- Jeffries, K. J., Fritz, J. B., and Braun, A. R. (2003). Words in melody: an H(2)15O PET study of brain activation during singing and speaking. *Neuroreport* 15, 749–754.
- Jurgens, U. (2002). Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* 26, 235–258.
- Kalmus, H., and Fry, D. B. (1980). On tone deafness (dysmelodia): frequency, development, genetics and musical background. *Ann. Hum. Genet.* 43, 369–382.
- Kinsella, G., Prior, M. R., and Murray, G. (1988). Singing ability after right and left sided brain damage. A research note. *Cortex* 24, 165–169.
- Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., and Lotze, M. (2007). Overt and imagined singing of an Italian aria. *Neuroimage* 36, 889–900.
- 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.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Levitin, D. J. (1994). Absolute memory for musical pitch: evidence from the production of learned melodies. *Percept. Psychophys.* 56, 414–423.
- Levitin, D. J., and Cook, P. R. (1996). Memory for musical tempo: additional evidence that auditory memory is absolute. *Percept. Psychophys.* 58, 927–935.
- Lo, Y. L., and Fook-Chong, S. (2004). Ipsilateral and contralateral motor inhibitory control in musical and vocalization tasks. *Exp. Brain Res.* 159, 258–262.
- Loui, P., Alsop, D., and Schlaug, G. (2009). Tone deafness: a new disconnection syndrome? *J. Neurosci.* 29, 10215–10220.
- Loui, P., Guenther, F., Mathys, C., and Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Curr. Biol.* 18, R331–R332.
- Mandell, J., Schulze, K., and Schlaug, G. (2007). Congenital amusia: an auditory-motor feedback disorder. *Restor. Neurol. Neurosci.* 25, 323–334.
- Marin, O. S. M., and Perry, D. W. (1999). “Neurological aspects of music perception and performance,” in *Psychology of Music*, ed. D. Deutsch (San Diego: Academic Press), 653–724.
- Mendes, A. P., Rothman, H. B., Sapienza, C., and Brown, W. S. Jr. (2003). Effects of vocal training on the acoustic parameters of the singing voice. *J. Voice* 17, 529–543.
- Mithen, S. (2006). *The Singing Neanderthals*. Cambridge, MA: Harvard University Press.
- Moore, R. E., Estis, J., Gordon-Hickey, S., and Watts, C. (2008). Pitch discrimination and pitch matching abilities with vocal and nonvocal stimuli. *J. Voice* 22, 399–407.
- Murayama, J., Kashiwagi, T., Kashiwagi, A., and Mimura, M. (2004). Impaired pitch production and preserved rhythm production in a right brain-damaged patient with amusia. *Brain Cogn.* 56, 36–42.
- Murbe, D., Friedemann, P., Hofmann, G., and Sundberg, J. (2002). Significance of auditory and kinesthetic feedback to singers’ pitch control. *J. Voice* 16, 44–51.
- Murry, T. (1990). Pitch-matching accuracy in singers and nonsingers. *J. Voice* 4, 317–321.
- Murry, T., and Zwiner, P. (1991). Pitch matching ability of experienced and inexperienced singers. *J. Voice* 5, 197–202.
- Ostwald, P. F. (1973). Musical behavior in early childhood. *Dev. Med. Child Neurol.* 15, 367–375.
- Özdemir, E., Norton, A., and Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage* 33, 628–635.
- Pa, J., and Hickok, G. (2008). A parietal-temporal sensory-motor integration area for the human vocal tract: evidence from an fMRI study of skilled musicians. *Neuropsychologia* 46, 362–368.
- Papoušek, H. (1996). “Musicality in infancy research: biological and cultural origins of early musicality,” in *Musical Beginnings*, eds I. Deliège and J. Sloboda (Oxford: Oxford University Press), 37–55.
- Patel, A. D. (2008). *Music, Language, and the Brain*. New York, NY: Oxford University Press.
- Patel, A. D., Peretz, I., Tramo, M., and Labèque, R. (1998). Processing prosodic and musical patterns: a neuropsychological investigation. *Brain Lang.* 61, 123–144.
- Paus, T. (2001). Primate anterior cingulate cortex: where motor control, drive and cognition interact. *Nat. Rev. Neurosci.* 2, 417–424.
- Peretz, I. (1993). “Auditory agnosia: a functional analysis,” in *Thinking in Sound. The Cognitive Psychology of Human Audition*, eds S. McAdams and E. Bigand (New York: Oxford University Press), 199–230.
- Peretz, I. (2001). Brain specialization for music: new evidence from congenital amusia. *Ann. N. Y. Acad. Sci.* 930, 189–192.
- Peretz, I., Ayotte, J., Zatorre, R., Mehler, J., Ahad, P., Penhune, V., and Jutras, B. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron* 33, 185–191.
- 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.
- Peretz, I., and Coltheart, M. (2003). Modularity of music processing. *Nat. Neurosci.* 6, 688–691.
- Peretz, I., Cummings, S., and Dubé, M.-P. (2007). The genetics of congenital amusia (tone deafness): a family-aggregation study. *Am. J. Hum. Genet.* 81, 582–588.
- Peretz, I., and Hyde, K. (2003). What is specific to music processing? Insights

- from congenital amusia. *Trends Cogn. Sci.* 7, 362–367.
- Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurize, G., and Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain* 117, 1283–1301.
- Peretz, I., and Zatorre, R. (2005). Brain organization for music processing. *Annu. Rev. Psychol.* 56, 89–114.
- Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., and Evans, A. C. (1999). Localization of cerebral activity during simple singing. *Neuroreport* 10, 3979–3984.
- Pfordresher, P. Q., and Brown, S. (2007). Poor-pitch singing in the absence of “tone-deafness”. *Music Percept.* 25, 95–115.
- Pfordresher, P. Q., and Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Atten. Percept. Psychophys.* 71, 1385–1398.
- Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., and Liotti, M. (2010). Imprecise singing is widespread. *J. Acoust. Soc. Am.* 128, 2182–2190.
- Phillips, K. H., and Aitchinson, R. E. (1997). The relationship of singing accuracy to pitch discrimination and tonal aptitude among third-grade students. *Contrib. Music Educ.* 24, 7–22.
- Prior, M., Kinsella, G., and Giese, J. (1990). Assessment of musical processing in brain-damaged patients: implications for laterality of music. *J. Clin. Exp. Neuropsychol.* 12, 301–312.
- Racette, A., Bard, C., and Peretz, I. (2006). Making non-fluent aphasics speak: sing along! *Brain* 129, 2571–2584.
- Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., and Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport* 11, 1997–2000.
- Roberts, E., and Davis, A. E. (1975). Poor pitch singing: the response of monotone singers to a program of remedial training. *J. Res. Music Educ.* 23, 227–239.
- Satoh, M., Takeda, K., and Kuzuhara, S. (2007). A cause of auditory agnosia with impairment of perception and expression of music: cognitive processing and tonality. *Eur. Neurol.* 58, 70–77.
- Schön, D., Lorber, B., Spacal, M., and Semenza, C. (2004). A selective deficit in the production of exact musical intervals following right-hemisphere damage. *Cogn. Neuropsychol.* 21, 773–784.
- Sloboda, J. A., Wise, K. J., and Peretz, I. (2005). Quantifying tone deafness in the general population. *Ann. N. Y. Acad. Sci.* 1060, 255–261.
- Small, A. R., and McCachern, F. L. (1983). The effect of male and female vocal modeling on pitch-matching accuracy of first-grade children. *J. Res. Music Educ.* 31, 227–233.
- Stewart, L., von Kriegstein, K., Dalla Bella, S., Warren, J. D., and Griffiths, T. D. (2009). “Disorders of musical cognition,” in *Oxford Handbook of Music Psychology*, S. Hallam, I. Cross, and M. Thaut (New York: Oxford University Press), 184–196.
- Suarez, R. O., Golby, A., Whalen, S., Satoh, S., Theodore, W. H., Kufta, C. D., Devinsky, O., Balish, M., and Bromfield, E. B. (2010). Contributions to singing ability by the posterior portion of the superior temporal gyrus of the non-language-dominant hemisphere: first evidence from subdural cortical stimulation, Wada testing, and fMRI. *Cortex* 46, 343–353.
- Sundberg, J. (1987). *The Science of the Singing Voice*. DeKalb, IL: Northern Illinois University Press.
- Sundberg, J. (1999). “The perception of singing,” in *The Psychology of Music*, ed. D. Deutsch (San Diego: Academic Press), 171–214.
- Terao, Y., Mizuno, T., Shindoh, M., Sakurai, Y., Ugawa, Y., Kobayashi, S., Nagai, C., Furubayashi, T., Arai, N., Okabe, S., Mochizuki, H., Hanajima, R., and Tsuji, S. (2006). Vocal amusia in a professional tango singer due to a right superior temporal cortex infarction. *Neuropsychologia* 44, 479–488.
- Ternström, S., Sundberg, J., and Collden, A. (1988). Articulatory F0 perturbations and auditory feedback. *J. Speech Hear. Res.* 31, 187–192.
- Tillmann, B., Schulze, K., and Foxton, J. M. (2009). Congenital amusia: a short-term memory deficit for nonverbal, but not verbal sounds. *Brain Cogn.* 71, 259–264.
- Titze, I. R. (1994). *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice Hall.
- Trehub, S. E., and Trainor, L. J. (1999). Singing to infants: lullabies and play songs. *Adv. Infancy Res.* 12, 43–77.
- Tremblay-Champoux, A., Dalla Bella, S., Phillips-Silver, J., Lebrun, M.-A., and Peretz, I. (in press). Singing proficiency in congenital amusia: Imitation helps. *Cogn. Neuropsychol.*
- Turkeltaub, P. E., Eden, G. F., Jones, K. M., and Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage* 16, 765–780.
- Vurma, A., and Ross, J. (2006). Production and perception of musical intervals. *Music Percept.* 23, 331–344.
- Wallin, N. L., Merker, B., and Brown, S. (2000). *The Origins of Music*. Cambridge, MA: MIT Press.
- Watts, C., Moore, R., and McCaghren, K. (2005). The relationship between vocal pitch-matching skills and pitch discrimination skills in untrained accurate and inaccurate singers. *J. Voice* 19, 534–543.
- Watts, C., Murphy, J., and Barnes-Burroughs, K. (2003). Pitch-matching accuracy of trained singers, untrained participants with talented singing voices, and untrained participants with non-talented singing voices in conditions of varying feedback. *J. Voice* 17, 187–196.
- Watts, C. R., and Hall, M. D. (2008). Timbral differences on vocal pitch-matching accuracy. *Logoped. Phoniatr. Vocol.* 33, 74–82.
- Welch, G. F. (2006). “Singing and vocal development,” in *The Child as Musician: A Handbook of Musical Development*, ed. G. McPherson (New York: Oxford University Press), 311–329.
- Welch, G. H. (1979). Poor pitch singing: a review of the literature. *Psychol. Music* 7, 50–58.
- Wildgruber, D., Ackermann, H., Klose, U., Kardatzki, B., and Grodd, W. (1996). Functional lateralization of speech production at primary motor cortex: a fMRI study. *Neuroreport* 7, 2791–2795.
- Williamson, V. J., Baddeley, A. D., and Hitch, G. J. (2010a). Musicians’ and nonmusicians’ short-term memory for verbal and musical sequences: comparing phonological similarity and pitch proximity. *Mem. Cognit.* 38, 163–175.
- Williamson, V. J., McDonald, C., Deutsch, D., Griffiths, T. D., and Stewart, L. (2010b). Faster decline of pitch memory over time in congenital amusia. *Adv. Cogn. Psychol.* 6, 15–22.
- Wilson, S. J., Abbott, D. F., Lusher, D., Gentle, E. C., and Jackson, G. D. (2010). Finding your voice: a singing lesson from functional imaging. *Hum. Brain Mapp.* doi: 10.1002/hbm.21173. [Epub ahead of print].
- Wise, K. J., and Sloboda, J. A. (2008). Establishing an empirical profile of self-defined “tone deafness”: perception, singing performance and self-assessment. *Music. Sci.* 12, 3–23.
- Yamadori, A., Osumi, Y., Masuhara, S., and Okubo, M. (1977). Preservation of singing in Broca’s aphasia. *J. Neurol. Neurosurg. Psychiatr.* 40, 221–224.
- Zarate, J. M., and Zatorre, R. J. (2008). Experience-related neural substrates involved in vocal pitch regulation during singing. *Neuroimage* 40, 1871–1887.
- Zurbruggen, E. L., Fontenot, D. L., and Meyer, D. E. (2006). Representation and execution of vocal motor programs for expert singing of tonal melodies. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 944–963.

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: 02 March 2011; accepted: 01 July 2011; published online: 14 July 2011.

Citation: Dalla Bella S, Berkowska M and Sowiński J (2011) Disorders of pitch production in tone deafness. *Front. Psychology* 2:164. doi: 10.3389/fpsyg.2011.00164

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

Copyright © 2011 Dalla Bella, Berkowska and Sowiński. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



Transfer of training between music and speech: common processing, attention, and memory

Mireille Besson^{1*}, Julie Chobert¹ and Céline Marie²

¹ Institut de Neurosciences Cognitives de la Méditerranée, CNRS, Université de la Méditerranée, Marseille, France

² Auditory Development Laboratory, Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, ON, Canada

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Psyche Loui, Beth Israel Deaconess Medical Center/Harvard Medical School, USA
Antoine Shahin, The Ohio State University, USA

*Correspondence:

Mireille Besson, Institut de Neurosciences Cognitives de la Méditerranée, CNRS, Université de la Méditerranée, 31-Chemin Joseph Aiguier, 13402 Marseille CEDEX 20, France.
e-mail: mireille.besson@incm.cnrs-mrs.fr

After a brief historical perspective of the relationship between language and music, we review our work on transfer of training from music to speech that aimed at testing the general hypothesis that musicians should be more sensitive than non-musicians to speech sounds. In light of recent results in the literature, we argue that when long-term experience in one domain influences acoustic processing in the other domain, results can be interpreted as common acoustic processing. But when long-term experience in one domain influences the building-up of abstract and specific percepts in another domain, results are taken as evidence for transfer of training effects. Moreover, we also discuss the influence of attention and working memory on transfer effects and we highlight the usefulness of the event-related potentials method to disentangle the different processes that unfold in the course of music and speech perception. Finally, we give an overview of an on-going longitudinal project with children aimed at testing transfer effects from music to different levels and aspects of speech processing.

Keywords: transfer effects, music training, speech processing, passive listening, mismatch negativity, active discrimination, attention, working memory

A BRIEF HISTORICAL PERSPECTIVE

Over the centuries, many authors have been interested in the relationship between music and language and this rich and fruitful comparison has been examined from many different perspectives. Historically, one of the first was the question of the common or independent origin of music and language that was highlighted again more recently in the wonderful book “The origin of music” (Wallin et al., 2000).

Both Rousseau (1781/1993) and Darwin (1871/1981) were in favor of a common origin of music and language. In his book on the origin of language (1781/1993), Rousseau was a fervent advocate of the idea that the first languages were sung, not spoken and Darwin considered that music evolved from the love calls produced during the reproduction period to charm the persons from the opposite sex: “musical notes and rhythm were first acquired by the male or female progenitors of mankind for the sake of charming the opposite sex” (Darwin, 1871/1981, p. 336). Such a seduction function of the singing voice certainly persisted nowadays as can be seen from the cult devoted to la Callas or Michael Jackson... The philosopher Herbert Spencer (1820–1903) also favored a common origin of music and language, and proposed a physiological theory to explain their common primary function: express emotions (Spencer, 1857). To produce large intervals in the intonation of the voice or on a keyboard requires larger movements than to produce small intervals. There is thus a direct connection between emotion and movement: the more intense the feeling, the larger, and faster the movement.

This very same idea was previously developed by Descartes (1618/1987), who in the early part of his long career, wrote “L’abrégé de musique” (Compendium musicae, 1618), his principal

contribution to music theory. In this book, Descartes divided music into three basic components: the physical aspect of sound, the nature of sensory perception, and the ultimate effect of this perception on the listener. Interestingly, the nature of perception results from the effects on the mind of the “animated spirits” through the pineal gland. Thus, listening to music activates the animated spirits in the brain who in turn affected the mind. When music has a fast tempo, the animated spirits are highly excited: they rush into the nerves that in turn excite the muscles. And you end up beating the tempo with your feet... (see also Cross, 2011 for the history of the link between the auditory and motor systems in music). Cartesian dualism does not therefore imply for the mind and body to be entirely separated. Rather, there are some instances, such as emotions, that cannot be attributed only to the mind or only to the body but who likely emerge from their tight union.

By contrast, to these early views, musicologists from the nineteenth and twentieth centuries seemed to favor the hypothesis of a different origin of music and language, and often argued that both evolved independently. While it is generally considered that both music and language have a survival value¹ in the evolution of the human species, through the production of sounds allowing individuals to locate themselves in space and to warn each other about potential dangers, for instance (Levman, 1992), they nevertheless evolved as different systems. For Wallaschek (1891), music emerged

¹Within this consensus, the dissonant voice of Pinker claimed “As far as biological cause and effect are concerned, music is useless... music could vanish from our species and the rest of our lifestyle would be virtually unchanged” (*How the mind works*, 1998). He clearly stands in opposition to Claude Lévi-Strauss who wrote “If we could explain music, we may find the key of human thought” (cited by Gardner, 1983, p. 123).

from a basic need to download surplus of excessive energy through rhythmic productions. As noted by Brown (2003), synchronized movements to music are found in all cultures. Newman (1905/1969) also considered that music developed before and independently from language: “man certainly expresses his feeling in pure indefinite sound long before he had learned to agree with his fellows to attach certain meanings to certain stereotypes sounds” (p. 210).

While music and language both play a fundamental role in the organization of human societies, the main function of music and language differs. Music allows expressing emotions and thereby ensures social bonding (Boucouchie, 1993). Ethnomusicological research has illustrated the social function of music by showing that music is invested of natural and supra-natural powers in all human societies (Nadel, 1930). By contrast, language permits to communicate thoughts relevant to the current context, to tell stories that happened years ago and to project into the future to plan upcoming events. Over the course of human evolution, language lost the isomorphism between sound and meaning (i.e., disappearance of onomatopoeias) to become symbolic (through the phenomenon of double articulation; Levman, 1992).

THE ADVENT OF BRAIN IMAGING

More recent years have witnessed a renewal of interest in the language–music comparison largely due to the development of cognitive science and to the advent of brain imaging methods. From the end of the nineteenth century until the early seventies, knowledge of the brain anatomo-functional organization was mainly derived from neurology and neuropsychology. In short, the dominant view was that language was located in brain regions of the left hemisphere (i.e., Broca and Wernicke areas) that were specifically devoted to language processing. However, the use of positron emission tomography (PET), functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and electroencephalography (EEG), and event-related brain potentials (ERPs) led to two major discoveries. First, it became increasingly clear that language processing is largely distributed within the left hemisphere (see Vigneau et al., 2006 for a meta-analysis of fMRI data), involving more areas than the sole Broca and Wernicke regions, and that the right hemisphere also plays an important role in language perception and comprehension (e.g., Federmeier et al., 2008 based on ERP data). Second, it was demonstrated that some brain regions that had long been considered as language specific (e.g., Broca and Wernicke’s areas) are also activated by music processing (e.g., Maess et al., 2001; Levitin and Menon, 2003; Vuust et al., 2006; Abrams et al., 2010). This may not be so surprising when considering the similarities between language and music processing.

Both language and music are complex processing systems that entertain intimate relationships with attention, memory, and motor abilities. Moreover, neither language nor music can be considered as entities; rather they comprise several levels of processing: morphology, phonology, semantics, syntax and pragmatics in language and rhythm, melody, and harmony in music. Maybe most importantly, both speech and music are auditory signals that are sequential in nature (in contrast to visual information) and that unfold in time, according to the rules of syntax and harmony. Moreover, speech and musical sounds rely on the same acoustic parameters, frequency, duration, intensity, and timber.

Results of many experiments in the neuroscience of music using both behavioral and electrophysiological methods have shown that musicians are particularly sensitive to the acoustic structure of sounds. For instance, musical expertise decreases pitch discrimination thresholds for pure and harmonic tones (e.g., Spiegel and Watson, 1984; Kishon-Rabin et al., 2001; Micheyl et al., 2006; Bidelman and Krishnan, 2010; Bidelman et al., 2010; Strait et al., 2010), and increases discrimination accuracy for frequency and duration (e.g., Koelsch et al., 1999; Micheyl et al., 2006; Tervaniemi et al., 2006). Using more musical materials, it has also been shown that musical expertise increases sensitivity to pitch changes in melodic contours (e.g., Trainor et al., 1999; Fujioka et al., 2004) and that musicians recognize familiar melodies, and detect subtle variations of pitch, rhythm, and harmony within musical phrases faster, and more accurately than non-musicians (e.g., Besson and Faïta, 1995; Koelsch et al., 2002; Bidelman et al., 2010). Moreover, seminal studies using MRI, fMRI, or MEG have demonstrated that the development of the perceptual, cognitive, and motor abilities, through years of intensive musical practice in the case of professional musicians, largely influences brain anatomy and brain function (e.g., Elbert et al., 1995; Schlaug et al., 1995a,b; Amunts et al., 1997; Pantev et al., 1998; Keenan et al., 2001; Schmithorst and Wilke, 2002; Schneider et al., 2002, 2005; Gaser and Schlaug, 2003; Hutchinson et al., 2003; Luders et al., 2004; Bengtsson et al., 2005; Bermudez et al., 2009; Imfeld et al., 2009). Finally, results of experiments using a longitudinal approach, with non-musician adults or children being trained with music, have shown that the differences between musicians and non-musicians are more likely to result from musical training rather than from genetic predispositions for music (e.g., Lahav et al., 2007; Hyde et al., 2009; Moreno et al., 2009).

TRANSFER OF TRAINING HYPOTHESIS

Based on the functional overlap of brain structures involved in language and music processing, and based on the findings of musicians’ increased sensitivity to acoustic parameters that are similar for music and speech, we developed a research program aimed at studying transfer of training effects between music and speech. The general hypothesis is that musicians should be more sensitive than non-musicians to speech sounds. To test this hypothesis we used both behavioral percentage of errors (%err) and reaction times (RTs), and electrophysiological methods (ERPs). Overall, the results that we have obtained until now are in line with this hypothesis. We briefly summarize these results below and we discuss several possible interpretations of these findings.

INFLUENCE OF MUSICAL EXPERTISE ON PITCH AND METRIC PROCESSING IN SPEECH

In the first two studies, we used natural speech and we parametrically manipulated the pitch of sentence final words (suprasegmental changes) so that pitch variations were larger (easy to detect) or subtle (difficult to detect). Sentences were spoken either in the native language of the listener (Schön et al., 2004) or in a foreign language, unknown to participants (Marques et al., 2007). In both studies, results showed that musicians outperformed non-musicians only when the pitch variation on the final word was

difficult to detect. Analysis of the ERPs revealed an increased positivity (of the P3 family) to subtle pitch variations but only in musicians.

More recently, Marie et al. (in press a) examined the influence of musical expertise on lexical pitch (tone) and on segmental variations in a language, Mandarin Chinese, unfamiliar to the French participants. They listened to two sequences of four-monosyllabic Mandarin words that were either same or different. When different, one word of the sequence varied in tone (e.g., *qíng/qíng*) or in segmental cues (consonant or vowel; e.g., *bán/zán*). In line with previous behavioral data (Gottfried et al., 2004; Delogu et al., 2006, 2010; Lee and Hung, 2008), musicians detected tone and segmental variations better than non-musicians. Analysis of the ERPs showed no influence of musical expertise on the N1 component (perceptual processing, e.g., Rugg and Coles, 1995; Eggermont and Ponton, 2002) and P3a component (automatic orienting of attention, e.g., Squires et al., 1975; Escera et al., 2000). By contrast, the latency of the N2/N3 component to tone variations (categorization process, e.g., Fujioka et al., 2006; Moreno et al., 2009) and the amplitude and latency of the P3b component (decision processes; e.g., Duncan-Johnson and Donchin, 1977; Picton, 1992) to tone and segmental variations were larger/shorter in musicians than in non-musicians (see Figure 1).

Finally, we examined the influence of musical expertise on vowel duration and metric processing in natural speech (Magne et al., 2007; Marie et al., 2011). We used a specific time-stretching algorithm (Pallone et al., 1999) to create an unexpected lengthen-

ing of the penultimate syllable that disrupted the metric structure of words without modifying timber or frequency. Sentences final words were also semantically congruous or incongruous within the context. Participants performed two tasks in separate blocks of trials. In the metric task, they focused attention on the metric structure of final words to decide whether they were correctly pronounced or not. In the semantic task, they focused attention on the semantics of the sentence to decide whether the final word was expected within the context or not. In both tasks, musicians outperformed non-musicians (as measured by the percentage of errors). However, the pattern of ERP data differed between tasks. Independently of the direction of attention, the P2 component (perceptual processing) to syllabic lengthening was larger in musicians than in non-musicians. By contrast, the N400 effect (semantic processing) was not different in both groups (Marie et al., 2011).

TRANSFER OF TRAINING AND COMMON PROCESSING?

Taken together, these different results showed that, compared to non-musicians, musicians were more sensitive to supra-segmental manipulations of pitch (i.e., intonation at the sentence level) in their own language (Schön et al., 2004) as well as in a foreign language (Marques et al., 2007), to segmental and to tone variations in a foreign language in which these variations are linguistically relevant (Marie et al., in press a) and to the metric structure of words (Marie et al., 2011). These differences were reflected in the pattern of brain waves that also differed between musicians and non-musicians. Based on these results we can argue that, through years of musical practice, musicians have developed an increased sensitivity to acoustic parameters that are important for music, such as frequency and duration. As argued above based upon the results of longitudinal studies with non-musicians, such an increased sensitivity is more likely to result from musical training than from genetic predispositions for music. Moreover, Musacchia et al. (2007, 2008) have reported positive correlations between the number of years of musical practice and the strength of subcortical pitch encoding as well as with the amplitude of ERPs cortical components. This sensitivity would extend from music to speech possibly because processing frequency and duration in music and speech draw upon the same pool of neural resources (e.g., Patel, 2003, 2008; Kraus and Chandrasekaran, 2010). In other words, common processes are involved in both cases. For instance, recent fMRI data coming from a direct comparison of temporal structure processing in music and speech suggested that similar anatomical resources are shared by the two domains (Abrams et al., in press). However, if the processes are common (i.e., domain-general) to music and speech, is it appropriate to take the results as evidence for transfer effects from music to speech?

At issue is how to reconcile an explanation in terms of common processing with the hypothesis of transfer of training effects from music to speech processing. We argue that enhanced sensitivity to acoustic features that are common to music and speech, and that imply domain-general processes, allows musicians to construct more elaborated percepts of the speech signal than non-musicians. This, in turn, facilitates stages of speech processing that are speech-specific (i.e., not common to music and speech). For example, acoustic processing of rapidly changing auditory patterns is a prerequisite for speech processing that may be sub-served by the left

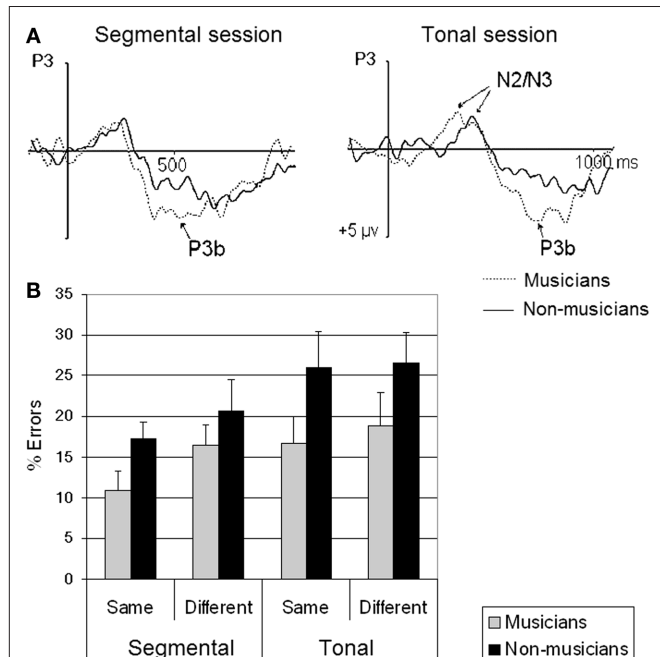


FIGURE 1 | (A) Difference-wave ERPs (different minus same final words) for the segmental (left) and tonal (right) sessions in musicians (dashed line) and non-musicians (solid line). **(B)** Percentage of errors for musicians (gray) and non-musicians (black) in the segmental and the tonal sessions and in the two experimental conditions (same and different words, with error bars). Adapted from Marie et al. (in press a).

planum temporale (Griffiths and Warren, 2002; Jancke et al., 2002; Hickok and Poeppel, 2007; Zaehle et al., 2008). This ability is necessary to hear formant transition (Bidelman and Krishnan, 2010) and to distinguish between phonemes (e.g., “b” and “p”). Correct phoneme discrimination is, in turn, a pre-requisite for correct word identification and for assessing word meaning, and children with language disorders often show temporal auditory processing difficulties (e.g., Overy, 2000; Tallal and Gaab, 2006; Gaab et al., 2007). In short, when long-term experience in one domain influences acoustic processing in the other domain, results can be interpreted as common acoustic processing. But when long-term experience in one domain influences the building-up of abstract and specific percepts in another domain, results are taken as evidence for transfer of training effects (see also Kraus and Chandrasekaran, 2010).

Similar conclusions were reached by Bidelman et al. (2009) from the results of an experiment specifically designed to directly compare the effects of linguistic and musical expertise on music and speech pitch processing. These authors presented homologs of musical intervals and of lexical tones to native Chinese, English musicians, and English non-musicians and recorded the brainstem frequency following response. Results showed that both pitch-tracking accuracy and pitch strength were higher in Chinese and English musicians as compared to English non-musicians. Thus, both linguistic and musical expertise similarly influenced the processing of pitch contour in music intervals and in Mandarin tones possibly because both draw into the same pool of neural resources. However, some interesting differences also emerged between Chinese and English musicians. While English musicians were more sensitive than Chinese to the parts of the stimuli that were similar to the notes of the musical scale, Chinese were more sensitive to rapid changes of pitch that were similar to those occurring in Mandarin Chinese. In line with the discussion above, the authors concluded that the “auditory brainstem is domain-general inasmuch as it mediates pitch encoding in both music and language” but that “pitch extraction mechanisms are not homogeneous for Chinese non-musicians and English musicians as they depend upon interactions between specific features of the input signal, their corresponding output representations and the domain of expertise of the listener” (p. 8).

Supporting an interpretation in terms of common processes in music and speech, recent results showed that non-musician native speakers of a quantity language, Finnish, in which duration is a phonemically contrastive cue, pre-attentively, and attentively processed the duration of harmonic sounds as efficiently as French musicians and better than French non-musicians (Marie et al., in press b; see Figure 2). Moreover, results of Marie et al. (in press a) also support an interpretation in terms of transfer effects by showing that musicians were more sensitive than non-musicians to segmental variations (consonant or vowel changes; e.g., *bán/zán*) that is, to abstract phonological representations derived from the processing of acoustic parameters. More generally, this interpretation is also in line with results showing a positive influence of musical skills on phonological processing (Anvari et al., 2002; Slevc and Miyake, 2006; Jones et al., 2009; Moreno et al., 2009). Along these lines, we recently found that musician children (with an average of 4 years of musical training) are more sensitive [larger mismatch negativity (MMNs), lower error rate, and shorter RTs;

see Figure 3] than non-musicians to syllabic duration (acoustic processing). Moreover, musician children were also sensitive to small differences in voice onset time (VOT; larger MMNs and shorter RTs for large than for small VOT deviants; see Figure 4). VOT is a fast temporal cue that allows to differentiate “ba” from “pa,” for instance, and that plays an important role in the development of phonological representations. By contrast, the MMNs and RTs recorded from non-musician children were not different for small and large differences in VOT which, in line with previous results by Phillips et al. (2000) with non-musician adults, was taken to indicate that non-musician children process all changes (whether large or small) as across-phonemic category changes (see Figure 4; Chobert et al., accepted). Finally, it is interesting to put these results in the perspective of previous findings by Musacchia et al. (2007) showing that Wave delta (~8 ms post-stimulus onset) of the brainstem evoked response that is related to the encoding of stimulus onset, and thereby necessary for the encoding of attack, and rhythm in both music and speech, is larger in musicians than in non-musicians. Thus, increased sensitivity at low-level of sensory processing may have strong consequences at higher level of perceptual and cognitive processing. In this respect, it becomes very interesting to simultaneously record brainstem and cortical evoked potentials, as recently done by Musacchia et al. (2008). They compared musicians and non-musicians and found strong correlations between brainstem measures (e.g., pitch encoding, F0), cortical measures (e.g., the slope of the P1–N1 components) and the level of performance on tests of tonal memory (Sheashore and MAT-3), thereby showing clearer and stronger brain–behavior relationships in musicians than in non-musicians.

In sum, the results reviewed above argue in favor of common processing of acoustic parameters such as frequency and duration in music and speech. Moreover, by showing that improved processing of these acoustic parameters has consequences at a higher level of speech processing (e.g., phonological level, lexical tone processing), they argue for positive transfer of training effects from musical expertise to speech processing. Of course, the challenge is then to try to disentangle the acoustic from the more abstract representations. This is not an easy task since these different aspects are strongly inter-mixed and interactive in speech perception and comprehension. Nevertheless, an interesting perspective for future research is to try to specify the upper limit for transfer effects, that is, whether musical expertise can influence phonological, semantic, syntactic, or pragmatic processing (e.g., Bigand et al., 2001; Patel, 2003, 2008; Koelsch et al., 2004, 2005; Poulin-Charronnat et al., 2005; Steinbeis and Koelsch, 2008; Fedorenko et al., 2009).

TRANSFER OF TRAINING AND ATTENTION

Because the results from our group reviewed above were most often obtained from designs in which participants were asked to focus attention on the sounds and because musicians often showed an overall facilitation compared to non-musicians, one could argue that these results reflect a general effect of focused attention (e.g., Fujioka et al., 2006; Moreno et al., 2009; Strait et al., 2010). For instance, in the Marie et al.’s (in press a) study musicians detected tone and segmental variations better than non-musicians in all experimental conditions. Similarly, in the Marie et al. (2011) study that manipulated the orientation of attention

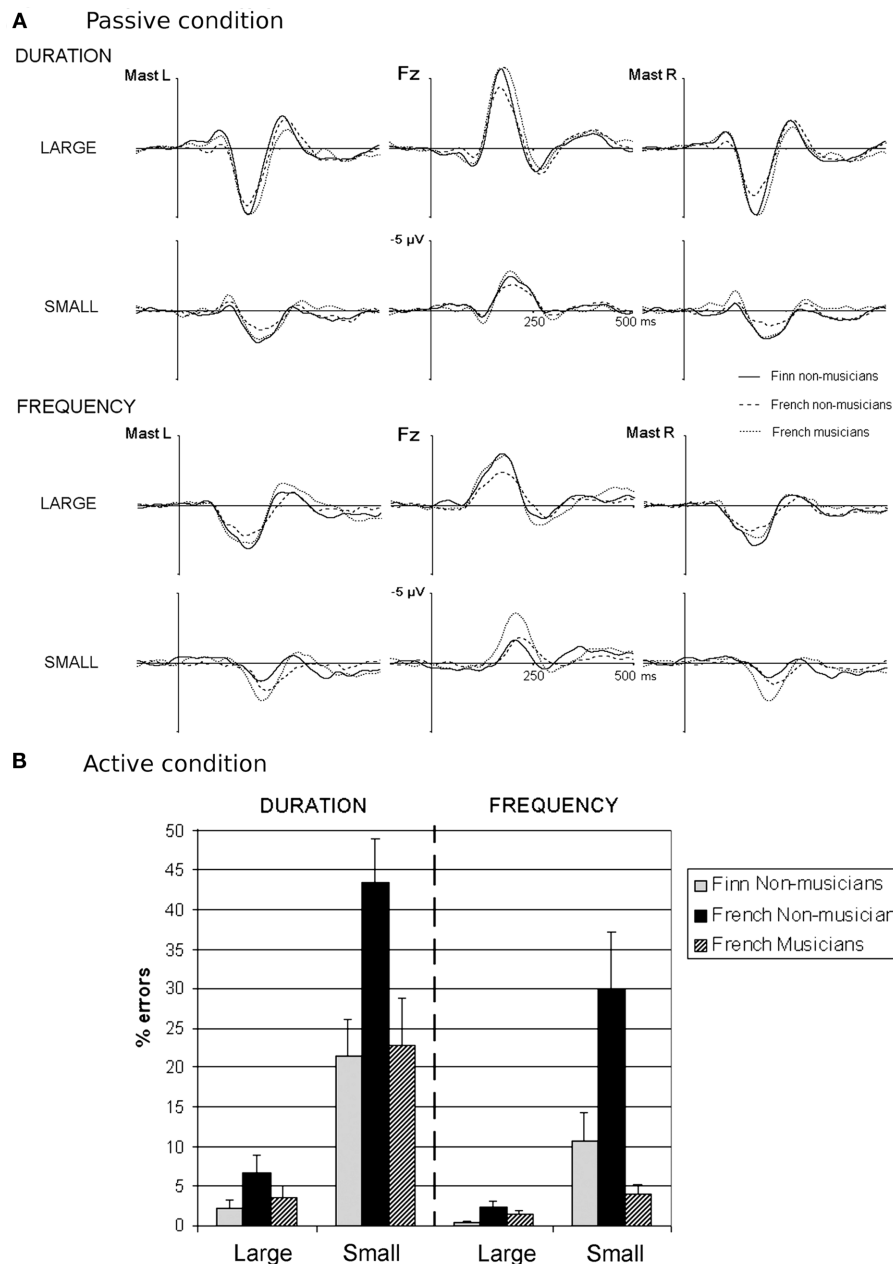
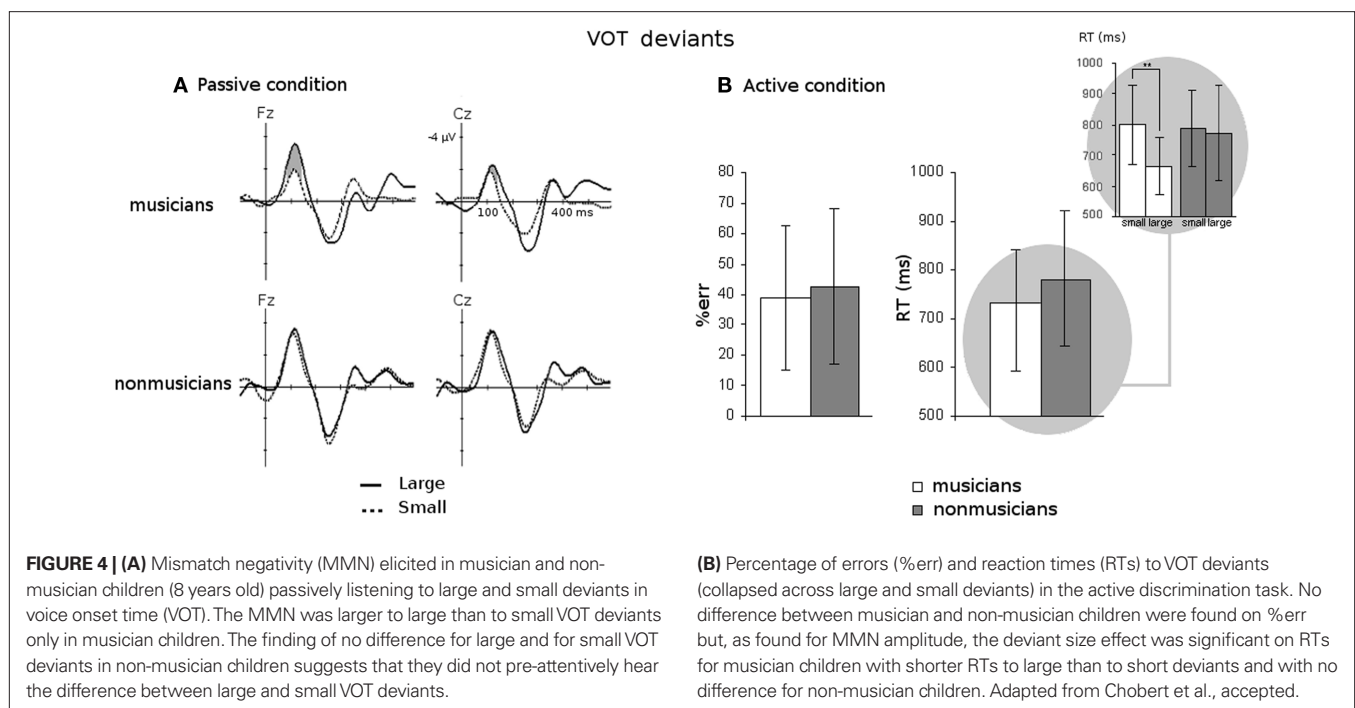
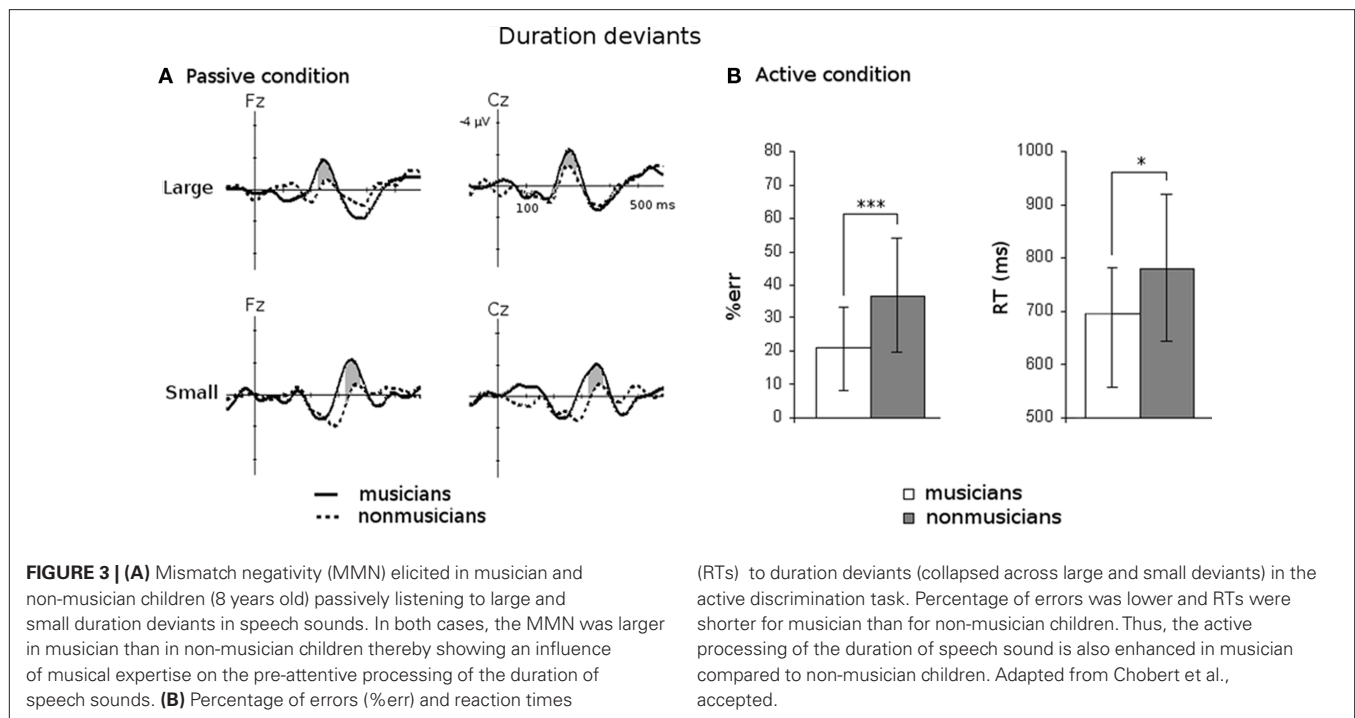


FIGURE 2 | (A) Mismatch negativity (MMN) in Finn non-musicians, French musicians, and French non-musicians passively listening to large and small duration and frequency deviants in harmonic sounds. The MMN to duration deviants was larger in Finn and in French musicians than in French non-musicians. However, only the French musicians showed an enhanced MMN to small frequency deviants. Thus, linguistic and musical expertise similarly influenced the pre-attentive processing of duration but not of frequency

deviants. **(B)** The percentage of errors to small duration and frequency deviants in the active discrimination task was lower for French musicians and Finn non-musicians than for French non-musicians. When attention is focused on the harmonic sounds, Finn non-musicians detected both Small Duration and Frequency deviants better than French non-musicians thereby showing a dissociation between the passive and active listening conditions for the frequency deviants. Adapted from Marie et al. (in press b).

toward the metric structure of words or toward the semantics of the sentence, musicians also outperformed non-musicians independently of the direction of attention. However, analysis of the ERPs is very informative relative to the influence of attention. In both studies, results showed no influence of musical expertise on the N1 component that is known to be particularly sensitive to

selective attention (e.g., Hillyard et al., 1973). Moreover, in the Marie et al.'s (2011) experiment, the P2 component elicited by unexpected syllabic lengthening was larger in musicians than in non-musicians (see also Atienza et al., 2002; Shahin et al., 2003; Bosnyak et al., 2007) and this differential effect occurred independently of the direction of attention. These results are in line



with previous findings from Baumann et al. (2008) showing that the effect of selective attention on the N1 and P2 component elicited by sine waves and harmonic sounds had a different time course and scalp distribution than the effect of musical expertise. Finally, when present (e.g., in Marie et al., in press a), the P3a component, that is taken to reflect the automatic orienting of

attention (Courchesne et al., 1975; Squires et al., 1975; Escera et al., 2000a), was of similar amplitude for both musicians and non-musicians.

Taken together, these findings indicate that the effect of musical expertise cannot be reduced to an attention effect. This is not to say, however, that attention plays no role in the results. Rather,

musicians may have developed increased abilities to focus attention on sounds and this ability may in turn help them to categorize the sounds and to make the relevant decision. Once again, analysis of the ERPs is revealing in this respect. The findings of shorter latency of the N2/N3 component (categorization process) to tone variations and of shorter latency and larger amplitude of the P3b component (decision process) to tone and segmental variations in musicians than in non-musicians in the Marie et al. (in press a) experiment were interpreted along these lines. Moreover, because metric and semantic incongruities were inter-mixed in both tasks used by Marie et al. (2011), the higher level of performance of musicians in the semantic task was interpreted as revealing less interferences (increased ability to focus attention) between the two dimensions (metric and semantic) for musicians than for non-musicians.

Finally, it is also important to note that many results in the neuroscience of music literature have revealed effects of musical expertise on measures of brain activity that are typically considered as reflecting pre-attentive processing. Fascinating results from the groups of Nina Kraus and Jack Gandour have shown an early influence of both musical and linguistic expertise on subcortical activity, as measured by the brainstem evoked responses. For instance, musicians show more robust representations of pitch contour of Mandarin tones than non-musicians, even if none of the participants spoke Mandarin (e.g., Wong et al., 2007). Conversely, native speakers of Mandarin Chinese show more accurate and more robust brainstem encoding of musical intervals than English non-musicians (Bidelman et al., 2011). Finally, the frequency following response (generated primarily in the inferior colliculus) to both speech and music stimuli is larger in musicians than in non-musicians (Musacchia et al., 2007; Bidelman et al., 2009). Thus, subcortical activity related to the encoding of pitch, whether in music or in speech, seems to be modulated by top-down influences related to musical or to linguistic expertise rather than only attention (for the effect of attention on FFR see Galbraith et al., 1998; Fritz et al., 2007).

Similarly, many results have shown that the amplitude of the MMN (Näätänen et al., 1978), typically considered as being pre-attentively generated, is larger in musicians than in non-musicians passively listening to harmonic or speech sounds (e.g., Koelsch et al., 1999; Nager et al., 2003; Tervaniemi et al., 2006). For instance, Koelsch et al. (1999) and more recently, Marie et al. (in press b) reported enhanced MMN to pitch changes in harmonic tones in musicians compared to non-musicians. There has been a long-lasting controversy about whether the MMN only reflects pre-attentive processing (e.g., Woldorff and Hillyard, 1991; Näätänen et al., 1993) and recent evidence suggests that while the MMN is pre-attentively generated, the amplitude of this component can be modulated by attention (Sussman et al., 2003; Loui et al., 2005). Moreover, dissociations between pre-attentive (as reflected by the MMN) and attentive (as reflected by ERP components such as the E(R)AN or the N2b) have been reported in several experiments (Sussman et al., 2004; Tervaniemi et al., 2009). For instance, in an experiment by Tervaniemi et al. (2009) designed to compare the processing of harmonic and speech sounds under ignore and attend conditions, the effect of musical expertise was only significant on the N2b component but not on the MMN.

TRANSFER OF TRAINING AND MEMORY

As mentioned above, language and music entertain strong relationships not only with attention but also with memory². The involvement of a working memory (WM) network in musical tasks has been reported in several brain imaging studies (e.g., Janata et al., 2002; Gaab and Schlaug, 2003; Schulze et al., 2009), with larger activation in musicians than in non-musicians. Moreover, common brain regions have been found to be activated during verbal and music short-term memory tasks (Ohnishi et al., 2001; Hickok et al., 2003; Brown et al., 2004; Brown and Martinez, 2007; Koelsch et al., 2009; Gordon et al., 2010; Schön et al., 2010). However, few studies have aimed at directly testing whether musicians show enhanced verbal memory abilities than non-musicians. Using behavioral measures, Chan et al. (1998) have shown better verbal memory in musicians than non-musicians. However, the level of education was a possible confound as it differed between the two groups. More recently, Tierney et al. (2008) reported that musicians can hold more information and/or for longer in auditory memory than non-musicians and positive correlations have been found between the amount of musical training and verbal WM (Brandler and Rammsayer, 2003 and Jakobson et al., 2003; but see Helmbold et al., 2005 for different results). Moreover, Franklin et al. (2008) also reported superior verbal WM performance (on reading span and operation span) when the criteria for selecting musicians and non-musicians were very well-controlled for. Importantly, they showed an improvement in long-term verbal memory in musicians that disappeared when the task did not allow for articulatory rehearsal. That different strategies can be used by musicians and non-musicians to perform the task at hand, was also suggested by very recent results from Williamson et al. (2010) in a study aimed at directly comparing short-term memory for verbal and musical pitch materials. These authors used immediate serial-recall tasks of four to eight letters or tones sequences and varied phonological and pitch proximity. First, and in line with previous results by Semal et al. (1996), they found that in both cases acoustic similarity was associated with decreased performance in non-musicians which was taken as evidence for “shared processing or overlap in verbal and musical short-term memory” (p. 172). Second, they found no pitch proximity effect in musicians (i.e., no decrease in recall performance for tones with similar compared to dissimilar pitches) which was taken to result from the use of multi-modal strategies (auditory, verbal, and tactile) in this group. Finally and directly related to our concerns, results of Experiment 3 showed that the phonological similarity effect (i.e., impaired performance for phonologically similar compared to dissimilar letters) was not significantly different for musicians and non-musicians thereby suggesting that the storage of verbal items in memory is not influenced by musical expertise.

By contrast, George and Coch (2011) used several subtest of the test of memory and learning (TOMAL, Reynolds and Voress, 2007) and found that musicians scored higher than

²While memory is taken here in its broad meaning including long-term, short term, working memory, in the following we only consider short-term/working memory but it is clear that neither language nor music can be perceived and/or understood without long-term memory (see Groussard et al., 2010 for music and long-term memory).

non-musicians across subtests (including digit forward and digit backward as well as letter forward and letter backward). However, the level of performance at each subtest was not detailed, and as mentioned above (and as acknowledged by the authors), the overall level of performance in the music group may be indicative of a general attention effect. Interestingly, the authors also reported that the P300 to deviant tones in an oddball paradigm was larger in amplitude and shorter in latency for musicians than non-musicians, which they interpreted as more efficient and faster updating of WM (Donchin and Coles, 1988) with increased musical expertise.

The level of performance in same-different tasks, in which participants have to judge whether two sequences of sounds are same or different, is also linked to the ability to maintain several sounds in WM. In this respect, results of Marie et al. (in press a) showing that musicians outperformed non-musicians in a same-different task on two successively presented sequences of four Mandarin monosyllabic words can also be interpreted as reflecting enhanced WM abilities in musicians compared to non-musicians. Moreover, the P300 component was also shorter and larger in musicians than in non-musicians which, as in the study of George and Coch (2011), can be taken as an index of faster and more efficient updating of WM in musicians than in non-musicians. However, even if the context updating hypothesis is a powerful and interesting interpretation of the functional interpretation of the P300 component (Donchin and Coles, 1988), other interpretations, based on the finding of strong correlations between P300 and RTs, also link P300 to decision processes (e.g., Renault et al., 1982; Picton, 1992). Thus, musicians can possibly make faster decisions and be more confident in their response than non-musicians. Similarly, when a tonal variation was included in the sequence, the latency of the N2 component that is typically related to categorization processes was shorter in musicians than in non-musicians. Thus, while the same/different task certainly recruits WM, the different aspects of the results cannot be entirely explained by enhanced WM in musicians compared to non-musicians. Finally, musicians also outperformed non-musicians in the on-line detection of unusual syllabic lengthening that did not specifically mobilize WM (Marie et al., 2011).

Taken together, these results give a somewhat mixed picture of whether musical expertise does influence verbal short-term memory. Further work is clearly needed to clarify the intricate connections between general cognitive functions such as attention and memory and the sensory, perceptive, and cognitive abilities that are shaped by musical training. In this perspective and to design well-controlled transfer experiments, it is necessary to include standardized tests of attention and memory with good reliability (e.g., Franklin et al., 2008; Strait et al., 2010) as well as perceptual tests (e.g., perceptual threshold and discrimination) and tests of general intelligence (e.g., Schellenberg and Peretz, 2008). Moreover, together with careful matching of the sample of participants, controlling for the materials (e.g., physical features and familiarity; Tervaniemi et al., 2009), the level of difficulty of the tasks at hand and the type of tasks to be used in a given experiment (e.g., Sadakata et al., 2010 for the use of identification tasks) are tricky issues that require careful consideration.

LONGITUDINAL STUDIES

In the final section of this review, we will briefly present an overview of an on-going research project with children³ in which we tried to take these remarks into account to test different facets of transfer effects from music to speech processing and to determine whether musical training can, together with speech therapy interventions, help children with dyslexia (Dys) to compensate their language deficits. As discussed above, the hope is that, by increasing the sensitivity to basic acoustic parameters such as pitch or duration, musical expertise will facilitate the building-up of higher-order phonological representations (e.g., phonemic categories) that are necessary for reading and that are known to be deficient in Dys (Swan and Goswami, 1997; Anvari et al., 2002; Foxton et al., 2003; Overy et al., 2003; Gaab et al., 2005; Tallal and Gaab, 2006; Santos et al., 2007). To this end, we used a longitudinal approach that has recently been used by several authors (e.g., Hyde et al., 2009; Moreno et al., 2009; Herdener et al., 2010) did to address the question of whether the effects of musical expertise result from intensive musical practice or from specific predispositions for music.

This longitudinal study started in September 2008 after we received all the agreements from the academy inspector, the local school authorities, the teachers, and the parents. Out of the 70 children who participated in the study, 37 were normal readers (NR) and 33 were children with Dys. All children had similar middle to low socioeconomic backgrounds (as determined from the profession of the parents) and none of the children, and none of their parents, had formal training in music or painting. All children were attending the third grade at the beginning of the experiment and the 37 children who remained at the end of the experiment (29 NR and 8 Dys) were attending the fifth grade. As can be expected from such a long-lasting experimental program (two school years from September 2008 until June 2010), the attrition rate was unfortunately very high for dyslexic children (76%; NR: 22%).

We used a Test 1 – Training – Test 2 – Training – Test 3 procedure. During Test 1 (September–October 2008) each child participated in two testing sessions, each lasting for 2 h. One included standardized neuropsychological assessments with subtests of the WISC IV (Wechsler, 2003, including Digit span (direct, reverse, total), similarities and symbols), the Raven matrices (Raven, 1962), the NEPSY (Korkman et al., 1998; including tests of visual and auditory attention, orientation, and visuo-motor abilities), the ODEDYS (Jacquier-Roux et al., 2005; with reading tests of regular and irregular words and of pseudo-words) and the Alouette test (Lefavrais, 1967) typically used to assess text reading abilities. Writing abilities were tested using a graphic tablet that allows measuring several parameters related to writing (e.g., pressure, velocity). Two perceptual tests were specifically designed to assess frequency and rhythmic thresholds using a just noticeable difference (JND; Grassi and Soranzo, 2009) procedure and one test aimed at testing the children's singing abilities. Finally, children were also tested using a simple RT task to assess speed of information processing and motor processing.

³Funded by the ANR-NEURO (#024-01) to Mireille Besson. Julie Chobert is supported by this grant and the project is part of her dissertation thesis.

The other session included four experiments specifically designed to test for the influence of musical training at different levels of speech perception and comprehension using both behavioral and/or electrophysiological data. In the first experiment, aimed at testing pre-attentive speech processing, we recorded the MMN while children were presented with the same materials (deviant syllables that differed in duration, frequency and VOT from the standard syllable) as in the experiment by Chobert et al., accepted described above. The aim was to determine whether musical training will induce a similar pattern of results (i.e., increased sensitivity to duration and VOT deviants) as found for musician children with an average of 4 years of musical training. In the second experiment, we examined the implicit learning of statistical regularities in a sung artificial language. Very recent results from our laboratory (François and Schön, in press) have shown that musical expertise in adults improves the learning of the musical and linguistic structure of a continuous artificial sung language. In children, we hypothesized that musical training should improve the segmentation, as revealed by subsequent recognition of items and this enhancement should translate into larger N400 to unfamiliar items. The third experiment aimed at testing the attentive perception of speech in noise. To this end we presented VCV patterns (ABA, APA, AVA, and ADA) from Ziegler et al. (2005) in an ABX design in silence or in noise. Based on previous results at the subcortical level (e.g., Parbery-Clark et al., 2009a,b) we expected musical training to facilitate the perception of these VCV patterns in noise. Finally, and in order to test for the upper limit of transfer effects, the last experiment⁴ aimed at determining whether musical training would influence speech comprehension. Children listened to semantically congruous and incongruous sentences presented at normal or fast (accelerated) speech rates. Again, we expected musical training to increase comprehension (as seen from the N400 effect), specifically at fast speech rate.

⁴The order in which the four experiments were presented was counter-balanced across children, except for the MMN experiment that was always presented first.

REFERENCES

- Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., and Menon, V. (2010). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cereb. Cortex*. doi: 10.1093/cercor/bhq198. [Epub ahead of print].
- Amunts, K., Schlaug, G., Jancke, L., Steinmetz, H., Schleicher, A., and Zilles, K. (1997). Hand skills covary with the size of motor cortex: a macrostructural adaptation. *Hum. Brain Mapp.* 5, 206–215.
- 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.
- Atienza, M., Cantero, J. L., and Dominguez-Marín, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learn. Mem.* 9, 138–150.
- Baumann, S., Meyer, M., and Jäncke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *J. Cogn. Neurosci.* 20, 1–12.
- Bengtsson, S. L., Nagy, Z., Skare, S., Forsman, L., Forssberg, H., and Ullén, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nat. Neurosci.* 8, 1148–1150.
- Bermudez, P., Lerch, J. P., Evans, A. C., and Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb. Cortex* 19, 1583–1596.
- Besson, M., and Faita, F. (1995). An event-related potential (ERP) study of musical expectancy: comparison of musicians with non-musicians. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 1278–1296.
- Bidelman, G. M., Gandour, J. T., and Krishnan, A. (2009). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* 29, 13165–13171.
- 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.
- Bidelman, G. M., and Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res.* 1355, 112–125.
- Bidelman, G. M., Krishnan, A., and Gandour, J. T. (2010). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *Eur. J. Neurosci.* 33, 530–538.
- Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D. A., and Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition* 81, 11–20.
- Bosnyak, D. J., Gander, P. E., and Roberts, L. E. (2007). Does auditory discrimination training modify representations in both primary and secondary auditory cortex? *Int. Congr. Ser.* 1300, 25–28.
- Boucouchie, A. (1993). *Le Langage Musical. Collections Les Chemins de la Musique*. Paris: Fayard.
- Brandner, S., and Rammsayer, T. H. (2003). Differences in mental abilities between musicians and non-musicians. *Psychol. Music* 31, 123–138.
- Brown, S. (2003). Biomusicology and three biological paradoxes about music. *Bull. Psychol. Arts* 4, 15–17.
- Brown, S., and Martinez, M. J. (2007). Activation of premotor vocal areas during musical discrimination. *Brain Cogn.* 63, 59–69.
- Brown, S., Martinez, M. J., and Parsons, L. M. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. *Neuroreport* 15, 2033–2037.

ACKNOWLEDGMENTS

The experiments reviewed in this paper were conducted at the “Institut de Neurosciences Cognitives de la Méditerranée” (INCM), CNRS, and Université de la Méditerranée, Marseille, France. This research was supported by a grant from the ANR-Neuro (#024-01) to Mireille Besson. Julie Chobert is PhD students also supported by the ANR-Neuro (#024-01).

⁵Music training was based on a combination of the Suzuki and Kodaly methods and painting training was based on the method developed by Arno Stern. Two artists involved in developing such abilities in children were hired for the duration of this project.

- Chan, A. S., Ho, Y. C., and Cheung, M. C. (1998). Music training improves verbal memory. *Nature* 396, 128.
- Courchesne, E., Hilliard, S. A., and Galambos, R. (1975) Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalogr. Clin. Neurophysiol.* 39, 131–143.
- Cross, I. (2001). Music, cognition, culture, and evolution. *Ann. N. Y. Acad. Sci.* 930, 28–42.
- Darwin, C. (1871/1981). “The descent of man and selection in relation to sex,” in *Facsimile of the First Edition*, Original Edn. (Princeton NJ: Princeton University Press). [John Murray, London].
- Delogu, F., Lampis, G., and Olivetti Belardinelli, M. (2006). Music-to-language transfer effect: may melodic ability improve learning of tonal languages by native nontonal speakers? *Cogn. Process.* 7, 203–207.
- Delogu, F., Lampis, G., and Olivetti Belardinelli, M. (2010). From melody to lexical tone: musical ability enhances specific aspects of foreign language perception. *Eur. J. Cogn. Psychol.* 22, 46–61.
- Descartes. (1618/1987). *Abrégé de Musique. Compendium musicae. Épiméthée*. Paris: Presses Universitaires de France.
- Donchin, E., and Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behav. Brain Sci.* 11, 357–374.
- Duncan-Johnson, C., and Donchin, E. (1977). On quantifying surprise, the variation of event-related potentials with subjective probability. *Psychophysiology* 14, 456–467.
- Eggermont, J., and Ponton, C. (2002). The neurophysiology of auditory perception: from single-units to evoked potentials. *Audiol. Neurotol.* 7, 71–99.
- Elbert, T., Pantev, C., Wienbruc, C., Rockstroh, B., and Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science* 270, 305–307.
- Escera, C., Alho, K., Schröger, E., and Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiol. Neurotol.* 5, 151–166.
- Federmeier, K. D., Wlotko, E. W., and Meyer, A. M. (2008). What’s “right” in language comprehension: event-related potentials reveal right hemisphere language capabilities. *Lang. Linguist. Compass* 2, 1–17.
- 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.
- Foxton, J. M., Talcott, J. B., Witton, C., Brace, H., McIntyre, F., and Griffiths, T. D. (2003). Reading skills are related to global, but not local, acoustic pattern perception. Brief communication. *Nat. Neurosci.* 6, 343–344.
- François, C., and Schön, D. (in press). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cereb. Cortex*. doi: 10.1093/cercor/bhr022 [Epub ahead of print].
- Franklin, M. S., Rattray, K., Moore, K. S., Moher, J., Yip, C., and Jonides, J. (2008). The effects of musical training on verbal memory. *Psychol. Music* 36, 353–365.
- Fritz, J. B., Elhilali, M., and Shamma, S. A. (2007). Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J. Neurophysiol.* 98, 2337–2346.
- Fujioka, T., Ross, B., Kakigi, R., Pantev, C., and Trainor, L. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain* 129, 2593–2608.
- Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., and Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *J. Cogn. Neurosci.* 16, 1010–1021.
- Gaab, N., Gabrieli, J. D. E., Deutsch, G. K., Tallal, P., and Temple, E. (2007). Neural correlates of rapid auditory processing are disrupted in children with developmental dyslexia and ameliorated with training: an fMRI study. *Restor. Neurol. Neurosci.* 25, 295–310.
- Gaab, N., and Schlaug, G. (2003). The effect of musicianship on pitch memory in performance matched groups. *Neuroimage* 14, 2291–2295.
- Gaab, N., Tallal, P., Kim, H., Lakshminarayanan, K., Archie, J. J., Glover, G. H., and Gabrieli, J. D. (2005). Neural correlates of rapid spectrotemporal processing in musicians and nonmusicians. *Ann. N. Y. Acad. Sci.* 1060, 82–88.
- Galbraith, G. C., Bhuta, S. M., Choate, A. K., Kitahara, J. M., and Mullen, T. A. Jr. (1998). Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport* 9, 1889–1893.
- Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.
- Gaser, C., and Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *J. Neurosci.* 23, 9240–9245.
- George, E. M., and Coch, D. (2011). Music training and working memory: an ERP study. *Neuropsychologia* 49, 1083–1094.
- Gordon, R., Schön, D., Magne, C., Astésano, C., and Besson, M. (2010). Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PLoS ONE* 5, 9889. doi: 10.1371/journal.pone.0009889
- Gottfried, T. L., Staby, A. M., and Ziemer, C. J. (2004). Musical experience and mandarin tone discrimination and imitation. *J. Acoust. Soc. Am.* 115, 2545.
- Grassi, M., and Soranzo, A. (2009). MLP: a MATLAB toolbox for rapid and reliable auditory threshold estimations. *Behav. Res. Methods* 41, 20–28.
- Griffiths, T. D., and Warren, J. D. (2002). The planum temporale as a computational hub. *Trends Neurosci* 25, 348–353.
- Groussard, M., La Joie, R., Rauchs, G., Landeau, B., Chételat, G., Viader, F., Desgranges, B., Eustache, F., and Platel, H. (2010). When music and long-term memory interact: effects of musical expertise on functional and structural plasticity in the hippocampus. *PLoS ONE* 5, 10: e13225. doi:10.1371/journal.pone.0013225
- Helmbold, N., Rammsayer, T., and Altenmüller, E. (2005). Differences in primary mental abilities between musicians and non-musicians. *J. Individ. Diff.* 26, 74–85.
- Herdener, M., Esposito, F., di Salle, F., Boller, C., Hilt, C. C., Habermeyer, B., Scheffler, K., Wetzel, S., Seifritz, E., and Cattapan-Ludewig, K. (2010). Musical training induces functional plasticity in human hippocampus. *J. Neurosci.* 30, 1377–1384.
- 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.
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science* 182, 177–180.
- Hutchinson, S., Lee, L. H. L., Gaab, N., and Schlaug, G. (2003). Cerebellar volume: gender and musicianship effects. *Cereb. Cortex* 13, 943–949.
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., and Schlaug, G. (2009). Musical training shapes structural brain development. *J. Neurosci.* 29, 3019–3025.
- Imfeld, A., Oechslin, M. S., Meyer, M., Loenneker, T., and Jancke, L. (2009). White matter plasticity in the corticospinal tract of musicians: a diffusion tensor imaging study. *Neuroimage* 46, 600–607.
- Jacquier-Roux, M., Valdois, S., and Zorman, M. (2005). *Odédys. Outil de de Pistage des Dyslexies*. Grenoble: Cogni-Sciences.
- Jakobson, L. S., Cuddy, L. L., and Kilgour, A. R. (2003). Time tagging: a key to musicians’ superior memory. *Music Percept.* 20, 307–313.
- Janata, P., Tillman, B., and Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cogn. Affect. Behav. Neurosci.* 2, 121–140.
- Jancke, L., Wustenberg, T., Scheich, H., and Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage* 15, 733–746.
- Jones, J. L., Lucker, J., Zalewski, C., Brewer, C., and Drayna, D. (2009). Phonological processing in adults with deficits in musical pitch recognition. *J. Commun. Disord.* 42, 226–234.
- Keenan, J. P., Halpern, A. R., Thangaraj, V., Chen, C., Edelman, R. R., and Schlaug, G. (2001). Absolute pitch and planum temporale. *Neuroimage* 14, 1402–1408.
- Kishon-Rabin, L., Amir, O., Vexler, Y., and Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *J. Basic Clin. Physiol. Pharmacol.* 12, 125–143.
- 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, 1565–1579.
- 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, 302–327.
- Koelsch, S., Schmidt, B. H., and Kansok, J. (2002). Influences of musical expertise on the ERAN: an ERP-study. *Psychophysiology* 39, 657–663.
- Koelsch, S., Schröger, E., and Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport* 10, 1309–1313.
- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Muller, K., and Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an fMRI study. *Hum. Brain Mapp.* 30, 859–873.
- Korkman, M., Kirk, U., and Kemp, S. (1998). *Manual for the NEPSY: A Developmental Neuropsychological Assessment*. San Antonio, TX: Psychological Testing Corporation.
- Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605.
- Lahav, A., Saltzman, E., and Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J. Neurosci.* 27, 308–314.
- Lee, C. H., and Hung, T. H. (2008). Identification of Mandarin tones by

- English speaking musicians and non-musicians. *J. Acoust. Soc. Am.* 124, 3235–3248.
- Lefavrais, P. (1967). *Le Test de l'Alouette*. Paris: Evangelical Christian Publishers Association.
- Levitin, D. J., and Menon, V. (2003). Musical structure is processed in “language” areas of the brain: a possible role for Brodmann area 47 in temporal coherence. *Neuroimage* 20, 2142–2152.
- Levman, B. (1992). The genesis of music and language. *Ethnomusicology* 36, 147–170.
- Loui, P., Greut-’t-Jong, T., Torpey, D., and Woldorff, M. (2005). Effects of attention on the neural processing of harmonic syntax in Western music. *Cogn. Brain Res.* 25, 678–87.
- Luders, E., Gaser, C., Jancke, L., and Schlaug, G. (2004). A voxel-based approach to gray-matter asymmetries. *Neuroimage* 22, 656–664.
- 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.
- Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., and Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken french: behavioral and electrophysiological evidence. *Cereb. Cortex* 17, 2659–2668.
- Marie, C., Delogu, F., Lampis, G., Olivetti, Belardinelli, M., and Besson, M. (in press a). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *J. Cogn. Neurosci.* doi: 10.1162/jocn.2010.21585
- Marie, C., Kujala, T., and Besson, M. (in press b). Musical and linguistic expertise influence preattentive and attentive processing of non-speech sounds. *Cortex* doi: 10.1016/j.cortex.2010.11.006
- Marie, C., Magne, C., and Besson, M. (2011). Musicians and the metric structure of words. *J. Cogn. Neurosci.* 23, 294–305.
- Marques, C., Moreno, S., Castro, S. L., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than non-musicians: behavioural and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463.
- Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (2006). Influence of musical learning and psychoacoustical training on pitch discrimination. *Hear. Res.* 219, 36–47.
- 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.
- Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci.* 104, 15894–15898.
- Musacchia, G., Strait, D., and Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and nonmusicians. *Hear. Res.* 241, 34–42.
- Näätänen, R., Gaillard, A. W. K., and Mäntysalo, S. (1978). Early selective attention effect on evoked potential reinterpreted. *Acta Psychol. (Amst.)* 42, 312–329.
- Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., and Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology* 30, 436–450.
- Nadel, S. (1930). The origin of music. *Musiq. Q.* 16, 531–546.
- Nager, W., Kohlmetz, C., Altenmüller, E., Rodriguez-Fornells, A., and Münte, T. F. (2003). The fate of sounds in conductor’s brains: an ERP study. *Cogn. Brain Res.* 17, 83–93.
- Newman, E. (1905/1969). “Herbert Spencer and the origin of music,” in *Musical Studies*, ed. Newman (New York: Haskell House Publishers), 189–216.
- Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., Katoh, A., and Imabayashi, E. (2001). Functional anatomy of musical perception in musicians. *Cereb. Cortex* 11, 754–760.
- Overly, K. (2000). Dyslexia, temporal processing and music: the potential of music as an early learning aid for dyslexic children. *Psychol. Music* 28, 218–229.
- Overly, K., Nicolson, R. I., Fawcett, A. J., and Clarke, E. F. (2003). Dyslexia and music: measuring musical timing skills. *Dyslexia* 9, 18–36.
- Pallone, G., Boussard, P., Daudet, L., Guillemain, P., and Kronland-Martinet, R. (1999). “A wavelet based method for audio-video synchronization in broadcasting applications,” in *Proceedings of the Conference on Digital Audio Effects DAFx-99 1999*, Trondheim, 59–62.
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature* 392, 811–814.
- Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009a). Musician enhancement for speech in noise. *Ear. Hear.* 30, 653–661.
- Parbery-Clark, A., Skoe, E., and Kraus, N. (2009b). Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* 25, 14100–14107.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681.
- Patel, A. D. (2008). *Music Language and the Brain*. Oxford, NY: Oxford University Press.
- Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., and Roberts, T. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *J. Cogn. Neurosci.* 12, 1038–1055.
- Picton, T. W. (1992). The P300 wave of the human event-related potential. *J. Clin. Neurophysiol.* 9, 456–479.
- Poulin-Charronnat, B., Bigand, E., Madurell, F., and Peereman, R. (2005). Musical structure modulates semantic priming in vocal music. *Cognition* 94, B67–B78.
- Raven, J. C. (1962). *Coloured Progressive Matrices, Sets A, AB, B*. London: H. K. Lewis.
- Renault, B., Ragot, R., Lesevre, N., and Remond, A. (1982). Onset and offset of brain events as indices of mental chronometry. *Science* 215, 1413–1415.
- Reynolds, C. R., and Vorell, J. (2007). *Test of Memory and Learning (TOMAL-2)*, 2nd Edn, Austin, TX: Pro-Ed.
- Rousseau, J. J. (1781/1993). *Essai Sur L’origine des Langues*. Paris: Flammarion.
- Rugg, M. D., and Coles, M. G. H. (1995). “The ERP and cognitive psychology: conceptual issues,” in *Electrophysiology of Mind: Event Related Brain Potentials and Cognition*, eds M. D. Rugg and M. G. H. Coles (Oxford: Oxford University Press), 27–39.
- Sadakata, M., van der Zanden, L., and Sekiyama, K. (2010). “Influence of musical training on perception of L2 speech,” in *11th Annual Conference of the International Speech Communication Association*, Makuhari, 118–121.
- Santos, A., Joly-Pottuz, B., Moreno, S., Habib, M., and Besson, M. (2007). Behavioural and event-related potential evidence for pitch discrimination deficit in dyslexic children: improvement after intensive phonetic intervention. *Neuropsychologia* 45, 1080–1090.
- Schellenberg, E. G. (2001). Music and nonmusical abilities. *Ann. N. Y. Acad. Sci.* 930, 355–371.
- Schellenberg, E. G., and Peretz, I. (2008). Music, language and cognition: Unresolved issues. *Trends Cogn. Sci.* 12, 45–46.
- Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F., and Steinmetz, H. (1995a). Increased corpus callosum size in musicians. *Neuropsychologia* 33, 1047–1055.
- Schlaug, G., Jäncke, L., Huang, Y., and Steinmetz, H. (1995b). In vivo evidence of structural brain asymmetry in musicians. *Science* 267, 699–671.
- Schmithorst, V. J., and Wilke, M. (2002). Differences in white matter architecture between musicians and non-musicians: a diffusion tensor imaging study. *Neurosci. Lett.* 321, 57–60.
- Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). Morphology of Heschl’s gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* 5, 688–694.
- Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., Dosch, H. G., Bleck, S., Stippich, C., Rupp, A. (2005). Structural and functional asymmetry of lateral Heschl’s gyrus reflects pitch perception preference. *Nat. Neurosci.* 8, 1241–1247.
- Schön, D., Gordon, R., Campagne, A., Magne, C., Astesano, C., Anton, J. L., and Besson, M. (2010). More evidence for similar cerebral networks in language, music and song perception. *Neuroimage* 51, 450–461.
- Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music facilitates pitch processing in language. *Psychophysiology* 41, 341–349.
- Schulze, K., Gaab, N., and Schlaug, G. (2009). Perceiving pitch absolutely: comparing absolute and relative pitch possessors in a pitch memory task. *BMC Neurosci.* 10, 106. doi: 10.1186/1471-2202-10-106
- Semal, C., Demany, L., Ueda, K., and Halle, P. A. (1996). Speech versus nonspeech in pitch memory. *J. Acoust. Soc. Am.* 100, 1132–1140.
- 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.
- Slevic, L. R., and Miyake, A. (2006). Individual differences in second-language proficiency: does musical ability matter? *Psychol. Sci.* 17, 675–681.
- Spencer, H. (1857). The origin and function of music. *Fraser’s Mag.* 56, 396–408.
- Spiegel, M. F., and Watson, C. S. (1984). Performance on frequency discrimination tasks by musicians and non-musicians. *J. Acoust. Soc. Am.* 76, 1690–1695.
- Squires, K. C., Squires, N. K., and Hillyard, S. A. (1975). Decision-related cortical potentials during an auditory signal detection task with cued observation intervals. *J. Exp. Psychol. Hum. Percept. Perform.* 1, 268–279.
- Steinbeis, N., and Koelsch, S. (2008). Comparing the processing of music and language meaning using EEG and

- fMRI provides evidence for similar and distinct neural representations. *PLoS One* 3, 2226. doi: 10.1371/journal.pone.0002226
- Strait, D. L., Kraus, N., Parbery-Clark, A., and Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear. Res.* 261, 22–29.
- Sussman, E., Kujala, T., Halmetoja, J., Lyytinen, H., Alku, P., and Näätänen, R. (2004). Automatic and controlled processing of acoustic and phonetic contrasts. *Hear. Res.* 190, 128–140.
- Sussman, E., Winkler, I., and Wang, W. (2003). MMN and attention: competition for deviance detection. *Psychophysiology* 40, 430–435.
- Swan, D., and Goswami, U. (1997). Phonological awareness deficits in developmental dyslexia and phonological representations hypothesis. *J. Exp. Child Psychol.* 66, 18–41.
- Tallal, P., and Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends Neurosci.* 29, 382–390.
- Tervaniemi, M., Castaneda, A., Knoll, M., and Uther, M. (2006). Sound processing in amateur musicians and nonmusicians: event-related potential and behavioral indices. *Neuroreport* 17, 1225–1228.
- Tervaniemi, M., Kruck, S., Baene, W. D., Schröger, E., Alter, K., and Friederici, A. D. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *Eur. J. Neurosci.* 30, 1636–1642.
- Tierney, A., Bergeson, T. R., and Pisoni, D. B. (2008). Effects of early musical experience on auditory sequence memory. *Empir. Musicol. Rev.* 3, 178–186.
- Trainor, L. J., Desjardins, R. N., and Rockel, C. (1999). A comparison of contour and interval processing in musicians and nonmusicians using event-related potentials. *Aust. J. Psychol.* 51, 147–153.
- Vigneau, M., Beaucois, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., and Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30, 1414–1432.
- Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., and Ostergaard, L. (2006). It don't mean a thing. Keeping the rhythm during polyrhythmic tension, activates language areas (BA47). *Neuroimage* 31, 832–841.
- Wallaschek, R. (1891). On the origins of music. *Mind* 16, 375–388.
- Wallin, N. L., Merker, B., and Brown, S. (2000). *The Origins of Music*. Cambridge: MIT Press.
- Wechsler, D. (2003). *WISC-IV: Technical and Interpretive Manual*. San Antonio, TX: The Psychological Corporation.
- Williamson, V. J., Baddeley, A. D., and Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: comparing phonological similarity and pitch proximity. *Mem. Cognit.* 38, 163–175.
- Woldorff, M. G., and Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalogr. Clin. Neurophysiol.* 79, 170–191.
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.
- Zaehle, T., Geiser, E., Alter, K., Jancke, L., and Meyer, M. (2008). Segmental processing in the human auditory dorsal stream. *Brain Res.* 1220, 179–190.
- Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F.-X., and Lorenzi, C. (2005). Deficits in speech perception explain language learning impairment. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14110–14115.

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: 10 March 2011; paper pending published: 05 April 2011; accepted: 29 April 2011; published online: 12 May 2011.

Citation: Besson M, Chobert J and Marie C (2011) Transfer of training between music and speech: common processing, attention, and memory. *Front. Psychology* 2:94. doi: 10.3389/fpsyg.2011.00094

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

Copyright © 2011 Besson, Chobert and Marie. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



Implicit memory in music and language

Marc Ettlinger¹, Elizabeth H. Margulis² and Patrick C. M. Wong^{1,3*}

¹ Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA

² Department of Music, University of Arkansas, Fayetteville, AR, USA

³ Department of Otolaryngology – Head and Neck Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Mathias S. Oechslin, University of Geneva, Switzerland

Stefan Elmer, University of Zurich, Switzerland

*Correspondence:

Patrick C. M. Wong, The Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Communication Neural Systems Research Group, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA.
e-mail: pwong@northwestern.edu

Research on music and language in recent decades has focused on their overlapping neurophysiological, perceptual, and cognitive underpinnings, ranging from the mechanism for encoding basic auditory cues to the mechanism for detecting violations in phrase structure. These overlaps have most often been identified in musicians with musical knowledge that was acquired explicitly, through formal training. In this paper, we review independent bodies of work in music and language that suggest an important role for implicitly acquired knowledge, implicit memory, and their associated neural structures in the acquisition of linguistic or musical grammar. These findings motivate potential new work that examines music and language comparatively in the context of the implicit memory system.

Keywords: language, music, implicit memory, artificial grammar learning

INTRODUCTION

Music has been called the universal language of mankind (Longfellow, 1835) reflecting longstanding curiosity on the relationship between music and language. Both share many traits including being perceived primarily through the auditory system, having similar acoustic attributes and reflecting analogous generative syntactic systems. This has led to decades of scientific research, exemplified by the papers included in this volume, exploring their overlapping neurophysiological, perceptual, and cognitive underpinnings. These range from the mechanism for encoding basic auditory cues (Wong et al., 2007; Kraus and Chandrasekaran, 2010), to mechanisms supporting acquisition (Slevc and Miyake, 2006; Schön et al., 2008) to the mechanism for detecting violations in predicted structure (Slevc et al., 2009).

Much of this research with respect to music has made use of trained musicians, in part to look for evidence that the cognitive and neural correlates of specialization for music are similar to the human specialization for language (e.g., Besson and Faita, 1995; Patel et al., 1998a; Maess et al., 2001; Schön et al., 2008; Kraus and Chandrasekaran, 2010). While using trained musicians has led to great strides in our understanding of how music is processed, it has obscured another important similarity between music and language: both may be acquired implicitly, without the aid of explicit instruction. In this paper, we review independent bodies of research exploring the role of implicitly acquired knowledge and associated neural structures in the acquisition of language and musical grammar.

We first consider the role of implicit memory in language by looking at both natural and artificial language learning studies. The studies discussed in the Section “*Implicit Memory and Language*” show that the implicit memory system plays an important

role in acquiring the grammar, or rules, of language at all levels of linguistics structure (Table 1). Similarly, implicit learning in music is found in the acquisition of rhythm, pitch, and melodic structures. The studies discussed in the Section “*Implicit Memory and Music*” suggest a potentially common learning mechanism shared by both music and language that allows for the acquisition of these complex systems without the need for instruction (Table 1).

The studies we discuss below help us understand this mechanism by highlighting the fact that both music and language involve expectation and the tracking of dependencies between sequential elements. Neurally, there is a significant three-way overlap of the brain structures implicated in implicit memory and those involved in learning language and learning music. This convergence encourages new work that juxtaposes music and language in the context of the implicit memory system. Given the known relationship between dopamine and the implicit memory system, we may also consider more directly the genomic and molecular bases of music and language abilities.

IMPLICIT MEMORY

Implicit memory is generally defined as acquired knowledge that is not available to conscious access (Schacter and Graf, 1986; Schacter, 1987). This contrasts with explicit memory, which is characterized by knowledge that involves conscious recollection, recall, or recognition. The majority of behavioral evidence for an implicit memory system is based on experiments wherein experience leads to altered performance on some task without participants being aware of having learned anything.

One type of implicit memory stems from perceptual learning, which involves changes to the perceptual system and to perceptual

Table 1 | Summary of representative neurological findings associating implicit memory, language, and music.

Modality	Method	Learning task	Brain region/ EEG component	References
Implicit Memory	Lesion	SRTT	BG, PFC	Vakil et al. (2000), Exner et al. (2002)
	Disorders	SRTT	BG	Reber and Squire (1994), Gabrieli et al. (1997)
	fMRI	SRTT	BG, PFC	Rauch et al. (1997), Koechlin et al. (2002)
	fMRI	Visual sequences	IFG, MFG	Doyon et al. (1997, 1998), Peigneux et al. (1999a)
	EEG	FSG	P600	Friederici et al. (2002)
Language	fMRI	WS	IFG, STG	McNealy et al. (2006)
	fMRI	WS + meaning	BG, IFG, STG	Mestres-Misse et al. (2008)
	Disorders	WS + rules	BG	De Diego-Balaguer et al. (2008)
	Disorders	FSG	BG	Reber and Squire (1994), Peigneux et al. (1999b), Reber et al. (2003)
	fMRI	FSG	BG, IFG	Lieberman et al. (2004), Petersson et al. (2004)
	fMRI	ALL	BG, IFG	Forkstam et al. (2006), Petersson et al. (in press)
	PET	ALL	BG, IFG	Moro et al. (2001)
	fMRI	Anticipation	IFG, MTG	Kiehl et al. (2002)
	EEG	Anticipation	P600	Kamide et al. (2003)
	Lesion	ID, NM	IFG	Sammiller et al. (2011)
Music	fMRI	Priming ID, M	IFG	Tillmann et al. (2003)
	MEG	ID, NM	IFG, premotor	Maess et al. (2001)
	EEG	ID, M + NM	P600	Besson and Faita (1995), Patel et al. (1998a)
	EEG	Passive ID, M + NM	P600	Koelsch and Jentschke (2008)
	EEG	ID, NM	Temporal/limbic	James et al. (2008)

Tasks: SRTT, serial reaction time test; FSG, finite state grammar; WS, word segmentation; ALL, artificial language learning; ID, incongruity detection; M, musicians, NM, non-musicians.

Regions: BG, basal ganglia (including striatum, caudate); PFC, prefrontal cortex; IFG, inferior frontal gyrus (including Broca's area); MFG, middle frontal gyrus, MTG, middle temporal gyrus.

categories (e.g., phonemes, chords) due to experience. For example, in one study (Wade and Holt, 2005), participants played a video game that involved navigating through a maze. A non-critical feature of the game was that certain non-speech auditory cues were associated with certain events. After playing the game, participants were better able to distinguish the sounds and reliably learned the sound-event patterns. Importantly, learning was qualitatively different, and in some cases better, than explicit training on these same patterns. While explicit attention has been shown to facilitate this sort of perceptual learning (e.g., Ranganath and Rainer, 2003) it is also well established that perceptual learning can be subliminal and implicit (Goldstone, 1998; Seitz and Watanabe, 2003).

Another type of implicit memory involves the implicit learning of sequences (e.g., sentences, melodies). A commonly used paradigm to test implicit memory for sequences is the serial reaction time test (SRTT; Nissen and Bullemer, 1987). In this test, participants are exposed to some stimuli (e.g., objects appearing sequentially at different points on a screen) cuing participants to respond (e.g., by indicating where the stimuli appear) as quickly as possible. While the sequences of stimuli appear to be random to the participant, embedded within the random sequences is a fixed pattern, repeatedly interspersed throughout the random sequences. Over the course of the experiment, response times and accuracy on the fixed sequences improves relative to the random sequences, presumably because the participants are learning this repeated sequence. Crucially, participants do not exhibit an improved

ability to explicitly recall this repeated sequence as compared to recalling random non-repeated sequences. The fact that participants show implicit learning without explicit knowledge suggests that these memory systems can operate independently, and that people can learn about the sequencing of some stimuli without being explicitly aware of it. While implicit memory is relevant for both sequence learning and category learning and both sequence learning and category learning are relevant to language and music, we focus primarily on implicit memory in the context of sequence learning.

A more specific kind of implicit sequence learning often discussed in the context of language and music is statistical learning (e.g., Saffran et al., 1996a). Statistical learning involves the same basic idea that participants can learn sequences without explicit awareness, but adds an additional component of tracking statistics over these sequences¹. For example, in a series of studies, Saffran et al. (1996a,b, 1999) showed that adults, children, and infants are able to track transitional probabilities between syllables and tones. Participants were exposed to seemingly random sequences of syllables obscuring consistent differences in the probability that certain syllables followed others (see below for more details). Participants were sensitive to these differences in transitional probability, and subsequent work has explored what types of statistics and what

¹“Statistical learning” is also sometimes used to refer to certain types of perceptual learning (e.g., Maye et al., 2002). Here, we use it to refer to sequence-based learning only.

types of dependencies can be implicitly tracked (Knowlton and Squire, 1996; Aslin et al., 1998; Gomez, 2002).

More recently, neurological studies have, for the most part, supported this dissociation between implicit and explicit memory systems for tasks like the SRTT (Curran, 1997) and other similar sensory-motor learning tasks. Evidence includes both lesion studies (Vakil et al., 2000; Exner et al., 2002; Peach and Wong, 2004) and functional imaging (Rauch et al., 1997; Koechlin et al., 2002) and implicate the striatum, and more specifically the caudate, in implicit learning. For example, Alzheimer's patients, characterized by degeneration of the medial temporal lobe, have little trouble with the SRTT despite exhibiting problems with declarative memory, while learning in the same task is impaired for people with diseases characterized by degeneration of the basal ganglia, including in Parkinson's patients (Reber and Squire, 1994; Jackson et al., 1995) and Huntington's patients (Gabrieli et al., 1997). More generally, the basal ganglia have been implicated in implicit learning across a number of different tasks (Squire and Knowlton, 2000; Eichenbaum and Cohen, 2001). In addition to implicit memory, the basal ganglia, and the caudate specifically, have also been implicated in motor learning (Knowlton et al., 1996), general learning plasticity (Graybiel, 2005) and learning from feedback (Packard and Knowlton, 2002). There is also some evidence that the inferior frontal gyrus, and in particular, Broca's area and its right homolog, are also involved in learning sequences (Doyon et al., 1997, 1998; Peigneux et al., 1999a). More generally, Broca's area has also been associated with a wide range of linguistic functions (Grodzinsky and Santi, 2008) including hierarchical processing (Musso et al., 2003), recursion, binding (Hagoort, 2005), and speech articulation (see Bookheimer, 2002 for a review).

Finally, by virtue of the fact that dopamine receptors are found in the basal ganglia, and in particular, the striatum (which includes the caudate, putamen, and nucleus accumbens), implicit memory has been associated with dopamine. This has been supported by studies showing that increasing dopamine levels in the brain can lead to improved implicit learning (de Vries et al., 2010b), that dopamine deficiencies, as in Parkinson patients, result in poor implicit learning though explicit learning is intact (Shohamy et al., 2009), and that dopaminergic neurons in primates show a burst of activity when learning implicitly (see Shohamy and Adcock, 2010 for a review).

IMPLICIT MEMORY AND LANGUAGE

Language learning shares a number of important similarities with the learning of sensory-motor sequences, which have been classically associated with implicit memory and which, as will be discussed below, are also implicated in acquiring a musical system. As with the tasks used in implicit learning experiments (e.g., the SRTT), people are often unaware of, or unable to articulate many of the rules of their language (Fodor, 1983). People can also learn language without any explicit instruction (Chomsky, 1957). This is particularly true before school age when children learn language with relative ease (Lenneberg, 1967), which has been argued to be, in part, due to children's good implicit memory capacity as compared to adults (DiGiulio et al., 1994; DeKeyser and Larson-Hall, 2005). Finally, certain aspects of linguistic knowledge, namely the rules of combination, may be represented probabilistically or

as information about the distributional relationships at different levels of linguistic structure (e.g., phonemes, morphemes, words, and sentences; Redington and Chater, 1997). This knowledge is generally not consciously accessible to speakers of a language and is similar in nature to the probabilistic knowledge acquired in implicit learning.

IMPLICITLY LEARNED ARTIFICIAL GRAMMARS

The use of implicitly learned distributional information for language learning has been demonstrated at many different levels of linguistic structure. For example, at the level of word segmentation, Saffran et al. (1996a) exposed 8-month-old infants to a stream of running speech consisting of four three-syllable words without any breaks or pauses indicating word-hood. Thus, the only cue to word segmentation was the transitional probability between syllables, where within-word transitional probability of syllables was 1.0 and between-word transitional probability of syllables was 0.33 (no word followed itself). Infants showed a significant ability to discriminate words from part-words (formed by combining the final syllable from one word with the first two syllables of another). Adults performed similarly (Saffran et al., 1996b) in what is argued to reflect implicit learning of word segmentation (Evans et al., 2009). Importantly, this ability is suggested to be domain general as it also applies to tones (Saffran et al., 1999 and below in the discussion on implicit memory and music) and visual stimuli (Fiser and Aslin, 2001).

Analogous behavior is also found with respect to the acquisition of phonotactics. Phonotactics are the restrictions on where phonemes can occur in a word in a language (e.g., English prohibits *ng* starting a word or *h* ending one). In one study (Onishi et al., 2002), adults briefly exposed to pseudo-words reflecting some non-English phonotactic generalization showed speeded repetition to words that adhered to the generalization as compared to words that did not.

Another study on implicit phonotactic learning (Dell et al., 2000) found that when participants are tasked with repeating sets of words reflecting some phonotactic generalization, their speech errors tend to reflect these newly learned generalizations, as is true of one's native language. The authors assessed the implicitness of learning using something they call the "ask-tell technique." This involved asking all participants whether they had noticed anything about the words they were pronouncing; the experimenters also told half the participants, explicitly, what the phonotactics would be before starting. Neither the uninformed nor informed participants were able to identify any regularities in the experimental materials. These results, in addition to the fact that the speech errors were not intentional, suggest that this learning is, in fact, implicit.

Another important component of learning the phonology of a language, acquiring phonological rules, has been shown to relate to non-linguistic implicit learning as well (Ettlinger et al., in press). In this study, participants took both an artificial grammar learning experiment and a test of implicit learning. The artificial grammar learning task involved exposure to words that reflected a set of rules for forming plural and diminutive variants (e.g., *dog*, *dogs*, *doggie*, *doggies*). The test of implicit learning was a modified version of the Tower of London task (Shallice, 1982). In this task,

participants were required to solve puzzles, increasing in difficulty, which involved virtually moving colored balls on three sticks to match a predetermined pattern. Embedded within the puzzles were repeated sequences of moves, and participants were asked to think through their moves before starting, to minimize the effects of motor coordination, unlike the SRTT. Implicit learning was measured by looking at the improvement in performance on the repeated sequences (Phillips et al., 1999). Results showed a strong correlation between learning the artificial language and performance on the Tower of London task, suggesting that implicit memory and language learning are linked.

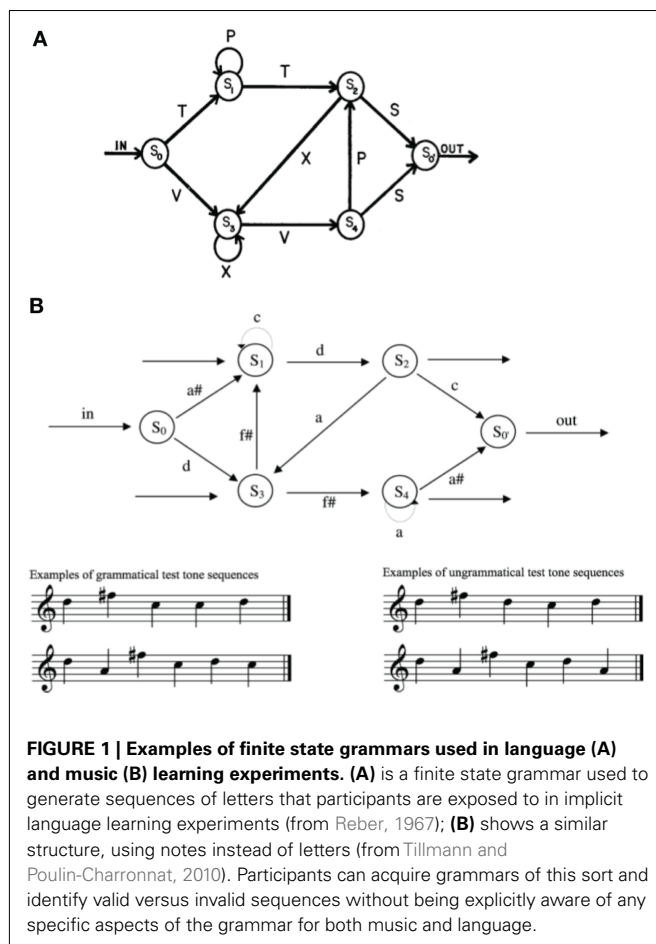
In another set of experiments exploring the possible implicit learning of syntactic structure, Reber (1967) taught participants an artificial finite state grammar for sequences of letters (**Figure 1**). After exposure to strings of letters generated by the grammar, participants were asked to judge the grammaticality of novel sequences of letters. Participants were able to successfully distinguish what constituted a valid sequence without being able to explicitly describe the rules of the grammar.

In addition to these associative studies, more concrete evidence on the role of implicit memory in language learning is provided by recent imaging studies.

McNealy et al. (2006) adapted a version of Saffran et al. (1996) word-segmentation paradigm for functional imaging by presenting participants with three speech streams: one containing

no regularities, one containing the statistical regularities as detailed above, and a third containing statistical regularities plus a standard phonetic word-segmentation cue. Greater activation was found in inferior frontal gyrus for the statistical cue and statistical cue plus phonetic cue conditions as compared to the random condition. Additional activation was found in the superior temporal gyrus, which is associated with the processing of speech (Geschwind, 1970). Another study found that when word meaning is introduced to this experimental paradigm, greater activation is also found in the basal ganglia, specifically the caudate (Mestres-Misse et al., 2008), plus the thalamus, which serves as a relay between subcortical (e.g., basal ganglia) and cortical networks and is involved in sensory perception (Steriade et al., 1997). This has led to the hypothesis that the basal ganglia, and therefore, presumably, implicit memory, is important for the integration of multiple information sources during the process of language learning (Rodriguez-Fornells et al., 2009). Similarly, patients with early stage Huntington's disease and striatal damage also do poorly on tasks of this sort (De Diego-Balaguer et al., 2008).

As with the SRTT and word segmentation, a fronto-striatal network is implicated in acquiring the finite state grammars described above. Alzheimer's and amnesic patients, with degeneration or lesions of the temporal cortices can still successfully learn artificial grammars of this sort while having trouble with more explicit language tasks (Reber and Squire, 1994; Knowlton and Squire, 1996; Reber et al., 2003). The ability of Parkinson's patients with degeneration of the basal ganglia to learn artificial grammars is less clear, however, with conflicting evidence present (Peigneux et al., 1999b; Witt et al., 2002). Similar findings are found using functional imaging, with the basal ganglia, and inferior frontal gyrus supporting the acquisition of implicit knowledge of an underlying pattern governing a sequence of letters, while the medial temporal lobe supports the recall of specific sequences (Lieberman et al., 2004; Petersson et al., 2004). Activation of the caudate is also found in a study of syntactic processing (Moro et al., 2001). In this latter study, participants were exposed to a version of Italian (the participants' native language) where all content words were replaced with pseudo-words, with function words left intact, which served to eliminate any semantic component of processing. Syntactic (word order), morphological (determiner agreement), and phonotactic violations were juxtaposed using PET. The results reflected Broca's and right IFG activation for the morphological and syntactic conditions, which has long been associated with syntactic processing (Embick et al., 2000; Grodzinsky, 2000) and may be part of a basal ganglia thalamocortical circuit (Ullman, 2006). Greater activation was also found in the left caudate, which is associated with implicit memory (see above). This result had been replicated a number of times with different types of artificial syntactic grammar, with activation consistently found in Broca's area and the caudate (Forkstam et al., 2006; Petersson et al., in press).



correlated with participants' ability to comprehend complex English sentences. Evans et al. (2009) looked at children with specific language impairment, ages five to seven, and showed that these children also performed worse on the word-segmentation task from Saffran et al. (1996) as compared to a control group with the same non-verbal IQ. They conclude that specific language impairment may not, in fact, be specific to language, but rather reflects an impairment of implicit learning, which is crucial for language learning but distinct from other measures of intelligence (see also Kaufman et al., 2010 for a similar view). Furthermore, performance on the word-segmentation task correlated with vocabulary size within each participant group, suggesting implicit learning facilitates word learning. Finally, research looking into language processing in more realistic settings has also considered language processing in noise (e.g., Wong et al., 2008, 2009a, 2010; Harris et al., 2009). In particular, Conway et al. (2010) showed a relationship between an ability to perceive speech in noise and implicit sensory-motor sequence learning. Participants who were good at an SRTT-like task were similarly good at perceiving sentences embedded in noise when the last word in the sentence had high-predictability (e.g., *Her entry should win first prize*), even when controlling for working memory and intelligence. The correlation disappeared for sentences ending in low-predictability words (e.g., *The arm is riding on the beach*).

This suggests that an important way in which implicit memory is related to language is through prediction and anticipation. A number of studies using eye-tracking (see Kamide, 2008 for a review) and event-related potentials (ERPs; see Van Berkum, 2008 for a review) have shown that people make significant use of context to facilitate processing. For example, participants look more often at a picture of beer than a doll when hearing the beginning of the sentence *the man will taste the...* (Kamide et al., 2003). A violation of an anticipated sentence completion will also yield a specific ERP response, either N400 for semantic incongruity or P600 for syntactic. The same ERP response is elicited on encountering anomalies in predicted outcome for artificial grammars similar to Reber (1967, above; Friederici et al., 2002) and music (Patel et al., 1998a).

Additional neural evidence comes from functional imaging, showing a significant overlap in the brain regions associated with implicit memory and language. As mentioned above, Broca's area has been implicated in implicit memory tasks, and Broca's area has a longstanding association with language learning and language processing (Embick et al., 2000; Grodzinsky, 2000; Sahin et al., 2009). Broca's area has also been implicated in prediction, and the expectations that yield the N400 or P600, above, result in activation of the bilateral inferior frontal gyrus (i.e., Broca's area, nearby regions and their right homologs) in addition to the middle temporal gyrus (Kiehl et al., 2002).

Thus, there is a wide range of similarities between language and implicit knowledge both in terms of their neural substrates (the fronto-striatal system) and in their cognitive underpinnings (sequential knowledge, expectation). These similarities have motivated myriad theories in linguistic processing positing that the dissociation between the words and rules of a language is homologous to the dissociation of explicit and implicit memory, respectively (Paradis, 1994, 2009; Pinker, 1999; Pinker and Ullman, 2002).

Evidence for this dissociation is discussed below, and is based on the idea that we can explicitly recall and conceptualize the words of our language, which is declarative in nature, whereas the application of the rules of language (when speaking naturally, as contrasted with attempting to adhere to a style guide, for example) is generally more difficult, if not impossible to articulate.

To conclude, there is extensive and convergent evidence for a close relationship between the cognitive and neurophysiological underpinnings of language learning and implicit memory. Language learning involves cognitive abilities that are generally learned implicitly, including tracking dependencies and developing expectations regarding adjacent linguistic structures. Language and implicit memory are also both supported by a set of neural structures including the anterior portion of the inferior frontal gyrus and the basal ganglia. As will be reviewed below, music shares many of these same associations with implicit memory and these shared associations are not restricted to musicians with formal musical training, but extend to everyday music listeners.

IMPLICIT MEMORY AND MUSIC

Although music is sometimes held to be the domain of specialists, its near-ubiquity in daily life, from mp3 players to Internet radio, cinema, and advertising, shows that affinity for music is widespread. Indeed, music has frequently been postulated by anthropologists to be a human universal, present in all known cultures (Blacking, 1973; Zatorre and Peretz, 2001). Although the ability to perform music skillfully is not evenly distributed and often relies on years of formal training, the ability to listen, process, and respond emotionally to music is shared across most of the population and seems to depend only on implicit exposure. For example, Bigand et al. (2005) showed that people with and without formal training responded largely interchangeably to non-vocal classical music. Other deep musical abilities in people without explicit training, such as the ability to perceive the relationship between a theme and its variations and to learn new compositional systems, are chronicled in Bigand and Poulin-Charronnat (2006). With little to no explicit training, how is it possible for people to develop the ability to represent and respond appropriately to the complex syntactic structures of music?

Desain and Honing (1999) demonstrate that even a seemingly simple and near-universal ability like tapping to a beat depends on complex internal representations of harmonic and syntactic musical structures. Indeed, research summarized in Krumhansl (1990) shows that implicit exposure to Western tonal music is sufficient for listeners to develop internal representations of the pitch relationships that music theorists hold to underlie tonality. Given a tonal context, such as a scale or chord progression, listeners without formal training can accurately judge how well a given continuation fits the established tonality.

One of the mechanisms by which passive exposure can ultimately yield sophisticated internal representations is statistical learning. Saffran et al. (1999) constructed long isochronous tone sequences out of 6 three-note "figures" repeated in random order, with no breaks or other indication of boundaries between the figures, and constrained so that the same figure never appeared twice in succession. When infants were exposed to this series

of tones over a 20-min period, they were able to abstract the constituent three-note figures, despite the fact that nothing but the reduced transition probabilities between them delineated the figures in the continuous stream of the musical surface. The infants, it seemed, had carefully tracked continuation probabilities in the sequence, despite the fact that their exposure to it was entirely passive. This ability to track common outcomes in musical repertoires may seem arbitrary, but in fact has been held by music theorists and psychologists since Meyer (1956) to form the basis of affective responses to music (see Huron and Margulis, 2010 for a summary). Continuations that are recognized, even implicitly, as unusual are thought to result in perceptions of special expressivity or esthetic charge. In this way, the ability to implicitly track statistics about continuations may form the fundamental scaffolding for the widespread ability to respond emotionally to music, even in the absence of formal training.

Implicit memory for music also reveals itself in various well-documented priming effects. Priming is generally defined as an implicit memory effect in which exposure to a stimulus influences responses to later stimulus without awareness of or an ability to recall the specific prime (Tulving et al., 1982). For example, Hutchins and Palmer (2008) showed that participants were more accurate in singing back the last tone of a short melody if that tone had appeared previously in the melody. Musical priming can also evidence itself in the form of faster and more accurate judgments about pitches or chords that are normative and expected given the tonal context. This kind of tonal priming has been documented in responses to melodic continuations (Margulis and Levine, 2006), and harmonic continuations (Bigand and Pineau, 1997) by listeners with no formal training. fMRI studies have implicated suppressed activity in bilateral inferior frontal regions of the brain during harmonic priming (Tillmann et al., 2000, 2003). It has even been documented in children (Schellenberg et al., 2005). Bharucha and Stoeckig (1986, 1987) provide evidence that harmonic priming is cognitive (based on the implicit abstraction of regularities in the musical environment) rather than sensory (based on psychoacoustic relationships) in nature. Tillmann et al. (2000) propose a self-organizing network model that can account for the kind of implicit learning of tonal structure revealed by priming studies. These priming effects are also observed to reflect the acquisition of musical grammars implicitly learned in the same fashion as in the implicit language learning experiments above (**Figure 1**; Tillmann and Poulin-Charronnat, 2010).

It is not only continuation statistics that listeners track implicitly. Duple and quadruple meters are more common than triple meters in Western music, and Brochard et al. (2003) confirmed that when presented with an ambiguous stimulus, listeners assume a binary division of the beat. Relatedly, the major mode is more common than the minor mode in Western music, and Huron (2006) confirmed that when presented with an ambiguous stimulus, listeners assume the major mode. And although absolute pitch perception is restricted to a tiny fraction of the population, Levitin (1994) demonstrated that ordinary listeners generally sing familiar songs within a semitone or so of their actual pitch level, suggesting that people have some implicit sense of pitch even in the absence of formal training on scales, producing notes, performing in key, or tuning an instrument. It is clear that mere exposure,

independent of formal training, or active use (such as performance or participation) is sufficient to engender highly structured and highly specific memory traces in ordinary listeners.

Implicit memory for music emerges consistently in preference effects. Halpern and O'Connor (2000) showed that although explicit recognition memory for melodies deteriorated with age, implicit memory was retained, in the form of elevated preference (the mere exposure effect first documented in Zajonc, 1968). A battery of studies over the past several decades (summarized nicely in Szpunar et al., 2004) illustrate that listeners' preference increases for music that has been encountered before. This effect is even stronger for music that is complex or ecologically valid (Bornstein, 1989). Halpern and Mullensiefen (2008) exploit this preference toward previously encountered music as a measure of implicit memory, showing that when melodies that are encountered in an exposure phase are later replayed in new timbres, participants continue to report increased liking for them, even when explicit memory of the music is obscured (i.e., the timbre change prevented them from recognizing explicitly that they had heard the excerpts before). Similarly, Peretz et al. (1998) found that explicit recognition memory was more susceptible to decay over time than implicit memory measured by elevated preference. They concluded that, in contrast with explicit memory, implicit memory as manifested in affective judgments operates obligatorily, in an automatic and unconscious fashion. Samson and Peretz (2005) further conclude, based on an analysis of patients with temporal lobe lesions on either the right or left side, that the right temporal lobe is more active in the formation of representations that underlie implicit musical memory, and the left temporal lobe is more active in processes related to explicit retrieval of musical memories.

In addition to the implicit learning of normative patterns in a particular musical style, many people are able to gain competence in more than one musical system through mere passive exposure, independent of any experience performing or producing the sound, as well as independent of any explicit instruction (formal musical training) about the style. Wong et al. (2009b) illustrate that passive exposure to the music from two cultures can result in the development of true bimusicals who approach both styles with affective and cognitive competence lacking in monomusicals of similar age and background. Wong et al. (in press) used structural equation modeling to investigate fMRI data from bimusical and monomusical listeners, finding more connectivity, and larger differentiation between the musical systems in bimusicals. These differences imply that even the implicit learning of multiple musical systems can result in fundamental changes to the way the brain approaches expressive sound.

Electrophysiological evidence also supports this conclusion. Violations of expected harmonic, melodic, and rhythmic patterns result in a late positive component (LPC) characteristic of the detection of an incongruity, even when the participants lacked formal training and were unable to explicitly identify the surprises (Besson and Faita, 1995). The elicitation of ERP components related to syntactic violations in music seem to be independent of the task relevance of unexpected chords, and provides strong evidence for important implicit components to musical ability (Koelsch et al., 2000). Patel et al. (1998b) were the first to show

that the P600 – a known marker of syntactic violations in language – extended to syntactic violations in music grammars that are abstracted implicitly by listeners. Generally, these responses have been found even when the musical exposure is entirely passive, as in Koelsch and Jentschke (2008), when participants were watching a silent movie. Koelsch (2010) emphasizes that the early right anterior negativity (ERAN) that emerges in response to syntactic violations in music depends on the long-term extraction of statistical regularities in music, not from short-term exposure to particular sequences.

Predictions based on these abstractions of musical syntax are thought to be localized in the premotor cortex and the inferior frontal gyrus (particularly Broca's area). Evidence for localization to the IFG comes from MEG (Maess et al., 2001), fMRI (Tillmann et al., 2003), and lesion studies (Sammler et al., 2011) exploring participants' responses to ungrammatical or incongruent musical stimuli (see Koelsch, 2006 for a review). There is also some evidence that the source of the ERP component responding to expectation violation may have origins in the right temporal–limbic areas, which is associated with affect and emotive processing (James et al., 2008).

The processing of syntactic violations in music has also been shown to interfere with the processing of syntactic violations in language, suggesting overlap for these two functions. When participants read garden path sentences while hearing chord progressions, they took longer to process syntactically unexpected words when they appeared at the same time as syntactically unexpected harmonies; however, no such interference occurred when the musical surprise was not syntactic in nature (e.g., when a chord sounded in a different timbre; Slevc et al., 2009). So, implicit memory seems to play an important role in syntactic processing in both language and music.

IMPLICIT MEMORY IN LANGUAGE AND MUSIC

We have reviewed above independent sets of empirical studies implicating the role of the implicit memory system in music and language, summarized in **Table 1**. In particular, we have discussed the fact that explicit training is not required for processing of language or music. It is important to note that these studies examined music or language alone. To ascertain common pathways in processing and/or representation, music and language should be examined in tandem. In terms of processing, studies could be conducted such as those performed by Patel and Slevc and colleagues (Patel et al., 1998b; Slevc et al., 2009) in which musical and linguistic stimuli were combined. However, it is preferable that everyday

music listeners should be examined to ascertain that the results are not due to formal musical training alone or trained musicians possessing a genetic difference.

Studies examining the dependence and independence of musical and linguistic functions sometimes yield conflicting results. In particular, the lesion literature favors independence while studies on neurologically normal subjects favor dependence. It is beyond the scope of this proposal to extensively discuss the nature of this debate, except to mention that a reconciliation has been proposed by imposing a distinction between representation and processing at least for syntax (Patel, 2008). In his Shared Syntactic Integration Resource Hypothesis, Patel (2003) postulates that while musical and linguistic syntactic representations are maintained separately, the processing of both musical and linguistic syntactic structures overlapped in neural resources. While the processing aspect of this hypothesis has much support (Patel et al., 1998b) and is conceivably more feasible to test, representations are difficult to examine. However, neural repetition-suppression/enhancement paradigms have been used recently to examine mental representations in humans (Grill-Spector et al., 2006) and can potentially be used to test whether musical and linguistic representations overlap in neural regions. More specifically related to the implicit memory system, we believe such experiments could be conducted with both music and language studied side-by-side.

Major divisions of the dopaminergic system contain neurons from the substantia nigra pars compacta and ventral tegmental area projecting to divisions of the striatum and prefrontal cortex, and other regions (see Seamans and Yang, 2004 for a review). As discussed above, these brain regions are also associated with the implicit memory system. Recent studies in humans, including pharmacological (de Vries et al., 2010b), molecular imaging (e.g., McNab et al., 2009), and genomic (e.g., Klein et al., 2007a,b) studies have examined the role of dopamine and related genes in a variety of implicit behaviors, such as acquiring an artificial grammar (de Vries et al., 2010a) and learning from feedback in a statistical learning paradigm (Klein et al., 2007b). Future research into the role of the implicit memory system in music and language could employ similar methods to more directly examine their potentially shared molecular neurobiological mechanisms.

ACKNOWLEDGMENTS

The authors would like to thank Lionel Newman for assistance in editing this manuscript. Support provided by grant T32 NS047987 to Marc Ettlinger and NIH grants R01DC008333, R21DC007468, and NSF BCS-1125144 to Patrick C. M. Wong.

REFERENCES

- Aslin, R. N., Saffran, J. R., and Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychol. Sci.* 9, 321–324.
- 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.
- Bharucha, J. J., and Stoeckig, K. (1986). Reaction time and musical expectancy: priming of chords. *J. Exp. Psychol. Hum. Percept. Perform.* 12, 403–410.
- Bharucha, J. J., and Stoeckig, K. (1987). Priming of chords: spreading activation or overlapping frequency spectra? *Percept. Psychophys.* 41, 519–524.
- Bigand, E., and Pineau, M. (1997). Global context effects on musical expectancy. *Percept. Psychophys.* 59, 1098–1107.
- Bigand, E., and Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition* 100, 100–130.
- Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., and Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: the effect of musical expertise and of the duration of the excerpts. *Cogn. Emot.* 19, 1113–1139.
- Blacking, J. (1973). *How Musical is Man?* Seattle: University of Washington Press.
- Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.* 25, 151–188.

- Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968–1987. *Psychol. Bull.* 106, 265–289.
- Brochard, R., Abecasis, D., Potter, D., Ragot, R., and Drake, C. (2003). The “ticktock” of our internal clock: direct brain evidence of subjective accents in isochronous sequences. *Psychol. Sci.* 14, 362–366.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Conway, C. M., Bauernschmidt, A., Huang, S. S., and Pisoni, D. B. (2010). Implicit statistical learning in language processing: word predictability is the key. *Cognition* 114, 356–371.
- Curran, T. (1997). Higher-order associative learning in amnesia: evidence from the serial reaction time task. *J. Cogn. Neurosci.* 9, 522–533.
- De Diego-Balaguer, R., Couette, M., Dolbeau, G., Durr, A., Youssov, K., and Bachoud-Levi, A. C. (2008). Striatal degeneration impairs language learning: evidence from Huntington’s disease. *Brain* 131, 2870–2881.
- de Vries, M. H., Barth, A. C. R., Maiworm, S., Knecht, S., Zwitserlood, P., and Flöel, A. (2010a). Electrical stimulation of Broca’s area enhances implicit learning of an artificial grammar. *J. Cogn. Neurosci.* 22, 2427–2436.
- de Vries, M. H., Ulte, C., Zwitserlood, P., Szymanski, B., and Knecht, S. (2010b). Increasing dopamine levels in the brain improves feedback-based procedural learning in healthy participants: an artificial-grammar-learning experiment. *Neuropsychologia* 48, 3193–3197.
- DeKeyser, R., and Larson-Hall, J. (2005). “What does the critical period really mean?” in *Handbook of Bilingualism: Psycholinguistic Approaches*, eds. J. F. Kroll and A. M. B. De Groot (New York, NY: Oxford University Press), 88–108.
- Dell, G. S., Reed, K. D., Adams, D. R., and Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 1355–1367.
- Desain, P., and Honing, H. (1999). Computational models of beat induction: the rule-based approach. *J. New Music Res.* 28, 29–42.
- Digiulio, D. V., Seidenberg, M., O’Leary, D. S., and Raz, N. (1994). Procedural and declarative memory: a developmental study. *Brain Cogn.* 25, 79–91.
- Doyon, J., Gaudreau, D., Laforce, R. Jr., Castonguay, M., Bedard, P. J., Bedard, F., and Bouchard, J. P. (1997). Role of the striatum, cerebellum, and frontal lobes in the learning of a visuomotor sequence. *Brain Cogn.* 34, 218–245.
- Doyon, J., Laforce, R. Jr., Bouchard, G., Gaudreau, D., Roy, J., Poirier, M., Bedard, P. J., Bedard, F., and Bouchard, J. P. (1998). Role of the striatum, cerebellum and frontal lobes in the automatization of a repeated visuomotor sequence of movements. *Neuropsychologia* 36, 625–641.
- Eichenbaum, H., and Cohen, N. J. (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain*. Oxford: Oxford University Press.
- Embick, D., Marantz, A., Miyashita, Y., O’Neil, W., and Sakai, K. L. (2000). A syntactic specialization for Broca’s area. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6150–6154.
- Ettlinger, M., Bradlow, A. R., and Wong, P. C. M. (in press). “The persistence and obliteration of opaque interactions,” in *Proceedings of the 45th Annual Meeting of the Chicago Linguistics Society*, ed. R. Bochnak, Chicago, IL: Chicago Linguistics Society.
- Evans, J. L., Saffran, J. R., and Roberts, K. (2009). Statistical learning in children with specific language impairment. *J. Speech Lang. Hear. Res.* 52, 321–335.
- Exner, C., Koschack, J., and Irl, E. (2002). The differential role of premotor frontal cortex and basal ganglia in motor sequence learning: evidence from focal basal ganglia lesions. *Learn. Mem.* 9, 376–386.
- Fiser, J., and Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol. Sci.* 12, 499–504.
- Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Forkstam, C., Hagoort, P., Fernandez, G., Ingvar, M., and Petersson, K. M. (2006). Neural correlates of artificial syntactic structure classification. *Neuroimage* 32, 956–967.
- Friederici, A. D., Steinhauer, K., and Pfeifer, E. (2002). Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 529–534.
- Gabrieli, J. D., Stebbins, G. T., Singh, J., Willingham, D. B., and Goetz, C. G. (1997). Intact mirror-tracing and impaired rotary-pursuit skill learning in patients with Huntington’s disease: evidence for dissociable memory systems in skill learning. *Neuropsychology* 11, 272–281.
- Geschwind, N. (1970). The organization of language and the brain. *Science* 170, 940–944.
- Goldstone, R. L. (1998). Perceptual learning. *Annu. Rev. Psychol.* 49, 585–612.
- Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychol. Sci.* 13, 431–436.
- Graybiel, A. M. (2005). The basal ganglia: learning new tricks and loving it. *Curr. Opin. Neurobiol.* 15, 638–644.
- Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci. (Regul. Ed.)* 10, 14–23.
- Grodzinsky, Y. (2000). The neurology of syntax: language use without Broca’s area. *Behav. Brain Sci.* 23, 1–21; discussion 21–71.
- Grodzinsky, Y., and Santi, A. (2008). The battle for Broca’s region. *Trends Cogn. Sci. (Regul. Ed.)* 12, 474–480.
- Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends Cogn. Sci. (Regul. Ed.)* 9, 416–423.
- Halpern, A. R., and Mullensiefen, D. (2008). Effects of timbre and tempo change on memory for music. *Q. J. Exp. Psychol. (Colchester)* 61, 1371–1384.
- Halpern, A. R., and O’Connor, M. G. (2000). Implicit memory for music in Alzheimer’s disease. *Neuropsychology* 14, 391–397.
- Harris, K. C., Dubno, J. R., and Keren, N. I. (2009). Speech recognition in younger and older adults: a dependency on low-level auditory cortex. *J. Neurosci.* 29, 6078–6087.
- Huron, D. (2006). *Sweet Anticipation: Music and The Psychology of Expectation*. Cambridge, MA: MIT Press.
- Huron, D., and Margulis, E. H. (2010). “Musical expectancy and thrills,” in *Handbook of Music and Emotion*, eds. P. N. Juslin and J. Sloboda (Oxford: Oxford University Press), 575–604.
- Hutchins, S., and Palmer, C. (2008). Repetition priming in music. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 693–707.
- Jackson, G. M., Jackson, S. R., Harrison, J., Henderson, L., and Kennard, C. (1995). Serial reaction time learning and Parkinson’s disease: evidence for a procedural learning deficit. *Neuropsychologia* 33, 577–593.
- James, C. E., Britz, J., Vuilleumier, P., Hauert, C. A., and Michel, C. M. (2008). Early neuronal responses in right limbic structures mediate harmony incongruity processing in musical experts. *Neuroimage* 42, 1597–1608.
- Kamide, Y. (2008). Anticipatory processes in sentence processing. *Lang. Linguist. Compass* 2, 647–670.
- Kamide, Y., Scheepers, C., and Altman, G. T. (2003). Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English. *J. Psycholinguist. Res.* 32, 37–55.
- Kaufman, S. B., Deyoung, C. G., Gray, J. R., Jimenez, L., Brown, J., and Mackintosh, N. (2010). Implicit learning as an ability. *Cognition* 116, 321–340.
- Kiehl, K. A., Laurens, K. R., and Liddle, P. F. (2002). Reading anomalous sentences: an event-related fMRI study of semantic processing. *Neuroimage* 17, 842–850.
- Klein, T. A., Neuman, J., Reuter, M., Hennig, J., Cramon, D., and Ullsperger, M. (2007a). Genetically determined differences in learning from errors. *Science* 318, 1642–1645.
- Klein, T. A., Neumann, J., Reuter, M., Hennig, J., Von Cramon, D. Y., and Ullsperger, M. (2007b). Genetically determined differences in learning from errors. *Science* 318, 1642–1645.
- Knowlton, B. J., Mangels, J. A., and Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science* 273, 1399–1402.
- Knowlton, B. J., and Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 169–181.
- Koechlin, E., Danek, A., Burnod, Y., and Grafman, J. (2002). Medial prefrontal and subcortical mechanisms underlying the acquisition of motor and cognitive action sequences in humans. *Neuron* 35, 371–381.
- Koelsch, S. (2006). Significance of Broca’s area and ventral premotor cortex for music-syntactic processing. *Cortex* 42, 518–520.
- Koelsch, S. (2010). “Unconscious memory representations underlying music-syntactic processing and processing of auditory oddballs,” in *Unconscious Memory Representations in Perception: Processes and Mechanisms in The Brain*, eds. I. Cziglar and I. Winkler (Herndon, VA: John Benjamins Publishing Co.), 209–244.
- Koelsch, S., Gunter, T., Friederici, A. D., and Schroger, E. (2000). Brain indices of music processing: “non-musicians” are musical. *J. Cogn. Neurosci.* 12, 520–541.
- Koelsch, S., and Jentschke, S. (2008). Short-term effects of processing musical syntax: an ERP study. *Brain Res.* 1212, 55–62.

- Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605.
- Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: John Wiley & Sons.
- Levitin, D. J. (1994). Absolute memory for musical pitch: evidence from the production of learned melodies. *Percept. Psychophys.* 56, 414–423.
- Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., and Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *J. Cogn. Neurosci.* 16, 427–438.
- Longfellow, H. W. (1835). *Outre-Mer; A Pilgrimage Beyond The Sea*. New York: Harper & Brothers.
- 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., and Levine, W. (2006). Timbre priming effects and expectation in melody. *J. New Music Res.* 35, 175–182.
- McNab, F., Varrone, A., Farde, L., Jucaite, A., Bystritsky, P., Forssberg, H., and Klingberg, T. (2009). Changes in cortical dopamine D1 receptor binding associated with cognitive training. *Science* 323, 800–802.
- McNealy, K., Mazziotta, J. C., and Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *J. Neurosci.* 26, 7629–7639.
- Mestres-Misse, A., Camara, E., Rodriguez-Fornells, A., Rotte, M., and Munte, T. F. (2008). Functional neuroanatomy of meaning acquisition from context. *J. Cogn. Neurosci.* 20, 2153–2166.
- Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.
- Misyak, J. B., Christiansen, M. H., and Tomblin, J. B. (2010). Sequential expectations: the role of prediction-based learning in language. *Top. Cogn. Sci.* 2, 138–153.
- Moro, A., Tettamanti, M., Perani, D., Donati, C., Cappa, S. F., and Fazio, F. (2001). Syntax and the brain: disentangling grammar by selective anomalies. *Neuroimage* 13, 110–118.
- Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Buchel, C., and Weiller, C. (2003). Broca's area and the language instinct. *Nat. Neurosci.* 6, 774–781.
- Nissen, M. J., and Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cogn. Psychol.* 19, 1–32.
- Onishi, K. H., Chambers, K. E., and Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition* 83, B13–B23.
- Packard, M. G., and Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.* 25, 563–593.
- Paradis, M. (1994). “Neurolinguistic aspects of implicit and explicit memory: implications for bilingualism,” in *Implicit and Explicit Learning of Second Languages*, ed. N. Ellis (London: Academic Press), 393–419.
- Paradis, M. (2009). *Neurolinguistic Aspects of Bilingualism*. Amsterdam: John Benjamins.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681.
- Patel, A. D. (2008). *Music, Language, and The Brain*. Oxford: Oxford University Press.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. J. (1998a). Processing syntactic relations in language and music: an event-related potential study. *J. Cogn. Neurosci.* 10, 717–733.
- Patel, A. D., Peretz, I., Tramo, M., and Labreque, R. (1998b). Processing prosodic and musical patterns: a neuropsychological investigation. *Brain Lang.* 61, 123–144.
- Peach, R. K., and Wong, P. C. M. (2004). Integrating the message level into treatment for agrammatism using story retelling. *Aphasiology* 18, 429–441.
- Peigneux, P., Maquet, P., Van Der Linden, M., Meulemans, T., Degueldre, C., Delfiore, G., Luxen, A., Cleeremans, A., and Franck, G. (1999a). Left inferior frontal cortex is involved in probabilistic serial reaction time learning. *Brain Cogn.* 40, 215–219.
- Peigneux, P., Meulemans, T., Van Der Linden, M., Salmon, E., and Petit, H. (1999b). Exploration of implicit artificial grammar learning in Parkinson's disease. *Acta Neurol. Belg.* 99, 107–117.
- Peretz, I., Gaudreau, D., and Bonnel, A. M. (1998). Exposure effects on music preference and recognition. *Mem. Cognit.* 26, 884–902.
- Petersson, K. M., Folia, V., and Hagoort, P. (in press). What artificial grammar learning reveals about the neurobiology of syntax. *Brain Lang.* doi: 10.1016/j.bandl.2010.08.003. [Epub ahead of print].
- Petersson, K. M., Forkstam, C., and Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cogn. Sci.* 28, 383–407.
- Phillips, L. H., Wynn, V., Gilhooly, K. J., Della Sala, S., and Logie, R. H. (1999). The role of memory in the Tower of London task. *Memory* 7, 209–231.
- Pinker, S. (1999). *Word and Rules: The Ingredients of Language*. New York: Basic Books.
- Pinker, S., and Ullman, M. T. (2002). The past and future of the past tense. *Trends Cogn. Sci. (Regul. Ed.)* 6, 456–463.
- Ranganath, C., and Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nat. Rev. Neurosci.* 4, 193–202.
- Rauch, S. L., Whalen, P. J., Savage, C. R., Curran, T., Kendrick, A., Brown, H. D., Bush, G., Breiter, H. C., and Rosen, B. R. (1997). Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. *Hum. Brain Mapp.* 5, 124–132.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *J. Mem. Lang.* 6, 855–863.
- Reber, P. J., Martinez, L. A., and Weintraub, S. (2003). Artificial grammar learning in Alzheimer's disease. *Cogn. Affect. Behav. Neurosci.* 3, 145–153.
- Reber, P. J., and Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learn. Mem.* 1, 217–229.
- Redington, M., and Chater, N. (1997). Probabilistic and distributional approaches to language acquisition. *Trends Cogn. Sci. (Regul. Ed.)* 1, 273–281.
- Rodriguez-Fornells, A., Cunillera, T., Mestres-Misse, A., and De Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3711–3735.
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.
- Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996b). Word segmentation: the role of distributional cues. *J. Mem. Lang.* 35, 606–621.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.
- Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., and Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* 326, 445–449.
- 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.
- Samson, S., and Peretz, I. (2005). Effects of prior exposure on music liking and recognition in patients with temporal lobe lesions. *Ann. N. Y. Acad. Sci.* 1060, 419–428.
- Schacter, D. L. (1987). Implicit memory: history and current status. *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 501–518.
- Schacter, D. L., and Graf, P. (1986). Preserved learning in amnesic patients: perspectives from research on direct priming. *J. Clin. Exp. Neuropsychol.* 8, 727–743.
- 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.
- Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., and Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition* 106, 975–983.
- Seamans, J., and Yang, C. R. (2004). The principal features and mechanisms of dopamine modulations in the prefrontal cortex. *Prog. Neurobiol.* 74, 1–57.
- Seitz, A. R., and Watanabe, T. (2003). Psychophysics: is subliminal learning really passive? *Nature* 422, 36.
- Shallice, T. (1982). Specific impairments of planning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 298, 199–209.
- Shohamy, D., and Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends Cogn. Sci. (Regul. Ed.)* 14.
- Shohamy, D., Myers, C. E., Hopkins, R. O., Sage, J., and Gluck, M. A. (2009). Distinct hippocampal and basal ganglia contributions to probabilistic learning and reversal. *J. Cogn. Neurosci.* 21, 1821–1833.
- Slevc, L. R., and Miyake, A. (2006). Individual differences in second-language proficiency: does musical ability matter? *Psychol. Sci.* 17, 675–681.
- 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.
- Squire, L. R., and Knowlton, B. J. (2000). “The medial temporal lobe, the hippocampus, and the memory systems of the brain,” in *The New Cognitive Neurosciences*, ed. M. S. Gazzaniga (Cambridge, MA: MIT Press), 765–780.

- Steriade, M., Jones, E. G., and McCormick, D. A. (1997). *Thalamus*. Amsterdam: Elsevier.
- Szpunar, K. K., Schellenberg, E. G., and Pliner, P. (2004). Liking and memory for musical stimuli as a function of exposure. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 370–381.
- Tillmann, B., Bharucha, J. J., and Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychol. Rev.* 107, 885–913.
- Tillmann, B., Janata, P., and Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Brain Res. Cogn. Brain Res.* 16, 145–161.
- Tillmann, B., and Poulin-Charronnat, B. (2010). Auditory expectations for newly acquired structures. *Q. J. Exp. Psychol. (Colchester)* 63, 1646–1664.
- Tulving, E., Schacter, D. L., and Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 8, 336–342.
- Ullman, M. T. (2006). Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex* 42, 480–485.
- Vakil, E., Kahan, S., Huberman, M., and Osimani, A. (2000). Motor and non-motor sequence learning in patients with basal ganglia lesions: the case of serial reaction time (SRT). *Neuropsychologia* 38, 1–10.
- Van Berkum, J. (2008). Understanding sentences in context: what brain waves can tell us. *Curr. Dir. Psychol. Sci.* 17, 376–380.
- Wade, T., and Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *J. Acoust. Soc. Am.* 118, 2618–2633.
- Witt, K., Nuhman, A., and Deuschl, G. (2002). Intact artificial grammar learning in patients with cerebellar degeneration and advanced Parkinson's disease. *Neuropsychologia* 40, 1534–1540.
- Wong, P. C. M., Chan, A. H. D., Roy, A., and Margulis, E. H. (in press). The bimusical brain is not two monomusical brains in one: evidence from musical affective processing. *J. Cogn. Neurosci.* doi: 10.1162/jocn_a_00105. [Epub ahead of print].
- Wong, P. C. M., Ettlinger, M., Shepard, J., Gunasekaran, G., and Dhar, S. (2010). Neuroanatomical characteristics and speech perception in noise in older adults. *Ear Hear.* 31, 471–479.
- Wong, P. C. M., Jin, J. X., Gunasekera, G. M., Abel, R., Lee, E. R., and Dhar, S. (2009a). Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia* 47, 693–703.
- Wong, P. C. M., Roy, A. K., and Margulis, E. H. (2009b). Bimusicalism: the implicit dual enculturation of cognitive and affective systems. *Music Percept.* 27, 81–88.
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.
- Wong, P. C. M., Uppunda, A. K., Parrish, T. B., and Dhar, S. (2008). Cortical mechanisms of speech perception in noise. *J. Speech Lang. Hear. Res.* 51, 1026–1041.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *J. Pers. Soc. Psychol.* 9, 1–27.
- Zatorre, R. J., and Peretz, I. (eds) (2001). *The Biological Foundations of Music*. New York: New York Academy of Sciences.

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: 16 February 2011; accepted: 15 August 2011; published online: 09 September 2011.

Citation: Ettlinger M, Margulis EH and Wong PCM (2011) Implicit memory in music and language. *Front. Psychology* 2:211. doi: 10.3389/fpsyg.2011.00211

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

Copyright © 2011 Ettlinger, Margulis and Wong. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



Toward a neural basis of music perception – a review and updated model

Stefan Koelsch*

Cluster of Excellence “Languages of Emotion,” Freie Universität Berlin, Berlin, Germany

Edited by:

Lutz Jäncke, University of Zurich,
Switzerland

Reviewed by:

Kristina Simonyan, Mount Sinai School
of Medicine, USA
Tom Ethofer, University Tübingen,
Germany

***Correspondence:**

Stefan Koelsch, Cluster of Excellence
“Languages of Emotion,” Freie
Universität Berlin, Habelschwerdter
Allee 45, 14195 Berlin, Germany.
e-mail: koelsch@cbs.mpg.de

Music perception involves acoustic analysis, auditory memory, auditory scene analysis, processing of interval relations, of musical syntax and semantics, and activation of (pre)motor representations of actions. Moreover, music perception potentially elicits emotions, thus giving rise to the modulation of emotional effector systems such as the subjective feeling system, the autonomic nervous system, the hormonal, and the immune system. Building on a previous article (Koelsch and Siebel, 2005), this review presents an updated model of music perception and its neural correlates. The article describes processes involved in music perception, and reports EEG and fMRI studies that inform about the time course of these processes, as well as about where in the brain these processes might be located.

Keywords: music, brain, fMRI, EEG, ERAN, semantics

1 INTRODUCTION

Music has been proven to be a valuable tool for the understanding of human cognition, human emotion, and their underlying brain mechanisms. Music is part of the human nature: It appears that throughout human history, in every human culture, people have played and enjoyed music. The oldest musical instruments discovered so far are around 30,000–40,000 years old (flutes made of vulture bones, found in the cave *Hohle Fels* in Geissenklösterle near Ulm in Southern Germany; Conard et al., 2009), but it is likely that already the first individuals belonging to the species *Homo sapiens* made music (about 100,000–200,000 years ago). Only humans learn to play musical instruments, and only humans play instruments cooperatively together in groups. It is assumed by some that human musical abilities played a key phylogenetical role in the evolution of language (e.g., Wallin et al., 2000), and that music-making behavior engaged and promoted evolutionarily important social functions (such as communication, cooperation, and social cohesion; Cross and Morley, 2008; Koelsch et al., 2010, these functions are summarized further below). Ontogenetically, newborns (who do not yet understand the syntax and semantics of words) are able to decode acoustic features of voices and prosodic features of languages (e.g., Moon et al., 1993), and it appears that infants' first steps into language are based in part on prosodic information (e.g., Jusczyk, 1999). Moreover, musical communication in early childhood (such as parental singing) plays a major role in the emotional, presumably also in the cognitive and social development of children (Trehub, 2003). Making music in a group is a tremendously demanding task for the human brain that elicits a large array of cognitive (and affective) processes, including perception, multimodal integration, learning, memory, action, social cognition, syntactic processing, and processing of meaning information. This richness makes music an ideal tool to investigate the workings of the human brain.

This review article presents an update of a previous model of music perception (Koelsch and Siebel, 2005) in which different stages of music perception are assigned to different modules (see

Figure 1; for neuroscientific investigations of music production see, e.g., Bangert and Altenmüller, 2003; Katahira et al., 2008; Herrojo-Ruiz et al., 2009, 2010; Maidhof et al., 2009, 2010). Note that these modules are thought of as entities that do *not* exclusively serve the music-perceptual processes described here; on the contrary: They also serve in part the processing of language, and – as will be illustrated in this review – the model presented here overlaps with models for language processing (for a discussion on the term modularity see also Fodor et al., 1991). The following sections will review research findings about the workings of these modules, thus synthesizing current knowledge into a framework for neuroscientific research in the field of music perception.

2 AUDITORY FEATURE EXTRACTION

Music perception begins with the decoding of acoustic information. Acoustic information is translated into neural activity in the cochlea, and progressively transformed in the auditory brainstem, as indicated by different neural response properties for the periodicity of sounds, timber (including roughness, or consonance/dissonance), sound intensity, and interaural disparities in the superior olivary complex and the inferior colliculus (Geisler, 1998; Sinex et al., 2003; Langner and Ochse, 2006; Pickles, 2008). It appears, notably, that already the dorsal cochlear nucleus projects into the reticular formation Koch et al. (1992). By virtue of these projections, loud sounds with sudden onsets lead to startle-reactions, and such projections perhaps contribute to our impetus to move to rhythmic music. Moreover, already the inferior colliculi can initiate flight and defensive behavior in response to threatening stimuli [even before the acoustic information reaches the auditory cortex (AC); Cardoso et al., 1994; Lamprea et al., 2002]. From the thalamus (particularly over the medial geniculate body) neural impulses are mainly projected into the AC (but note that the thalamus also projects auditory impulses into the amygdala and the medial orbito-frontal cortex; Kaas et al., 1999; LeDoux, 2000; Öngür and Price,

2000)¹. Importantly, The auditory pathway does not only consist of bottom-up, but also of top-down projections; nuclei such as the dorsal nucleus of the inferior colliculus presumably receive even more descending than ascending projections from diverse auditory cortical fields (Huffman and Henson, 1990).

During the last years, a number of studies investigated decoding of frequency information in the auditory brainstem using the frequency-following response (FFR; see also contribution by A. Patel, this volume). The FFR can be elicited preattentively, and is thought to originate mainly from the inferior colliculus (but note also that it is likely that the AC is at least partly involved in shaping the FFRs, e.g., by virtue of top-down projections to the inferior colliculus). Using FFRs, Wong et al. (2007) measured brainstem responses to three Mandarin tones that differed only in their (f0) pitch contours. Participants were amateur musicians and non-musicians, and results revealed that musicians had more accurate encoding of the pitch contour of the phonemes (as reflected in the FFRs) than non-musicians. This finding indicates that the auditory brainstem is involved in the encoding of pitch contours of speech information (vowels), and that the correlation between the FFRs and the properties of the acoustic information is modulated by musical training. Similar training effects on FFRs elicited by syllables with a dipping pitch contour have also been observed in native English speakers (non-musicians) after a training period of 14 days (with eight 30-min-sessions; Song et al., 2008). The latter results show the contribution of the brainstem in language learning, and its neural plasticity in adulthood².

A study by Strait et al. (2009) also reported musical training effects on the decoding of the acoustic features of an affective vocalization (an infant's unhappy cry), as reflected in auditory brainstem potentials. This suggests (a) that the auditory brainstem is involved in the auditory processing of communicated states of emotion (which substantially contributes to the decoding and understanding of *affective prosody*), and (b) that musical training can lead to a finer tuning of such (subcortical) processing.

2.1 ACOUSTICAL EQUIVALENCY OF "TIMBER" AND "PHONEME"

With regard to a comparison between music and speech, it is worth mentioning that, in terms of acoustics, there is *no* difference between a phoneme and the timber of a musical sound (and it is only a matter of convention that some phoneticians rather use terms such as "vowel quality" or "vowel color," instead of "timber")³.

¹The lateral nucleus of the amygdala receives impulses from the medial division of the medial geniculate body of the thalamus (as well as from associated regions of the posterior thalamus, and from auditory association cortex), the central nucleus of the amygdala is reciprocally connected with the brainstem. For a review see LeDoux (2000).

²For a study using FFRs to investigate effects of musical training on audiovisual integration of music as well as speech see Musacchia et al. (2007); for a study on the development of auditory brainstem responses to speech in 3- to 12-year-old children see Johnson et al. (2008).

³When two sounds are heard that match for pitch, loudness, duration, and location, and a difference can still be heard between the two sounds, that difference is called timber (e.g., Moore, 2008). For example: a clarinet, a saxophone, and a piano all play a c' at the same location, with the same loudness and the same duration. Each of these instruments has a unique sound quality. This difference is called timber, tone color, or simply sound quality. There are also many examples of timber differences in speech. For example, two vowels (such as /ā/ and /i/) spoken with the same loudness and same pitch differ from one another in timber.

Both are characterized by the two physical correlates of timber: spectrum envelope (i.e., differences in the relative amplitudes of the individual harmonics) and amplitude envelope (also sometimes called the amplitude contour or energy contour of the sound wave, i.e., the way that the loudness of a sound changes, particularly with regard to the on- and off-set of a sound)⁴. Aperiodic sounds can also differ in spectrum envelope (see, e.g., the difference between /j/ and /s/), and timber differences related to amplitude envelope play a role in speech, e.g., in the shape of the attack for /b/ vs. /w/ and /j/ vs. /tj/.

2.2 AUDITORY FEATURE EXTRACTION IN THE AUDITORY CORTEX

As mentioned above, auditory information is projected mainly via the subdivisions of the medial geniculate body into the primary auditory cortex [PAC, corresponding to Brodmann's area (BAs) 41] and adjacent secondary auditory fields (corresponding to BAs 42 and 52)⁵. These auditory areas perform that a more fine-grained, and more specific, analysis of acoustic features compared to the auditory brainstem. For example, Tramo et al. (2002) reported that a patient with bilateral lesion of the PAC (a) had normal detection thresholds for sounds (i.e., the patient could say whether there was a tone or not), but (b) had elevated thresholds for determining whether two tones had the same pitch or not (i.e., the patient had difficulties to detect fine-grained frequency differences between two subsequent tones), and (c) had markedly increased thresholds for determining the pitch direction (i.e., the patient had great difficulties in saying whether the second tone was higher or lower in pitch than the first tone, *even though* he could tell that both tones differed⁶). Note that the AC is also involved in a number of other functions, such as auditory sensory memory, extraction of inter-sound relationships, discrimination, and organization of sounds as well as sound patterns, stream segregation, automatic change detection, and multisensory integration (for reviews see Hackett and Kaas, 2004; Winkler, 2007; some of these functions are also mentioned further below).

Moreover, the (primary) AC is involved in the transformation of acoustic features (such as frequency information) into percepts (such as pitch height and pitch chroma)⁷: Lesions of the (right) PAC result in a loss of the ability to perceive residue pitch in both animals (Whitfield, 1980) and humans (Zatorre, 1988), and neurons in the anterolateral region of the PAC show responses to a missing fundamental frequency (Bendor and

⁴For example, sudden or slow attack or decay, such as in the sounds of plucked vs. bowed stringed instruments.

⁵For a detailed description of primary auditory "core," and secondary auditory "belt" fields, as well as their connectivity, see Kaas and Hackett (2000). With regard to the functional properties of primary and secondary auditory fields, a study by Petkov et al. (2006) showed that, in the macaque monkey, all of the PAC core areas, and most of the surrounding belt areas, show a tonotopic structure, with the tonotopic structure being clearest in the field A1; the tonotopic organization in the (rostral) field R seems weaker than in A1, but stronger than in RT (which is a third field of the PAC, located rostrally of the field R). The majority of belt areas appears to show a tonotopic structure comparable to that of R and RT, and only few belt areas seem to show only weak, or no, tonotopic structure (Petkov et al., 2006).

⁶For similar results obtained from patients with (right) PAC lesions see Johnsrude et al. (2000) and Zatorre (2001).

⁷For example, a sound with the frequencies 200, 300, and 400 Hz is transformed into the pitch percept of 100 Hz.

Wang, 2005). Moreover, magnetoencephalographic data indicate that response properties in the PAC depend on whether or not a missing fundamental of a complex tone is perceived (Patel and Balaban, 2001; data were obtained from humans). Note, however, that combination tones emerge already in the cochlea, and that the periodicity of complex tones is coded in the spike pattern of auditory brainstem neurons; therefore, different mechanisms contribute to the perception of residue pitch on at least three different levels (basilar membrane, brainstem, and AC)⁸. However, the studies by Zatorre (1988) and Whitfield (1980) suggest that, compared to the brainstem or the basilar membrane, the AC plays the a more prominent role for the transformation of acoustic features into auditory percepts (such as the transformation of information about the frequencies of a complex sound, as well as about the periodicity of a sound, into a pitch percept).

Warren et al. (2003) report that changes in pitch chroma involve auditory regions anterior of the PAC (covering parts of the planum polare) more strongly than changes in pitch height. Conversely, changes in pitch height appear to involve auditory regions posterior of the PAC (covering parts of the planum temporale) more strongly than changes in pitch chroma (Warren et al., 2003). Moreover, with regard to functional differences between the left and the right PAC, as well as neighboring auditory association cortex, several studies suggest that the left AC has a higher resolution of temporal

information than the right AC, and that the right AC has a higher spectral resolution than the left AC (Zatorre et al., 2002; Hyde et al., 2008; Perani et al., 2010).

Finally, the AC also prepares acoustic information for further conceptual and conscious processing. For example, with regard to the meaning of sounds, just a short single tone can sound, for example, “bright,” “rough,” or “dull.” That is, single tones are already capable of conveying meaning information (this is indicated by the line connecting the module “Feature Extraction II” and “Meaning” in **Figure 1**; processing of musical meaning will be dealt with further below).

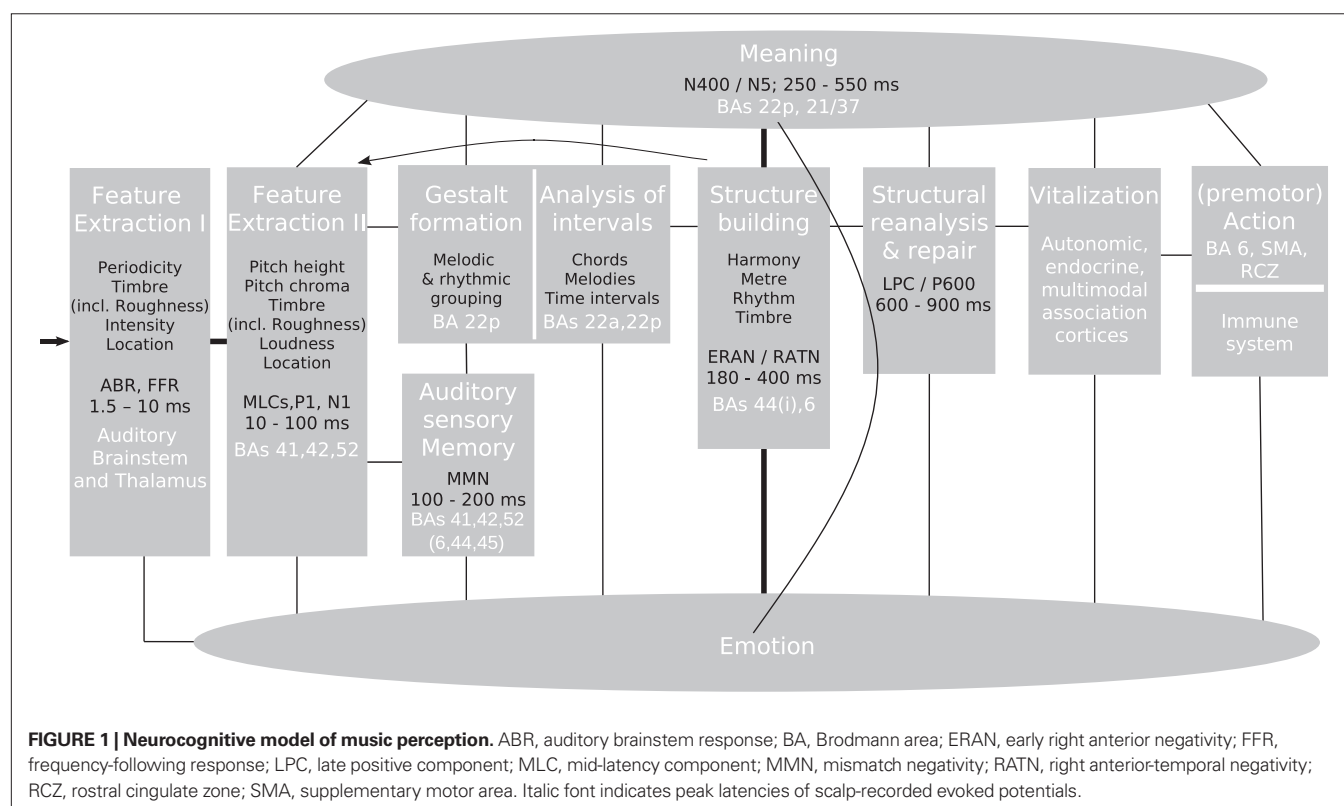
Operations within the (primary and adjacent) AC related to auditory feature analysis are reflected in electrophysiological recordings in ERP components that have latencies of about 10–100 ms, particularly middle-latency responses, including the P1, and the later “exogenous” N1 component (for effects of musical training on feature extraction as reflected in the N1 see, e.g., Pantev et al., 2001).

3 ECHOIC MEMORY AND GESTALT FORMATION

While auditory features are extracted, the acoustic information enters the auditory sensory memory (or “echoic memory”), and representations of auditory Gestalten (Griffiths and Warren, 2004; or “auditory objects”) are formed (**Figure 1**). Operations of the auditory sensory memory are at least partly reflected electrically in the mismatch negativity (MMN, e.g., Näätänen et al., 2001). The MMN has a peak latency of about 100–200 ms⁹, and most presumably

⁸Note that responses in the PAC related to the perception of missing fundamental frequencies in the studies by Bendor and Wang (2005) and Patel and Balaban (2001) are presumably in part due to the periodicity information about the missing fundamental frequency coded in the spike pattern of collicular neurons.

⁹For a study differentiating peak latencies of the sensorial and the cognitive part of the MMN see Maess et al. (2007).



receives its main contributions from neural sources located in the PAC and adjacent auditory (belt) fields, with additional (but smaller) contributions from frontal cortical areas (Giard et al., 1990; Alho et al., 1996; Alain et al., 1998; Opitz et al., 2002; Liebenthal et al., 2003; Molholm et al., 2005; Rinne et al., 2005; Maess et al., 2007; Schonwiesner et al., 2007; for a review see Deouell, 2007). These frontal areas appear to include (ventral) premotor cortex (BA 6), dorsolateral prefrontal cortex near and within the inferior frontal sulcus, and the posterior part of the inferior frontal gyrus (BAs 45 and 44). The frontal areas are possibly involved due to their role in attentional processes, sequencing, and working memory (WM) processes (see also Schonwiesner et al., 2007)^{10,11}.

Auditory sensory memory operations are indispensable for music perception; therefore, practically all MMN studies are inherently related to, and relevant for, the understanding of the neural correlates of music processing. As will be outlined below, numerous MMN studies have contributed to this issue (a) by investigating different response properties of the auditory sensory memory to musical and speech stimuli, (b) by using melodic and rhythmic patterns to investigate auditory Gestalt formation, and/or (c) by studying effects of long- and short-term musical training on processes underlying auditory sensory memory operations. Especially the latter studies have contributed substantially to our understanding of neuroplasticity¹², and thus to our understanding of the neural basis of learning. A detailed review of these studies goes beyond the scope of this article (for reviews see Tervaniemi and Huotilainen, 2003; Tervaniemi, 2009). Here, suffice it to say that MMN studies showed effects of long-term musical training on the pitch discrimination of chords (Koelsch et al., 1999)¹³, on temporal acuity (Rammsayer and Altenmüller, 2006), on the temporal window of integration (Rüsseler et al., 2001), on sound localization changes (Tervaniemi et al., 2006a), and on the detection of spatially peripheral sounds (Rüsseler et al., 2001). Moreover, using MEG, an MMN study from Menning et al. (2000) showed effects of 3-weeks auditory musical training on the pitch discrimination of tones.

Auditory oddball paradigms were also used to investigate processes of melodic and rhythmic grouping of tones occurring in tone patterns (such grouping is essential for auditory Gestalt formation, see also Sussman, 2007), as well as effects of musical long-term training on these processes. These studies showed effects of musical training (a) on the processing of melodic patterns (Tervaniemi et al., 1997, 2001; Fujioka et al., 2004; Zuijen et al., 2004, in these studies, patterns consisted of four or five tones), (b) on the encoding of the number of elements in a tone pattern (Zuijen et al., 2005), and (c) on the processing of patterns consisting of two voices (Fujioka et al., 2005).

Finally, several MMN studies investigated differences between the processing of musical and speech information. These studies report larger right-hemispheric responses to chord deviants than

to phoneme deviants (Tervaniemi et al., 1999, 2000), and different neural generators of the MMN elicited by chords compared to the MMN generators of phonemes (Tervaniemi et al. 1999); similar results have also been shown for the processing of complex tones compared to phonemes (Tervaniemi et al., 2006b, 2009, the latter study also reported effects of musical expertise on chord and phoneme processing). These lateralization effects are presumably due to the different requirements of spectral and temporal processing posed by the chords and the phonemes used as stimuli in that study.

The formation of auditory Gestalten entails processes of perceptual separation, as well as processes of melodic, rhythmic, timbral, and spatial grouping. Such processes have been summarized under the concepts of auditory scene analysis and auditory stream segregation (Bregman, 1994). Grouping of acoustic events follows Gestalt principles such as similarity, proximity, and continuity (for acoustic cues used for perceptual separation and auditory grouping see Darwin, 1997, 2008). In everyday life, such operations are not only important for music processing, but also, for instance, for separating a speaker's voice during a conversation from other sound sources in the environment. That is, these operations are important because their function is to recognize and to follow acoustic objects, and to establish a cognitive representation of the acoustic environment. Knowledge about neural mechanisms of auditory stream segregation, auditory scene analysis, and auditory grouping is still relatively sparse (for reviews see Griffiths and Warren, 2002, 2004; Carlyon, 2004; Nelken, 2004; Scott, 2005; Shinn-Cunningham, 2008; Winkler et al., 2009). However, it appears that the planum temporale (which is part of the auditory association cortex) is a crucial structure for auditory scene analysis and stream segregation (Griffiths and Warren, 2002), particularly due to its role for the processing of pitch intervals and sound sequences (Zatorre et al., 1994; Patterson et al., 2002).

4 ANALYSIS OF INTERVALS

Presumably closely linked to the stage of auditory Gestalt formation is a stage of a more fine-grained analysis of intervals, which includes (i) a more detailed processing of the pitch relations between the tones of a chord (required to determine whether a chord with a specific root is a major or minor chord, played in root position, inversion, etc., see also below), or between the tones of a melody; and possibly (ii) a more detailed processing of temporal intervals. With regard to chords, such an analysis is required to specify the "chord form," that is, whether a chord is a major or a minor chord, and whether a chord is played in root position, or in an inversion; note that chords with the same root can occur in several versions (e.g., all of the following chords have the root c: c–e–g, c–e–flat–g, e–g–c, g–c–e, c–e–g–b–flat, e–g–b–flat–c, etc.)¹⁴. The neural correlates of such processes are not known, but it is likely that both temporal and (inferior) prefrontal regions contribute to such processing; with regard to the processing of melodies, lesion data suggest that the analysis of the contour of a melody (which is part of the auditory Gestalt formation) particularly relies on the posterior part of the right superior temporal gyrus (STG), whereas

¹⁰Moreover, activation of the posterior part of the inferior frontal gyurs (including BA 44) in auditory oddball paradigms (as, e.g., reported in studies by Opitz et al., 2002; Doeller et al., 2003) is perhaps elicited when an auditory stimulus also triggers hierarchical processing of pitch relations.

¹¹For a study localizing the generators of the abstract-feature MMN see Korzyukov et al. (2003).

¹²That is, to changes in neuronal structure and function due to experience.

¹³But note that superior attentive pitch discrimination accuracy is not always reflected in the MMN (Tervaniemi et al., 2005).

¹⁴Listening to a Bach chorale, for example, and specifying on-line for each chord whether it is a major or a minor chord, and specifying the inversion in which it is played, illustrates that this is not a trivial perceptual process.

the processing of more detailed interval information appears to involve both posterior and anterior areas of the supratemporal cortex bilaterally (Liegeois-Chauvel et al., 1998; Patterson et al., 2002; Peretz and Zatorre, 2005)¹⁵.

Melodic and temporal intervals appear to be processed independently, as suggested by the observation that brain damage can interfere with the discrimination of pitch relations but spare the accurate interpretation of time relations, and vice versa (Di Pietro et al., 2004; Peretz and Zatorre, 2005). However, these different types of interval processing have so far not been dissociated in functional neuroimaging studies.

5 WHEN INTELLIGENCE COMES INTO PLAY: PROCESSING MUSICAL SYNTAX

The following section deals with the processing of major–minor tonal syntax, particularly with regard to chord functions (i.e., with regard to harmony; for explanation of chord functions see **Figure 2A**). Other aspects of tonal syntax are rhythm, meter, melody (including voice leading), and possibly timber (these aspects will not be dealt with here due to lack of neurophysiological data).

¹⁵Patients with frontal lesions were not tested in that study. For the processing of chords, more prominent frontal involvement might come into play in the region of (superior) BA 44 (perhaps area 44d according to Amunts et al., 2010) and BA 45, analogous to the activation of these areas during the processing of word form (Longoni et al., 2005).

One possible theoretical description of the syntax of major–minor tonal music is the classical theory of harmony as formulated, e.g., by Rameau (1722), Piston (1948/1987), Schönberg (1969), and Riemann (1971). These descriptions primarily deal with the derivation of chord functions, less so with how chord functions are chained into longer sequences.

5.1 FORMAL DESCRIPTIONS OF MUSICAL SYNTAX

Before describing neural correlates of music-syntactic processing, a theoretical basis of musical syntax will be outlined briefly. Heinrich Schenker was the first theorist to deal systematically with the principles underlying (large-scale) structures; this included his thoughts on the *Ursatz*, which implicitly assume (a) a hierarchical structure (such as the large-scale tonic-dominant-tonic structure of the sonata form), (b) that (only) certain chord functions can be omitted, and (c) recursion (e.g., a sequence in one key which is embedded in yet another sequence with yet another different key, etc.). Schenker's thoughts were formalized by the approaches of the "Generative Theory of Tonal Music" (GTTM) by Lerdahl and Jackendoff (1999), which, however, does not provide generative rules (in contrast to what its name says). Nevertheless, the advance of the GTTM lies in the description of Schenker's *Urlinie* as tree-structure (in this regard, GTTM uses terms such as *time-span reduction* and *prolongation*). These tree-structures also parallel the

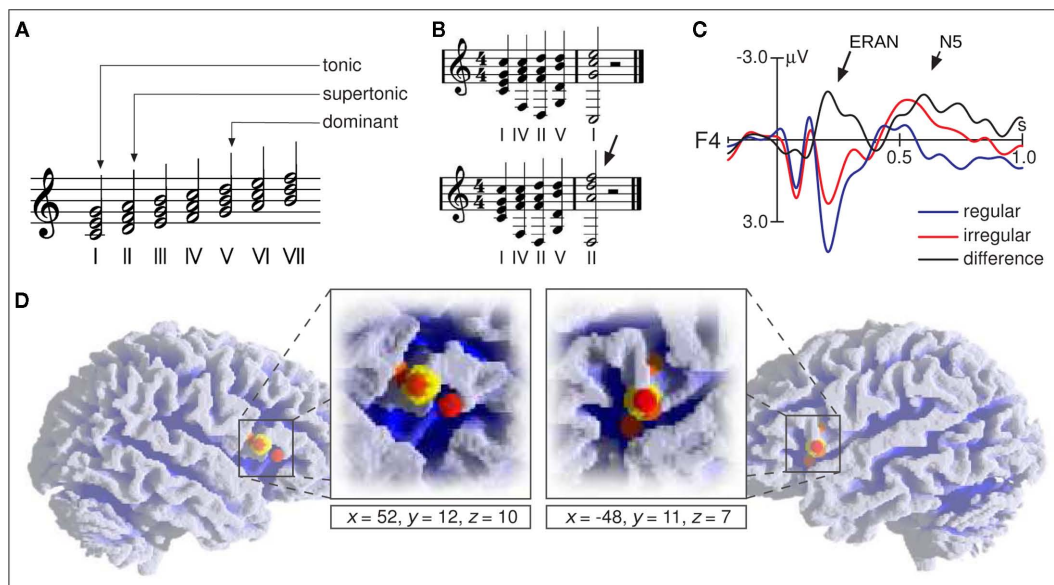


FIGURE 2 | (A) Examples of chord functions: The chord built on the first scale tone is denoted as the tonic, the chord on the second tone as the supertonic, and the chord on the fifth tone as the dominant. **(B)** The dominant–tonic progression represents a regular ending of a harmonic sequence (top), the dominant–supertonic progression is less regular and unacceptable as a marker of the end of a harmonic progression (bottom sequence, the arrow indicates the less regular chord). **(C)** ERPs elicited in a passive listening condition by the final chords of the two sequence types shown in **(B)**. Both sequence types were presented in pseudorandom order equiprobably in all 12 major keys. Brain responses to irregular chords clearly differ from those to regular chords (best to be seen in the black difference wave, regular

subtracted from irregular chords). The first difference between the two waveforms is maximal around 200 ms after the onset of the fifth chord (ERAN, indicated by the long arrow) and taken to reflect processes of music-syntactic analysis. The ERAN is followed by an N5 taken to reflect processes of harmonic integration (short arrow). **(D)** Activation foci (small spheres) reported by functional imaging studies on music-syntactic processing using chord sequence paradigms (Koelsch et al., 2002, 2005a; Maess et al., 2001; Tillmann et al., 2003) and melodies (Janata et al., 2002a). Large yellow spheres show the mean coordinates of foci (averaged for each hemisphere across studies, coordinates refer to standard stereotaxic space). Reprinted from Koelsch and Siebel (2005).

description of linguistic syntax using tree-structures (see also, e.g., Patel, 2008), although Lerdahl and Jackendoff's GTTM at least implicitly implies that musical and linguistic syntax have only little to do with each other (except perhaps with regard to certain prosodic aspects).

Up to now, no neurophysiological investigation has tested whether individuals perceive music cognitively according to tree-structures; similarly, behavioral studies on this topic are extremely scarce (but see Cook, 1987; Bigand et al., 1996; Lerdahl and Krumhansl, 2007). Note that GTTM is an analytical (not a generative) model; Lerdahl's tonal pitch space (TPS) theory (Lerdahl, 2001) attempts to provide algorithms for this analytical approach, but the algorithms are often not sharp and precise (and need subjective "corrections"). However, TPS provides the very interesting approach to model tension–resolution patterns from tree-structures (e.g., Lerdahl and Krumhansl, 2007).

A more recent approach to model tonal harmony with explicit generative rules according to tree-structures is the *Generative Syntax Model* (GSM) from Rohrmeier (2007) (see also Rohrmeier, 2011). Rohrmeier's GSM distinguishes between four structural levels with increasing abstraction: a level with the surface structure (in terms of the naming of the chords, e.g., C, G, C), a scale-degree structure level (e.g., I, V, I), a functional–structural level (e.g., t, d, t), and a phrase-structure level specifying tonic regions, dominant regions, and subdominant regions (see also Figure 3). Each of these levels is described with generative rules of a phrase-structure grammar,

the generation beginning on the highest (phrase-structure) level, and propagating information (in the process of derivation) through the sub-ordinate levels. This opens the possibility for the recursive derivation of complex sequences on both a functional and a scale-degree structural level (e.g., "dominant of the dominant of the dominant"). The actual tonality of a tree is propagated from the head ("knot") into the sub-ordinate branches. The tree-structure reflects (a) the formal arrangement of the piece, (b) the phrase structure, (c) the functional aspects of partial phrases and chords, (d) the relation of key-regions, and (e) the degree of relative stability (which is determined by the sub-ordination rules between the branches).

The GSM combines three principles:

- (1) Given (a) that musical elements (e.g., a tone or a chord) always occur in relation to other musical elements (and not alone), and (b) that each element (or each group of elements) has a functional relation either to a preceding or to a subsequent element, the simplest way of representing these relations is a tree-structure. This implies that relations of elements that are functionally related, but not adjacent, can be represented by a tree-structure. For example, in a sequence which consists of several chords and which begins and ends on a tonic chord, the first and the last tonic are not adjacent (because there are other chord functions between them), but the last tonic picks up (and "prolongates") the first tonic; this can be represented in a tree-structure in a way that the first and the last tonic chords build beginning- and end-points of the tree, and that the relations of the other chord functions can be represented by dendritic ramification ("prolongating" chords can extend the tonal region, and "progressing" chords determine the progression of tonal functions). This also implies that adjacent elements do not necessarily have a direct structural relation (for example, an initial tonic chord can be followed by a secondary dominant to the next chord: the secondary dominant then has a direct relation to the next chord, but not to the preceding tonic chord). It is not possible that a chord does not have a functional relation to a preceding, nor to a subsequent chord: If in our example the secondary dominant would not be followed by an (implicit) tonic, the sequence would be ungrammatical (with regard to major–minor tonal regularities). Compared to the classical theory of harmony, this GSM approach has the advantage that chord functions are not determined by their tonality (or the derivation from other chords), but by their functional position within the sequence.
- (2) The tree-structure indicates which elements can successively be omitted in a way that the sequence still sounds correct. For example, secondary dominants can be omitted, or all chords between first and last tonic chord can be omitted, and a sequence still sounds correct. In a less trivial case, there is a dominant in a local phrase between two tonic chords; the tree-structure then indicates that on the next higher level all chords except the two tonic and the dominant chords can be omitted. This principle is a consequence of the "headedness," in which each knot in the syntax tree dominates sub-ordinate branches. That is, the tree-structure provides a weighting of

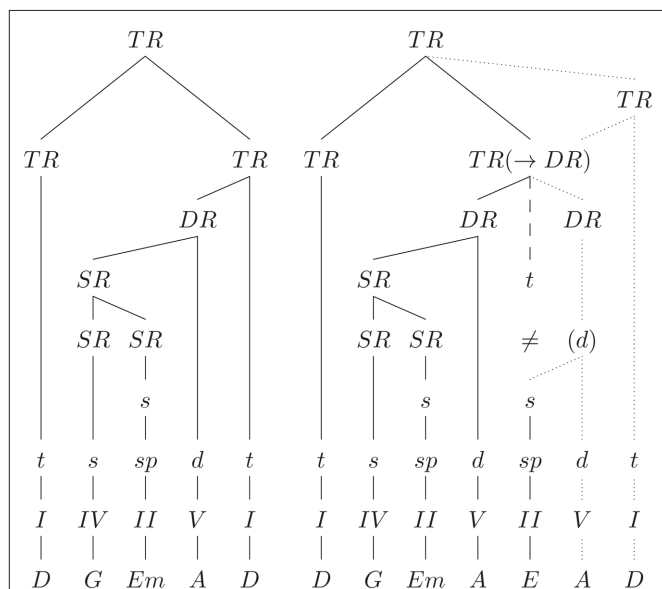


FIGURE 3 | Tree-structures (according to the GSM) for the sequences shown in Figure 2B, ending on a regular tonic (left), and on a supertonic (right). Dashed line: expected structure (\neq : the tonic chord is expected, but a supertonic is presented); dotted lines: a possible solution for the integration of the supertonic. $TR(\rightarrow DR)$ indicates that the supertonic can still be integrated, e.g., if the expected tonic region is re-structured into dominant region. TR, tonic region; DR, dominant region; SR, subdominant region. Lower-case letters indicate chord functions (functional–structural level), Roman numerals indicate the scale-degree structure, and the bottom row indicates the surface structure in terms of the naming of the chords.

the elements, leading to an objectively derivable deep structure (another advantage compared to the classical theory of harmony).

- (3) Chord functions are a result of their position within the branches of the tree. For example, a pre-dominant is the subordinate branch preceding a dominant, or there are stable and instable dominants which can be differentiated based on how deeply they are located in the syntax tree. Similarly, a half-cadence is a case in which a dominant is reached as stable local endpoint of a phrase, which must be followed by a phrase that ends on a tonic, thus resolving the open dominant (with regard to the deep structure).

A detailed description of the GSM goes beyond the scope of this article (for details see Rohrmeier, 2011). As mentioned before, the GSM is relatively new, and due to the lack of usable tree-models, previous studies investigating neurophysiological correlates of music-syntactic processing have so far simply utilized the classical theory of harmony as a syntactic rule system: According to the classical theory of harmony, chord functions are arranged within harmonic sequences according to certain regularities (Riemann, 1971, was the first to refer to such regularity-based arrangement as musical syntax). As described above, this regularity-based arrangement implies long-distance dependencies involving hierarchical organization (also referred to as phrase-structure grammar). However, it cannot be excluded that “irregular” chord functions such as the final supertonic in **Figure 2B** are detected as irregular based on a finite state grammar (according to which the tonic is the most regular chord after a I–IV–II–V progression); as will be illustrated below, processing of such chord functions presumably involves processing of both finite state and phrase-structure grammar.

5.2 NEURAL CORRELATES OF MUSIC-SYNTACTIC PROCESSING

Neurophysiological studies using EEG and MEG showed that music-syntactically irregular chord functions (such as the final supertonic shown in **Figure 2B**) elicit brain potentials with negative polarity that are maximal at around 150–350 ms after the onset of an irregular chord, and have a frontal/fronto-temporal scalp distribution, often with right-hemispheric weighting (see also **Figure 2C**). In experiments with isochronous, repetitive stimulation, this effect is maximal at around 150–200 ms over right anterior electrodes, and denoted as early right anterior negativity, or ERAN (for a review see Koelsch, 2009b). In experiments in which the position of irregular chords within a sequence is not known (and thus unpredictable), the negativity often has a longer latency, and a more anterior-temporal distribution (also referred to as right anterior-temporal negativity, or RATN; Patel et al., 1998; Koelsch and Mulder, 2002). The ERAN elicited by irregular tones of melodies has a shorter peak latency than the ERAN elicited by irregular chord functions (Koelsch and Jentschke, 2010).

Functional neuroimaging studies using chord sequence paradigms (Maess et al., 2001; Koelsch et al., 2002, 2005a; Tillmann et al., 2003) and melodies (Janata et al., 2002a) suggest that music-syntactic processing involves the pars opercularis of the inferior frontal gyrus (corresponding to BA 44) bilaterally, but with right-hemispheric weighting (see yellow spheres in **Figure 2D**). It seems likely that the involvement of (inferior) BA 44 (perhaps area 44v

according to Amunts et al., 2010) in music-syntactic processing is due to the hierarchical processing of (syntactic) information: This part of Broca's area is involved in the hierarchical processing of syntax in language (e.g., Friederici et al., 2006; Makuuchi et al., 2009), the hierarchical processing of action sequences (e.g., Koechlin and Jubault, 2006; Fazio et al., 2009), and possibly also in the processing of hierarchically organized mathematical formulas and termini (Friedrich and Friederici, 2009, although activation in the latter study cannot clearly be assigned to BA 44 or BA 45). These findings suggest that at least some cognitive operations of music-syntactic and language-syntactic processing (and neural populations mediating such operations) overlap, and are shared with the syntactic processing of actions, mathematical formulas, and other structures based on long-distance dependencies involving hierarchical organization (phrase-structure grammar).

However, it appears that inferior BA 44 is not the only structure involved in music-syntactic processing: additional structures include the superior part of the pars opercularis (Koelsch et al., 2002), the anterior portion of the STG (Koelsch et al., 2002, 2005a), and ventral premotor cortex (PMCv; Parsons, 2001; Janata et al., 2002b; Koelsch et al., 2002, 2005a). The PMCv probably contributes to the processing of music-syntactic information based on finite state grammar: activations of PMCv have been reported in a variety of functional imaging studies on auditory processing using musical stimuli, linguistic stimuli, auditory oddball paradigms, pitch discrimination tasks, and serial prediction tasks, underlining the importance of these structures for the sequencing of structural information, the recognition of structure, and the prediction of sequential information (Janata and Grafton, 2003). With regard to language, Friederici (2004) reported that activation foci of functional neuroimaging studies on the processing of long-distance hierarchies and transformations are located in the posterior IFG (with the mean of the coordinates reported in that article being located in the inferior pars opercularis), whereas activation foci of functional neuroimaging studies on the processing of local structural violations are located in the PMCv (see also Friederici et al., 2006; Makuuchi et al., 2009; Opitz and Kotz, 2011). Moreover, patients with lesion in the PMCv show disruption of the processing of finite state, but not phrase-structure grammar (Opitz and Kotz, 2011).

That is, in the above-mentioned experiments that used chord sequence paradigms to investigate the processing of harmonic structure, the music-syntactic processing of the chord functions probably involved processing of both finite state and phrase-structure grammar. The music-syntactic analysis involved a computation of the harmonic relation between a chord function and the context of preceding chord functions (phrase-structure grammar). Such a computation is more difficult (and less common) for irregular than for regular chord functions (as illustrated by the branch with the dashed line in the right panel of **Figure 3**), and this increased difficulty is presumably reflected in a stronger activation of (inferior) BA 44 in response to irregular chords. In addition, the local transition probability from the penultimate to the final chord is lower for the dominant–supertonic progression than for the dominant–tonic progression (finite state grammar), and the computation of the (less predicted) lower-probability progression is presumably reflected in a stronger activation of PMCv in response to irregular

chords. The stronger activation of both BA 44 and PMCV appears to correlate with the perception of a music-syntactically irregular chord as “unexpected.”

Note that the ability to process phrase-structure grammar is available to humans, whereas non-human primates are apparently not able to master such grammars (Fitch and Hauser, 2004). Thus, it is highly likely that only humans can adequately process music-syntactic information at the phrase-structure level. It is also worth noting that numerous studies showed that even “non-musicians” (i.e., individuals who have not received formal musical training) have a highly sophisticated (implicit) knowledge about musical syntax (e.g., Tillmann et al., 2000). Such knowledge is presumably acquired during listening experiences in everyday life.

5.3 INTERACTIONS BETWEEN LANGUAGE- AND MUSIC-SYNTACTIC PROCESSING

As mentioned above, hierarchical processing of syntactic information from different domains (such as music and language) requires contributions from neural populations located in BA 44. However, it is still possible that, although such neural populations are located in the same brain area, entirely different (non-overlapping) neural populations serve the syntactic processing of music and language *within the same area*. That is, perhaps the neural populations mediating language-syntactic processing in BA 44 are different from neural populations mediating music-syntactic processing in the same area. Therefore, the strongest evidence for shared neural resources for the syntactic processing of music and language stems from experiments that revealed interactions between music-syntactic and language-syntactic processing (Koelsch et al., 2005b; Steinbeis

and Koelsch, 2008b; Fedorenko et al., 2009; Slevc et al., 2009). In these studies, chord sequences were presented simultaneously with visually presented sentences (for an example see **Figure 4**). Two of these studies (Koelsch et al., 2005b; Steinbeis and Koelsch, 2008b) used EEG, and three different sentence types: The first type was a syntactically correct sentence in which the occurrence of the final noun was semantically highly probable. The other two sentence types were modified versions of the first sentence type: Firstly, a sentence with a gender disagreement between the last word (noun) on the one hand, and the prenominal adjective as well as the definite article that preceded the adjective on the other; such gender disagreements elicit a left anterior negativity (LAN; Gunter et al., 2000). Secondly, a sentence in which the final noun was semantically less probable; such “low-cloze probability” words elicit an N400 reflecting the semantic processing of words (see also section on musical semantics below). Each chord sequence consisted of five chords (and each chord was presented together with a word, see **Figure 4**), one chord sequence type ending on a music-syntactically regular, and the other type ending on an incorrect chord function.

Both studies (Koelsch et al., 2005b; Steinbeis and Koelsch, 2008b) showed that the ERAN elicited by irregular chords interacted with the LAN elicited by linguistic (morpho-syntactic) violations (Koelsch et al., 2005b; Steinbeis and Koelsch, 2008b): The LAN elicited by words was reduced when the syntactically irregular word was presented simultaneously with a music-syntactically irregular chord (compared to when the irregular word was presented with a regular chord). In the study from Koelsch et al. (2005b) a control experiment was conducted in which the same sentences were presented simultaneously with sequences of single tones. The

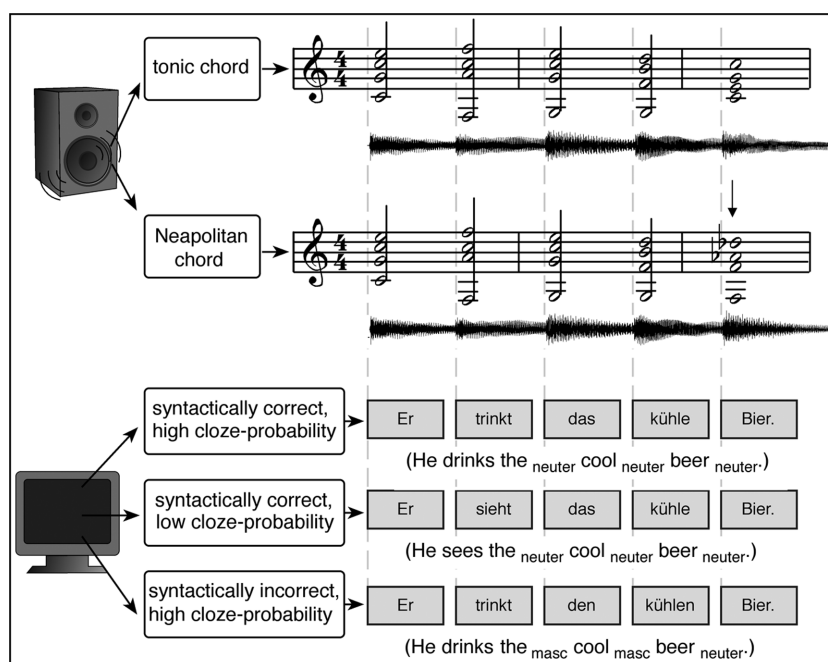


FIGURE 4 | Examples of experimental stimuli used in the studies by Koelsch et al. (2005b) and Steinbeis and Koelsch (2008b). Top: examples of two chord sequences in C major, ending on a regular (upper row) and an irregular chord (lower row, the irregular chord is indicated by the arrow). Bottom: examples of the three different sentence types. Onsets of chords (presented auditorily) and words (presented visually) were synchronous. Reprinted from Steinbeis and Koelsch (2008b).

tone sequences ended either on a standard tone or on a frequency deviant. The physical mismatch negativity (phMMN) elicited by the frequency deviants did not interact with the LAN (in contrast to the ERAN), indicating that the processing of auditory oddballs (as reflected in the phMMN) does not consume resources related to syntactic processing¹⁶. In addition, the study by Steinbeis and Koelsch (2008b) also showed that the amplitude of the ERAN was reduced when participants processed a syntactically wrong word (**Figure 5A**)¹⁷. Notably, the ERAN was not affected when words were semantically incongruous (**Figure 5B**), showing that the interaction between ERAN and LAN is specific for syntactic processing (the double interaction between ERAN and syntax, as well as between N5 and semantics, is shown in **Figure 5C**).

The findings of these EEG studies were corroborated by two behavioral studies (Fedorenko et al., 2009; Slevc et al., 2009), as well as by a patient study by Patel et al. (2008). The latter study showed that individuals with Broca's aphasia also show impaired music-syntactic processing in response to out-of-key

chords occurring in harmonic progressions (note that all patients had Broca's aphasia, but only some of them had a lesion that included Broca's area).

It is also interesting to note that there are hints for an interaction between the ERAN and the *early left anterior negativity* (ELAN, an ERP component taken to reflect the processing of phrase structure violations during language perception; Friederici, 2002): In a study by Maidhof and Koelsch (2011), participants were presented simultaneously with spoken sentences and chord sequences. Thus, the experiment was similar to the studies by Koelsch et al. (2005b) and Steinbeis and Koelsch (2008b), except that (a) sentences were presented auditorily, (b) that syntactic violations were phrase-structure violations (not morpho-syntactic violations), and (c) that the attention of listeners was directed to either the speech or the music. In that study, the amplitude of the ERAN was slightly smaller when elicited during the presentation of phrase-structure violations occurring in sentences (compared to when elicited during correct sentences); this effect was, however, only marginally significant. Thus, it appears that music-syntactic processing of an irregular chord function (as reflected in the ERAN) also interacts with the processing of the phrase-structure of a sentence, although this interaction seems less obvious than the interaction between ERAN and LAN (i.e., between music-syntactic processing of chord functions and the processing of morpho-syntactic information in language).

In summary, neurophysiological studies show that music- and language-syntactic processes engage overlapping resources (presumably located in the inferior fronto-lateral cortex), and evidence showing that these resources underlie music- and language-syntactic processing is provided by experiments showing interactions between ERP components reflecting music- and language-syntactic processing (in particular LAN and ERAN). Importantly, such interactions are observed in the absence of interactions between LAN and MMN, i.e., in the absence between language-syntactic and acoustic deviance processing (reflected in the MMN), and in the absence of interactions between the ERAN and the N400 (i.e., in the absence of music-syntactic and language-semantic processing). Therefore, the reported interactions between LAN and ERAN are syntax-specific, and cannot be observed in response to any kind of irregularity.

5.4 PROCESSING OF PHRASE BOUNDARIES

During auditory music or language perception, the recognition of phrase boundaries helps to decode the syntactic (phrase) structure of a musical phrase, as well as of a sentence (the melodic and rhythmic features that mark a phrase boundary in a spoken sentence are part of the *speech prosody*). The processing of phrase boundaries is reflected in a *closure positive shift* (CPS), during both the processing of musical phrase boundaries (Knösche et al., 2005), and the processing of intonational phrase boundaries during speech perception (Steinhauer et al., 1999; although perhaps with a slightly different latency). The first study on the perception of musical phrase boundaries by Knösche et al. (2005) investigated musicians only, and a subsequent study also reported a CPS in non-musicians (although the CPS had a considerably smaller amplitude than the CPS observed in musicians; Neuhaus et al., 2006). The latter study also reported larger CPS amplitudes for longer pauses as well as longer boundary tones (which make the phrase boundary more

¹⁶Whether the abstract-feature MMN consumes such resources remains to be investigated.

¹⁷The fact that the ERAN was not reduced by the syntactically incorrect words in the study by Koelsch et al. (2005b) is presumably due to the task: Participants focused more attention on the musical stimulus in the study by Steinbeis and Koelsch (2008b), leading to a larger ERAN amplitude (and thus to a better signal-to-noise ratio).

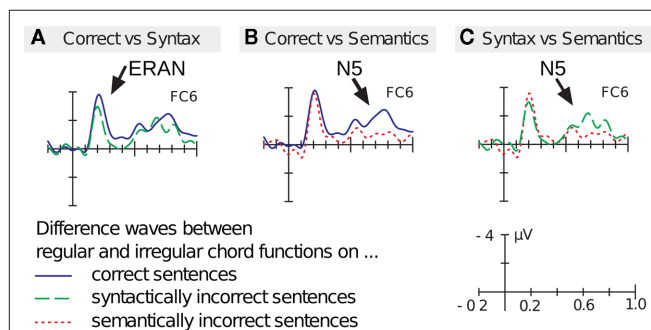


FIGURE 5 | Grand-average ERPs elicited by the stimuli shown in Figure 4.

Participants monitored whether the sentences were (syntactically and semantically) correct or incorrect; in addition, they had to attend to the timber of the chord sequences and to detect infrequently occurring timber deviants. ERPs were recorded on the final chords/words and are shown for the different word conditions (note that only difference waves are shown). **(A)** The solid (blue) difference wave shows ERAN (indicated by the arrow) and N5 elicited on syntactically and semantically correct words. The dashed (green) difference wave shows ERAN and N5, elicited when chords are presented on morpho-syntactically incorrect (but semantically correct) words. Under the latter condition, the ERAN (but not the N5) is reduced. **(B)** The solid (blue) difference wave is identical to the solid difference wave of **(A)**, showing the ERAN and the N5 (indicated by the arrow) elicited on syntactically and semantically correct words. The dotted (red) difference wave shows ERAN and N5, elicited when chords are presented on semantically incorrect (but morpho-syntactically correct) words. Under the latter condition, the N5 (but not the ERAN) is reduced. **(C)** shows the direct comparison of the difference waves in which words were syntactically incorrect (dashed, green line) or semantically incorrect (dotted, red line). These ERPs show that the ERAN is influenced by the morpho-syntactic processing of words, but not by the semantic processing of words. By contrast, the N5 is influenced by the semantic processing of words, but not by the morpho-syntactic processing of words. Data from Steinbeis and Koelsch (2008b).

salient). Both studies (Knösche et al., 2005; Neuhaus et al., 2006) used EEG as well as MEG, showing that the CPS can be observed with both methods. In a third study, Chinese and German musicians performed a categorization task with Chinese and Western music (with both groups being familiar with Western music, but only the Chinese group being familiar with Chinese music; Nan et al., 2006). Both groups showed CPS responses to both types of music (with no significant difference between groups). The exact contributions of the processes reflected in the CPS to syntactic processing remain to be specified.

Using fMRI, a study by Meyer et al. (2004) suggests that the processing of the prosodic aspects of the sentences used in the study by Steinhauer et al. (1999) involves premotor cortex of the (right) Rolandic operculum, (right) AC located in the planum temporale, as well as the anterior insula (or perhaps deep frontal operculum) and the striatum bilaterally¹⁸.

A very similar activation pattern was also observed for the processing of the melodic contour of spoken main clauses (Meyer et al., 2002). That is, similar to the right-hemispheric weighting of activations observed in functional neuroimaging studies on music perception, a right-hemispheric weighting of (neo-cortical) activations is also observed for the processing of speech melody.

6 STRUCTURAL REANALYSIS AND REPAIR

Following a syntactic anomaly (or an unexpected syntactic structure), processes of structural reanalysis and repair may be engaged. It appears that these processes are reflected in the ERP as positive potentials that are maximal around 600–900 ms (often referred to as P600, or *late positive component*, LPC; Besson and Schön, 2001).

The P600/LPC can be elicited by irregular melody-tones (e.g., Besson and Macar, 1986; Verleger, 1990; Paller et al., 1992; Besson and Faita, 1995; Besson et al., 1998; Miranda and Ullman, 2007; Peretz et al., 2009), as well as by irregular chords (e.g., Patel et al., 1998). It seems that the P600/LPC can only be elicited when individuals attend to the musical stimulus, and that the LPC is partly connected to processes of the conscious detection of music-structural incongruities. In an experiment by Besson and Faita (1995), in which incongruous melody-endings had to be detected, the LPCs had a greater amplitude, and a shorter latency, in musicians compared to “non-musicians,” presumably because musicians were more familiar with the melodies than non-musicians (thus, musicians could detect the irregular events more easily). Moreover, the LPC had a larger amplitude for familiar melodies than for novel melodies (presumably because incongruous endings of familiar phrases were easier to detect), and for non-diatonic than for diatonic endings. Diatonic incongruities terminating unfamiliar melodies did not elicit an LPC (presumably because they are hardly to detect), whereas non-diatonic incongruities did (they were detectable for participants by the application of tonal rules).

It is remarkable that in *all* of the mentioned studies on melodies (Besson and Macar, 1986; Verleger, 1990; Paller et al., 1992; Besson and Faita, 1995; Besson et al., 1998; Brattico et al., 2006; Miranda and Ullman, 2007; Peretz et al., 2009), the unexpected tones also elicited an early frontal negative ERP (emerging around the N1

peak, i.e., around 100 ms after stimulus onset). This ERP effect resembles the ERAN (or, more specifically, an early component of the ERAN; for details see Koelsch and Jentschke, 2010), although it presumably overlapped in part with a subsequent N2b due to the detection of irregular or unexpected tones. That is, earlier processes underlying the detection of structural irregularities (as partly reflected in the ERAN) are followed by later processes of structural reanalysis (as reflected in the P600/LPC) when individuals attend to the musical stimulus and detect structural incongruities. It remains to be specified whether the P600/LPC is a late P3, and how the processes of structural reanalysis and repair are possibly related to context-updating (see also Donchin and Coles, 1998; Polich, 2007).

Using sentences and chord sequences, Patel et al. (1998) compared ERPs elicited by “syntactic incongruities” in language and music within subjects (harmonic incongruities were taken as grammatical incongruity in music). Task-relevant (target) chords within homophonic musical phrases were manipulated, so that the targets were either within the key of a phrase, or out-of-key (from a “nearby,” or a “distant” key). Both musical and linguistic structural incongruities elicited P600/LPC potentials which were maximal at posterior sites and statistically indistinguishable. Moreover, the degree of a structural anomaly (moderate or high) was reflected in the amplitude of the elicited positivities. Therefore, the results indicated that the P600 reflects more general knowledge-based structural (re-)integration (and/or reanalysis) during the perception of rule-governed sequences. The intersection between the processes of music-syntactic and language-syntactic (re-)integration and repair is referred to as the *shared syntactic integration resource hypothesis* (SSIRH), which states that music and language rely on shared, limited processing resources that activate separable syntactic representations (Patel, 2003). The SSIRH also states that, whereas the linguistic and musical knowledge systems may be independent, the system used for online structural integration may be shared between language and music (see also Fedorenko et al., 2009). This system might be involved in the integration of incoming elements (words in language, and tones/chords in music) into evolving structures (sentences in language, harmonic sequences in music). Beyond the SSIRH, however, the previous section on music-syntactic processing also showed early interactions (between the ERAN and LAN), indicating that processing of music- and language-syntactic information intersects also at levels of morpho-syntactic processing, phrase-structure processing, and possibly word-category information.

7 PROCESSING MEANING IN MUSIC

Music is a means of communication, and during music listening, meaning emerges through the interpretation of (musical) information. Previous accounts on musical meaning can be summarized with regard to three fundamentally different classes of meaning emerging from musical information: extra-musical meaning, intra-musical meaning, and musicogenic meaning (for an extensive review see Koelsch, 2011). Extra-musical meaning refers to meaning emerging from reference to the extra-musical world (Meyer, 1956, referred to this class of musical meanings as “designative meaning”). This class of meaning comprises three dimensions: musical meaning due to iconic, indexical, and symbolic sign qualities (Karbusicky, 1986).

¹⁸Sentences were stripped of phonological information by filtering, and sounded like hummed sentences.

- (a) Iconic musical meaning emerges from common patterns or forms, such as musical sound patterns that resemble sounds of objects, or qualities of objects. For example, acoustic events may sound “warm,” “round,” “sharp,” or “colorful,” and a musical passage may sound, e.g., “like a bird,” or “like a thunderstorm.” In linguistics, this sign quality is also referred to as *onomatopoeic*.
- (b) Indexical musical meaning emerges from action-related patterns (such as movements and prosody) that index the presence of a psychological state, for example an emotion, or an intention. Juslin and Laukka (2003) compared in a meta analysis the acoustical signs of emotional expression in music and speech, finding that the acoustic properties that code emotional expression in speech are highly similar to those coding these expressions in music. With regard to intentions, an fMRI study by Steinbeis and Koelsch (2008c) showed that listeners automatically engage social cognition during listening to music, in an attempt to decode the intentions of the composer or performer (as indicated by activations of the cortical *theory-of-mind* network). That study also reported activations of posterior temporal regions implicated in semantic processing (Lau et al., 2008), presumably because the decoding of intentions has meaning quality.
Cross (2008) refers to this dimension of musical meaning as “motivational–structural” due to the relationship between affective–motivational states of individuals on the one side, and the structural–acoustical characteristics of (species-specific) vocalizations on the other.
- (c) Symbolic musical meaning emerges from explicit (or conventional) extra-musical associations (e.g., any national anthem). Note that the meaning of the majority of words is due to symbolic meaning. Cross and Morley (2008) refers to this dimension of musical meaning as “culturally enactive,” emphasizing that symbolic qualities of musical practice are shaped by (and shape) culture.

Musical semantics, notably, extends beyond the relations between concepts and the (extra-musical) world in that musical meaning also emerges from the reference of one musical element to another musical element, that is, from intra-musical combinations of formal structures. The term “intra-musical” was also used by Budd (1996), other theorists have used the terms “formal meaning” (Alperson, 1994), “formal significance” (Davies, 1994), or “embodied meaning” (Meyer, 1956). Similar distinctions as the one drawn here between extra- and intra-musical meaning have been made by several theorists (for a review see Koopman and Davies, 2001).

Processing of musical meaning information is reflected in (at least) two negative ERP components: The N400 and the N5. The N400 reflects processing of meaning in both language and music, the N5 has so far only been observed for the processing of musical information. In the following I will present evidence showing that, with regard to the processing of musical meaning, the N400 reflects processing of extra-musical meaning, and the N5 processing of intra-musical meaning.

The N400 has been used to investigate processing of musical meaning (or “musical semantics”) in semantic priming paradigms (Koelsch et al., 2004; Steinbeis and Koelsch, 2008a, 2011; Daltrozzo

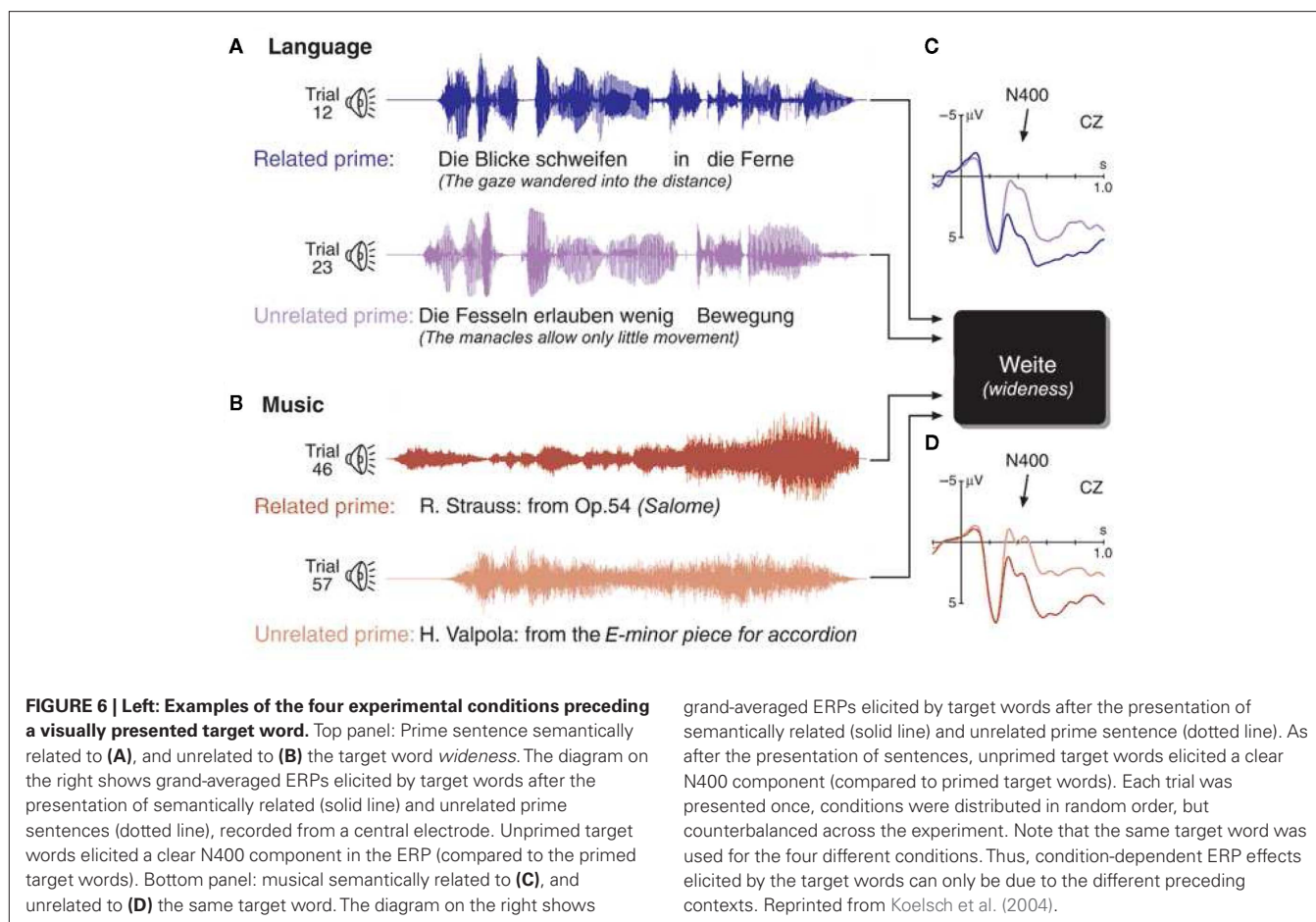
and Schön, 2009a; Grieser-Painter and Koelsch, 2011). In an initial study (Koelsch et al., 2004), sentences and musical excerpts were presented as prime stimuli. The prime stimuli were semantically either related or unrelated to a target word that followed the prime stimulus (see top of **Figure 6**). For example, the sentence “The gaze wandered into the distance” primes the word “wideness” (semantically related), rather than the word “narrowness” (semantically unrelated). Analogously, certain musical passages, for example from a symphony by Mozart, prime the word “angel,” rather than the word “scallywag.”

In the language condition (i.e., when target words followed the presentation of sentences), unrelated words elicited a clear N400 effect (this is a classical semantic priming effect). This semantic priming effect was also observed when target words followed musical excerpts. That is, target words that were semantically unrelated to a preceding musical excerpt also elicited a clear N400. The N400 effects did not differ between the language condition (in which the target words followed sentences) and the music condition (in which the target words followed musical excerpts), neither with respect to amplitude nor with respect to latency or scalp distribution. A source analysis localized the main sources of the N400 effects, in both conditions, in the posterior part of the medial temporal gyrus bilaterally (BA 21/37), in proximity to the superior temporal sulcus. These regions have been implicated in the processing of semantic information during language processing (Lau et al., 2008).

The N400 effect in the music condition demonstrates that musical information can have a systematic influence on the semantic processing of words. The N400 effects did not differ between the music and the language condition, indicating that musical and linguistic priming can have the same effects on the semantic processing of words. That is, the data demonstrated that music can activate representations of meaningful concepts (and that, thus, music is capable of transferring considerably more meaningful information than previously believed), and that the cognitive operations that decode meaningful information while listening to music can be identical to those that serve semantic processing during language perception. The N400 effect was observed for both abstract and concrete words, showing that music can convey both abstract and concrete semantic information. Moreover, effects were also observed when emotional relationships between prime and target words were balanced, indicating that music does not only transfer emotional information.

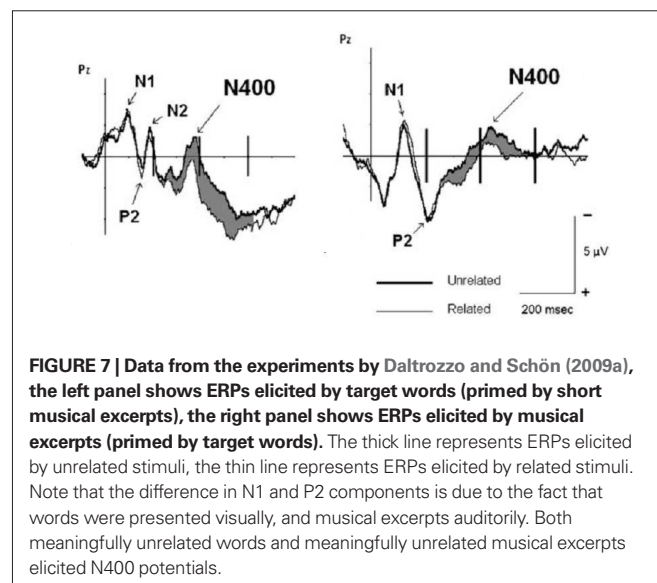
The priming of meaning by the musical information was due to (a) iconic sign qualities, i.e., common patterns or forms (such as succeeding interval steps priming the word staircase), (b) indexical sign quality, such as the suggestion of a particular emotion due to the resemblance to movements and prosody typical for that emotion (e.g., saxophone tones sounding like derisive laughter), or (c) symbolic sign quality due to meaning inferred by explicit extra-musical associations (e.g., a church anthem priming the word devotion). Unfortunately, it was not investigated in that study whether N400 responses differed between these dimensions of musical meaning (and, so far, no subsequent study has investigated this). However, the results still allow to conclude that processing of extra-musical meaning is associated with N400 effects.

Due to the length of musical excerpts (~10 s), musical information could not be used as target stimulus, thus only words were used as target stimuli (and the N400 was only elicited on words, not on musical information). Hence, a question emerging from that study



grand-averaged ERPs elicited by target words after the presentation of semantically related (solid line) and unrelated prime sentence (dotted line). As after the presentation of sentences, unprimed target words elicited a clear N400 component (compared to primed target words). Each trial was presented once, conditions were distributed in random order, but counterbalanced across the experiment. Note that the same target word was used for the four different conditions. Thus, condition-dependent ERP effects elicited by the target words can only be due to the different preceding contexts. Reprinted from Koelsch et al. (2004).

was whether musical information can also elicit N400 responses. One study addressing this issue (Daltrozzo and Schön, 2009a) used short musical excerpts (duration was ~1 s) that could be used either as primes (in combination with word targets), or as targets (in combination with word primes). **Figure 7** shows that when the musical excerpts were used as primes, meaningfully unrelated target words elicited an N400 (compared to related target words, as in the study from Koelsch et al., 2004). Importantly, when musical excerpts were used as target stimuli (and words as primes), an N400 was observed in response to excerpts that the participants rated as meaningfully unrelated to the preceding target word (compared to excerpts that were rated as related to the preceding word). This was the first evidence that musical information can also elicit N400 responses. Note that the musical excerpts were composed for the experiment, and thus not known by the participants. Therefore, musical meaning was not due to symbolic meaning, but due to indexical (e.g., “happy”) and iconic (e.g., “light”) meaning. In the data analysis used in that study (Daltrozzo and Schön, 2009a), the relatedness of prime–target pairs was based on the relatedness judgments of participants. In another article (Daltrozzo and Schön, 2009b) the author showed that even if the data are analyzed based on the un/relatedness of prime–target pairs as pre-defined by the experimenters, a significant N400 was elicited (in the latter study a lexical decision task was used, whereas in the former study participants were asked to judge the conceptual relatedness of prime and targets).



Further studies investigated the processing of musical meaning using only single chords (Steinbeis and Koelsch, 2008a, 2011) or single tones (Grieser-Painter and Koelsch, 2011). These studies showed that even single chords and tones can prime the meaning of words

(in both musicians and non-musicians), and that even chords (in musicians) and tones (in non-musicians and therefore presumably also in musicians) can elicit N400 effects.

INTRA-MUSICAL MEANING AND THE N5

The previous section dealt with the N400 and extra-musical meaning. However, musical meaning can also emerge from intra-musical references, that is, from the reference of one musical element to at least one other musical element (for example, a G major chord is usually perceived as the tonic in G major, as the dominant in C major, and – in its first inversion – possibly as a Neapolitan sixth chord in F# minor). This section will review empirical evidence for this hypothesis, and describe an electrophysiological correlate of the processing of intra-musical meaning (the so-called N5, or N500). The N5 was described first in reports of experiments using chord sequence paradigms with music-syntactically regular and irregular chord functions (Koelsch, 2000), in which the ERAN was usually followed by a late negativity, the N5 (see also **Figure 2C**). Initially (Koelsch, 2000) the N5 was proposed to reflect processes of harmonic integration, reminiscent of the N400 reflecting semantic integration of words, and therefore proposed to be related to the processing of musical meaning, or musical semantics, although the type of musical meaning had remained unclear.

One reason for this proposition was a remarkable resemblance between N5 and N400: *Firstly*, similar to the N400 elicited by open-class words, which declines toward the end of sentences (Van Petten and Kutas, 1990), the amplitude of the N5 (elicited by regular chords) declined toward the end of a chord sequence (Koelsch, 2000). That is, during sentence processing, a semantically correct final open-class word usually elicits a rather small N400, whereas the open-class words preceding this word elicit larger N400 potentials. This is due to the semantic expectedness of words, which is rather unspecific at the beginning of a sentence, and which becomes more and more specific toward the end of the sentence (where people already have a hunch of what the last word will actually be). Thus, a smaller amount of semantic integration is required at the end of a sentence, reflected in a smaller N400. If the last word is semantically unexpected, then a large amount of semantics is required, which is reflected in a larger amplitude of the N400. Similarly, the N5 is not only elicited by irregular chords, but also by regular chords, and the amplitude of the N5 decreases with each progressing chord of the chord sequence. A small N5 elicited by the (expected) final chord of a chord sequence presumably reflects that only a small amount of harmonic integration is required at this position of a chord sequence.

Secondly, at the same position within a chord sequence (or a sentence) the N5 (or the N400) is modulated by the degree of fit with regard to the previous harmonic (or semantic) context. A wealth of studies has shown that irregular chords (for reviews see Koelsch, 2004, 2009b) and irregular tones of melodies (e.g., Miranda and Ullman, 2007; Koelsch and Jentschke, 2010) evoke larger N5 potentials than regular ones (see also **Figure 2C**). As mentioned above, the ERAN is related to syntactic processing. In addition, the experiment from Steinbeis and Koelsch (2008b) also showed that the N5 reflects processing of meaning information: In that study, the N5 interacted with the N400 elicited by words with low semantic cloze probability. The N5 was smaller when elicited on words that were semantically less probable (“He sees the cold beer”) compared to when elicited on words that were semantically highly probable (“He drinks the cold

beer”; see **Figure 5B**). Importantly, the N5 did not interact with the LAN (i.e., the N5 did not interact with the syntactic processing of words), indicating that the N5 is not simply modulated by any type of deviance, or incongruency, but that the N5 is specifically modulated by neural mechanisms underlying semantic information processing. That is, the N5 potential can be modulated by semantic processes, namely by the activation of lexical representations of words with different semantic fit to a previous context. This modulation indicates that the N5 is related to the processing of meaning. Note that the harmonic relation between the chord functions of a harmonic sequence is an intra-musical reference (i.e., a reference of one musical element to another musical element, and not a reference to anything belonging to the extra-musical world). Therefore, we have reason to believe that the N5 reflects the processing of intra-musical meaning.

The neural generators of the N5 have remained elusive. This is in part due to the difficulty that in most experiments the N5 follows the ERAN (but see also, Poulin-Charronnat et al., 2006), making it difficult to differentiate the neural correlates of N5 and ERAN in experiments using functional magnetic resonance imaging. The N5 usually has a clear frontal scalp distribution, thus the scalp distribution of the N5 is more anterior than that of the N400, suggesting at least partly different neural correlates. Perhaps the N5 originates from combined sources in the temporal lobe (possibly overlapping with those of the N400 in BAs 21/37) and the frontal lobe (possibly in the posterior part of the inferior frontal gyrus). This needs to be specified, for example using EEG source localization in a study that compares an auditory N4 with an auditory N5 within subjects.

The third class of musical meaning (musicogenic meaning) emerges from individual responses to musical information, particularly movement, emotional responses, and self-related memory associations. This dimension of musical meaning, as well as details about the extra- and intra-musical dimensions of musical meaning, are discussed in more detail elsewhere (Koelsch, 2011).

8 HOW THE BODY REACTS TO MUSIC

The present model of music perception also takes the potential “vitalization” of an individual into account: vitalization entails activity of the autonomic nervous system (i.e., regulation of sympathetic and parasympathetic activity) along with the cognitive integration of musical and non-musical information. Non-musical information comprises associations evoked by the music, as well as emotional (e.g., happy), and bodily reactions (e.g., tensioned or relaxed). The *subjective feeling* (e.g., Scherer, 2005) requires conscious awareness, and therefore presumably involves multimodal association cortices such as parietal association cortices in the region of BA 7 (here, the musical percept might also become conscious; Block, 2005). Effects of music perception on activity of the autonomic nervous system have mainly been investigated by measuring electrodermal activity and heart rate, as well as the number and intensity of reported “shivers” and “chills” (Sloboda, 1991; Blood and Zatorre, 2001; Khalfa et al., 2002; Panksepp and Bernatzky, 2002; Grewe et al., 2007a,b; Lundqvist et al., 2009; Orini et al., 2010).

Vitalizing processes can, in turn, have an influence on processes within the immune system. Effects of music processing on the immune system have been assessed by measuring variations of (salivary) immunoglobulin A concentrations (e.g., McCraty et al., 1996; Hucklebridge et al., 2000; Kreutz et al., 2004). Interestingly, effects

on the immune system have been suggested to be tied to motor activity such as singing (Kreutz et al., 2004) or dancing (Quiroga Murcia et al., 2009, see also **Figure 1**, rightmost box). With regard to music perception, it is important to note that there might be overlap between neural activities of the late stages of perception and those related to the early stages of action (such as premotor functions related to action planning; Janata et al., 2002b; Rizzolatti and Craighero, 2004).

Music perception can interfere with action planning in musicians (Drost et al., 2005a,b), and premotor activity can be observed during the perception of music (a) in pianists listening to piano pieces (Haueisen and Knösche, 2001), (b) in non-musicians listening to song (Callan et al., 2006), (c) in non-musicians who received 1 week of piano training and listened to the trained piano melody (Lahav et al., 2007; for a detailed review see Koelsch, 2009a). For neuroscience studies related to music production see, e.g., Bangert and Altenmüller (2003), Katahira et al. (2008), Maidhof et al. (2009), Kamiyama et al. (2010), Maidhof et al. (2010).

Movement induction by music perception in the way of dancing, singing, tapping, hopping, swaying, head-nodding, etc., along with music is a very common experience (Panksepp and Bernatzky, 2002); such movements also serve social functions, because synchronized movements of different individuals represent coordinated social activity. Notably, humans have a need to engage in social activities; emotional effects of such engagement include fun, joy, and happiness, whereas exclusion from this engagement represents an emotional stressor, and has deleterious effects on health. Humans making music is an activity involving several social functions, which were recently summarized as the “Seven Cs” (Koelsch, 2010): (1) When we make music, we make *contact* with other individuals (preventing from social isolation). (2) Music automatically engages *social cognition* (Steinbeis and Koelsch, 2008c). (3) Music engages *co-pathy* in the sense that inter-individual emotional states become more homogenous (e.g., reducing anger in one individual, and depression or anxiety in another), thus promoting inter-individual understanding and decrease of conflicts. (4) Music involves *communication* (notably, for infants and young children, musical communication during parent–child singing of lullabies and play-songs is important for social and emotional regulation, as well as for social, emotional, and cognitive development; Trehub, 2003; Fitch, 2006). (5) Music making also involves coordination of movements (requiring the capability to synchronize movements to an external beat; see also Kirschner and Tomasello, 2009; Overy and Molnar-Szakacs, 2009; Patel et al., 2009). The coordination of movements in a group of individuals appears to be associated with pleasure (for example, when dancing together), even in the absence of an explicit shared goal (apart from deriving pleasure from concerted movements). (6) Performing music also requires *cooperation* (involving a shared goal, and increasing inter-individual trust); notably, engaging in cooperative behavior is an important potential source of pleasure Rilling et al. (2002). (7) As an effect, music leads to increased *social cohesion* of a group (Cross and Morley, 2008), fulfilling the “need to belong” (Baumeister and Leary, 1995), and the motivation to form and maintain interpersonal attachments. Social cohesion also strengthens the confidence in reciprocal care (see also the caregiver hypothesis; Trehub, 2003), and the confidence that opportunities to engage with others in the mentioned social

functions will also emerge in the future. Music seems to be capable of engaging all of the “Seven Cs” at the same time, which is presumably part of the emotional power of music. These evolutionarily advantageous social aspects of music-making behavior represent one origin for the evolution of music-making behavior in humans.

Action induction by music perception is accompanied by neural impulses in the reticular formation (in the brainstem; for example, for the release of energy to move during joyful excitement). It is highly likely that connections also exist between the reticular formation and structures of the auditory brainstem (as well as between reticular formation and the AC; Levitt and Moore, 1979), and that the neural activity of the reticular formation therefore also influences the processing of (new) incoming acoustic information.

9 MUSIC PERCEPTION AND MEMORY

The modules presented in **Figure 1** are associated with a variety of memory functions (for a review see Jäncke, 2008). For example, the auditory sensory memory (along with Gestalt formation) is connected with both WM (Berti and Schröger, 2003) and long-term memory (Näätänen et al., 2001, see above for information about brain structures implicated in auditory sensory memory). Structure building requires WM as well as a long-term store for syntactic regularities, and processing of meaning information is presumably tied to a mental “lexicon” (containing conceptual–semantic knowledge), as well as to a musical “lexicon” containing knowledge about timbers, melodic contours, phrases, and musical pieces (Peretz and Coltheart, 2003). However, the details about interconnections between the different modules and different memory functions remain to be specified.

Neuroimaging studies suggest that the phonological loop of verbal WM and the “tonal loop” of WM for pitch strongly overlap in non-musicians (Hickok et al., 2003; Koelsch et al., 2009; Schulze et al., 2011a), involving PMCV (encroaching Broca’s area), dorsal premotor cortex, the planum temporale, inferior parietal lobe, the anterior insula, subcortical structures (basal ganglia and thalamus), as well as the cerebellum. A study by Schulze et al. (2011b) showed that, in contrast to non-musicians, musicians use specific neural subcomponents of WM only during verbal (right insular cortex) or only during tonal WM (right globus pallidus, right caudate nucleus, and left cerebellum). These results revealed the existence of two WM systems in musicians: A phonological loop supporting rehearsal of phonological information, and a tonal loop supporting rehearsal of tonal information. Note that differences between non-musicians and musicians for tonal WM (and between verbal and tonal WM within musicians), were mainly related to structures involved in controlling, programming, and planning of actions, thus presumably reflecting differences in action-related sensorimotor coding of verbal and tonal information. That is, verbal and tonal WM rely strongly on action-related coding (and in this regard sensorimotor coding forms one basis for cognition; for a study on strategic WM processes in musicians see Schulze et al., 2011a).

The knowledge about musical long-term memory is still rather limited. Functional neuroimaging data suggest that access to musical semantic memory involves the (left) middle temporal gyrus, and that musical semantic representations (that is, parts of a musical lexicon) are stored in (left) anterior-temporal areas (Groussard et al., 2009, 2010). In addition, Watanabe et al. (2008) reported that

retrieval of musical information involves the hippocampal formation and the inferior frontal gyrus. Further systematic research in these areas is needed to differentiate memory operations from operations of the modules described in **Figure 1**; it is important to bear in mind that, especially in functional imaging experiments, both types of operations usually co-occur.

10 MUSIC AND LANGUAGE

The previous sections provided a number of examples for cognitive processes (and their neural correlates) underlying the processing of music as well as of (spoken) language. Decoding of both music and speech information requires a fine-grained analysis of the spectral and temporal features of acoustic information (boxes “Feature Extraction” in **Figure 1**). The section on auditory feature extraction mentioned the acoustic equivalence of timber and phoneme, and the *identification* of “phonemes” in language is thus presumably paralleled by the identification of “timbers” in music. However, the segmentation of phonemic information during language perception usually requires a higher temporal resolution compared to the music (because timbral information in music usually does not change as rapidly as phonemic information in language). This probably leads to the left-hemispheric weighting for the segmentation of phonemes during language perception, whereas segmentation of spectral information such as melodic information of speech prosody or musical melodies engages the right AC more strongly than the left AC.

Adequate processing of both music and speech also requires auditory sensory memory, and auditory scene analysis/auditory stream segregation (see boxes “Auditory sensory memory” and “Gestalt formation” in **Figure 1**), particularly in noisy environments (which are more common than the quiet auditory environments typical for experiments on language or music perception)¹⁹. The processes mediating the identification of chords (e.g., whether a chord is a major or a minor chord, whether a chord is presented in root position or in inversion, etc.; see box “Analysis of intervals” in **Figure 1**) perhaps parallel those underlying the identification of word form (due to both words and chords having a stem, or “root,” from which different versions emerge); however, this remains to be specified.

The section on music-syntactic processing (see also box “Syntactic structure building” in **Figure 1**) presented evidence for interactions between the processing of music- and language-syntactic information at levels of morpho-syntactic processing, phrase-structure processing, and possibly word-category information (see the evidence for the interaction between the ERAN and LAN, and the provisional evidence for an interaction between ERAN and ELAN). Beyond these early interactions between music- and language-syntactic processes, it appears that cognitive and neural resources are also shared during later stages of syntactic integration (see box “Structural reanalysis and repair” in **Figure 1**). The notion of shared resources during these later stages is also referred to as SSIRH (Patel, 2003).

Both speech and music perception involve premotor coding [e.g., Liberman and Mattingly, 1985; Koelsch, 2009a, see also box “(premotor) Action” in **Figure 1**], and both music and language give rise to affective processes: With regard to language, the perception

of affective prosody (e.g., Ethofer et al., 2009; Wittfoth et al., 2010) as well as the affective contents of words (e.g., Herbert et al., 2009; Vö et al., 2009) can elicit emotional responses; the neural correlates of music-evoked emotions have been reviewed elsewhere (e.g., Koelsch, 2010).

Finally, this review illustrated that communication of meaning is not exclusively a linguistic domain, but that music can also convey meaningful information: The interpretation of extra-musical sign qualities of musical information can prime representations of meaningful concepts, and such priming can emerge from different cognitive levels (e.g., from auditory feature extraction, or from a more fine-grained analysis of intervals, see also **Figure 1**). Moreover, structural relations give rise to intra-musical meaning (suggested to be reflected in the N500).

Corresponding to these shared processes, there is considerable resemblance, overlap, and interaction between ERP components (and their neural generators) elicited during the perception of music or language: Processing of musical as well as of speech information evokes (a) FFRs originating from the auditory brainstem, (b) P1, N1, and P2 potentials originating from the AC, (c) MMN potentials originating from temporal and frontal cortical areas, (d) potentials of early syntactic processing that interact with each other (ERAN/LAN), and receive main contributions from the (inferior) pars opercularis (BA 44i), (e) potentials of syntactic (re)integration/reanalysis and repair (P600), and (f) N400 effects reflecting semantic processing (probably originating from posterior temporal and inferior frontal cortex).

These overlaps and shared cognitive/neural processes call for more integrative models that synthesize the cognitive processes underlying music and language perception into a common theoretical framework. For example, a comparison of the present model (**Figure 1**) with the language perception model by Friederici (2002) shows several analogies, such as “Feature Extraction”/“Primary acoustic analysis,” “Analysis of intervals (chords, melodies, time intervals)”/“Identification of word form,” “Syntactic structure building” (both models), “Reanalysis & repair” (both models), or “Meaning”/“Semantic relations.” Additional overlaps include WM for verbal and tonal information (as described in the previous section), and processing of rhythm in speech and music (which is not included in this review, but see, e.g., Grahn and Brett, 2007; Schmidt-Kassow and Kotz, 2009).

The illustrated overlaps between music and language also show that “language” and “music” are different aspects of the same domain, or two poles of a rather continuous dimension, rather than being two strictly separate domains (I refer to this continuum as the “music-language-continuum”). Spoken language has rhythm, melody, and timber, and the illusory transformation from speech to song described by Diana Deutsch (“sometimes they behave so strangely”; Deutsch et al., 2011) shows that humans can perceive spoken language also as song (although individuals often do not realize this in everyday life). Once an individual puts emphasis in his/her utterances, the speech becomes more song-like²⁰, and many art-forms, such as Rap-music or Recitatives, are both song and speech. In addition, as described in this review, music is often structured

¹⁹For a study showing topological shifts in the activation patterns of language perception due to speech degradation (noise-band vocoding) see Obleser et al. (2011).

²⁰Martin Luther King’s speeches are a nice example of how it is often difficult to say whether someone is singing or speaking.

according to a syntactic system, and during music listening, meaning emerges from the interpretation of musical information. Therefore, any clear-cut distinction between music and language (and thus also any pair of separate definitions for language and music) is likely to be inadequate, or incomplete, and a rather artificial construct²¹.

As mentioned already above, even individuals without formal musical training show sophisticated abilities with regard to the decoding of musical information, the acquisition of knowledge about musical syntax, the processing of musical information according to that knowledge, and the understanding of music. This finding supports the notion that musicality is a natural ability

of the human brain. Such musical abilities are important for the acquisition and the processing of language: infants acquire information about word and phrase boundaries (possibly even about word meaning) in part through different types of prosodic cues. Moreover, tonal languages rely on a meticulous decoding of pitch information, and both tonal and non-tonal languages require an accurate analysis of speech prosody to decode structure and meaning of speech. The assumption of an intimate connection between music and speech is corroborated by the reviewed findings of overlapping and shared neural resources for music and language processing in both adults and children. These findings suggest that the human brain, particularly at an early age, does not treat language and music as strictly separate domains, but rather treats language as a special case of music.

²¹Similar to the distinction between perception and action, or motor and sensory processes, which do not occur in isolation in a living individual.

REFERENCES

- Alain, C., Woods, D. L., and Knight, R. T. (1998). A distributed cortical network for auditory sensory memory in humans. *Brain Res.* 812, 23–37.
- Alho, K., Tervaniemi, M., Huottilainen, M., Lavikainen, J., Tiitinen, H., Ilmoniemi, R. J., Knuutila, J., and Näätänen, R. (1996). Processing of complex sounds in the human auditory cortex as revealed by magnetic brain responses. *Psychophysiology* 33, 369–375.
- Alpers, P. (1994). *What is Music? An Introduction to the Philosophy of Music*. Pennsylvania: State University Press.
- Amunts, K., Lenzen, M., Friederici, A., Schleicher, A., Morosan, P., Palomero-Gallagher, N., and Zilles, K. (2010). Broca's region: novel organizational principles and multiple receptor mapping. *PLoS Biol.* 8, e1000489. doi: 10.1371/journal.pbio.1000489
- Bangert, M., and Altenmüller, E. (2003). Mapping perception to action in piano practice: a longitudinal de- EEG study. *BMC Neurosci.* 4, 26. doi: 10.1186/1471-2202-4-26
- Baumeister, R., and Leary, M. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol. Bull.* 117, 497–497.
- Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature* 436, 1161–1165.
- Berti, S., and Schröger, E. (2003). Working memory controls involuntary attention switching: evidence from an auditory distraction paradigm. *Eur. J. Neurosci.* 17, 1119–1122.
- 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.
- Besson, M., Faita, F., Peretz, I., Bonnel, A. M., and Requin, J. (1998). Singing in the brain: independence of lyrics and tunes. *Psychol. Sci.* 9, 494–498.
- Besson, M., and Macar, F. (1986). Visual and auditory event-related potentials elicited by linguistic and non-linguistic incongruities. *Neurosci. Lett.* 63, 109–114.
- Besson, M., and Schön, D. (2001). "Comparison between language and music," in *The Biological Foundations of Music*, Vol. 930, eds R. J. Zatorre and I. Peretz (New York: The New York Academy of Sciences), 232–258.
- Bigand, E., Parncutt, R., and Lerdahl, J. (1996). Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Percept. Psychophys.* 58, 125–141.
- Block, N. (2005). Two neural correlates of consciousness. *Trends Cogn. Sci. (Regul. Ed.)* 9, 46–52.
- Blood, A., and Zatorre, R. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11818.
- Brattico, E., Tervaniemi, M., Naatanen, R., and Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. *Brain Res.* 1117, 162–174.
- Bregman, A. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: The MIT Press.
- Budd, M. (1996). *Values of Art*. London: Penguin Books.
- Callan, D., Tsytarev, V., Hanakawa, T., Callan, A., Katsuhara, M., Fukuyama, H., and Turner, R. (2006). Song and speech: brain regions involved with perception and covert production. *Neuroimage* 31, 1327–1342.
- Cardoso, S., Coimbra, N., and Brandão, M. (1994). Defensive reactions evoked by activation of NMDA receptors in distinct sites of the inferior colliculus. *Behav. Brain Res.* 63, 17–24.
- Carlyon, R. (2004). How the brain separates sounds. *Trends Cogn. Sci. (Regul. Ed.)* 8, 465–471.
- Conard, N., Malina, M., and Münzel, S. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature* 460, 737–740.
- Cook, N. (1987). The perception of large-scale tonal closure. *Music Percept.* 197–205.
- Cross, I. (2008). The evolutionary nature of musical meaning. *Music. Sci.* 179–200.
- Cross, I., and Morley, I. (2008). "The evolution of music: theories, definitions and the nature of the evidence," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 61–82.
- Daltrozzo, J., and Schön, D. (2009a). Conceptual processing in music as revealed by N400 effects on words and musical targets. *J. Cogn. Neurosci.* 21, 1882–1892.
- Daltrozzo, J., and Schön, D. (2009b). Is conceptual processing in music automatic? An electrophysiological approach. *Brain Res.* 1270, 88–94.
- Darwin, C. (1997). Auditory grouping. *Trends Cogn. Sci. (Regul. Ed.)* 1, 327–333.
- Darwin, C. (2008). Listening to speech in the presence of other sounds. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1011.
- Davies, S. (1994). *Musical Meaning and Expression*. Ithaca, NY: Cornell University Press.
- Deouell, L. (2007). The frontal generator of the mismatch negativity revisited. *J. Psychophysiol.* 21, 188.
- Deutsch, D., Henthorn, T., and Lapidis, R. (2011). Illusory transformation from speech to song. *J. Acoust. Soc. Am.* 129, 2245–2252.
- Di Pietro, M., Laganaro, M., Leemann, B., and Schnider, A. (2004). Receptive amusia: temporal auditory processing deficit in a professional musician following a left temporo-parietal lesion. *Neuropsychologia* 42, 868–877.
- Doeller, C., Opitz, B., Mecklinger, A., Krick, C., Reith, W., and Schröger, E. (2003). Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *Neuroimage* 20, 1270–1282.
- Donchin, E., and Coles, M. G. H. (1998). Context updating and the p300. *Behav. Brain Sci.* 21, 152.
- Drost, U., Rieger, M., Brass, M., Gunter, T., and Prinz, W. (2005a). Action-effect coupling in pianists. *Psychol. Res.* 69, 233–241.
- Drost, U., Rieger, M., Brass, M., Gunter, T., and Prinz, W. (2005b). When hearing turns into playing: movement induction by auditory stimuli in pianists. *Q. J. Exp. Psychol. A* 58, 1376–1389.
- Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., and Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *J. Cogn. Neurosci.* 21, 1255–1268.
- Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A., Pozzo, T., Calzolari, F., Granieri, E., and Fadiga, L. (2009). Encoding of human action in Broca's area. *Brain* 132, 1980.
- 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.
- Fitch, W. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215.
- Fitch, W., and Hauser, M. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377.
- Fodor, J., Mann, V., and Samuel, A. (1991). "Panel discussion: the mod-

- ularity of speech and language,” in *Modularity and the Motor Theory of Speech Perception: Proceedings of a Conference to Honor Alvin M. Liberman*, eds Ignatius G. Mattingly and Michael Studdert-Kennedy, (Hillsdale, NJ: Lawrence Erlbaum Associates), 359.
- Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci. (Regul. Ed.)* 6, 78–84.
- Friederici, A. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends Cogn. Sci. (Regul. Ed.)* 8, 245–247.
- Friederici, A., Bahlmann, J., Heim, S., Schubotz, R., and Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2458.
- Friedrich, R., and Friederici, A. (2009). Mathematical logic in the human brain: syntax. *PLoS ONE* 4, e5599. doi: 10.1371/journal.pone.0005599
- Fujioka, T., Trainor, L., Ross, B., Kakigi, R., and Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *J. Cogn. Neurosci.* 16, 1010–1021.
- Fujioka, T., Trainor, L., Ross, B., Kakigi, R., and Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J. Cogn. Neurosci.* 17, 1578–1592.
- Geisler, C. (1998). *From Sound to Synapse: Physiology of the Mammalian Ear*. New York, NY: Oxford University Press.
- Giard, M., Perrin, F., and Pernier, J. (1990). Brain generators implicated in processing of auditory stimulus deviance: a topographic ERP study. *Psychophysiology* 27, 627–640.
- Grahn, J., and Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* 19, 893–906.
- Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2007a). Emotions over time: synchronicity and development of subjective, physiological, and facial affective reactions of music. *Emotion* 7, 774–788.
- Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2007b). Listening to music as a re-creative process: physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Percept.* 24, 297–314.
- Grieser-Painter, J., and Koelsch, S. (2011). Can out-of-context musical sounds convey meaning? an ERP study on the processing of meaning in music. *Psychophysiology* 48, 645–655.
- Griffiths, T., and Warren, J. (2002). The planum temporale as a computational hub. *Trends Neurosci.* 25, 348–353.
- Griffiths, T., and Warren, J. (2004). What is an auditory object? *Nat. Rev. Neurosci.* 5, 887–892.
- Groussard, M., Viader, F., Hubert, V., Landeau, B., Abbas, A., Desgranges, B., Eustache, F., and Platel, H. (2010). Musical and verbal semantic memory: two distinct neural networks? *Neuroimage* 49, 2764–2773.
- Groussard, M., Viader, F., Landeau, B., Desgranges, B., Eustache, F., and Platel, H. (2009). Neural correlates underlying musical semantic memory. *Ann. N. Y. Acad. Sci.* 1169, 278–281.
- Gunter, T., Friederici, A., and Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J. Cogn. Neurosci.* 12, 556–568.
- Hackett, T. A., and Kaas, J. (2004). “Auditory cortex in primates: functional subdivisions and processing streams,” in *The Cognitive Neurosciences*, ed. M. S. Gazzaniga (Cambridge: MIT Press), 215–232.
- Haueisen, J., and Knösche, T. (2001). Involuntary motor activity in pianists evoked by music perception. *J. Cogn. Neurosci.* 13, 786–792.
- Herbert, C., Ethofer, T., Anders, S., Junghofer, M., Wildgruber, D., Grodd, W., and Kissler, J. (2009). Amygdala activation during reading of emotional adjectives – an advantage for pleasant content. *Soc. Cogn. Affect. Neurosci.* 4, 35.
- Herrojo-Ruiz, M., Jabusch, H., and Altenmüller, E. (2009). Detecting wrong notes in advance: neuronal correlates of error monitoring in pianists. *Cereb. Cortex* 19, 2625.
- Herrojo-Ruiz, M., Strübing, F., Jabusch, H. C., and Altenmüller, E. (2010). EEG oscillatory patterns are associated with error prediction during music performance and are altered in musician’s dystonia. *Neuroimage* 55, 1791–1803.
- 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.
- Hucklebridge, F., Lambert, S., Clow, A., Warburton, D., Evans, P., and Sherwood, N. (2000). Modulation of secretory immunoglobulin A in saliva; response to manipulation of mood. *Biol. Psychol.* 53, 25–35.
- Huffman, R., and Henson, O. (1990). The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus. *Brain Res. Rev.* 15, 295–323.
- Hyde, K., Peretz, I., and Zatorre, R. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia* 46, 632–639.
- Janata, P., Birk, J., Van Horn, J., Leman, M., Tillmann, B., and Bharucha, J. (2002a). The cortical topography of tonal structures underlying Western music. *Science* 298, 2167.
- Janata, P., Tillmann, B., and Bharucha, J. (2002b). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cogn. Affect. Behav. Neurosci.* 2, 121.
- Janata, P., and Grafton, S. T. (2003). Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nat. Neurosci.* 6, 682–687.
- Jäncke, L. (2008). Music, memory and emotion. *J. Biol.* 7, 21.
- Johnson, K., Nicol, T., Zecker, S., and Kraus, N. (2008). Developmental plasticity in the human auditory brainstem. *J. Neurosci.* 28, 4000.
- Johnsrude, I., Penhune, V., and Zatorre, R. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* 123, 155.
- Jusczyk, P. (1999). How infants begin to extract words from speech. *Trends Cogn. Sci. (Regul. Ed.)* 3, 323–328.
- Juslin, P., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814.
- Kaas, J., and Hackett, T. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11793.
- Kaas, J., Hackett, T., and Tramo, M. (1999). Auditory processing in primate cerebral cortex. *Curr. Opin. Neurobiol.* 9, 164–170.
- Kamiyama, K., Katahira, K., Abba, D., Hori, K., and Okanoya, K. (2010). Music playing and memory trace: evidence from event-related potentials. *Neurosci. Res.* 67, 334–340.
- Karbusicky, V. (1986). *Grundriß der musikalischen Semantik*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Katahira, K., Abba, D., Masuda, S., and Okanoya, K. (2008). Feedback-based error monitoring processes during musical performance: an ERP study. *Neurosci. Res.* 61, 120–128.
- Khalfa, S., Isabelle, P., Jean-Pierre, B., and Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neurosci. Lett.* 328, 145–149.
- Kirschner, S., and Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child. Psychol.* 102, 299–314.
- Knösche, T., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O., and Friederici, A. D. (2005). Perception of phrase structure in music. *Hum. Brain Mapp.* 24, 259–273.
- Koch, M., Lingenhöhl, K., and Pilz, P. (1992). Loss of the acoustic startle response following neurotoxic lesions of the caudal pontine reticular formation: possible role of giant neurons. *Neuroscience* 49, 617–625.
- Koechlin, E., and Jubault, T. (2006). Broca’s area and the hierarchical organization of human behavior. *Neuron* 50, 963–974.
- Koelsch, S. (2000). *Brain and Music – A Contribution to the Investigation of Central Auditory Processing with a New Electrophysiological Approach*. Leipzig: Risse.
- Koelsch, S. (2004). *Spatio-temporal Aspects of Processing Syntax and Semantics in Music*. Habilitation thesis, University of Leipzig.
- Koelsch, S. (2009a). Music-syntactic processing and auditory memory: similarities and differences between ERAN and MMN. *Psychophysiology* 46, 179–190.
- Koelsch, S. (2009b). A neuroscientific perspective on music therapy. *Ann. N. Y. Acad. Sci.* 1169, 374–384.
- Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends Cogn. Sci. (Regul. Ed.)* 14, 131–137.
- Koelsch, S. (2011). Towards a neural basis of processing musical semantics. *Phys. Life Rev.* (in press).
- Koelsch, S., Fritz, T., Schulze, K., Alsop, D., and Schlaug, G. (2005a). Adults and children processing music: an fMRI study. *Neuroimage* 25, 1068–1076.
- Koelsch, S., Gunter, T., Wittfoth, M., and Sammler, D. (2005b). Interaction between syntax processing in language and in music: an ERP study. *J. Cogn. Neurosci.* 17, 1565–1577.
- Koelsch, S., Gunter, T. C., Cramon, D. Y. von, Zysset, S., Lohmann, G., and Friederici, A. D. (2002). Bach speaks: a cortical “language-network” serves the processing of music. *Neuroimage* 17, 956–966.
- Koelsch, S., and Jentschke, S. (2010). Differences in electric brain responses to melodies and chords. *J. Cogn. Neurosci.* 22, 2251–2262.
- Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T. C., and Friederici, A. D. (2004). Music, language, and meaning: brain signatures of semantic processing. *Nat. Neurosci.* 7, 302–307.
- Koelsch, S., and Mulder, J. (2002). Electric brain responses to inappropriate harmonies during listening to expressive music. *Clin. Neurophysiol.* 113, 862–869.
- Koelsch, S., Offermanns, K., and Franzke, P. (2010). Music in the treatment of affective disorders: an exploratory

- investigation of a new method for music-therapeutic research. *Music Percept.* 27, 307–316.
- Koelsch, S., Schröger, E., and Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport* 10, 1309.
- 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.
- Koelsch, S., and Siebel, W. (2005). Towards a neural basis of music perception. *Trends Cogn. Sci. (Regul. Ed.)* 9, 578–584.
- Koopman, C., and Davies, S. (2001). Musical meaning in a broader perspective. *J. Aesthet. Art Criticism* 59, 261–273.
- Korzyukov, O., Winkler, I., Gumenyuk, V., and Alho, K. (2003). Processing abstract auditory features in the human auditory cortex. *Neuroimage* 20, 2245–2258.
- Kreutz, G., Bongard, S., Rohrmann, S., Hodapp, V., and Grebe, D. (2004). Effects of choir singing or listening on secretory immunoglobulin A, cortisol, and emotional state. *J. Behav. Med.* 27, 623–635.
- Lahav, A., Saltzman, E., and Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J. Neurosci.*, 27, 308–314.
- Lamprea, M., Cardenas, F., Vianna, D., Castilho, V., Cruz-Morales, S., and Brandão, M. (2002). The distribution of fos immunoreactivity in rat brain following freezing and escape responses elicited by electrical stimulation of the inferior colliculus. *Brain Res.* 950, 186–194.
- Langner, G., and Ochse, M. (2006). The neural basis of pitch and harmony in the auditory system. *Music. Sci.* 10, 185.
- Lau, E., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the n400. *Nat. Rev. Neurosci.* 9, 920–933.
- LeDoux, J. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23, 155–184.
- Lerdahl, F. (2001). *Tonal Pitch Space*. New York, NY: Oxford University Press.
- Lerdahl, F., and Jackendoff, R. (1999). *A Generative Theory of Music*. Cambridge: MIT.
- Lerdahl, F., and Krumhansl, C. (2007). Modeling tonal tension. *Music Percept.* 24, 329–366.
- Levitt, P., and Moore, R. (1979). Origin and organization of brainstem catecholamine innervation in the rat. *J. Comp. Neurol.* 186, 505–528.
- Liberman, A., and Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36.
- Liebenthal, E., Ellingson, M., Spanaki, M., Prieto, T., Ropella, K., and Binder, J. (2003). Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. *Neuroimage* 19, 1395–1404.
- Liegeois-Chauvel, C., Peretz, I., Babiak, M., Laguiton, V., and Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain* 121, 1853–1867.
- Longoni, F., Grande, M., Hendrich, V., Kastrau, F., and Huber, W. (2005). An fMRI study on conceptual, grammatical, and morpho-phonological processing. *Brain Cogn.* 57, 131–134.
- Lundqvist, L., Carlsson, F., Hilmersson, P., and Juslin, P. (2009). Emotional responses to music: experience, expression, and physiology. *Psychol. Music* 37, 61.
- Maess, B., Jacobsen, T., Schröger, E., and Friederici, A. (2007). Localizing pre-attentive auditory memory-based comparison: magnetic mismatch negativity to pitch change. *Neuroimage* 37, 561–571.
- 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.
- Maidhof, C., and Koelsch, S. (2011). Effects of selective attention on syntax processing in music and language. *J. Cogn. Neurosci.* [Epub ahead of print].
- Maidhof, C., Rieger, M., Prinz, W., and Koelsch, S. (2009). Nobody is perfect: ERP effects prior to performance errors in musicians indicate fast monitoring processes. *PLoS ONE* 4, e5032. doi: 10.1371/journal.pone.0005032
- Maidhof, C., Vavatzanidis, N., Prinz, W., Rieger, M., and Koelsch, S. (2010). Processing expectancy violations during music performance and perception: an ERP study. *J. Cogn. Neurosci.* 22, 2401–2413.
- Makuuchi, M., Bahlmann, J., Anwander, A., and Friederici, A. (2009). Segregating the core computational faculty of human language from working memory. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8362.
- McCraty, R., Atkinson, M., Rein, G., and Watkins, A. (1996). Music enhances the effect of positive emotional states on salivary IgA. *Stress Med.* 12, 167–175.
- Menning, H., Roberts, L., and Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *Neuroreport* 11, 817.
- Meyer, L. (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.
- Meyer, M., Alter, K., Friederici, A., Lohmann, G., and Cramon, D. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum. Brain Mapp.* 17, 73–88.
- Meyer, M., Steinhauer, K., Alter, K., Friederici, A., and Cramon, D. (2004). Brain activity varies with modulation of dynamic pitch variances in sentence melody. *Brain Lang.* 89, 277–289.
- Miranda, R., and Ullman, M. (2007). Double dissociation between rules and memory in music: an event-related potential study. *Neuroimage* 38, 331–345.
- Molholm, S., Martinez, A., Ritter, W., Javitt, D., and Foxe, J. (2005). The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. *Cereb. Cortex* 15, 545.
- Moon, C., Cooper, R., and Fifer, W. (1993). Two-day-olds prefer their native language. *Infant Behav. Dev.* 16, 495–500.
- Moore, B. (2008). *An Introduction to the Psychology of Hearing*, 5th Edn. Bingley: Emerald.
- Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15894.
- Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., and Winkler, I. (2001). Primitive intelligence in the auditory cortex. *Trends Neurosci.* 24, 283–288.
- Nan, Y., Knösche, T., and Friederici, A. (2006). The perception of musical phrase structure: a cross-cultural ERP study. *Brain Res.* 1094, 179–191.
- Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Curr. Opin. Neurobiol.* 14, 474–480.
- Neuhaus, C., Knösche, T., and Friederici, A. (2006). Effects of musical expertise and boundary markers on phrase perception in music. *J. Cogn. Neurosci.* 18, 472–493.
- Obleser, J., Meyer, L., and Friederici, A. (2011). Dynamic assignment of neural resources in auditory comprehension of complex sentences. *Neuroimage* 56, 2310–2320.
- Öngür, D., and Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* 10, 206.
- Opitz, B., and Kotz, S. (2011). Ventral premotor cortex lesions disrupt learning of sequential grammatical structures. *Cortex*. (in press).
- Opitz, B., Rinne, T., Mecklinger, A., Cramon, D., and Schröger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *Neuroimage* 15, 167–174.
- Orini, M., Bailón, R., Enk, R., Koelsch, S., Mainardi, L., and Laguna, P. (2010). A method for continuously assessing the autonomic response to music-induced emotions through HRV analysis. *Med. Biol. Eng. Comput.* 48, 423–433.
- Overy, K., and Molnar-Szakacs, I. (2009). Being together in time: musical experience and the mirror neuron system. *Music Percept.* 26, 489–504.
- Paller, K. A., McCarthy, G., and Wood, C. C. (1992). Event-related potentials elicited by deviant endings to melodies. *Psychophysiology* 29, 202–206.
- Panksepp, J., and Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behav. Processes* 60, 133–155.
- Pantev, C., Roberts, L. E., Schulz, M., Engelen, A., and Ross, B. (2001). Timbre-specific enhancement of auditory cortical representation in musicians. *Neuroreport* 12, 169–174.
- Parsons, L. (2001). Exploring the functional neuroanatomy of music performance, perception, and comprehension. *Ann. N. Y. Acad. Sci.* 930, 211–231.
- Patel, A. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681.
- Patel, A. (2008). *Music, Language, and the Brain*. New York, NY: Oxford University Press.
- Patel, A., and Balaban, E. (2001). Human pitch perception is reflected in the timing of stimulus-related cortical activity. *Nat. Neurosci.* 4, 839–844.
- Patel, A., 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.
- Patel, A., Iversen, J., Bregman, M., and Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr. Biol.* 19, 827–830.
- Patel, A., Iversen, J., Wassenaar, M., and Hagoort, P. (2008). Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology* 22, 776–789.
- Patterson, R., Uppenkamp, S., Johnsrude, I., and Griffiths, T. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36, 767–776.

- Perani, D., Saccuman, M., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., Baldoli, C., and Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4758.
- Peretz, I., Brattico, E., Järvenpää, M., and Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain* 132, 1277.
- Peretz, I., and Coltheart, M. (2003). Modularity of music processing. *Nat. Neurosci.* 6, 688–691.
- Peretz, I., and Zatorre, R. (2005). Brain organization for music processing. *Annu. Rev. Psychol.* 56, 89–114.
- Petkov, C., Kayser, C., Augath, M., and Logothetis, N. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol.* 4, e215. doi: 10.1371/journal.pbio.0040215
- Pickles, J. (2008). *An Introduction to the Physiology of Hearing*, 3rd Edn. Bingley: Emerald.
- Piston, W. (1948/1987). *Harmony*. New York: Norton.
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148.
- Poulin-Charronnat, B., Bigand, E., and Koelsch, S. (2006). Processing of musical syntax tonic versus subdominant: an event-related potential study. *J. Cogn. Neurosci.* 18, 1545–1554.
- Quiroga Murcia, C., Bongard, S., and Kreutz, G. (2009). Emotional and neurohumoral responses to dancing tango argentino. *Music Med.* 1, 14.
- Rameau, J.-P. (1722). *Traité de l'harmonie réduite à ses principes naturels*. Paris: J. B. C. Ballard.
- Rammsayer, T., and Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Percept.* 24, 37–48.
- Riemann, H. (1877/1971). *Musikalische syntaxis: Grundriss einer harmonischen satzbildungslehre*. Niederwalluf: Sändig.
- Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. *Neuron* 35, 395–405.
- Rinne, T., Degerman, A., and Alho, K. (2005). Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: an fMRI study. *Neuroimage* 26, 66–72.
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.
- Rohrmeier, M. (2007). “A generative grammar approach to diatonic harmonic structure,” in *Proceedings of the 4th Sound and Music Computing Conference*, Lefkada, 97–100.
- Rohrmeier, M. (2011). A generative grammar approach to tonal harmony. *J. Math. Music* 5.
- Rüsseler, J., Altenmüller, E., Nager, W., Kohlmetz, C., and Münte, T. (2001). Event-related brain potentials to sound omissions differ in musicians and non-musicians. *Neurosci. Lett.* 308, 33–36.
- Scherer, K. (2005). What are emotions? And how can they be measured? *Soc. Sci. Inform.* 44, 695.
- Schmidt-Kassow, M., and Kotz, S. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *J. Cogn. Neurosci.* 21, 1693–1708.
- Schönberg, A. (1969). *Structural functions of harmony*. New York: Norton.
- Schonwiesner, 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.
- Schulze, K., Mueller, K., and Koelsch, S. (2011a). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *Eur. J. Neurosci.* 33, 189–196.
- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., and Koelsch, S. (2011b). Neuroarchitecture of verbal and tonal working memory in non-musicians and musicians. *Hum. Brain Mapp.* 32, 771–783.
- Scott, S. (2005). Auditory processing—speech, space and auditory objects. *Curr. Opin. Neurobiol.* 15, 197–201.
- Shinn-Cunningham, B. (2008). Object-based auditory and visual attention. *Trends Cogn. Sci. (Regul. Ed.)* 12, 182–186.
- Sinex, D., Guzik, H., Li, H., and Henderson Sabes, J. (2003). Responses of auditory nerve fibers to harmonic and mistuned complex tones. *Hear. Res.* 182, 130–139.
- Slevc, L., Rosenberg, J., and Patel, A. (2009). Making psycholinguistics musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychon. Bull. Rev.* 16, 374.
- Sloboda, J. A. (1991). Music structure and emotional response: some empirical findings. *Psychol. Music* 19, 110–120.
- Song, J., Skoe, E., Wong, P., and Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *J. Cogn. Neurosci.* 20, 1892–1902.
- 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.
- Steinbeis, N., and Koelsch, S. (2008c). Understanding the intentions behind man-made products elicits neural activity in areas dedicated to mental state attribution. *Cereb. Cortex* 19, 619–623.
- Steinbeis, N., and Koelsch, S. (2011). Affective priming effects of musical sounds on the processing of word meaning. *J. Cogn. Neurosci.* 23, 604–621.
- Steinhauer, K., Alter, K., and Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat. Neurosci.* 2, 191–196.
- Strait, D., Kraus, N., Skoe, E., and Ashley, R. (2009). Musical experience and neural efficiency – effects of training on subcortical processing of vocal expressions of emotion. *Eur. J. Neurosci.* 29, 661–668.
- Sussman, E. (2007). A new view on the MMN and attention debate: the role of context in processing auditory events. *J. Psychophysiol.* 21, 164–175.
- Tervaniemi, M. (2009). Musicians – same or different? *Ann. N. Y. Acad. Sci.* 1169, 151–156.
- Tervaniemi, M., Castaneda, A., Knoll, M., and Uther, M. (2006a). Sound processing in amateur musicians and nonmusicians: event-related potential and behavioral indices. *Neuroreport* 17, 1225.
- Tervaniemi, M., Szameitat, A., Kruck, S., Schroger, E., Alter, K., De Baene, W., and Friederici, A. D. (2006b). From air oscillations to music and speech: functional magnetic resonance imaging evidence for fine-tuned neural networks in audition. *J. Neurosci.* 26, 8647.
- Tervaniemi, M., and Huotilainen, M. (2003). The promises of change-related brain potentials in cognitive neuroscience of music. *Ann. N. Y. Acad. Sci.* 999, 29–39.
- Tervaniemi, M., Ilvonen, T., Karma, K., Alho, K., and Näätänen, R. (1997). The musical brain: brain waves reveal the neurophysiological basis of musicality in human subjects. *Neurosci. Lett.* 226, 1–4.
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., and Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Exp. Brain Res.* 161, 1–10.
- Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., and Friederici, A. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *Eur. J. Neurosci.* 30, 1636–1642.
- 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.
- Tervaniemi, M., Medvedev, S., Alho, K., Pakhomov, S., Roudas, M., Zuijten, T., and Näätänen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Hum. Brain Mapp.* 10, 74–79.
- Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R., and Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learn. Mem.* 8, 295.
- Tillmann, B., Bharucha, J., and Bigand, E. (2000). Implicit learning of tonality: a self-organized approach. *Psychol. Rev.* 107, 885–913.
- Tillmann, B., Janata, P., and Bharucha, J. (2003). Activation of the inferior frontal cortex in musical priming. *Brain Res. Cogn. Brain Res.* 16, 145–161.
- Tramo, M., Shah, G., and Braid, L. (2002). Functional role of auditory cortex in frequency processing and pitch perception. *J. Neurophysiol.* 87, 122.
- Trehub, S. (2003). The developmental origins of musicality. *Nat. Neurosci.* 6, 669–673.
- Van Petten, C., and Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Mem. Cognit.* 18, 380–393.
- Verleger, R. (1990). P3-evoking wrong notes: unexpected, awaited, or arousing? *Int. J. Neurosci.* 55, 171–179.
- Võ, M., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M., and Jacobs, A. (2009). The Berlin affective word list reloaded (BAWL-R). *Behav. Res. Methods* 41, 534–538.
- Wallin, N. L., Merker, B., and Brown, S. (eds). (2000). *The Origins of Music*. Cambridge, MA: MIT Press.
- Warren, J. D., Uppenkamp, S., Patterson, R. D., and Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10038–10042.
- Watanabe, T., Yagishita, S., and Kikyo, H. (2008). Memory of music: roles

- of right hippocampus and left inferior frontal gyrus. *Neuroimage* 39, 483–491.
- Whitfield, I. (1980). Auditory cortex and the pitch of complex tones. *J. Acoust. Soc. Am.* 67, 644.
- Winkler, I. (2007). Interpreting the mismatch negativity. *J. Psychophysiol.* 21, 147–163.
- Winkler, I., Denham, S., and Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci. (Regul. Ed.)* 13, 532–540.
- Wittfoth, M., Schröder, C., Schardt, D., Dengler, R., Heinze, H., and Kotz, S. (2010). On emotional conflict: interference resolution of happy and angry prosody reveals valence-specific effects. *Cereb. Cortex* 20, 383.
- Wong, P., Skoe, E., Russo, N., Dees, T., and Kraus, N. (2007). Musical experience shapes human brain-stem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.
- Zatorre, R. (1988). Pitch perception of complex tones and human temporal-lobe function. *J. Acoust. Soc. Am.* 84, 566–572.
- Zatorre, R. (2001). Neural specializations for tonal processing. *Ann. N. Y. Acad. Sci.* 930, 193–210.
- Zatorre, R., Belin, P., and Penhune, V. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci. (Regul. Ed.)* 6, 37–46.
- Zatorre, R., Evans, A., and Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* 14, 1908.
- Zuijlen, T., Sussman, E., Winkler, I., Näätänen, R., and Tervaniemi, M. (2004). Grouping of sequential sounds—an event-related potential study comparing musicians and nonmusicians. *J. Cogn. Neurosci.* 16, 331–338.
- Zuijlen, T., Sussman, E., Winkler, I., Näätänen, R., and Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity—a mismatch negativity study comparing musicians and non-musicians. *Brain Res. Cogn. Brain Res.* 23, 270–276.
- 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.

Received: 13 March 2011; paper pending published: 14 April 2011; accepted: 13 May 2011; published online: 09 June 2011.

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

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

Copyright © 2011 Koelsch. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



The interplay between musical and linguistic aptitudes: a review

Riia Milovanov^{1,2,3,4*} and Mari Tervaniemi^{3,5,6}

¹ Department of Music, University of Jyväskylä, Jyväskylä, Finland

² Department of Teacher Training, University of Turku, Turku, Finland

³ Finnish Center of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Jyväskylä, Finland

⁴ Centre for Cognitive Neuroscience, University of Turku, Turku, Finland

⁵ Cognitive Brain Research Unit, Cognitive Science, Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland

⁶ Department of Psychology, University of Jyväskylä, Jyväskylä, Finland

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Mireille Besson, CNRS, Institut de Neurosciences Cognitives de la Méditerranée, France

Thomas Jacobsen, Helmut-Schmidt-University, Germany

*Correspondence:

Riia Milovanov, Department of Teacher Training, University of Turku, Assistentinkatu 5, Turku 20014, Finland.
e-mail: riia.milovanov@utu.fi

According to prevailing views, brain organization is modulated by practice, e.g., during musical or linguistic training. Most recent results, using both neuropsychological tests and brain measures, revealed an intriguing connection between musical aptitude and second language linguistic abilities. A significant relationship between higher musical aptitude, better second language pronunciation skills, accurate chord discrimination ability, and more prominent sound-change-evoked brain activation in response to musical stimuli was found. Moreover, regular music practice may also have a modulatory effect on the brain's linguistic organization and alter hemispheric functioning in those who have regularly practised music for years. These findings, together with their implications, will be introduced and elaborated in our review.

Keywords: musical aptitude, linguistic abilities, seashore test, event-related potentials

INTRODUCTION

Musical training can modulate brain functions in a holistic manner, causing even structural changes in the brain morphology. When comparing adult musicians and non-musicians, differences have been found in the auditory, motor, somatosensory, and visuospatial areas at the cortical level as well as in the cerebellum (Elbert et al., 1995; Pantev et al., 1998; Schneider et al., 2002; Gaser and Schlaug, 2003; Hutchinson et al., 2003; for a review see Jäncke, 2009). In addition to this, musicians seem to possess a larger anterior part of the corpus callosum than non-musicians (Schlaug et al., 1995). Also, several other neural projections appear to be structurally modulated by increased myelinization in musicians (Bengtsson et al., 2005).

As Magne et al. (2006) have pointed out, such structural differences naturally have functional implications. However, it is far from evident how or why these anatomical changes occur. For instance, the role of the age at which practising a musical instrument was initiated and its effects on brain plasticity has not been widely investigated; however, pioneering evidence points out that the younger the subjects begin playing an instrument, the greater are the neuroplastic effects on the brain (Schlaug et al., 1995; Pantev et al., 1998). Correspondingly, it is not yet certain how much the specific demands set up by a given musical genre or instrument can modulate the underlying brain functions (Tervaniemi, 2009).

Neurofunctional accounts of music listening indicate the contribution of a network consisting of fronto-temporal and temporal areas in both the right and left hemispheres. With linguistic functions, the same areas have been shown to be activated with comprehension and production of language. As in music practising,

the starting age seems to play a crucial role in foreign language acquisition skills. It is also recognized in foreign language learning that the pronunciation skills are more fluent and better when the second language has been started prior puberty (Johnson and Newport, 1989).

A number of studies indicate an overlap of the behavioral and neural resources between language and music. However, data indicating the lateralization of speech and music functions in the brain hemispheres (and thus differential neural architectures for speech and music) have also been obtained. According to those findings, specific neurocognitive functions intrinsic to music are governed by the right hemisphere and linguistic functions by the left hemisphere (Zatorre et al., 2002; Tervaniemi and Hugdahl, 2003).

Nevertheless, the conceptualization of hemispheric specialization is not straightforward, and the involvement of other higher-order processes, such as memory functions, should also be considered. Any domain-specific process (such as encoding of temporal acoustic properties of speech or a music sound, detecting pitch differences, etc.) also includes the involvement of both long-term and short-term memory functions and, naturally also working memory functions. For instance, linguistic status, context effects, and learning as well as attention can modulate early sensory cortices' functioning, which can therefore effect early processing via top-down effects. In other words, top-down executive processes within auditory pathways may have an important role in explaining the perception processes of speech and music stimuli (Zatorre and Gandour, 2007). Additionally, individual variations must also be borne in mind when discussing brain functioning.

For instance, different learning strategies could possibly reflect differential involvement of brain organization and may result in different neural representation of the material to be learnt. To sum up, learning and processing musical and linguistic material is composed of different neurofunctional levels with individual variations, ranged from higher to lower order, and each associated with different levels of neural organization.

Only a few studies examine the possible effects of linguistic skills on improving musical skills. It has been shown, though, that speaking Mandarin (which is one of the tonal languages) can help an individual acquire absolute pitch (Deutsch et al., 2006). Very recently, Elmer et al. (2011) asked whether long-term language training influences the functional architecture that supports the discrimination of non-verbal stimulus attributes, and whether such a functional reorganization favors the behavioral performance. They found evidence for a modulation of fronto-parietal brain regions as a function of language expertise and, further, that language training modulates brain activity in those regions involved in the top-down regulation of auditory functions.

In parallel, there is an increasing body of evidence indicating that practising music affects language skills at both the cognitive and neural levels. Magne et al. (2006) tested the hypothesis that musical training facilitates pitch processing not only in music but also in language. Their results show that musician children detected incongruities in both music and language better than non-musician children did. The differences between the two groups of participants were also seen in auditory cortex functioning: early negative components in music and late positive components in language were elicited in musician children, while no such components were present with the non-musician children. Marques et al. (2007) examined further the effects of musical expertise on pitch perception in a foreign language unknown to the participants. The results showed that adult musicians perceived pitch variations better than non-musicians and that a late positivity developed around 300 ms earlier in musicians than in non-musicians. Moreno et al. (2009) investigated the effect of musical training on linguistic pitch processing and reading skills. The results showed that musical training not only improved pitch processing in speech but also the reading of irregular words.

The roles of musical aptitude and musical training have recently gained greater attention as a contributing factor in the individual differences in language learning in general (Anvari et al., 2002; Slevc and Miyake, 2006; Tallal and Gaab, 2006; Besson et al., 2007; Marques et al., 2007). Recent results, using both neuropsychological tests and brain measures, revealed an intriguing relationship between musical aptitude and linguistic abilities. These data, acquired in the context of the doctoral thesis project of the first author (Milovanov, 2009), will be introduced and elaborated below. In sum, it was found that individuals with more musical aptitude were better equipped to discriminate phonemic minimal pair contrasts and pronounce second language (English) phonemes than individuals with less musical aptitude. Moreover, the more musical individuals showed enhanced activation in the auditory cortex, reflected by the mismatch negativity (MMN) when compared to less musical individuals.

INTERPLAY BETWEEN MUSICAL APTITUDE AND SECOND LANGUAGE PRONUNCIATION SKILLS: AN EMPIRICAL ENDEAVOR

The relationship between musical aptitude and linguistic abilities, particularly in terms of second language pronunciation skills and phoneme discrimination skills, was examined in a large project (Milovanov et al., 2008, 2009, 2010; Milovanov, 2009). It consisted of both behavioral and brain measures of Finnish-speaking children (age range 10–12 years) or young adults (age range 20–29 years). Based on prior behavioral investigations (Milovanov et al., 2004), the hypothesis was that musical aptitude and linguistic skills would interact both on the attentive and preattentive levels of neurocognitive processing.

Linguistic capacities were determined by Wechsler (1999) or Wechsler (1992), and a pronunciation test, in which English phonemes that are typically difficult for Finnish speakers were read onto a minidisc player after a native speaker model (Milovanov et al., 2008, 2010). A discrimination task of phonemic minimal pairs was also conducted. Moreover, an analogous discrimination task of musical minimal pairs, namely chords that differ slightly in pitch was executed. The musical aptitude test used in our experiments measures the accuracy and threshold of auditory discrimination (Seashore et al., 1960a,b). The test consists of Pitch, Loudness, Rhythm, Time, Timbre, and Tonal memory tasks.

The laterality effects among children and adult groups were investigated by means of a dichotic listening (DL) test (Hugdahl and Andersson, 1986). In DL, two different auditory stimuli are presented to the participant simultaneously, exactly at the same time, one to each ear, by using a set of headphones. Participants are asked to attend to one [Forced-Right (FR) or Forced-Left (FL)], or both of the auditory stimuli and are asked to report what they have heard (Hugdahl, 1998). Here, special attention was paid to FR and FL conditions in order to determine whether these two are differentially affected by musico-linguistic abilities and age (Milovanov et al., 2007).

We also investigated the ability to preattentively discriminate between slight changes in sounds: more specifically, in pitch with regard to music stimuli (Milovanov et al., 2008) and in duration with regard to music and speech stimuli (Milovanov et al., 2009). To this end, event-related brain potentials were performed in accordance with the MMN paradigm. The MMN component of the event-related potentials (ERPs) is an index of the precision which the auditory cortex has when modeling the auditory environment of any listener (Näätänen et al., 2001).

INTERPLAY BETWEEN MUSICAL APTITUDE AND SECOND LANGUAGE PRONUNCIATION SKILLS: NOVEL FINDINGS

In Milovanov et al. (2010) we compared the relationship between foreign language pronunciation and musical aptitude with three adult test groups of Finnish learners of English: non-musical university students, choir members, and English philology students. The subjects were tested on the production of English phonemes and on a discrimination task of phonemic minimal pairs. Their musical aptitude was determined by the Seashore musicality test. It was found that performance on the English phoneme discrimination test was not connected with their English phonemic production ability. Moreover, the phonemic discrimination ability

did not differ between the three test groups. Yet, performance on the English pronunciation test was better for subjects with musical aptitude than with less musical aptitude.

In Milovanov et al. (2007), the aim was to determine the effects of age and musical aptitude on phonemic processing in a forced-attention DL paradigm. Subjects differing in musical ability, as tested with the Seashore musical aptitude subtests, listened to consonant–vowel (CV) syllables presented dichotically under three different attention instructions: Non-Forced, FR, and FL conditions. The results showed that musical aptitude and age interacted with the ability to use attention to modulate a bottom-up laterality effect. Only adults who performed well in the Seashore musical aptitude test and practised music regularly showed more accurate left ear monitoring skills when listening to Finnish CV-syllables.

Therefore, based on the result described above, it can be inferred that those subjects with more advanced musical aptitude use the right hemisphere more in language processing in comparison with the non-musical subjects. Also, the musical subjects may possibly pay more attention to the musical components of language than the non-musical subjects.

Milovanov et al. (2008) examined the relationship between musical aptitude and second language pronunciation skills. Twenty children with advanced English pronunciation skills had better musical skills as measured by the Seashore musicality test than 20 children with less accurate English pronunciation skills. The individual Seashore subtests indicated that the participants with advanced pronunciation skills were superior to the participants with less-advanced pronunciation skills in pitch discrimination ability, timbre, sense of rhythm, and sense of tonality. The results in time or loudness subtests did not differ significantly between groups.

The children's ability to produce correct English phonemes with no direct equivalents in Finnish was further determined (Milovanov et al., in preparation). To ensure that all the children had an adequate amount of pretraining in the pronunciation of English, they received an 8-week course of pronunciation training of English including phonemic discrimination exercises (5–7 min/5 times a week, slightly modified from Baker, 1981). Moreover, two behavioral discrimination tests were conducted, both before and after the pronunciation training period. First, the children were required to distinguish the phonemic dissimilarities between English and Finnish through triplets based on minimal pair contrasts of the phonemes, e.g., (a) jeep–jeep–cheap, (b) they–day–they. After that, a standard C-major triad, C, and its two mistuned modifications C2% and C4% were investigated following the principles of the phonemic discrimination test.

The advanced pronunciation group outperformed the participants with less-advanced pronunciation skills in terms of a higher number of correct answers and a smaller number of mistakes in both the music and phonemic discrimination tests before and after the training measured by the Error rate. For the less-advanced pronunciation group, triad contrasts were more difficult than the phonemic contrasts both before and after the training. Both test groups marginally improved their phonemic discrimination skills after the training. Interestingly, the training period did not only develop the participants' linguistic skills but also strengthened

their discrimination skills for musical sounds, especially in the participants with advanced pronunciation skills.

We also investigated whether children with a more advanced performance in foreign language production represent musical sound features more readily in the preattentive level of neural processing compared with children with less-advanced production skills. Sound processing accuracy was examined by means of ERP recordings. The accuracy of the auditory cortex in representing musical sounds (C-major chord, C, and mistuned C-major chords of 150 ms, C2% and C4%) was examined by means of ERP recordings in the MMN paradigm. The ERP data accompany the results of the behavioral tests: the advanced pronunciation group showed larger fronto-central MMN with the music stimuli than the less-advanced pronunciation group. The MMN lateralization pattern did not differ between the test conditions nor the test groups. Taken together, the results could imply that musical and phonemic skills may partly be based on shared neural mechanisms.

In Milovanov et al. (2009) we aimed at specifically comparing the duration discrimination skills between both music and vowel conditions. The subjects were the same as described in Milovanov et al. (2008). The standard speech stimulus consisted of a 250 ms long monaural recording of the Finnish vowel/ö/and the shorter deviant speech stimulus/ö/with a 150-ms duration. The music sound was a violin tone C4, fundamental frequency 261.3 Hz. The standard and deviant durations of the violin tones were equivalent to those of the speech sounds.

The subjects with advanced pronunciation skills and greater musical aptitude were able to preattentively distinguish the duration difference in both conditions more effectively than the less-advanced pronunciation group with less musical aptitude. Only the advanced pronunciation group showed an MMN lateralization effect: larger MMN amplitudes were found above the right hemisphere, as also in the majority of previous linguistic MMN-studies, in the advanced pronunciation group. Moreover, ERP data show that, irrespective of general musical aptitude, duration changes from 250 to 150 ms are more prominently and accurately processed in music than in speech sounds. However, one must bear in mind that reducing the duration of speech sounds from 250 to 150 ms may not necessarily have the same perceptual consequences as reducing the duration of a violin sound by the same amount since speech and music sounds are known to differ in their perceptual attributes (Tervaniemi et al., 2009).

DISCUSSION AND CONCLUSION

Our review aims at introducing recent evidence about whether musical aptitude could be a crucial factor in learning foreign language pronunciation and discrimination skills. To this end, we introduce empirical evidence about the facilitating role of musical aptitude on foreign language pronunciation acquisition and phonemic and listening discrimination skills. This evidence was obtained from children and adult subjects without any neurological disorders or learning disabilities using various behavioral and brain recording paradigms.

Within the current framework, it seems that music and language are closely related neurocognitive systems (Patel, 2008). Music is one of the oldest, and most basic, socio-cognitive domains of the human species (Koelsch, 2005). Primate vocalizations are

determined by music-like features, such as pitch, timbre, and rhythm. Fine-grained temporal processing is fundamental to both speech and language (Alcock et al., 2000). Tallal et al. (1991) have already proposed that the underlying deficit which leads to language disturbance is control and processing of timing skills. In addition, Overy (2003) is of the view that musical training develops temporal processing abilities, which are also relevant to phonological segmentation skills. Most recently, Goswami and colleagues have emphasized the role of musical processing for speech processing and the readings skill development: they found a connection between metrical processing, rise-time perception, and reading development in a large sample of normal and dyslexic readers (Huss et al., 2011).

However, the role of possibly shared neural mechanisms between linguistic and musical functions is still unsettled. This is the case even though there is evidence that musical training improves sensory encoding of dynamically changing sounds, which, in turn, helps with linguistic coding. Koelsch (2005) is of the view that the human brain processes music and language with overlapping cognitive mechanisms in overlapping cerebral structures. This view promotes the apparent relationship between music and language which seems to be present from the very early stages of life. Music and language seem to grow from a common source ever since birth, and musical elements aid learning linguistic functions such as sound patterns and meaning (Fernald, 1989) and sound patterns and syntax (Jusczyk and Krumhansl, 1993).

Musical training seems to drive adaptive plasticity in speech processing networks. Kraus and Chandrasekaran (2010) raise an important point that years of processing pitch, duration, and timbre in a fine-grained way in music may enhance their processing in the context of speech. Patel (2011) proposes that music-driven adaptive plasticity in speech processing occurs because five essential conditions are met; there is overlap in the brain networks that process acoustical features used in both speech and music. Moreover, music places higher demands on the networks than does speech, in terms of the precision of processing. The role of positive emotion brought about by musical activities should not be underestimated. In addition to this, practising music requires constant repetition and focused attention. Patel points out that when these conditions are met neural plasticity drives the networks in question to function with higher precision than needed for ordinary speech communication and yet, since speech shares these networks with music, speech processing benefits.

In the present studies, a significant relationship was found between musical aptitude and second language linguistic skills, independent of verbal intelligence. It was discovered that participants with advanced musical aptitude also had advanced foreign language pronunciation skills. Based on the present results, it is proposed that linguistic skills, both in production and discrimination, are interconnected with perceptual musical skills. It is proposed that the musical subjects are able to more efficiently process the musical features in both speech and music, both attentively and preattentively.

The subjects with advanced pronunciation skills and greater musical aptitude were able to preattentively process the duration difference in the vowel and violin sounds more effectively than the less-advanced pronunciation group with less musical aptitude.

Moreover, both triad changes evoked significant MMN responses in both participant groups, although without significant amplitude differences between the groups. Still, the participants with advanced pronunciation skills showed larger fronto-central MMN responses with the C4% triad condition than the participants with less accurate pronunciation skills. The result could imply that musical and linguistic skills may partly be based on shared neural mechanisms. Alternatively, one could also speculate whether the superior performance in musical and linguistic tasks draws on the same source, namely improved auditory discrimination ability. To illuminate this issue still further, testing of more basic auditory processing skills, not directly related to speech or to music sound listening, should be conducted.

In general, the neurocognitive approach to music and brain processing is still in its infancy. However, the brain research methods still provide us with additional understanding concerning how to relate language and music knowledge to brain function and how phonemic training affects the brain's linguistic and musical organization. The results will give some insight into those neurocognitive mechanisms which are utilized during the task of acquiring a second language or learning to play an instrument. However, results from the current studies suggest that early exposure, even when it is minimal, is of crucial importance in causing neuroanatomical changes. Moreover, even minimal exposure to music practising may help in linguistic functions. The question remains whether certain kinds of reorganization processes of different brain functions, based on constant practise, are possible.

Knowledge regarding the facilitation provided by music in behavioral and neural learning is still quite limited, yet it is recognized. Both musical aptitude and practising may have the capacity to facilitate the learning of academic skills (Schön et al., 2004, 2008; Magne et al., 2006; Besson et al., 2007; Santos et al., 2007). Even if the potential of music in boosting cognitive faculties, other than linguistic ones, such as motor abilities or social skills is recognized, currently the empirical evidence on this is very scarce. Schellenberg and Peretz (2008) point out that the impact of musical training can go far beyond boosting linguistic abilities, extending to more general skills and functions such as working memory, mathematical skills, and spatial abilities.

Based on this, the paradigms in future studies should enable one to differentiate between the underlying factors for the enhancement music is shown to cause in linguistic or other cognitive skills: in the most optimal paradigms, we would learn whether the boosting effects caused by music directly facilitate linguistic and other skills or whether this boosting is actually caused by improved executive functions (attentional skills, working memory; Schellenberg and Peretz, 2008; Besson et al., 2011).

To sum up, the results presented here in this review article provide further evidence that musical aptitude and linguistic skills are interconnected, and that music and language skills are worth examining together. As Patel (2008) puts it, comparing these two domains helps us to explore how the mind makes sense out of sound. More longitudinal studies are needed to resolve the effectiveness of the transfer effects of music skills into other domains. In these endeavors, the participant groups should be matched at the outset with regard to various neural and behavioral measures of auditory processing and tested for linguistic skills before and after musical training.

REFERENCES

- Alcock, K., Passingham, R. E., Watkins, K., and Vargha-Khadem, F. (2000). Pitch and timing abilities in inherited speech and language impairment. *Brain Lang.* 75, 34–46.
- Anvari, S., Trainor, L., Woodside, J., and Levy, B. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *J. Exp. Child Psychol.* 83, 111–130.
- Baker, A. (1981). *Ship or Sheep. An Intermediate Pronunciation Course*. New York, NY: Cambridge University Press.
- Bengtsson, S., Nagy, Z., Skare, S., Forsman, L., Forssberg, H., and Ullén, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nat. Neurosci.* 8, 1148–1150.
- 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
- Besson, M., Schön, D., Moreno, S., Santos, A., and Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restor. Neurol. Neurosci.* 25, 399–410.
- Deutsch, D., Henthorn, T., Marvin, E. W., and Xu, H.-S. (2006). Absolute pitch among American and Chinese conservatory students: prevalence differences, and evidence for a speech-related critical period. *J. Acoust. Soc. Am.* 119, 719–722.
- Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., and Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science* 270, 305–307.
- Elmer, S., Meyer, M., Marra, L., and Jäncke, L. (2011). Intensive language training and attention modulate the involvement of frontoparietal regions during a non-verbal auditory discrimination task. *Eur. J. Neurosci.* 34, 165–175.
- Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Dev.* 60, 1497–1510.
- Gaser, C., and Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *J. Neurosci.* 23, 9240–9245.
- Hugdahl, K. (Ed.). (1998). *Handbook of Dichotic Listening*. Chichester: John Wiley & Sons.
- Hugdahl, K., and Andersson, L. (1986). The "forced-attention paradigm" in dichotic listening to children: neural and behavioral evidence. *Brain Res.* 1194, 81–89.
- Milovanov, R., Pietilä, P., Tervaniemi, M., and Esquef, P. A. A. (2010). Foreign language pronunciation skills and musical aptitude: a study of Finnish adults with higher education. *Learn. Individ. Differ.* 20, 56–60.
- Milovanov, R., Tervaniemi, M., and Gustafsson, M. (2004). "The impact of musical aptitude in foreign language acquisition," in *Proceedings of the 8th International Conference on Music Perception and Cognition (ICMPC8)*, eds S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, and P. Webster (Adelaide: Causal Productions).
- Milovanov, R., Tervaniemi, M., Takio, F., and Hämäläinen, H. (2007). Modification of dichotic listening (DL) performance by musico-linguistic abilities and age. *Brain Res.* 1156, 168–173.
- 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.
- Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., and Winkler, I. (2001). "Primitive intelligence" in the auditory cortex. *Trends Neurosci.* 24, 283–288.
- Overly, K. (2003). Dyslexia and music. From timing deficits to musical intervention. *Ann. N. Y. Acad. Sci.* 999, 497–505.
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L., and Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature* 392, 811–814.
- 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
- Santos, A., Joly-Pottuz, B., Moreno, S., Habib, M., and Besson, M. (2007). Behavioural and event-related potential evidence for pitch discrimination deficit in dyslexic children: improvement after intensive phonetic intervention. *Neuropsychologia* 45, 1080–1090.
- Schellenberg, E. G., and Peretz, I. (2008). Music, language, and cognition: unresolved issues. *Trends Cogn. Sci. (Regul. Ed.)* 12, 45–46.
- Schlaug, G., Jancke, L., Huang, Y., Staiger, J. F., and Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia* 33, 1047–1055.
- Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* 5, 688–694.
- Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., and Kollinsky, R. (2008). Songs as an aid for language acquisition. *Cognition* 106, 975–983.
- Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349.
- Seashore, C. E., Lewis, D., and Saetveit, J. G. (1960a). *Seashore Measures of Musical Talents Manual*, 2nd Edn. New York: Psychological Corporation.
- Seashore, C. E., Lewis, D., and Saetveit, J. G. (1960b). *Seashore Measures of Musical Talents CD, Digitally Remastered Version by Esquef, P. (2003)*. Espoo: Helsinki University of Technology.
- Slevc, L. R., and Miyake, A. (2006). Individual differences in second-language proficiency – does musical ability matter? *Psychol. Sci.* 17, 675–681.
- Tallal, P., and Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends Neurosci.* 29, 382–390.
- Tallal, P., Sainburg, R. L., and Jernigan, T. (1991). The neuropathology of developmental dysphasia: behavioural, morphological, and physiological evidence for a pervasive temporal processing disorder. *Read. Writ.* 3, 363–377.
- Tervaniemi, M. (2009). Musicians – same or different? *Ann. N. Y. Acad. Sci.* 1169, 151–156.
- Tervaniemi, M., and Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Res. Rev.* 43, 231–246.
- Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., and Friederici, A. D. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *Eur. J. Neurosci.* 30, 1636–1642.
- Wechsler, D. (1992). *Wechsler Adult Intelligence Scale Revised (WAIS-R)*. Helsinki: Psykologien kustannus Oy.

- Wechsler, D. (1999). *Wechsler Intelligence Scales for Children*, 3rd Edn (WISC-III). Helsinki: Psykologien Kustannus Oy.
- 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.
- Zatorre, R. J., and Gandour, J. T. (2007). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1087–1104.
- 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: 01 April 2011; accepted: 20 October 2011; published online: 21 November 2011.
- Citation: Milovanov R and Tervaniemi M (2011) The interplay between musical and linguistic aptitudes: a review. *Front. Psychology* 2:321. doi: 10.3389/fpsyg.2011.00321
- This article was submitted to *Frontiers in Auditory Cognitive Neuroscience*, a specialty of *Frontiers in Psychology*. Copyright © 2011 Milovanov and Tervaniemi. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



Why would musical training benefit the neural encoding of speech? The OPERA hypothesis

Aniruddh D. Patel*

Department of Theoretical Neurobiology, The Neurosciences Institute, San Diego, CA, USA

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Kai Alter, Newcastle University, UK
Stefan Koelsch, Freie Universität Berlin, Germany

***Correspondence:**

Aniruddh D. Patel, Department of Theoretical Neurobiology, The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121, USA.
e-mail: apatel@nsi.edu

Mounting evidence suggests that musical training benefits the neural encoding of speech. This paper offers a hypothesis specifying why such benefits occur. The “OPERA” hypothesis proposes that such benefits are driven by adaptive plasticity in speech-processing networks, and that this plasticity occurs when five conditions are met. These are: (1) **Overlap**: there is anatomical overlap in the brain networks that process an acoustic feature used in both music and speech (e.g., waveform periodicity, amplitude envelope), (2) **Precision**: music places higher demands on these shared networks than does speech, in terms of the precision of processing, (3) **Emotion**: the musical activities that engage this network elicit strong positive emotion, (4) **Repetition**: the musical activities that engage this network are frequently repeated, and (5) **Attention**: the musical activities that engage this network are associated with focused attention. According to the OPERA hypothesis, when these conditions are met neural plasticity drives the networks in question to function with higher precision than needed for ordinary speech communication. Yet since speech shares these networks with music, speech processing benefits. The OPERA hypothesis is used to account for the observed superior subcortical encoding of speech in musically trained individuals, and to suggest mechanisms by which musical training might improve linguistic reading abilities.

Keywords: music, speech, neural plasticity, neural encoding, hypothesis

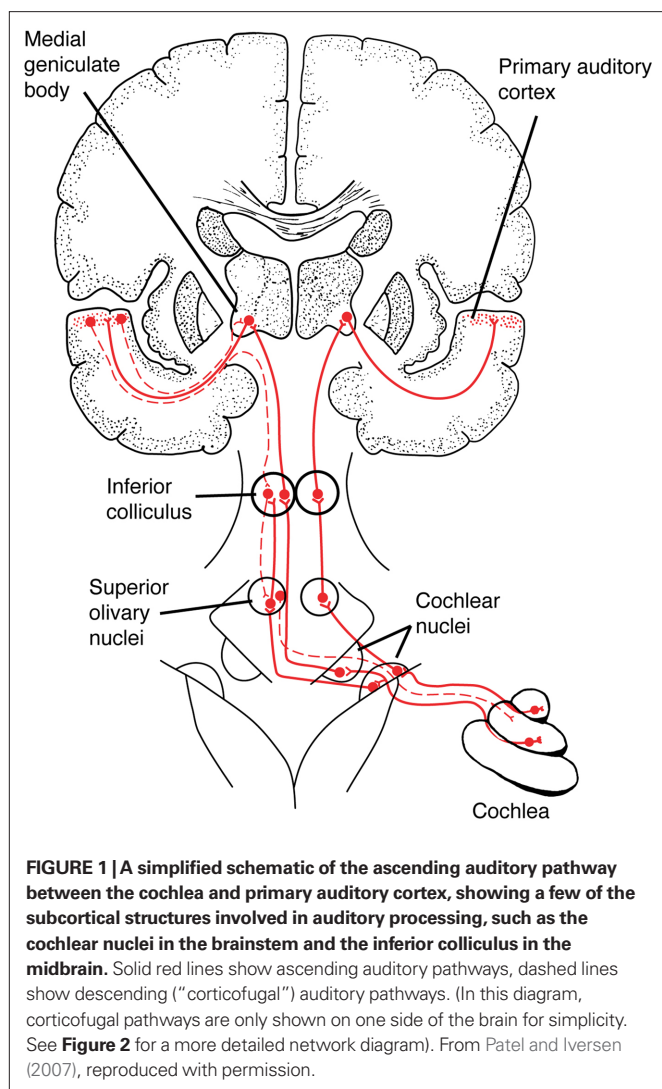
INTRODUCTION

Recent EEG research on human auditory processing suggests that musical training benefits the neural encoding of speech. For example, across several studies Kraus and colleagues have shown that the neural encoding of spoken syllables in the auditory brainstem is superior in musically trained individuals (for a recent overview, see Kraus and Chandrasekaran, 2010; this is further discussed in *The Auditory Brainstem Response to Speech: Origins and Plasticity*). Kraus and colleagues have argued that experience-dependent neural plasticity in the brainstem is one cause of this enhancement, based on the repeated finding that the degree of enhancement correlates significantly with the amount of musical training (e.g., Musacchia et al., 2007, 2008; Wong et al., 2007; Strait et al., 2009). They suggest that plasticity in subcortical circuits could be driven by descending (“corticofugal”) neural projections from cortex onto these circuits. There are many such projections in the auditory system (exceeding the number of ascending fibers), providing a potential pathway for cortical signals to tune subcortical circuits (Winer, 2006; Kral and Eggermont, 2007; **Figure 1**).

The arguments of Kraus and colleagues have both theoretical and practical significance. From a theoretical standpoint, their proposal contradicts the view that auditory processing is strictly hierarchical, with hardwired subcortical circuits conveying neural signals to cortical regions in a purely feed-forward fashion. Rather, their ideas support a view of auditory processing involving rich two-way interactions between subcortical and cortical regions, with structural malleability at both levels (cf. the “reverse hierarchy theory” of auditory processing, Ahissar et al., 2009). From a practical standpoint, their proposal suggests that the neural encoding of speech

can be enhanced by non-linguistic auditory training (i.e., learning to play a musical instrument or sing). This has practical significance because the quality of brainstem speech encoding has been directly associated with important language skills such as hearing in noise and reading ability (Banai et al., 2009; Parbery-Clark et al., 2009). This suggests that musical training can influence the development of these skills in normal individuals (Moreno et al., 2009; cf. Strait et al., 2010; Parbery-Clark et al., 2011). Furthermore, as noted by Kraus and Chandrasekaran (2010), musical training appears to strengthen the same neural processes that are impaired in individuals with certain speech and language processing problems, such as developmental dyslexia or hearing in noise. This has clear clinical implications for the use of music as a tool in language remediation.

Kraus and colleagues have provided a clear hypothesis for *how* musical training might influence the neural encoding of speech (i.e., via plasticity driven by corticofugal projections). Yet, from a neurobiological perspective, *why* would musical training drive adaptive plasticity in speech processing networks in the first place? Kraus and Chandrasekaran (2010) point out that both music and speech use pitch, timing, and timbre to convey information, and suggest that years of processing these cues in a fine-grained way in music may enhance their processing in the context of speech. The current “OPERA” hypothesis builds on this idea and makes it more specific. It proposes that music-driven adaptive plasticity in speech processing networks occurs because five essential conditions are met. These are: (1) **Overlap**: there is overlap in the brain networks that process an acoustic feature used in both speech and music, (2) **Precision**: music places higher demands on these networks than does speech, in terms of the precision of processing,



(3) **Emotion**: the musical activities that engage this network elicit strong positive emotion, (4) **Repetition**: the musical activities that engage this network are frequently repeated, and (5) **Attention**: the musical activities that engage this network are associated with focused attention. According to the OPERA hypothesis, when these conditions are met neural plasticity drives the networks in question to function with higher precision than needed for ordinary speech communication. Yet, since speech shares these networks with music, speech processing benefits.

The primary goals of this paper are to explain the OPERA hypothesis in detail (section The Conditions of the OPERA Hypothesis), and to show how it can be used to generate predictions to drive new research (section Putting OPERA to Work: Musical Training and Linguistic Reading Skills). The example chosen to illustrate the prediction-generating aspect of OPERA concerns relations between musical training and linguistic reading skills (cf. Goswami, 2010). One motivation for this choice is to show that the OPERA hypothesis, while developed on the basis of research on subcortical auditory processing, can also be applied to the cortical processing of acoustic features of speech.

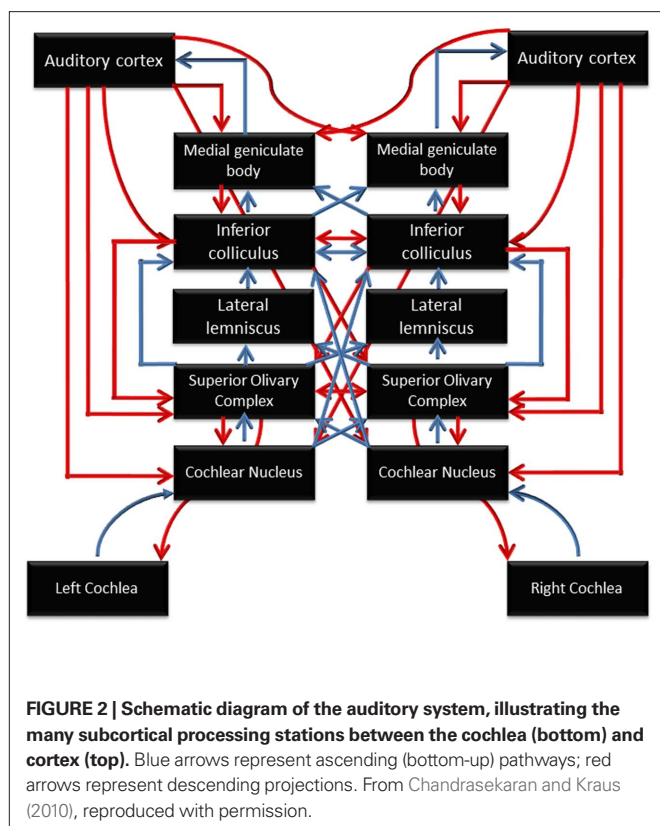
Prior to these core sections of the paper, section "The Auditory Brainstem Response to Speech: Origins and Plasticity" provides background on the auditory brainstem response to speech and evidence for neural plasticity in this response. Following the two core sections, section "Musical vs. Linguistic Training for Speech Sound Encoding" addresses the question of the relative merits of musical vs. linguistic training for improving the neural encoding of speech.

THE AUDITORY BRAINSTEM RESPONSE TO SPEECH: ORIGINS AND PLASTICITY

As noted by Chandrasekaran and Kraus (2010), "before speech can be perceived and integrated with long-term stored linguistic representations, relevant acoustic cues must be represented through a neural code and delivered to the auditory cortex with temporal and spectral precision by subcortical structures." The auditory pathway is notable for complexity of its subcortical processing, which involves many distinct regions and a rich network of ascending and descending connections (**Figure 2**). Using scalp-recorded EEG, population-level neural responses to speech in subcortical regions can be recorded non-invasively from human listeners (for a tutorial, see Skoe and Kraus, 2010). The brainstem response to the spoken syllable /da/ is shown in **Figure 3**, and the details of this response near syllable onset are shown in **Figure 4**. (Note that the responses depicted in these figures were obtained by averaging together the neural responses to thousands of repetitions of the syllable /da/, in order to boost the signal-to-noise ratio in the scalp-recorded EEG).

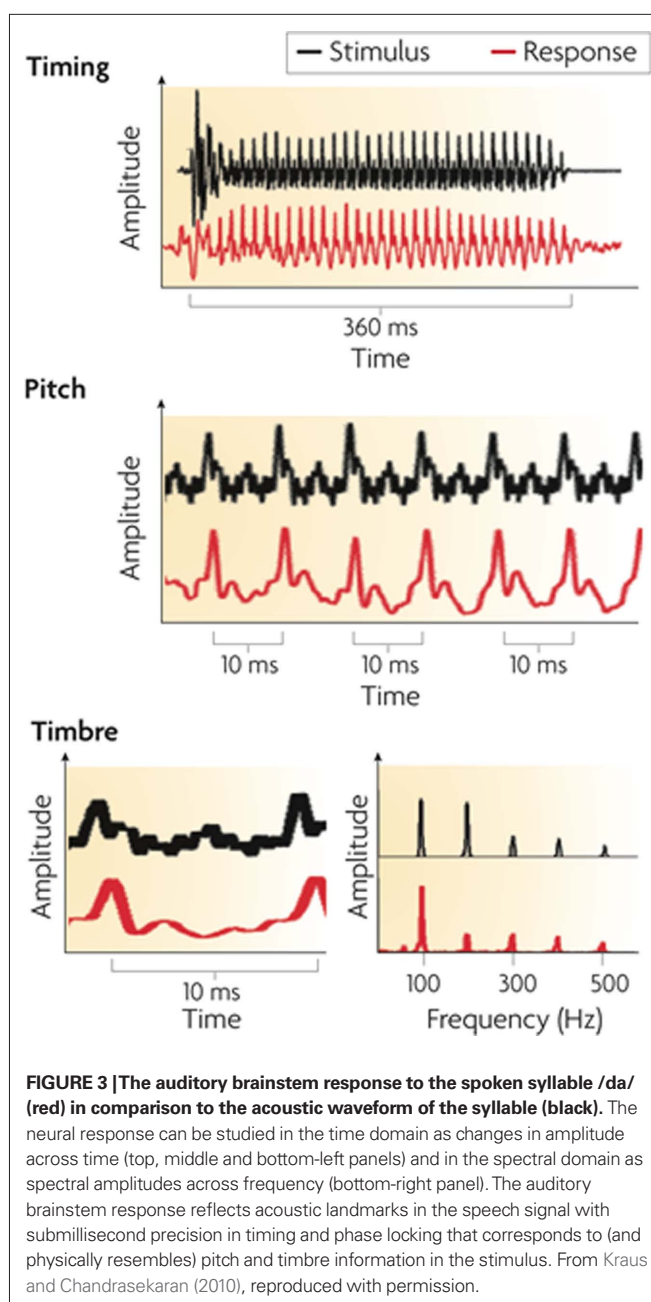
Broadly speaking, the response can be divided into two components: a transient onset response (commencing 5–10 ms after syllable onset, due to neural delays) and an ongoing component known as the frequency-following response (FFR). The transient onset response reflects the synchronized impulse response of neuronal populations in a number of structures, including the cochlea, brainstem, and midbrain. This response shows electrical peaks which are sensitive to the release burst of the /d/ and the formant transition into the vowel. The FFR is the summed responses of electrical currents from structures including cochlear nucleus, superior olivary complex, lateral lemniscus, and the inferior colliculus (Chandrasekaran and Kraus, 2010). The periodicities in the FFR reflect the synchronized component of this population response, arising from neural phase locking to speech waveform periodicities in the auditory nerve and propagating up the auditory pathway (cf. Cariani and Delgutte, 1996). The form of the FFR reflects the transition period between burst onset and vowel, and the vowel itself, via reflection of the fundamental frequency (F0) and some of the lower-frequency harmonics of the vowel.

The auditory brainstem response occurs with high reliability in all hearing subjects (Russo et al., 2004), and can be recorded in passive listening tasks (e.g., while the listener is watching a silent movie or even dozing). Hence it is thought to reflect sensory, rather than cognitive processing (though it may be shaped over longer periods of time via cognitive processing of speech or music, as discussed below). Since the response resembles the acoustic signal in several respects (e.g., in its temporal and spectral structure), correlation measures can be used to quantify the similarity of the neural response to the spoken sound, and hence to quantify the quality of speech sound encoding by the brain (Russo et al., 2004).



As noted previously, the quality of this encoding is correlated with important real-world language abilities, such as hearing in noise and reading. For example, Banai et al. (2009) found that the latency of specific electrical peaks in the auditory brainstem response to /da/ predicted word reading scores, when the effects of age, IQ, and the timing of click-evoked brainstem responses were controlled via partial correlation.

Several recent studies have found that the quality of subcortical speech sound encoding is significantly greater in musically trained individuals (e.g., Musacchia et al., 2007; Wong et al., 2007; Lee et al., 2009; Parbery-Clark et al., 2009; Strait et al., 2009). To take one example, Parbery-Clark et al. (2009) examined the auditory brainstem response in young adults to the syllable /da/ in quiet and in background babble noise. When listening to “da” in noise, musically trained listeners showed shorter-latency brainstem responses to syllable onset and formant transitions relative to their musically untrained counterparts, as well as enhanced representation of speech harmonics and less degraded response morphology. The authors suggest that these differences indicate that musicians had more synchronous neural responses to the syllable. To take another example, Wong et al. (2007) examined the FFR in musically trained and untrained individuals. Prior work had shown that oscillations in the FFR track the F0 and lower harmonics of the voice dynamically over the course of a single syllable (Krishnan et al., 2005). In the study of Wong et al. (2007), native English speakers unfamiliar with Mandarin listened passively to the Mandarin syllable /mi/ (pronounced “me”) with three different lexical tones. The salient finding was that the quality of F0 tracking was superior in musically trained individuals (Figure 5). It is worth noting that musician–non-musician differences in brainstem



encoding in this study, and the study of Parbery-Clark et al. (2009), were evident in the absence of attention to the speech sounds or any overt behavioral task.

One interpretation of the above findings is that they are entirely due to innate anatomical or physiological differences in the auditory systems of musicians and non-musicians. For example, the shorter latencies and stronger FFR responses to speech in musicians could be due to increased synchrony among neurons that respond to harmonically complex sounds, and this in turn could reflect genetically mediated differences in the number or spatial distribution of synaptic connections between neurons. Given the heritability of certain aspects of cortical structure (e.g., cortical thickness, which may in turn reflect the amount of arborization of dendrites, Panizzon et al., 2009), it is plausible that

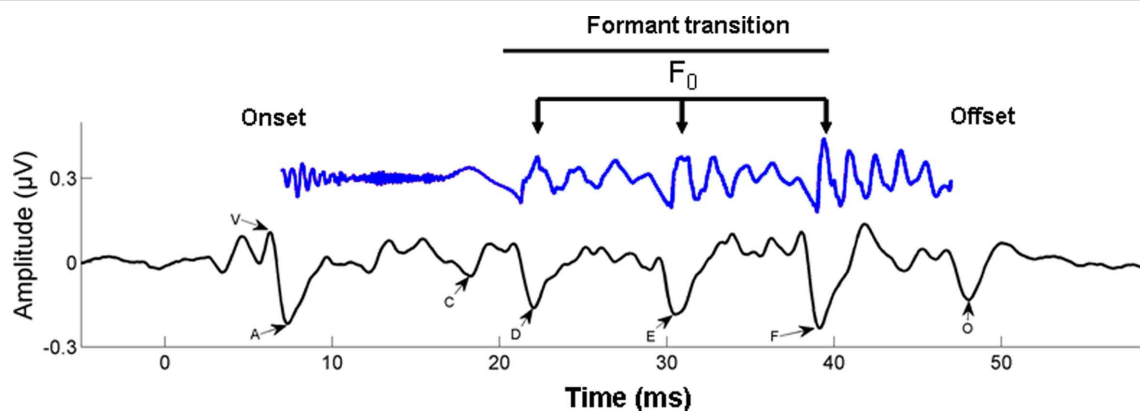


FIGURE 4 | Time-amplitude waveform of a 40-ms synthesized speech stimulus /da/ is shown in blue (time shifted by 6 ms to be comparable with the neural response). The first 10 ms of the syllable are characterized by the onset burst of the consonant /d/; the following 30 ms are the formant transition to the vowel /a/. The time-amplitude waveform of the time-locked brainstem response to the 40-ms /da/ is shown below the stimulus, in black. The onset response (V) begins 6–10 ms following the stimulus, reflecting the

time delay to the auditory brainstem. The start of the formant transition period is marked by wave C, marking the change from the burst to the periodic portion of the syllable, that is, the vowel. Waves D, E, and F represent the periodic portion of the syllable (frequency-following response) from which the fundamental frequency (F0) of the stimulus can be extracted. Finally, wave O marks stimulus offset. From Chandrasekaran and Kraus (2010), reproduced with permission.

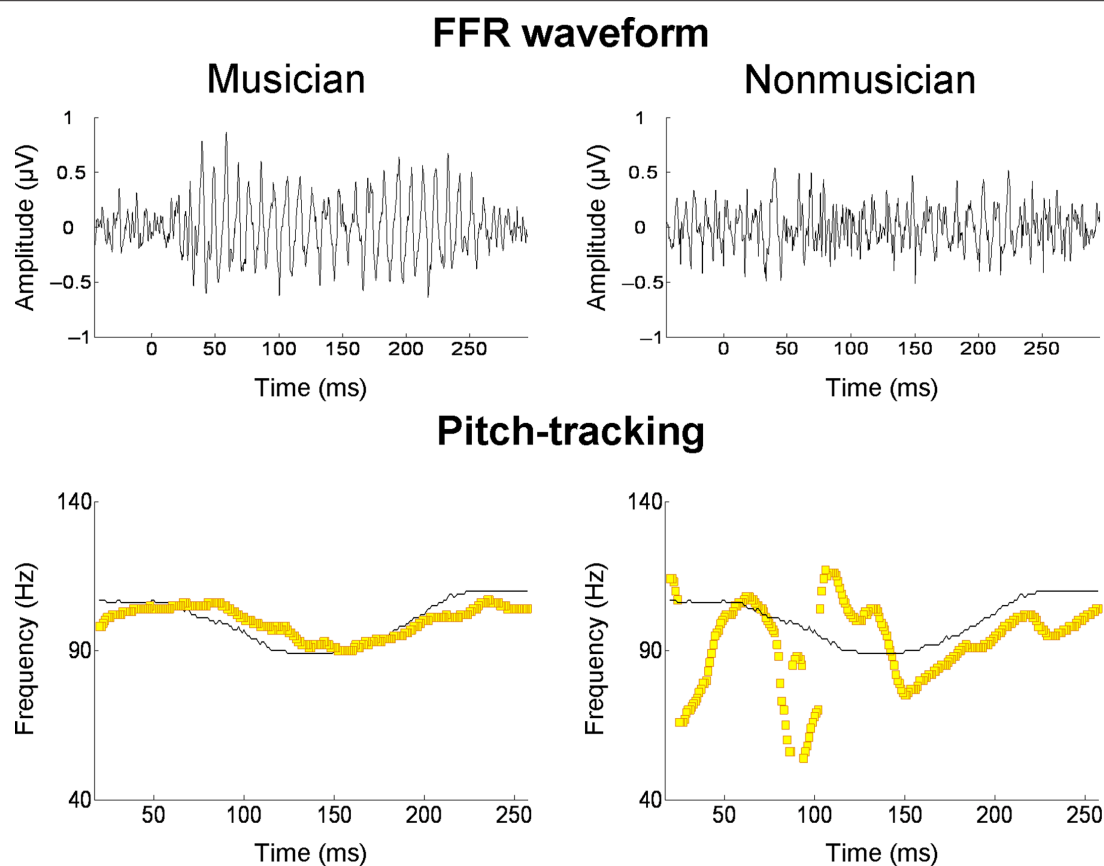


FIGURE 5 | Frequency-following responses to the spoken syllable /mi/ with a “dipping” (tone 3) pitch contour from Mandarin. The top row shows FFR waveforms of a musician and non-musician subject; the bottom row shows the fundamental frequency of the voice (thin black line) and the trajectories (yellow

lines) of the FFR's primary periodicity, from the same two individuals. For the musician, the FFR waveform is more periodic and its periodicity tracks the time-varying F0 contour of the spoken syllable with greater accuracy. From Wong et al. (2007), reproduced with permission.

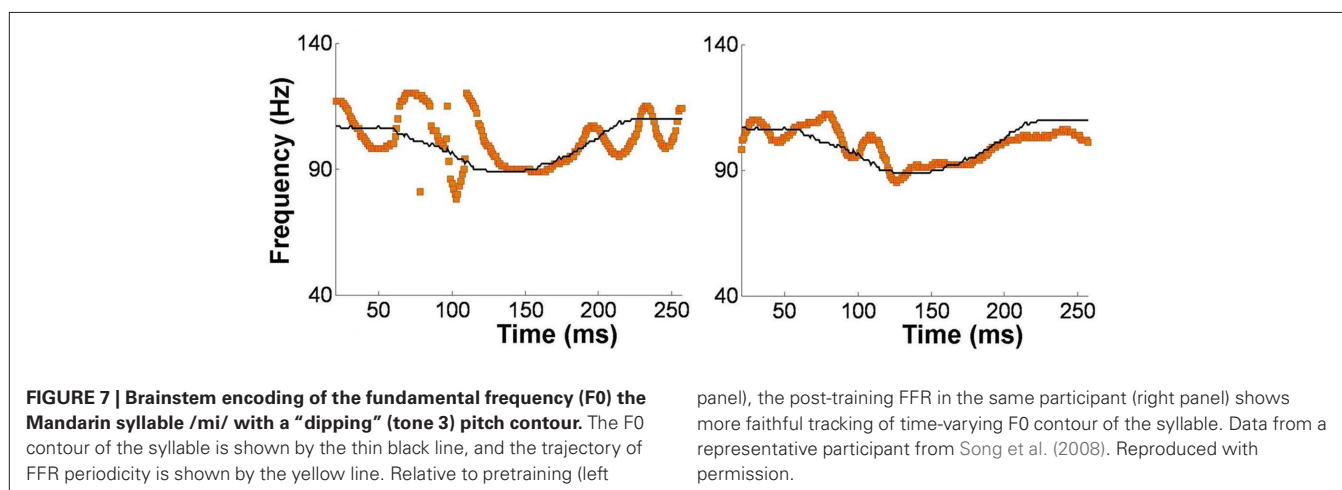
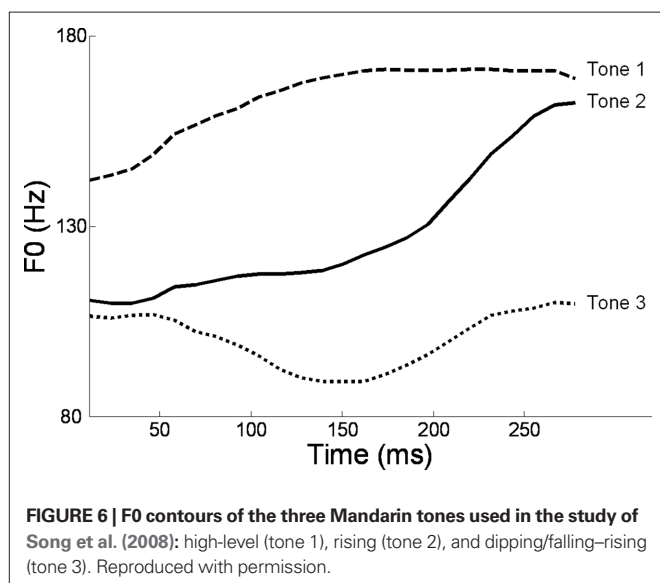
some proportion of the variance in brainstem responses to speech is due to genetic influence on subcortical neuroanatomy. In other words, it is plausible that some individuals are simply born with sharper hearing than others, which confers an advantage in processing any kind of complex sound. If sharper hearing makes an individual more likely to pursue musical training, this could explain the positive association between musical training and the superior neural encoding of speech sounds, in the absence of any experience-dependent brainstem plasticity.

Alternatively, musical training could be a *cause* of the enhanced neural encoding of speech sounds. That is, experience-dependent neural plasticity due to musical training might cause changes in brain structure and/or function that result in enhanced subcortical speech processing. There are several reasons to entertain this possibility. First, there is recent evidence that the brainstem encoding of speech sounds shows training-related neural plasticity, even in adult listeners. This was demonstrated by Song et al. (2008), who had native English speakers learn a vocabulary of six nonsense syllables, each paired with three lexical tones based on Mandarin tones 1–3 (Figure 6). (The participants had no prior knowledge

of tone languages.) For example, the syllable “pesh” meant glass, pencil, or table depending on whether it was spoken with a level, rising, or falling–rising tone. Participants underwent a training program in which each syllable, spoken with a lexical tone, was heard while viewing a picture of the word’s meaning. Quizzes were given periodically to measure word learning. The FFR to an untrained Mandarin syllable (/mi/) with tones 1–3 was measured before and after training. The salient finding was that the quality of FFR tracking of fundamental frequency (F0) of Mandarin tones improved with training. Compared to pretraining measures, participants showed increased energy at the fundamental frequency of the voice and fewer F0 tracking errors (Figure 7). Song et al. (2008) note that this enhancement likely reflects enhanced synchronization of neural firing to stimulus F0, which could in turn result from more neurons firing at the F0 rate, more synchronous firing, or some combination of these two. A striking aspect of these findings is the small amount of training involved in the study: just eight 30-min sessions spread across 2 weeks. These results show that the adult auditory brainstem response to speech is surprisingly malleable, exhibiting plasticity following relatively brief training.

The second reason to consider a role for musical training in enhanced brainstem responses to speech is that several studies which report superior brainstem encoding of speech in musicians also report that the degree of enhancement correlates significantly with the amount of musical training (e.g., Musacchia et al., 2007, 2008; Wong et al., 2007; Lee et al., 2009; Strait et al., 2009). Finally, the third reason is that longitudinal brain-imaging studies have shown that musical training causes changes in auditory cortical structure and function, which are correlated with increased auditory acuity (Hyde et al., 2009; Moreno et al., 2009). As noted previously there exist extensive corticofugal (top-down) projections from cortex to all subcortical auditory structures. Evidence from animal studies suggests that activity in such projections can tune the response patterns of subcortical circuits to sound (see Tzounopoulos and Kraus, 2009 for a review, cf. Kral and Eggermont, 2007; Suga, 2008; Bajo et al., 2010).

Hence the idea that musical training benefits the neural encoding of speech is neurobiologically plausible, though longitudinal randomized controlled studies are needed to establish this with



certainty. To date, randomized controlled studies of the influence of musical training on auditory processing have focused on cortical rather than subcortical signals (e.g., Moreno et al., 2009). Hopefully such work will soon be extended to include subcortical measures. Many key questions remain to be addressed, such as how much musical training is needed before one sees benefits in speech encoding, how large and long-lasting such benefits are, whether cortical changes precede (and cause) subcortical changes, and the relative effect of childhood vs. adult training. Even at this early stage, however, it is worth considering why musical training would benefit the neural encoding of speech.

THE CONDITIONS OF THE OPERA HYPOTHESIS

The OPERA hypothesis aims to explain why musical training would lead to adaptive plasticity in speech-processing networks. According to this hypothesis, such plasticity is engaged because five essential conditions are met by music processing. These are: overlap, precision, emotion, repetition, and attention (as detailed below). It is important to note that music processing does not automatically meet these conditions. Rather, the key point is that music processing has the potential to meet these conditions, and that by specifying these conditions, OPERA opens itself to empirical testing. Specifically, musical activities can be designed which fulfill these conditions, with the clear prediction that they will lead to enhanced neural encoding of speech. Conversely, OPERA predicts that musical activities not meeting the five conditions will not lead to such enhancements.

Note that OPERA is agnostic about the particular cellular mechanisms involved in adaptive neural plasticity. A variety of changes in subcortical circuits could enhance the neural encoding of speech, including changes in the number and spatial distribution of synaptic connections, and/or changes in synaptic efficacy, which in turn can be realized by a broad range of changes in synaptic physiology (Edelman and Gally, 2001; Schnupp et al., 2011, Ch. 7). OPERA makes no claims about precisely which changes are involved, or precisely how corticofugal projections are involved in such changes. These are important questions, but *how* adaptive plasticity is manifested in subcortical networks is a distinct question from *why* such plasticity is engaged in the first place. The OPERA hypothesis addresses the latter question, and is compatible with a range of specific physiological mechanisms for neural plasticity.

The remainder of this section lists the conditions that must be met for musical training to drive adaptive plasticity in speech processing networks, according to the OPERA hypothesis. For the sake of illustration, the focus on one particular acoustic feature shared by speech and music (periodicity), but the logic of OPERA applies to any acoustic feature important for both speech and music, including spectral structure, amplitude envelope, and the timing of successive events. (In section “Putting OPERA to Work: Musical Training and Linguistic Reading Skills,” where OPERA is used to make predictions about how musical training might benefit reading skills, the focus will be on amplitude envelope).

OVERLAP

For musical training to influence the neural encoding of speech, an acoustic feature important for both speech and music perception must be processed by overlapping brain networks. For example,

periodicity is an important feature of many spoken and musical sounds, and contributes to the perceptual attribute of pitch in both domains. At the subcortical level, the auditory system likely uses similar mechanisms and networks for encoding periodicity in speech and music, including patterns of action potential timing in neurons in the auditory pathway between cochlea and inferior colliculus (Cariani and Delgutte, 1996; cf. **Figure 2**). As noted in section “The Auditory Brainstem Response to Speech: Origins and Plasticity,” the synchronous aspect of such patterns may underlie the FFR.

Of course, once the pitch of a sound is determined (via mechanisms in subcortical structures and perhaps in early auditory cortical regions, cf. Bendor and Wang, 2005), then it may be processed in different ways depending on whether it occurs in a linguistic or musical context. For example, neuroimaging of human pitch perception reveals that when pitch makes lexical distinctions between words (as in tone languages), pitch processing shows a left hemisphere bias, in contrast to the typical right-hemisphere dominance for pitch processing (Zatorre and Gandour, 2008). The normal right-hemisphere dominance in the cortical analysis of pitch may reflect enhanced spectral resolution in right-hemisphere auditory circuits (Zatorre et al., 2002), whereas left hemisphere activations in lexical tone processing may reflect the need to interface pitch information with semantic and/or syntactic information in language.

The larger point is that *perceptual attributes* (such as pitch) should be conceptually distinguished from *acoustic features* (such as periodicity). The cognitive processing of a perceptual attribute (such as pitch) can be quite different in speech and music, reflecting the different patterns and functions the attribute has in the two domains. Hence some divergence in the cortical processing of that attribute is to be expected, based on whether it is embedded in speech or music (e.g., Peretz and Coltheart, 2003). On the other hand, the basic encoding of acoustic features underlying that attribute (e.g., periodicity, in the case of pitch) may involve largely overlapping subcortical circuits. After all, compared to auditory cortex, with its vast numbers of neurons and many functionally specialized subregions (Kaas and Hackett, 2000), subcortical structures have far fewer neurons, areas, and connections and hence less opportunity to partition speech and music processing into different circuits. Hence the idea that the sensory encoding of periodicity (or other basic acoustic features shared by speech and music) uses overlapping subcortical networks is neurobiologically plausible.

PRECISION

Let us assume that OPERA condition 1 is satisfied, that is, that a shared acoustic feature in speech and music is processed by overlapping brain networks. OPERA holds that for musical training to influence the neural encoding of speech, music must place higher demands on the nervous system than speech does, in terms of the precision with the feature must be encoded for adequate communication to occur. This statement immediately raises the question of what constitutes “adequate communication” in speech and music. For the purposes of this paper, adequate communication for speech is defined as conveying the semantic and propositional content of spoken utterances, while for music,

it is defined as conveying the structure of musical sequences. Given these definitions, how can one decide whether music places higher demands on the nervous system than speech, in terms of the precision of encoding of a given acoustic feature? One way to address this question is to ask to what extent a perceiver *requires* detailed information about the patterning of that feature in order for adequate communication to occur. Consider the feature of periodicity, which contributes to the perceptual attribute of pitch in both speech and music. In musical melodies, the notes of melodies tend to be separated by small intervals (e.g., one or two semitones, i.e., approximately 6 or 12% changes in pitch, cf. Vos and Troost, 1989), and a pitch movement of just one semitone (~6%) can be structurally very important, for example, when it leads to a salient, out-of-key note that increases the complexity of the melody (e.g., a C# note in the key of C, cf. Eerola et al., 2006). Hence for music perception, detailed information about pitch is important. Indeed, genetically based deficits in fine-grained pitch processing are thought to be one of the important underlying causes of musical tone deafness or “congenital amusia” (Peretz et al., 2007; Liu et al., 2010).

How crucial is detailed information about spoken pitch patterns to the perception of speech? In speech, pitch has a variety of linguistic functions, including marking emphasis and phrase boundaries, and in tone languages, making lexical distinctions between words. Hence there is no doubt that pitch conveys significant information in language. Yet the crucial question is: to what extent does a perceiver *require* detailed information about pitch patterns for adequate communication to occur? One way to address this question is to manipulate the pitch contour of natural sentences and see how this impacts a listener’s ability to understand the semantic and propositional content of the sentences. Recent research on this topic has shown that spoken language comprehension is strikingly robust to manipulations of pitch contour. Patel et al. (2010) measured the intelligibility of natural Mandarin Chinese sentences with intact vs. flattened (monotone) pitch contours (the latter created via speech resynthesis, with pitch fixed at the mean fundamental frequency of the sentence). Native speakers of Mandarin found the monotone sentences just as intelligible as the natural sentences when heard in a quiet background. Hence despite the complete removal of all details of pitch variation, the sentences were fully intelligible. How is this possible? Presumably listeners used the remaining phonetic information and their knowledge of Mandarin to guide their perception in a way that allowed them to infer which words were being said. The larger point is that spoken language comprehension is remarkably robust to lack of detail in pitch variation, and this presumably relaxes the demands placed on high-precision encoding of periodicity patterns.

It seems likely that this sort of robustness is not just limited to the acoustic feature of periodicity. One reason that spoken language comprehension is so robust is that it involves integrating multiple cues, some of which provide redundant sources of information regarding a sound’s phonological category. For example, in judging whether a sound within a word is a /b/ or a /p/, multiple acoustic cues are relevant, including voice onset time (VOT), vowel length, fundamental frequency (F0; i.e., “microintonational” perturbations of F0, which behave differently after voiced vs. voiceless stop consonants), and first and second formant patterns (i.e.,

their frequencies and rates of change; Stevens, 1998). While any individual cue may be ambiguous, when cues are integrated and interpreted in light of the current phonetic context, they provide a strong pointer to the intended phonological category being communicated (Toscano and McMurray, 2010; Toscano et al., 2010). Furthermore, when words are heard in sentence context, listeners benefit from multiple knowledge sources (including semantics, syntax, and pragmatics) which provide mutually interacting constraints that help a listener identify the words in the speech stream (Mattys et al., 2005).

Hence one can hypothesize that the use of context-based cue integration and multiple knowledge sources in word recognition helps relax the need for a high degree of precision in acoustic analysis of the speech signal, at least in terms of what is needed for adequate communication (cf. Ferreira and Patson, 2007). Of course, this is not to say that fine phonetic details are not relevant for linguistic communication. There is ample evidence that listeners are sensitive to such details when they are available (e.g., McMurray et al., 2008; Holt and Idemaru, 2011). Furthermore, these details help convey rich “indexical” information about speaker identity, attitude, emotional state, and so forth. Nevertheless, the question at hand is what demands are placed on the nervous system of a perceiver in terms of the precision of acoustic encoding for basic semantic/propositional communication to occur (i.e., “adequate” linguistic communication). How do these demands compare to the demands placed on a perceiver for adequate musical communication?

As defined above, adequate musical communication involves conveying the structure of musical sequences. Conveying the structure of a musical sequence involves playing the intended notes with appropriate timing. Of course, this is not to say that a few wrong notes will ruin musical communication (even professionals make occasional mistakes in complex passages), but by and large the notes and rhythms of a performance need to adhere fairly closely to a specified model (such as a musical score, or to a model provided by a teacher) for a performance to be deemed adequate. Even in improvisatory music such as jazz or the classical music of North India, a performer must learn to produce the sequence of notes they intend to play, and not other, unintended notes. Crucially, this requirement places fairly strict demands on the regulation of pitch and timing, because intended notes (whether from a musical score, a teacher’s model, or a sequence created “on the fly” in improvisation) are never far in pitch or duration from unintended notes. For example, as mentioned above, a note just one semitone away from an intended note can make a perceptually salient change in a melody (e.g., when the note departs from the prevailing musical key). Similarly, a note that is just a few hundred milliseconds late compared to its intended onset time can make a salient change in the rhythmic feel of a passage (e.g., the difference between an “on beat” note and a “syncopated” note). In short, conveying musical structure involves a high degree of precision, due to the fact that listeners use fine acoustic details in judging the structure of the music they are hearing. Furthermore, listeners use very fine acoustic details in judging the expressive qualities of a performance. Empirical research has shown that expressive (vs. deadpan) performances involve subtle, systematic modifications to the duration, intensity, and (in certain instruments) pitch of notes relative to

their nominal score-based values (Repp, 1992; Clynes, 1995; Palmer, 1997). Listeners are quite sensitive to these inflections and use them in judging the emotional force of musical sequences (Bhatara et al., 2011). Indeed, neuroimaging research has shown that expressive performances containing such inflections are more likely (than deadpan performances) to activate limbic and paralimbic brain areas associated with emotion processing (Chapin et al., 2010). This helps explain why musical training involves learning to pay attention to the fine acoustic details of sound sequences and to control them with high-precision, since these details matter for the aesthetic and emotional qualities of musical sequences.

Returning to our example of pitch variation in speech vs. music, the above paragraph suggests that the adequate communication of melodic music *requires* the detailed regulation and perception of pitch patterns. According to OPERA, this puts higher demands on the sensory encoding of periodicity than does speech, and helps drive experience-dependent plasticity in subcortical networks that encode periodicity. Yet since speech and music share such networks, speech processing benefits (cf. the study of Wong et al., 2007, described in section The Auditory Brainstem Response to Speech: Origins and Plasticity).

Yet would this benefit in the neural encoding of voice periodicity have any consequences for real-world language skills? According to the study of Patel et al. (2010) described above, the details of spoken pitch patterns are not essential for adequate spoken language understanding, even in the tone language Mandarin. So why would superior encoding of voice pitch patterns be helpful to a listener? One answer to this question is suggested by another experimental manipulation in the study of Patel et al. (2010). In this manipulation, natural vs. monotone Mandarin sentences were embedded in background babble noise. In this case, the monotone sentences were significantly less intelligible to native listeners (in the case where the noise was as loud as the target sentence, monotone sentences were 20% less intelligible than natural sentences). The precise reasons why natural pitch modulation enhances sentence intelligibility in noise remain to be determined. Nevertheless, based on these results it seems plausible that a nervous system which has high-precision encoding of vocal periodicity would have an advantage in understanding speech in a noisy background, and recent empirical data support this conjecture (Song et al., 2010). Combining this finding with the observation that musicians show superior brainstem encoding of voice F0 (Wong et al., 2007) leads to the idea that musically trained listeners should show better speech intelligibility in noise.

In fact, it has recently been reported that musically trained individuals show superior intelligibility for speech in noise, using standard tests (Parbery-Clark et al., 2009). This study also examined brainstem responses to speech, but rather than focusing on the encoding of periodicity in syllables with different pitch contour, it examined the brainstem response to the syllable /da/ in terms of latency, representation of speech harmonics, and overall response morphology. When /da/ was heard in noise, all of these parameters were enhanced in musically trained individuals compared to their untrained counterparts. This suggests that musical training influences the encoding of the temporal onsets and spectral details of speech, in addition to the encoding of voice periodicity, all of which may contribute to enhanced speech perception in noise (cf. Chandrasekaran et al., 2009).

This section has argued that music perception is often likely to place higher demands on the encoding of certain acoustic features than does speech perception, at least in terms of what is needed for adequate communication in the two domains. However, OPERA makes no *a priori* assumption that influences between musical and linguistic neural encoding are unidirectional. Hence an important question for future work is whether certain types of linguistic experience with heightened demands in terms of auditory processing (e.g., multilingualism, or learning a tone language) can impact the neural encoding of music (cf. Bidelman et al., 2011). This is an interesting question, but is not explored in the current paper. Instead, the focus is on explaining why musical training would benefit the neural encoding of speech, given the growing body of evidence for such benefits and the practical importance of such findings.

EMOTION, REPETITION, AND ATTENTION

For musical training to enhance the neural encoding of speech, the musical activities that engage speech-processing networks must elicit strong positive emotion, be frequently repeated, and be associated with focused attention. These factors work in concert to promote adaptive plasticity, and are discussed together here.

As noted in the previous section, music can place higher demands on the nervous system than speech does, in terms of the precision of encoding of a particular acoustic feature. Yet this alone is not enough to drive experience-dependent plasticity to enhance the encoding of that feature. There must be some (internal or external) motivation to enhance the encoding of that feature, and there must be sufficient opportunity to improve encoding over time. Musical training has the potential to fulfill these conditions. Accurate music performance relies on accurate perception of the details of sound, which is presumably associated in turn with high-precision encoding of sound features. Within music, accurate performance is typically associated with positive emotion and reward, for example, via internal satisfaction, praise from others, and from the pleasure of listening to well-performed music, in this case music produced by oneself or the group one is in. (Note that according to this view, the particular emotions expressed by the music one plays, e.g., joy, sadness, tranquility, etc., are not crucial. Rather, the key issue is whether the musical experience as a whole is emotionally rewarding.) Furthermore, accurate performance is typically acquired via extensive practice, including frequent repetitions of particular pieces. The neurobiological association between accurate performance and emotional rewards, and the opportunity to improve with time (e.g., via extensive practice) create favorable conditions for promoting plasticity in the networks that encode acoustic features.

Another factor likely to promote plasticity is focused attention on the details of sound during musical training. Animal studies of auditory training have shown that training-related cortical plasticity is facilitated when sounds are actively attended vs. experienced passively (e.g., Fritz et al., 2005; Polley et al., 2006). As noted by Jagadeesh (2006), attention “marks a particular set of inputs for special treatment in the brain,” perhaps by activating cholinergic systems or increasing the synchrony of neural firing, and this modulation plays an important role in inducing neural plasticity (cf. Kilgard and Merzenich, 1998; Thiel, 2007; Weinberger, 2007). While animal studies of attention and plasticity have focused on cortical circuits, it is worth recalling that the auditory system has rich

cortico-subcortical (corticofugal) connections, so that changes at the cortical level have the potential to influence subcortical circuits, and hence, the basic encoding of sound features (Schofield, 2010).

Focused attention on sound may also help resolve an apparent “bootstrapping” problem raised by OPERA: if musical training is to improve the precision of auditory encoding, how does this process get started? For example, focusing on pitch, how does the auditory system adjust itself to register finer pitch distinctions if it cannot detect them in the first place? One answer might be that attention activates more neurons in the frequency (and perhaps periodicity) channels that encode pitch, such that more neurons are recruited to deal with pitch, and hence acuity improves (P. Cariani, pers. comm.; cf. Recanzone et al., 1993; Tervaniemi et al., 2009).

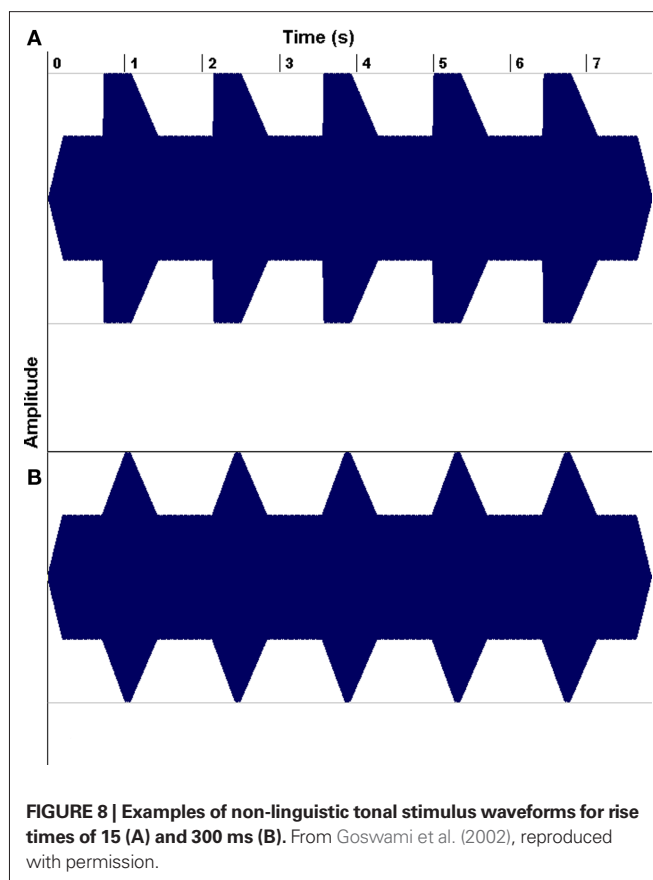
It is worth noting that the emotion, repetition, and attention criteria of OPERA are falsifiable. Imagine a child who is given weekly music lessons but who dislikes the music he or she is taught, who does not play any music outside of lessons, and who is not very attentive to the music during lessons. In such circumstances, OPERA predicts that musical training will not result in enhanced neural encoding of speech.

PUTTING OPERA TO WORK: MUSICAL TRAINING AND LINGUISTIC READING SKILLS

There is growing interest in links between musical training, auditory processing, and linguistic reading skills. This stems from two lines of research. The first line has shown a relationship between musical abilities and reading skills in normal children, for example, via correlational and experimental studies (e.g., Anvari et al., 2002; Moreno et al., 2009). For example, Moreno et al. (2009) assigned normal 8-year olds to 6 months of music vs. painting lessons, and found that after musical (but not painting) training, children showed enhanced reading abilities and improved auditory discrimination in speech, with the latter shown by both behavioral and neural measures (scalp-recorded cortical EEG).

The second line of research has demonstrated that a substantial portion of children with reading problems have auditory processing deficits, leading researchers to wonder whether musical training might be helpful for such children (e.g., Overy, 2003; Tallal and Gaab, 2006; Goswami, 2010). For example, Goswami has drawn attention to dyslexics' impairments in discriminating the rate of change of amplitude envelope at a sound's onset (its “rise time”; **Figure 8**), and have shown that such deficits predict reading abilities even after controlling for the effects of age and IQ. Of course, rise-time deficits are not the only auditory processing deficits that have been associated with dyslexia (Tallal and Gaab, 2006; Vandermosten et al., 2010), but they are of interest to speech–music studies because they suggest a problem with encoding amplitude envelope patterns, and amplitude envelope is an important feature in both speech and music perception.

In speech, the amplitude envelope is the relatively slow (syllable-level) modulation of overall energy within which rapid spectral changes take place. Amplitude envelopes play an important role as cues to speech rhythm (e.g., stress) and syllable boundaries, which in turn help listeners segment words from the flow of speech (Cutler, 1994). Envelope fluctuations in speech have most of their energy in the 3–20 Hz range, with a peak around 5 Hz (Greenberg, 2006), and experimental manipulations have shown that envelope modulations



in this range are necessary for speech intelligibility (Drullman et al., 1994; cf. Ghitza and Greenberg, 2009). In terms of connections to reading, Goswami has suggested that problems in envelope perception during language development could result in less robust phonological representations at the syllable-level, which would then undermine the ability to consciously segment syllables into individual speech sounds (phonemes). Phonological deficits are a core feature of dyslexia (if the dyslexia is not primarily due to visual problems), likely because reading requires the ability to segment words into individual speech sounds in order to map the sounds onto visual symbols. In support of Goswami's ideas about relations between envelope processing and reading, a recent meta-analysis of studies measuring dyslexics' performance on non-speech auditory tasks and on reading tasks reported that amplitude modulation and rise time discrimination were linked to developmental dyslexia in 100% of the studies that they reviewed (Hämäläinen et al., in press).

Amplitude envelope is also an important acoustic feature of musical sounds. Extensive psychophysical research has shown that envelope is one of the major contributors to a sound's musical timbre (e.g., Caclin et al., 2005). For example, one of the important cues that allows a listener to distinguish between the sounds of a flute and a French horn is the amplitude envelope of their musical notes (Strong and Clark, 1967). Furthermore, the slope of the amplitude envelope at tone onset is an important cue to the perceptual attack time of musical notes, and thus to the perception of rhythm and timing in sequences of musical events (Gordon, 1987). Given that envelope is an important acoustic feature in both speech and music,

the OPERA hypothesis leads to the prediction that musical training which relies on high-precision envelope processing could benefit the neural processing of speech envelopes, via mechanisms of adaptive neural plasticity, *if the five conditions of OPERA are met*. Hence the rest of this section is devoted to examining these conditions in terms of envelope processing in speech and music.

The first condition is that envelope processing in speech and music relies on overlapping neural circuitry. What is known about amplitude envelope processing in the brain? Research on this question in humans and other primates has focused on cortical responses to sound. It is very likely, however, that subcortical circuits are involved in envelope encoding, so that measures of envelope processing taken at the cortex reflect in part the encoding capacities of subcortical circuits. Yet since extant primate data come largely from cortical studies, those will be the focus here.

Primate neurophysiology research suggests that the temporal envelope of complex sound is represented by population-level activity of neurons in auditory cortex (Nagarajan et al., 2002). In human auditory research, two theories suggest that slower temporal modulations in sounds (such as amplitude envelope) are preferentially processed by right-hemisphere cortex (Zatorre et al., 2002; Poeppel, 2003). Prompted by these theories, Abrams et al. (2008) recently examined cortical encoding of envelope using scalp EEG. They collected neural responses to an English sentence (“the young boy left home”) presented to healthy 9- to 13-year-old children. (The children heard hundreds of repetitions of the sentence in one ear, while watching a movie and hearing the soundtrack quietly in the other ear.) The EEG response to the sentence was averaged across trials and low-pass filtered at 40 Hz to focus on cortical responses and how they might reflect the amplitude envelope of the sentence. Abrams et al. discovered that the EEG signal at temporal electrodes over both hemispheres tracked the amplitude envelope of the spoken sentence. (The authors argue that the neural EEG signals they studied are likely to arise from activity in secondary auditory cortex.) Notably, the quality of tracking, as measured by cross-correlating the EEG waveform with the speech envelope, was far superior in right-hemisphere electrodes (approx. 100% superior), in contrast to the usual left hemisphere dominance for spoken language processing (Hickok and Poeppel, 2007; **Figure 9**). (In

light of Goswami’s ideas, described above, it is worth noting that in a subsequent study Abrams et al. [2009] repeated their study with good vs. poor readers, aged 9–15 years. The good readers showed a strong right-hemisphere advantage for envelope tracking in multiple measures [e.g., higher cross-correlation, shorter-latency between neural response and speech input], while poor readers did not. Furthermore, empirical measures of neural envelope tracking predicted reading scores, even after controlling for IQ differences between good and poor readers).

Thus it appears that right-hemisphere auditory cortical regions (and very likely, subcortical inputs to these regions) are involved in envelope processing in speech. Are the same circuits involved in envelope processing in music? This remains to be tested directly, and could be studied using an individual differences approach. For example, one could use the design of Abrams et al. (2008) to examine cortical responses to the amplitude envelope of spoken sentences and musical melodies in the same listeners, with the melodies played using instruments that have sustained (vs. percussive) notes, and hence envelopes reminiscent of speech syllables. If envelope processing mechanisms are shared between speech and music, then one would predict that individuals would show similar envelope-tracking quality across the two domains. While such data are lacking, there are indirect indications of a relationship between envelope processing in music and speech. In music, one set of abilities that should depend on envelope processing are rhythmic abilities, because such abilities depend on sensitivity to the timing of musical notes, and envelope is an important cue for the perceptual onset and duration of event onsets in music (Gordon, 1987). Hence if musical and linguistic amplitude envelopes are processed by overlapping brain circuits, one would expect (somewhat counterintuitively) a relationship between musical rhythmic abilities and reading abilities. Is there any evidence for such a relationship? In fact, Goswami and colleagues have reported that dyslexics have problems with musical rhythmic tasks (e.g., Thompson and Goswami, 2008; Huss et al., 2011), and have shown that their performance on these tasks correlates with their reading skills, after controlling for age and IQ. Furthermore, normal 8-year-old children show positive correlations between performance on rhythm discrimination tasks (but not pitch discrimination tasks) and reading tasks, even after factoring out effects of age, parental education, and the number of hours children spend reading per week (Corrigall and Trainor, 2010).

Based on such findings, let us assume for the sake of argument that the first condition of OPERA is satisfied, that is, that envelope processing in speech and music utilizes overlapping brain networks. It is not required that such networks are entirely overlapping, simply that they overlap to a significant degree at some level of processing (e.g., subcortical, cortical, or both). With such an assumption, one can then turn to the next component of OPERA and consider what sorts of musical tasks would promote high-precision envelope processing. In ordinary musical circumstances, the amplitude envelope of sounds is an acoustic feature relevant to timbre, and while timbre is an important attribute of musical sound, it (unlike pitch) is rarely a primary structural parameter for musical sequences, at least in Western melodic music (Patel, 2008). Hence in order to encourage high-precision envelope processing in music, it may be necessary

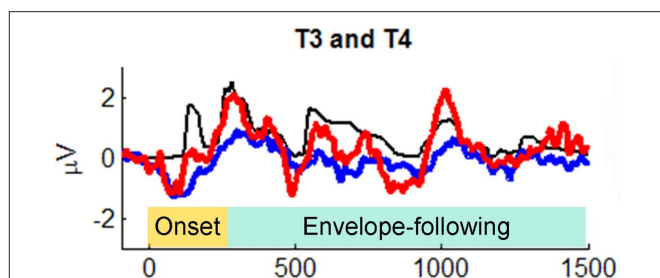


FIGURE 9 | Grand average cortical responses from temporal electrodes T3 and T4 (red: right hemisphere, blue: left hemisphere) and broadband speech envelope (black) for the sentence “the young boy left home.” Ninety-five milliseconds of the prestimulus period is plotted. The speech envelope was shifted forward in time 85 ms to enable comparison to cortical responses. From Abrams et al. (2008), reproduced with permission.

to devise novel musical activities, rather than relying on standard musical training. For example, using modern digital technology in which keyboards can produce any type of synthesized sounds, one could create synthesized sounds with similar spectral content but slightly different amplitude envelope patterns, and create musical activities (e.g., composing, playing, listening) which rely on the ability to distinguish such sounds. Note that such sounds need not be musically unnatural: acoustic research on orchestral wind-instrument tones has shown, for example, that the spectrum of the flute, bassoon, trombone, and French horn are similar, and that listeners rely on envelope cues in distinguishing between these instruments (Strong and Clark, 1967). The critical point is that success at the musical activity should *require* high-precision processing of envelope patterns.

For example, if working with young children, one could create computer-animated visual characters who “sing” with different “voices” (i.e., their “songs” are non-verbal melodies in which all tones have a particular envelope shape). After learning to associate different visual characters and their voices, one could then do games where novel melodies are heard without visual cues and the child has to guess which character is singing. Such “envelope training” games could be done adaptively, so that at the start of training, the envelope shapes of the different characters are quite different, and then once successful discrimination is achieved, new characters are introduced whose voices are more similar in terms of envelope cues.

According to OPERA, if this sort of training is to have any effect on the precision of envelope encoding by the brain, the musical tasks should be associated with strong positive emotion, extensive repetition, and focused attention. Fortunately, music processing is known to have a strong relationship to the brain’s emotion systems (Koelsch, 2010) and it seems plausible that the musical tasks could be made pleasurable (e.g., via the use of attractive musical sounds and melodies, and stimulating rewards for good performance). Furthermore, if they are sufficiently challenging, then participants will likely want to engage in them repeatedly and with focused attention.

To summarize, the OPERA hypothesis predicts that musical training which requires high-precision amplitude envelope processing will benefit the neural encoding of amplitude envelopes in speech, via mechanisms of neural plasticity, if the five conditions of OPERA are met. Based on research showing relationships between envelope processing and reading abilities (e.g., Goswami, 2010), this in turn may benefit linguistic reading skills.

MUSICAL VS. LINGUISTIC TRAINING FOR SPEECH SOUND ENCODING

As described in the previous section, the OPERA hypothesis leads to predictions for how musical training might benefit specific linguistic abilities via enhancements of the neural encoding of speech. Yet this leads to an obvious and important question. Why attempt to improve the neural encoding of speech by training in other domains? If one wants to improve speech processing, would it not be more effective to do acoustic training in the context of speech? Indeed, speech-based training programs that aim to improve sensitivity to acoustic features of speech are now widely used (e.g., the Fast ForWord program, cf. Tallal and Gaab, 2006). It

is important to note that the OPERA hypothesis is not proscriptive: it says nothing about the relative merits of musical vs. linguistic training for speech sound encoding. Instead, it tries to account for why musical training would benefit speech sound encoding in the first place, given the growing empirical evidence that this is indeed the case. Hence the relative efficacy of musical vs. linguistic training for speech sound encoding is an empirical question which can only be resolved by direct comparison in future studies. A strong motivation for conducting such comparisons is the demonstration that non-linguistic auditory perceptual training generalizes to linguistic discrimination tasks that rely on acoustic cues similar to those trained in the non-linguistic context (Lakshminarayanan and Tallal, 2007).

While direct comparisons of musical vs. linguistic training have yet to be conducted, it is worth considering some of the potential merits of music-based training. First, musical activities are often very enjoyable, reflecting the rich connections between music processing and the emotion systems of the brain (Koelsch, 2010; Salimpoor et al., 2011). Hence it may be easier to have individuals (especially children) participate repeatedly in training tasks, particularly in home-based tasks which require voluntary participation. Second, if an individual is experiencing a language problem, then a musical activity may not carry any of the negative associations that have developed around the language deficits and language-based tasks. This increases the chances of associating auditory training with strong positive emotions, which in turn may facilitate neural plasticity (as noted in section Emotion, Repetition, and Attention). Third, speech is acoustically complex, with many acoustic features varying at the same time (e.g., amplitude envelope, harmonic spectrum), and its perception automatically engages semantic processes that attempt to extract conceptual meaning from the signal, which draws attention away from the acoustic details of the signal. Musical sounds, in contrast, can be made acoustically relatively simple, with variation primarily occurring along specific dimensions that are the focus of training. For example, if the focus is on training sensitivity to amplitude envelope (cf. section Putting OPERA to work: musical training and linguistic reading skills), sounds can be created in which the spectral content is simple and stable, and in which the primary differences between tones are in envelope structure. Furthermore, musical sounds do not engage semantic processing, leaving the perceptual system free to focus more attention on the details of sound. This ability of musical sounds to isolate particular features for attentive processing could lower the overall complexity of auditory training tasks, and hence make it easier for individual to make rapid progress in increasing their sensitivity to such features, via mechanisms of neural plasticity.

A final merit of musical training is that it typically involves building strong sensorimotor links between auditory and motor skills (i.e., the sounds one produces are listened to attentively, in order to adjust performance to meet a desired model). Neuroscientific research suggests that sensorimotor musical training is a stronger driver of neural plasticity in auditory cortex than purely auditory musical training (Lappe et al., 2008). Hence musical training provides an easy, ecologically natural route to harness the power of sensorimotor processing to drive adaptive neural plasticity in the auditory system.

CONCLUSION

The OPERA hypothesis suggests that five essential conditions must be met in order for musical training to drive adaptive plasticity in speech processing networks. This hypothesis generates specific predictions, for example, regarding the kinds of musical training that could benefit reading skills. It also carries an implication regarding the notion of “musical training.” Musical training can involve different skills depending on what instrument and what aural abilities are being trained. (For example, learning to play the drums places very different demands on the nervous system than learning to play the violin.) The OPERA hypothesis states that the benefits of musical training depend on the particular acoustic features emphasized in training, the demands that music places on those features in terms of the precision of processing, and the degree of emotional reward, repetition and attention associated with musical activities. According to this hypothesis, simply giving an individual music lessons may not result in any benefits for speech processing. Indeed, depending on the acoustic feature being trained (e.g., amplitude envelope), learning a standard musical instrument may not be an effective way to enhance neural processing of that feature. Instead, novel

digital instruments may be necessary, which allow controlled manipulation of sound features. Thus OPERA raises the idea that in the future, musical activities aimed at benefiting speech processing should be purposely shaped to optimize the effects of musical training.

From a broader perspective, OPERA contributes to the growing body of research aimed at understanding the relationship between musical and linguistic processing in the human brain (Patel, 2008). Understanding this relationship will likely have significant implications for how we study and treat a variety of language disorders, ranging from sensory to syntactic processing (Jentschke et al., 2008; Patel et al., 2008).

ACKNOWLEDGMENTS

Supported by Neurosciences Research Foundation as part of its research program on music and the brain at The Neurosciences Institute, where ADP is the Esther J. Burnham Senior Fellow. I thank Dan Abrams, Peter Cariani, Bharath Chandrasekaran, Jeff Elman, Lori Holt, John Iversen, Nina Kraus, Stephen McAdams, Stefanie Shattuck-Hufnagel, and the reviewers of this paper for insightful comments.

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.
- Abrams, D. A., Nicol, T., Zecker, S., and Kraus, N. (2009). Abnormal cortical processing of the syllable rate of speech in poor readers. *J. Neurosci.* 29, 7686–7693.
- Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 285–299.
- Anvari, S., 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.
- Bajo, V. M., Nodal, F. R., Moore, D. R., and King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* 13, 252–260.
- Banai, K., Hornik, J., Skoe, E., Nicol, T., Zecker, S., and Kraus, N. (2009). Reading and subcortical auditory function. *Cereb. Cortex* 19, 2699–2707.
- Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature* 436, 1161–1165.
- Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., and Levitin, D. J. (2011). Perception of emotional expression in musical performance. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 921–934.
- 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.
- Cacclin, A., McAdams, S., Smith, B. K., and Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. *J. Acoust. Soc. Am.* 118, 471–482.
- 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.
- Chandrasekaran, B., Hornik, J., Skoe, E., Nicol, T., and Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. *Neuron* 64, 311–319.
- Chandrasekaran, B., and Kraus, N. (2010). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47, 236–246.
- Chapin, H., Jantzen, K., Kelso, S. J. A., Steinberg, F., and Large, E. (2010). Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLoS ONE* 5: e13812. doi: 10.1371/journal.pone.0013812
- Clynes, M. (1995). Microstructural musical linguistics: composers' pulses are liked most by the best musicians. *Cognition* 55, 269–310.
- Corrigall, K., and Trainor, L. (2010). “The predictive relationship between length of musical training and cognitive skills in children,” in *Paper presented at the 11th Intl. Conf. on Music Perception & Cognition (ICMPC11)*, August 2010, Seattle, WA.
- Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua* 92, 81–104.
- Drullman, R., Festen, J. M., and Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* 95, 1053–1064.
- Edelman, G. M., and Gally, J. (2001). Degeneracy and complexity in biological systems. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13763–13768.
- Eerola, T., Himberg, T., Toivainen, P., and Louhivuori, J. (2006). Perceived complexity of western and African folk melodies by western and African listeners. *Psychol. Music* 34, 337–371.
- Ferreira, F., and Patson, N. D. (2007). The ‘good enough’ approach to language comprehension. *Lang. Linguist. Compass* 1, 71–83.
- Fritz, J., Elhilali, M., and Shamma, S. (2005). Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear. Res.* 206, 159–176.
- Ghitza, O., and Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66, 113–126.
- Gordon, J. W. (1987). The perceptual attack time of musical tones. *J. Acoust. Soc. Am.* 82, 88–105.
- Goswami, U., Thompson, J., Richardson, U., Stainthorpe, R., Hughes, D., Rosen, S., and Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: a new hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10911–10916.
- Goswami, U. (2010). A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* 15, 3–10.
- Greenberg, S. (2006). “A multi-tier framework for understanding spoken language,” in *Listening to Speech: An Auditory Perspective*, eds S. Greenberg and W. A. Ainsworth (Mahwah, NJ: Erlbaum), 411–433.
- Hämäläinen, J. A., Salminen, H. K., and Leppänen, P. H. T. (in press). Basic auditory processing deficits in dyslexia: Review of the behavioural, event-related potential and magnetoencephalographic evidence. *J. Learn. Disabil.*
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Holt, L. L., and Idemaru, K. (2011). “Generalization of dimension-based statistical learning of speech,” in *Proc. 17th Intl. Cong. Phonetic Sci. (ICPhS XVII)*, Hong Kong, China.
- 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.
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. E., and Schlaug, G. (2009). Musical training shapes structural brain development. *J. Neurosci.* 29, 3019–3025.
- Jagadeesh, B. (2006). “Attentional modulation of cortical plasticity,” in *Textbook of Neural Repair and Rehabilitation:*

- Neural Repair and Plasticity*, Vol. 1. eds M. Selzer, S. E. Clarke, L. G. Cohen, P. W. Duncan, and F. H. Gage (Cambridge: Cambridge University Press), 194–205.
- Jentschke, S., Koesch, S., Sallat, S., and Friederici, A. (2008). Children with specific language impairment also show impairment of music-syntactic processing. *J. Cogn. Neurosci.* 20, 1940–1951.
- Kaas, J., and Hackett, T. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11793–11799.
- Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends Cogn. Sci.* 14, 131–137.
- Kilgard, M., and Merzenich, M. (1998). Cortical reorganization enabled by nucleus basalis activity. *Science* 279, 1714–1718.
- Kral, A., and Eggermont, J. (2007). What's to lose and what's to learn: development under auditory deprivation, cochlear implants and limits of cortical plasticity. *Brain Res. Rev.* 56, 259–269.
- Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605.
- Krishnan, A., Xu, Y., Gandour, J., and Ciani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res. Cogn. Brain Res.* 25, 161–168.
- Lakshminarayanan, K., and Tallal, P. (2007). Generalization of non-linguistic auditory perceptual training to syllable discrimination. *Restor. Neurol. Neurosci.* 25, 263–272.
- Lappe, C., Herholz, S. C., Trainor, L. J., and Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *J. Neurosci.* 8, 9632–9639.
- Lee, K. M., Skoe, E., Kraus, N., and Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *J. Neurosci.* 29, 5832–5840.
- Liu, F., Patel, A. D., Fourcin, A., and Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification, and imitation. *Brain* 133, 1682–1693.
- Mattys, S. L., White, L., and Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *J. Exp. Psychol. General* 134, 477–500.
- McMurray, B., Aslin, R., Tanenhaus, M., Spivey, M., and Subik, D. (2008). Gradient sensitivity to within-category variation in speech: implications for categorical perception. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1609–1631.
- 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.
- Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15894–15898.
- Musacchia, G., Strait, D. L., and Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and nonmusicians. *Hear. Res.* 241, 34–42.
- Nagarajan, S. S., Cheung, S. W., Bedenbaugh, P., Beitel, R. E., Schreiner, C. E., and Merzenich, M. M. (2002). Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. *J. Neurophysiol.* 87, 1723–1737.
- Overly, K. (2003). Dyslexia and music: from timing deficits to musical intervention. *Ann. N. Y. Acad. Sci.* 999, 497–505.
- Palmer, C. (1997). Music performance. *Ann. Rev. Psychol.* 48, 115–138.
- Panizzon, M. W., Fennema-Notestine, C., Eyer, L. T., Jernigan, T. L., Prom-Wormley, E., Neale, M., Jacobson, K., Lyons, M. J., Grant, M. D., Franz, C. E., Xian, H., Tsuang, M., Tsuang, M., Fischl, B., Seidman, L., Dale, A., and Kremen, W. S. (2009). Distinct genetic influences on cortical surface area and cortical thickness. *Cereb. Cortex* 19, 2728–2735.
- Parbery-Clark, A., Skoe, E., and Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* 29, 14100–14107.
- Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., and Kraus, N. (2011). Musical training 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: Oxford University Press.
- Patel, A. D., and Iversen, J. R. (2007). The linguistic benefits of musical abilities. *Trends Cogn. Sci.* 11, 369–372.
- Patel, A. D., Iversen, J. R., Wassenaar, M., and Hagoort, P. (2008). Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology* 22, 776–789.
- Patel, A. D., Xu, Y., and Wang, B. (2010). “The role of F0 variation in the intelligibility of Mandarin sentences,” in *Proc. Speech Prosody 2010*, May 11–14, 2010, Chicago, IL, USA.
- Peretz, I., and Coltheart, M. (2003). Modularity of music processing. *Nat. Neurosci.* 6, 688–691.
- Peretz, I., Cummings, S., and Dubé, M.-P. (2007). The genetics of congenital amusia (or tone-deafness): a family aggregation study. *Am. J. Hum. Genet.* 81, 582–588.
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun.* 41, 245–255.
- Polley, D. B., Steinberg, E. E., and Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *J. Neurosci.* 26, 4970–4982.
- Recanzone, G. H., Schreiner, C. C., and Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* 13, 87–103.
- Repp, B. H. (1992). Diversity and commonality in music performance: an analysis of timing microstructure in Schumann's “Träumerei.” *J. Acoust. Soc. Am.* 92, 2546–2568.
- Russo, N., Nicol, T., Musacchia, G., and Kraus, N. (2004). Brainstem response to speech syllables. *Clin. Neurophysiol.* 115, 2021–2030.
- Salimpoor, V., 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.
- Schnupp, J., Nelken, I., and King, A. (2011). *Auditory Neuroscience: Making Sense of Sound*. Cambridge, MA: MIT Press.
- Schofield, B. R. (2010). Projections from auditory cortex to midbrain cholinergic neurons that project to the inferior colliculus. *Neuroscience* 166, 231–240.
- Skoe, E., and Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear Hear.* 31, 302–324.
- Song, J., Skoe, E., Banai, K., and Kraus, N. (2010). Perception of speech in noise: neural correlates. *J. Cogn. Neurosci.* 23, 2268–2279.
- Song, J. H., Skoe, E., Wong, P. C., and Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *J. Cogn. Neurosci.* 10, 1892–1902.
- Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.
- Strait, D. L., Kraus, N., Parbery-Clark, A., and Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear. Res.* 261, 22–29.
- Strait, D. L., Skoe, E., Kraus, N., and Ashley, R. (2009). Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *Eur. J. Neurosci.* 29, 661–668.
- Strong, W., and Clark, M. C. (1967). Perturbations of synthetic orchestral wind-instrument tones. *J. Acoust. Soc. Am.* 41, 277–285.
- Suga, N. (2008). Role of corticofugal feedback in hearing. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 194, 169–183.
- Tallal, P., and Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends Neurosci.* 29, 382–390.
- Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., and Friederici, A. D. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *Eur. J. Neurosci.* 30, 1636–1642.
- Thiel, C. M. (2007). Pharmacological modulation of learning-induced plasticity in human auditory cortex. *Restor. Neurol. Neurosci.* 25, 435–443.
- Thompson, J. M., and Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *J. Physiol. Paris* 102, 120–129.
- Toscano, J. C., and McMurray, B. (2010). Cue integration with categories: weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cogn. Sci.* 34, 434–464.
- Toscano, J. C., McMurray, B., Dennhardt, J., and Luck, S. J. (2010). Continuous perception and graded categorization: electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychol. Sci.* 21, 1532–1540.
- Tzounopoulos, T., and Kraus, N. (2009). Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron* 62, 463–469.
- Vandermosten, M., Boets, B., Luts, H., Poelmans, H., Golestani, N., Wouters, J., and Ghesquière, P. (2010). Adults with dyslexia are impaired in categorizing speech and nonspeech sounds on the basis of temporal cues. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10389–10394.
- Vos, P. G., and Troost, J. M. (1989). Ascending and descending melodic

- intervals: statistical findings and their perceptual relevance. *Music Percept.* 6, 383–396.
- Weinberger, N. (2007). Auditory associative memory and representational plasticity in the primary auditory cortex. *Hear. Res.* 229, 54–68.
- Winer, J. (2006). Decoding the auditory corticofugal systems. *Hear. Res.* 212, 1–8.
- 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.
- 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.
- Zatorre, R. J., and Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Phil. Trans. R. Soc. B* 363, 1087–1104.
- 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.
- Received: 16 March 2011; paper pending published: 05 April 2011; accepted: 12 June 2011; published online: 29 June 2011.
- Citation: Patel AD (2011) Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychology* 2:142. doi: 10.3389/fpsyg.2011.00142
- This article was submitted to *Frontiers in Auditory Cognitive Neuroscience*, a specialty of *Frontiers in Psychology*. Copyright © 2011 Patel. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.



Neurophysiological influence of musical training on speech perception

Antoine J. Shahin*

Department of Otolaryngology-Head and Neck Surgery, The Ohio State University, Columbus, OH, USA

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Erich Schröger, University of Leipzig, Germany

Nina Kraus, Northwestern University, USA

***Correspondence:**

Antoine J. Shahin, Department of Otolaryngology-Head and Neck Surgery, The Ohio State University, 915 Olentangy River Road, Columbus, OH 43212, USA.
e-mail: shahin.5@osu.edu

Does musical training affect our perception of speech? For example, does learning to play a musical instrument modify the neural circuitry for auditory processing in a way that improves one's ability to perceive speech more clearly in noisy environments? If so, can speech perception in individuals with hearing loss (HL), who struggle in noisy situations, benefit from musical training? While music and speech exhibit some specialization in neural processing, there is evidence suggesting that skills acquired through musical training for specific acoustical processes may transfer to, and thereby improve, speech perception. The neurophysiological mechanisms underlying the influence of musical training on speech processing and the extent of this influence remains a rich area to be explored. A prerequisite for such transfer is the facilitation of greater neurophysiological overlap between speech and music processing following musical training. This review first establishes a neurophysiological link between musical training and speech perception, and subsequently provides further hypotheses on the neurophysiological implications of musical training on speech perception in adverse acoustical environments and in individuals with HL.

Keywords: auditory cortex, speech perception, musical training, hearing loss, speech in noise, EEG, MEG, neuroplasticity

The brain is astoundingly neuroplastic, and studying neurophysiological alterations due to musical training is an increasingly popular method to assess the precise mechanisms underlying neuroplasticity. Furthermore, it is particularly interesting to understand how neuroplastic changes due to musical training may affect other auditory functions, namely speech perception. It has been repeatedly shown that musical training modifies the auditory neural circuitry in a way that allows for enhanced musical skills (e.g., processing musical timbre, pitch contour, and rhythm). However, the extent to which enhanced auditory representations and processing due to musical training transfers to other auditory functions, especially speech perception, is an issue of exciting debate and a rich topic to explore. In order for enhanced auditory skills due to musical training to transfer to speech perception, musical training must target the neural mechanisms that underlie speech perception and modify them in a way to enhance their function.

From a psychoacoustics perspective, music and speech processing share several mechanisms. To extract meaning from a piece of music or ongoing speech, the auditory system must encode pitch or voice cues (to identify a musical note or speaker) as well as rhythm (the unfolding of words and chords over time in a discourse or symphony), and segregate a voice in the midst of background noise or a musical instrument within a symphony orchestra. Thus, the neural mechanisms involved in processing music should overlap with and influence speech mechanisms (and vice versa). Furthermore, musical training should recruit resources associated with speech processing, thereby enhancing the neural mechanisms underlying speech perception (and vice versa). This article reviews neurophysiological evidence supporting an influence of musical training on speech perception at the sensory level, and discusses whether such transfer could facilitate speech perception in individuals with hearing loss (HL).

METHODOLOGIES FOR ASSESSING AUDITORY STRUCTURE AND FUNCTION

Effects of musical training on brain attributes have been largely studied using non-invasive imaging methods, primarily electroencephalography (EEG), magnetoencephalography (MEG), magnetic resonance imaging (MRI), and functional MRI (fMRI). However, the electrophysiological approaches (EEG and MEG) will be emphasized in this review.

Electroencephalography and MEG detect the brain's electrical and magnetic activity, respectively, at the scalp with a millisecond resolution. The amplitude of the event-related potential (ERP as in EEG) or field (ERF as in MEG) or of the auditory-evoked potential/field (AEP/AEF) indicates the size of the activated neural population and/or the trial-to-trial phase synchrony of neural firing when evoked by sensory stimuli (e.g., sound). The latency of the ERP/ERF reflects the timing of neural activation. Combined with the topography of the neural response (i.e., its position on the scalp), these ERP/ERF signatures provide specifics on the function and origin of activity (Pantev et al., 2001; Musacchia et al., 2008). MRI and fMRI, in contrast, probe the entire brain as opposed to scalp activity. MRI provides fine details of brain structure due to differences in magnetic properties of brain tissues (i.e., white matter versus gray matter). fMRI reveals the hemodynamic response associated with neural activity, and the foci of this activity are then superimposed on MRI images to determine the precise neuroanatomical structure(s) generating the neural activity. Thus, both MRI and fMRI reveal spatially well-defined neuroanatomical and functional loci undergoing structural and functional changes, for example due to musical training (Ohnishi et al., 2001; Hyde et al., 2009).

PREREQUISITES FOR FUNCTIONAL TRANSFER FOLLOWING ACOUSTICAL TRAINING

Acoustical features of speech and music are represented by a hierarchical auditory network. Low-level regions [i.e., brain stem and primary auditory cortex (A1)] encode simple acoustical features, such as sound onset and pitch. More complex features, such as spectrotemporal combinations that represent speech or musical timbre, recruit higher-level processes in non-primary auditory cortex (NPAC). Studies have provided evidence for neural networks, at the level of the auditory cortex and beyond, that are relatively specialized for processing either music or speech (Peretz et al., 1994; Tervaniemi et al., 1999, 2006; Zatorre et al., 2002; Rogalsky et al., 2011). However, functional and structural overlap for speech and music processing along the sensory (Wong et al., 2007; Sammler et al., 2009; Rogalsky et al., 2011) and cognitive (Sammler et al., 2009; Schulze et al., 2011) levels exist. A prerequisite for the transfer of auditory function to speech perception following musical training is the fostering of greater overlap between the neural mechanisms associated with speech and music perception. To put this into perspective consider the following example: since a sound's temporal information, which is fundamental to speech perception, is favorably processed in the left auditory cortex (Zatorre and Belin, 2001; Zatorre et al., 2002), musicians may be inclined to regularly use left-hemisphere resources to assess temporal relationships between musical segments (Bever and Chiarello, 2009). This in turn may induce neuroplastic modifications in the left auditory cortex that would support greater temporal processing for speech stimuli in musicians than in non-musicians.

From an electrophysiological perspective (EEG/MEG), musical training or training in the speech domain, such as learning to discriminate slight differences in the fundamental frequency (f_0) of vowels (Reinke et al., 2003) or voice onset time (VOT; Tremblay et al., 2001), has been linked to neuroplastic modifications of the same neural components associated with processing acoustical features common to music and speech. Thus, both forms of training may lead to an increase in overlap of the neural mechanisms that underlie speech and music processing, and thus a transfer of auditory function between the two domains may be possible. A caveat of this assumption is that even if training in speech or music leads to modification of the same neural component(s) measured in EEG/MEG, one cannot necessarily assume that the observed neuroplastic changes occurred in overlapping neural populations. The use of fMRI or source localization techniques in EEG/MEG (Freeman and Nicholson, 1975; Scherg, 1990; Hamalainen and Ilmoniemi, 1994; Tervaniemi et al., 1999) would elucidate the extent of such neural overlap. However, if training in one domain (e.g., music) influences the same neural processes in the other domain (e.g., speech), emerging as a neuroplastic change in the same EEG/MEG components, then an overlap in processing and thus a transfer of function can be inferred. In this review, I will first present evidence demonstrating that musical training and training in the speech domain may target the same neurophysiological components along the auditory pathway. Next, I will show evidence of musical training-related modulations of these same components during speech processing (i.e., transfer of auditory function). Finally, I will discuss how musical training might enhance speech perception in adverse acoustical environments and in individuals with HL.

MUSICAL TRAINING AND TRAINING IN THE SPEECH DOMAIN TARGET SHARED NEURAL MECHANISMS

BRAIN STEM

Training in both musical and speech domains affect brain stem processing, as evidenced by training-related changes in the EEG frequency following response (FFR). The FFR reflects the resonance (amplitude and phase-alignment) of brain stem neuronal firing with a sound's f_0 . Therefore, a more robust FFR following training in music and speech is indicative of an enhanced representation of f_0 at the brain stem (Musacchia et al., 2007; Krishnan et al., 2009). Indeed, the FFR is more robust in musicians for music sounds than in non-musicians (Musacchia et al., 2007). Likewise, in the realm of speech learning, the FFR is enhanced in English speakers for Mandarin tones, following training on tonal speech (Song et al., 2008). Taken together, tonal training, either in music or speech, elicits neuroplastic modifications in the brain stem FFR, which implies that an inter-domain functional transfer at the brain stem level is probable.

PRIMARY AUDITORY CORTEX (A1)

While training in both speech and musical domains target the FFR, there is less evidence for such functional overlap at the level of A1, as indexed by the middle latency response (MLR). The MLR, which is measured in EEG and MEG, is believed to reflect thalamo-cortical input to and processing in A1 (Hall, 2006). The MLR consists of Na/Pa/Nb/P1 components (see **Figure 1**; Note that the P1 is also referred to as the P50) and typically occurs between 19 and 50 ms following sound onset. Larger MLR components were shown to index enhanced pitch and rhythm encoding in musicians relative to non-musicians for pure tones and music sounds (N19m–P30m – magnetic counterparts of the electrical Na–Pa: Schneider et al., 2002; P1: Shahin et al., 2004; Neuhaus and Knösche, 2008). Furthermore, a training-related increase in P1 amplitude may reflect the acoustical feature binding mechanisms that integrate rhythmic and pitch patterns into

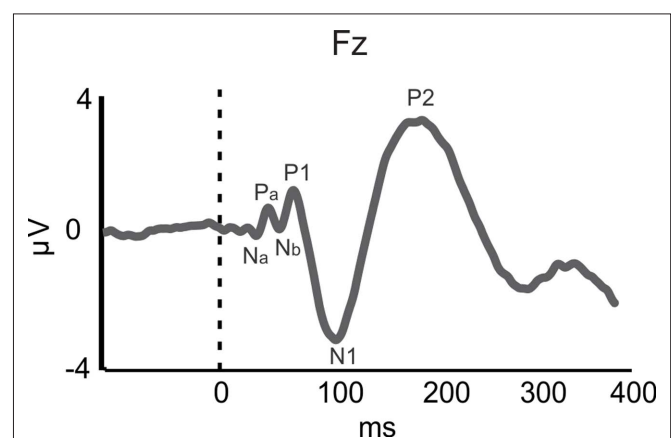


FIGURE 1 | Auditory-evoked potentials (AEPs). AEP waveform for piano tone at the frontal channel Fz for 34 participants (averaged over musicians and non-musicians). The Na–Pa–Nb–P1 complex represents the middle latency response (MLR) originating in primary auditory cortex (A1). The N1 and P2 components represent activity originating in the surrounding belt areas of A1. Based on data from Shahin et al. (2003).

a coherent melody (Neuhaus and Knösche, 2008). However, a recent study challenged the relation between the P1 and musical expertise, such that the P1 was smaller in amplitude and delayed in musicians than in non-musicians for harmonic tones (Nikjeh et al., 2009).

In the realm of speech perception and training in the speech domain, the functional role of the MLR is less clear. Since we are more exposed to voiced than unvoiced speech in naturalistic situations and musical training affects the MLR, one would expect that voiced stimuli would elicit larger MLR amplitude than unvoiced stimuli, which lack pitch information. However, Hertrich et al. (2000) showed that the M50 (the magnetic counterpart of the electrical P1) was reduced for periodic (voiced) compared to aperiodic (unvoiced) speech-like stimuli. Instead, the M100 (the magnetic counterpart of the electrical N1), which has sources in NPAC, was larger for periodic than non-periodic speech-like stimuli (Hertrich et al., 2000). This suggests that experience-based enhancements in pitch processing of speech may commence at a higher cortical level than that of music. Likewise, discrimination training on VOT in young adults resulted in smaller P1 amplitudes but larger N1 and P2 AEPs (Tremblay and Kraus, 2002). Taken together, these results imply that higher cortical regions (i.e., NPAC) favor the processing of more spectrotemporally complex signals, such as speech, consistent with animal (Rauschecker et al., 1995), and human fMRI (Rauschecker et al., 1995; Patterson et al., 2002) results. Thus, whereas enhanced MLR may reflect neuroplastic modifications in temporal (i.e., rhythm) and spectral (i.e., pitch) processing following musical training, analogous MLR effects in the speech domain do not seem to be related to expertise. Instead, expertise-related effects in the speech domain appear to emerge later, in the N1 (~100 ms) and P2 (~180 ms; **Figure 1**) AEPs/AEFs.

NON-PRIMARY AUDITORY CORTEX

Because of the complex spectrotemporal structure of music and speech sounds, their processing may be favored in NPAC. Thus, the possibility of inter-domain functional transfer is more likely to occur in later auditory processing stages (i.e., N1 and P2) instead of the MLR.

The N1 and P2 are thought to originate in the region surrounding A1 (Shahin et al., 2003; Bosnyak et al., 2004), including belt and parabelt regions of the superior temporal gyrus (Hackett et al., 2001). These regions are collectively referred to here as NPAC. While the N1 and P2 code for low-level acoustical features such as sound onset and pitch, they also represent higher-level sound features brought about by the spectrotemporal complexity of speech and music. **Figure 2** shows the P2 AEP as a function of sound complexity. Musicians and non-musicians were presented with four tones varying in spectral complexity. Three of the tones were piano tones (C4 note) which contained the fundamental (f_0) and 8 (Piano 8), 2 (Piano 2), or no (Piano 0) harmonics, and a pure tone (Pure) which had the same f_0 but not envelope of the piano tone. Notice that as more harmonics were added to the piano tone (**Figure 2A**), P2 increased in amplitude [**Figure 2C**; 2 (group) \times 4 (tone) ANOVA; group main effect $F_{(1,14)} = 14$, $p < 0.005$; tone main effect $F_{(3,42)} = 35$, $p < 0.0001$, interaction $F_{(3,42)} = 6$, $p < 0.001$], especially in musicians [*post hoc* Fisher's least significant difference (LSD) test $p < 0.05$]. Also, the P2 amplitude in response to the Piano 0 tone

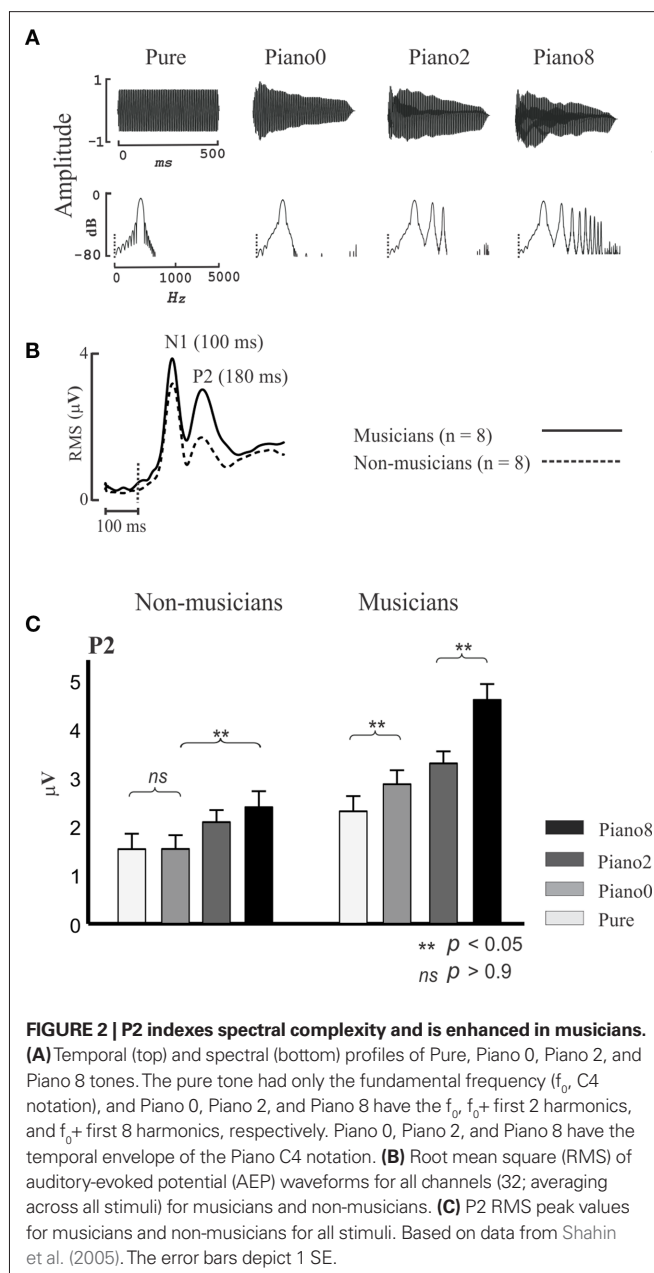
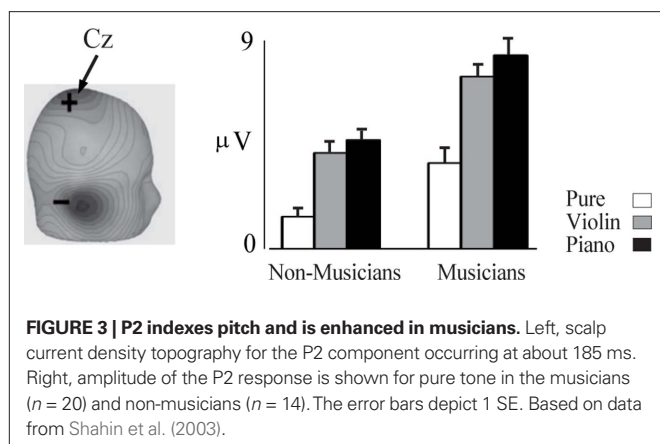


FIGURE 2 | P2 indexes spectral complexity and is enhanced in musicians. (A) Temporal (top) and spectral (bottom) profiles of Pure, Piano 0, Piano 2, and Piano 8 tones. The pure tone had only the fundamental frequency (f_0 , C4 notation), and Piano 0, Piano 2, and Piano 8 have the f_0 , $f_0 +$ first 2 harmonics, and $f_0 +$ first 8 harmonics, respectively. Piano 0, Piano 2, and Piano 8 have the temporal envelope of the Piano C4 notation. (B) Root mean square (RMS) of auditory-evoked potential (AEP) waveforms for all channels (32; averaging across all stimuli) for musicians and non-musicians. (C) P2 RMS peak values for musicians and non-musicians for all stimuli. Based on data from Shahin et al. (2005). The error bars depict 1 SE.

was significantly larger than the P2 for the pure tone in musicians ($p < 0.02$, LSD test) but this difference was not evident in non-musicians. Pure and Piano 0 tones differed only in the shape of their temporal envelopes, with Piano 0's envelope matching the temporal envelope of the piano C4 note, thus implicating the P2 in coding the temporal onset of the learned sound. Finally, although the P2 amplitude for the pure tone tended to be larger in musicians than in non-musicians in this study ($p < 0.07$ LSD test), the same contrast was highly significant when professional musicians were used in a prior study (**Figure 3**). P2 enhancement for pure tones in musicians represents enhanced pitch encoding, even in a non-timbre context, and hence may index basic transfer of auditory function due to musical training. In sum, the P2 is associated with coding the temporal (temporal onset) and spectral (harmonics) features of



music sounds, and these functions are refined by musical training. Like the P2, the N1 was also associated with increased number of harmonics and with coding the timbre in ERP/ERF studies (Pantev et al., 2001; Seither-Preisler et al., 2003; Meyer et al., 2006).

More pertinent to the current topic is the finding that acoustical training in the speech domain has been shown to evoke larger P2s for the same acoustical features that led to P2 augmentation following musical training. Short-term training on vowel discrimination tasks based on slight f_0 manipulations elicited a larger P2 responses in the trained group compared to the untrained group (Reinke et al., 2003), in the same way that pure tones matched in f_0 to music tones evoked a larger P2s in musicians than in non-musicians (Shahin et al., 2003). Also, P2 enhancement was reported following VOT discrimination training (Tremblay et al., 2001, 2009; Tremblay and Kraus, 2002) in the same manner that musicians exhibited larger P2s to the temporal onset of learned music sounds (Figure 2). These increases in P2 amplitude observed following either musical training or training in the speech domain may index a greater overlap of neural populations, via increased neuronal recruitment associated with processing spectral or temporal features of music and speech. In short, if musical training or training in the speech domain can foster neuroplastic changes observed in the same brain response (i.e., P2) associated with processing the same acoustical feature, enhanced speech processing due to musical training is likely.

It is worth noting that the N1's and P2's relationship to coding the spectrotemporal combination of acoustical features (timbre) may suggest a feature binding role for these AEP components. In support of this hypothesis, Neuhaus and Knösche (2008) showed the interdependency of the neuroplastic N1 and P2, as well as the earlier P1 component, on pitch and rhythm processing. They presented musicians and non-musicians with melodies, either ordered by pitch and duration, or randomized in pitch, duration, or both. They found a general effect in which the amplitudes of the P1, N1, and P2 AEPs were augmented as the level of disorder in pitch and/or duration increased in both musicians and non-musicians (Neuhaus and Knösche, 2008). Furthermore, in musicians but not in non-musicians, the P1 and P2 amplitudes were influenced (augmented) by pitch randomization only when the duration was ordered, suggesting that the integration of pitch and rhythm processing is fostered by musical training. Thus, the results imply that an increase

in these AEP amplitudes may reflect increased neural effort associated with binding spectrotemporal features into a coherent melody. Feature binding may be facilitated by selective attention (Treisman and Gelade, 1980) via modulation of early auditory cortical activity. However, increased spectrotemporal acuity observed in musicians (as discussed previously) must be a precursor of enhanced feature binding, whether automatic or under attentive conditions. In short, interdependence of pitch and rhythmic processing and thus binding of these acoustical features may be fostered by musical training at the level of A1 and NPAC (P1, N1, and P2 AEPs). A remaining question is whether enhancements in feature binding mechanisms due to musical training are specific to the learned auditory objects (Pantev et al., 2001) or are also transferable to speech perception.

NEUROPHYSIOLOGICAL EVIDENCE FOR THE INFLUENCE OF MUSICAL TRAINING ON SPEECH PERCEPTION

The previous section highlighted neurophysiological evidence supporting targeted modification of the same neural components (FFR and P2) by musical training and training in the speech domain. These findings do not provide evidence for inter-modal functional transfer, but rather offer a starting point in formulating hypotheses regarding the likelihood of such transfer. In this section, I will present findings showing that musical training can impact speech perception (i.e., inter-modal functional transfer), through neuroplastic changes in the same components that were modified by training in either the music or speech modality (previous section).

MUSICAL TRAINING'S INFLUENCE ON PITCH AND RHYTHM ENCODING IN SPEECH PERCEPTION

Superior pitch and rhythm encoding for speech in musicians, compared to non-musicians, is reflected at the brain stem and auditory cortex.

Brain stem

At the brain stem level, the FFR was shown to be more robust in musicians than in non-musicians for both music and speech (/da/) stimuli (Musacchia et al., 2007, 2008). Furthermore, enhanced pitch tracking (FFR) for Mandarin speech at the brain stem was seen in English-speaking (unfamiliar with tonal languages) musicians compared to non-musicians during passive listening to Mandarin words (Wong et al., 2007). The aforementioned neurophysiological evidence was consistent with better pitch discrimination abilities in musicians compared to non-musicians (Musacchia et al., 2007, 2008; Wong et al., 2007). More precise pitch encoding of speech sounds in musicians at the brain stem and under passive listening conditions (Wong et al., 2007) suggests that functional transfer of auditory mechanisms to speech perception following musical training can occur even at a relatively low level of the auditory pathway. However, corticofugal feedback from the cortex cannot be discounted, especially in attentive conditions (Musacchia et al., 2008).

Auditory cortex

The aforementioned enhanced FFR effect reported by Musacchia et al. (2008) for the speech syllable/da/was accompanied by earlier latencies and larger amplitudes of the components of the P1–N1–P2–N2 complex, which have origins in A1 and NPAC. The P2

enhancement in musicians compared to non-musicians is noteworthy. Musacchia et al.'s (2008) results are consistent with the neurophysiological consideration that if musical training and training in the speech domain impact a shared neural system similarly – as observed via P2 enhancement following musical (Pantev et al., 1998; Shahin et al., 2003) and vowel and VOT discrimination training (Tremblay and Kraus, 2002; Reinke et al., 2003) – then functional transfer to speech perception following musical training can be inferred. A more recent study showed a similar P2 effect that implicated rhythm processing as the mechanism affected by functional transfer following musical training. Specifically, Marie et al. (2010) demonstrated that the neuroplastic P2, previously shown to reflect temporal acuity (Tremblay et al., 2001; Figure 2), is associated with the musicians' greater ability to encode metric structure in speech. Marie et al. (2010) varied syllabic length of the last word in a collection of sentences and measured ERPs while musicians and non-musicians judged whether the last word in a sentence was well-pronounced. They found that compared to non-musicians, musicians exhibited larger P2 amplitudes, coupled with greater accuracy, for metrically incongruous than for congruous words (Marie et al., 2010). While this P2 enhancement is consistent with a transfer effect associated with coding the temporal acuity of speech, it may also be partly attributed to increased perceptual effort to integrate the sequential pattern into a coherent segment (Neuhaus and Knösche, 2008). In sum, these studies demonstrate that P2 enhancement, an index of musical training (Shahin et al., 2003), for coding the metric structure in speech coincides with enhanced temporal (Tremblay et al., 2001; Marie et al., 2010) acuity in speech processing. Thus, a transfer effect for spectrotemporal features, and possibly the neural mechanisms involved in binding them, can be facilitated by musical training.

In sum, there is preliminary neurophysiological evidence, exhibited in several obligatory auditory components (e.g., FFR, P2), supporting a functional transfer to speech perceptual mechanisms following musical training. Moving forward, it is important to weigh the implications of such neurophysiological transfer to speech perception in naturalistic acoustical environments and in individuals with HL.

MUSICAL TRAINING'S INFLUENCE ON SPEECH PERCEPTION IN ADVERSE ACOUSTICAL ENVIRONMENTS

One important aspect of speech processing related to everyday auditory experience, especially for populations with HL, is the ability to comprehend speech in acoustically adverse environments. For example, in a cocktail party, a listener must segregate her friend's voice from a concurrent multi-talker background to understand his story. Musicians are better at concurrent sound segregation (Zendel and Alain, 2009), which entails enhanced pitch discrimination ability (Tervaniemi et al., 2005), working memory function (Chan et al., 1998), and selective attention (Strait et al., 2010). Thus, musically trained individuals should demonstrate improved perception of speech in everyday noisy environments. Indeed, when contrasting speech comprehension abilities of musicians and non-musicians with comparable hearing status, musicians perform better in noisy backgrounds (Parbery-Clark et al., 2009b). Parbery-Clark et al. (2009b) administered the hearing-in-noise test (HINT), the quick speech-in-noise (QuickSIN) test, frequency discrimination

tests, and working memory tests to musicians and non-musicians. HINT tests speech perception in a background speech-shaped noise, while QuickSIN tests speech perceptual ability in a multi-talker environment. Musicians outperformed non-musicians in all tasks. Furthermore, musicians' performance on all tests, except for HINT, correlated with years of practice. These results suggest that enhanced working memory, selective attention, and frequency discrimination abilities as a function of musical expertise play a significant role when the target speech is embedded within a multi-talker environment (QuickSIN) versus when background noise is environmental (HINT).

The aforementioned behavioral outcomes seen in musicians (Parbery-Clark et al., 2009b) were subsequently supported by electrophysiological findings. Parbery-Clark et al. (2009a) showed that the brain stem response morphology was less influenced by background noise in musicians than in non-musicians. They examined the latency and amplitude of two auditory brain stem response (ABR) peaks representing the onset of the consonant–vowel pair/da/ and the formant transition (/d/to/a/) in musicians and non-musicians and in quiet and multi-talker babble (Parbery-Clark et al., 2009a). When comparing results of noisy and quiet conditions, the authors found that the latencies of the ABR peaks were less delayed and the amplitudes were more preserved in musicians than in non-musicians. The authors concluded that the degradative effects of noise on the neural processing of sounds can be limited with musical training.

It is not clear whether musical training can enhance other forms of auditory scene analysis. For example, does musical training boost one's ability to perceive degraded speech, as in phonemic restoration (PR)? In PR, a word with a missing segment can be perceived as continuous, provided that the missing segment is masked by another sound, such as a cough or white noise (Warren, 1970). The ability to restore speech in PR requires a dynamic mechanism in which the missing spectrotemporal structure is interpolated (also known as perceptual “filling-in”) from the masking entity. In addition, A1's usual sensitivity to the onsets/offsets of missing segments (Riecke et al., 2009; Shahin et al., 2009) is suppressed, resulting in the illusory perception of continuous speech. This filling-in process relies on higher-level mechanisms, such as Gestalt processing and access to prior knowledge (e.g., template matching in memory; Shahin et al., 2009). The question then becomes, how would musical training affect PR? Musicians may be expected to perform worse (i.e., fail to fill-in the degraded speech signal) than non-musicians on a PR task, given their enhanced concurrent sound segregation abilities (Zendel and Alain, 2009) and stronger sensitivity to acoustical onsets (Musacchia et al., 2008) and offsets of missing segments. However, since PR strongly relies on top-down influences, it can be argued that musical training may benefit PR.

Two analogies to PR include restoration of the missing fundamental (f_0) and of the missing beat in a metric structure (Bendor and Wang, 2005; Snyder and Large, 2005). Representations of an omitted f_0 in the auditory cortex (Bendor and Wang, 2005) are preserved due to interpolation from the remaining harmonic structure of the sound. Yet this skill (restoration of f_0) has been shown to be more robust in musicians than in non-musicians (Preisler, 1993; Seither-Preisler et al., 2007). However, restoration of a missing beat within a metric structure is more closely related to PR,

since like PR it involves filling-in in the temporal domain (Snyder and Large, 2005). Snyder and Large (2005) showed that when a tone was omitted from a rhythmic train of tones, auditory gamma band activity continued through the gap as if the missing tone was physically present, thus preserving the cortical representations of the rhythmic tone pattern. This temporal filling-in is likely related to enhanced Gestalt integration, expectancy, and template matching in long-term memory, which are all processes important in PR and are enhanced in musicians compared to non-musicians (Besson and Ffytche, 1995; Fujioka et al., 2005; Lenz et al., 2008; Shahin et al., 2008). An experiment assessing the sensitivity of A1 to the onsets and offsets of missing speech segments in musically trained and non-trained individuals should determine how musical training influences PR. Reduced sensitivity of A1 (i.e., reduced MLR amplitudes) in musicians than in non-musicians to onsets/offsets of missing segments would support the premise that musical training enhances PR, thereby providing convincing evidence of the transfer of auditory function to speech perception following musical training.

IMPACT OF MUSICAL TRAINING ON SPEECH PERCEPTION IN INDIVIDUALS WITH HEARING LOSS

As stated earlier, making sense of degraded speech in noisy environments is one of the most critical challenges for individuals with HL. Although individuals with HL (e.g., presbycusis) can typically understand speech in quiet settings (e.g., a one-on-one conversation with little or no background noise), degraded speech or the addition of background noise has a disastrous effect on their speech comprehension. The etiology of HL varies considerably among individuals. In general, HL in the aging population (presbycusis) is characterized by difficulty hearing high-frequency sounds, while individuals with cochlear implants (CIs) must make sense of CI signals that are severely limited in spectral details. If musical training modifies the neural circuitry of speech processing, then how would such modification affect speech perception experience in individuals with HL?

Loss of sensitivity to high-frequency sounds in the aging population leads to a reduced ability to perceive certain sounds, such as the phonemes *s*, *sh*, and *ch*. For example, assuming intact low-frequency perception, the sentence *Trisha cherishes her friend's memories* may sound as *Tri**a**eri**e* her friend** memorie**, where the asterisks indicate unintelligible sounds. Thus, similar to PR, individuals with HL must temporally fill-in representations of degraded speech. If musical training enhances PR mechanisms as discussed above, then musical training could serve as a possible prevention/intervention strategy in individuals with presbycusis. Neurophysiologically, this can be assessed by examining A1 sensitivity to onsets/offsets of degraded fricatives (e.g., *s*, *sh*) in musician and non-musician groups of older individuals with similar HL etiology.

In the case of individuals with CIs, the problem is related to how musical training can improve perception of the spectrally degraded signal outputted by the CI. CIs convert a speech signal into a few (usually eight or more) channels that represent several frequency bands of noise. The signals outputted by CIs can be imitated closely, though not perfectly, by noise-vocoded speech. In vocoded speech, the spectral structure of speech is replaced by noise bands, resulting in a highly degraded spectral structure (i.e., pitch quality), but the

overall amplitude (i.e., temporal envelope) of the sound is preserved to a good degree. It is not surprising that CI users rely on temporal information (rhythm) to a greater extent than a sound's spectral information (pitch; Gfeller and Lansing, 1991). Musical pitch perception in CI users has been shown to be highly correlated with lexical pitch perception, suggesting shared neural mechanisms for processing pitch in music and speech in CI users (Wang et al., 2011). Given that musical training refines one's ability to resolve fine pitch differences in normal-hearing individuals, it would be plausible that CI users who receive musical training may show improvements in pitch interval perception compared to non-musically trained CI individuals. Indeed, a recent study revealed that musically trained children with congenital/prelingual deafness fitted with CIs exhibit better pitch identification than non-trained children (Chen et al., 2011). Furthermore, pitch identification accuracy in the trained children was correlated with the duration of musical training. However, the musical training-related functional transfer to speech perception seen in normal-hearing individuals (discussed above) has not been demonstrated in CI users. A recent longitudinal study using behavioral measures found that while musically trained (over a 2-year period) CI children performed better on pitch discrimination tasks than children who did not receive music lessons, the two groups performed equally on speech discrimination tasks (Yucel et al., 2009). Additional studies using electrophysiological measures may be necessary, as changes due to musical training, especially over short time periods, may emerge more quickly in subcortical and cortical assessments, preceding observed behavioral improvements. This can be examined using pitch-sensitive auditory-evoked components (P2 AEP, Reinke et al., 2003) to assess pitch perception fidelity in speech and speech in noise following musical training in individuals with CIs. Longer-term musical training should also be considered in HL populations.

Another practical approach to explore involves assessing the effects of musical training on audio-visual (AV) speech processing in HL populations. Some individuals with HL enhance their perception of spoken language by relying on visual cues (e.g., lip-reading). In a behavioral study, Kaiser et al. (2003) contrasted the ability to identify spoken words in healthy-hearing individuals and in individuals with CIs. They found that all subjects performed better in the AV task, followed by the auditory-only and finally the visual-only task. Also, the healthy-hearing participants outperformed the CI users on the auditory-only condition, while both groups performed at roughly the same level on the AV condition (Kaiser et al., 2003). Furthermore, healthy-hearing musicians and non-musicians seem to process AV speech differently. Recent findings in normal-hearing populations have shown that pitch encoding at the brain stem and auditory cortex is enhanced in musicians compared to non-musicians for AV speech (Musacchia et al., 2007, 2008). FFR responses and wave d (~18 ms) of the ABR were larger for musicians than for non-musicians for speech stimuli presented in an AV context, and FFR enhancement correlated with years of musical training (Musacchia et al., 2007). Also, the cortical P1–N1–P2–N2 complex was larger in musicians than in non-musicians for the consonant–vowel/da/presented in an AV context. Note that enhancements of all the components within the P1–N1–P2–N2 complex have been previously shown to represent manifestations of musical training (Pantev et al., 1998; Schneider et al., 2002; Shahin

et al., 2003; Fujioka et al., 2006). These ABR and cortical enhancements in musicians may reflect superior sensitivity to onsets of speech sounds when modulated by meaningful visual cues, as suggested by Musacchia et al. (2007). Heightened neural sensitivity to sound onsets would improve the identification of word boundaries, a crucial process for word segmentation and thus particularly useful for individuals with HL. Multisensory musical training in a HL population (i.e., watching the instructor's hands during a piano lesson), may not only promote neuroplastic changes in the auditory cortex, but also enhance neurophysiological mechanisms underlying multisensory integration.

Finally, one could ask why musical training? For example, would CI users develop enhanced pitch identification skills following speech discrimination training (Reinke et al., 2003) in the same way following musical training (Chen et al., 2011)? One advantage is that musical training, or even passive listening to music, can provide an emotional experience (Jancke, 2008) unmatched by other acoustical training methods. In a targeted auditory training program, CI users might process sound features more efficiently when enjoying the task, for example, by selecting his/her preferred genre of music (Looi and She, 2010). Targeted auditory training programs have been shown to significantly improve CI users' level of speech recognition (Fu and Galvin, 2008).

GENERAL DISCUSSION

Thus far, I have provided evidence supporting the idea that enhanced basic acoustic (e.g., frequency or rhythm) processing in the brain stem (FFR) and NPAC (P2 AEP) following musical training may be transferable to speech processing. After developing this concept, I then discussed the possibility of applying musical training to individuals with HL to improve their speech perception. These possibilities are exciting, but some limitations of the qualifications of the presented concepts and additional considerations must be discussed here.

The vast majority of studies associating neuroplasticity with musical training have used the musician's brain as a model (Jancke, 2002). A shortcoming of this approach is that distinguished brain attributes seen in musicians may be due to a genetic predisposition, rather than neuroplastic adaptation. Several lines of evidence indicate that this genetic hypothesis is not sufficient to explain these neuroplastic changes observed in musicians. First, enhanced brain responses in musicians are correlated with age of commencement of music lessons, reinforcing the view that musical training is indeed a factor in neuroplastic modifications (Pantev et al., 1998; Schneider et al., 2002; Wong et al., 2007; Lee et al., 2009). Second, short-term acoustical training has been shown to induce neuroplastic effects in the same neural responses that typically distinguish musicians from non-musicians (N1m: Pantev et al., 1998; P2: Shahin et al., 2003). For example, participants showed increased N1m and P2 amplitudes after short-term frequency discrimination training (Menning et al., 2000; Bosnyak et al., 2004). Third, enhanced brain responses in musicians have also been found to be specific for the instrument of practice, implying a neuroplastic effect (Pantev et al., 2001; Neuhaus et al., 2006; Shahin et al., 2008). Finally, several longitudinal studies comparing brain responses in children before and after musical training further underscore the role of musical training on brain plasticity (Fujioka et al., 2006; Shahin

et al., 2008; Hyde et al., 2009; Moreno et al., 2009), with one study providing inconclusive results (Shahin et al., 2004). In Shahin et al. (2004), musically trained children showed larger AEP responses than non-musically trained children before and after commencement of musical training. However, other studies showed enlarged late AEFs/AEPs (N250/N300) in children who took music lessons relative to non-trained children after 1 year (Fujioka et al., 2006), or to art-trained children after 6 months (Moreno et al., 2009) of training. In the study by Moreno et al. (2009), children were randomly assigned to music or art lessons, further strengthening the argument that musical training rather than a genetic predisposition is the primary agent of neuroplastic changes seen in musicians.

Assessing transfer of auditory function following musical training may not be limited to evidence of neuroplastic modifications along the sensory pathway. Processing syntactic violations of musical or speech segments, organizing percepts in working memory, and selecting relevant percepts require the recruitment of higher-level cognitive processes. Syntactic violation processing in music is linked to the early right anterior negativity (ERAN) localized to the left (Broca's region) and right inferior frontal gyri (Maess et al., 2001). Broca's region is also associated with detecting linguistic syntactic violations (Friederici, 2002). The ERAN and ELAN (left-hemisphere speech counterpart) were shown to be larger, and thus more developed, in musically trained children compared to non-trained children following detection of syntactic irregularities in music and speech (Jentschke and Koelsch, 2009). Musically trained children also tended to perform better (with faster reaction time and increased accuracy) than untrained children. Hence, shared mechanisms, as well as a transfer effect, for syntactic violation processing following musical training can be inferred.

It is worth noting that some higher-level auditory functions, such as those associated with the mismatch negativity (MMN), may be specialized for either music or speech processing (Tervaniemi et al., 2000; Tervaniemi and Huotilainen, 2003) and thus may be less susceptible, than FFR and P2 for example, to inter-domain transfer following musical training. The MMN (150–250 ms) is a pre-attentive cognitive process, which has likely sources in A1 and NPAC, as well as inferior frontal cortex (Tervaniemi et al., 2001; Hall, 2006), used to assess the detection of deviancy in an otherwise predictable sequence of events encoded in sensory memory (Naatanen et al., 2007). A larger MMN in musicians than in non-musicians may indicate stronger sensory representations and/or a better ability to predict the next sound in a patterned musical sequence (Koelsch et al., 1999; Fujioka et al., 2004). Currently, the evidence supporting functional overlap for deviant-detection processes associated with the MMN between speech and music domains and the influence of musical training on such overlap are inconclusive (Tervaniemi et al., 1999; Lidji et al., 2009, 2010). A recent study compared the MMN response dynamics for pure tone, music, and speech stimuli in musicians and non-musicians and found that the MMN occurred earlier in musicians than in non-musicians for all types of stimuli (Nikjeh et al., 2009). This suggests that musical training may lead to faster pre-attentive deviant detection, even for speech stimuli. However, there were no group differences in MMN amplitude for any of the stimuli. This is consistent with Tervaniemi et al. (2009) who similarly failed to find amplitude differences in the MMN between musicians and non-musicians for speech stimuli with

deviants in either frequency or duration. However, the MMN was larger in musicians than in non-musicians when participants paid attention to the speech sounds (Tervaniemi et al., 2009). These two studies (Nikjeh et al., 2009; Tervaniemi et al., 2009) suggest that while the MMN mechanisms (at least when considering its amplitude dynamics) may be specialized for music and speech processing, musical training can shape attentional processes (see paragraph below) that could facilitate functional inter-domain transfer for these same neural mechanisms.

One important consideration when assessing transfer effects involves how musical training shapes selective attention mechanisms, a necessary process in noisy or crowded auditory scene situations (Parbery-Clark et al., 2009b; Kerlin et al., 2010). Enhanced selective attention abilities may: (1) impact how relevant and irrelevant signals are organized in working memory (Sreenivasan and Jha, 2007), and (2) act as an auditory gain function in which the relevant acoustical signal intensity is dialed up and the interfering noise/talker is dialed down (Kerlin et al., 2010). The acoustic feature selectively attended to during training shapes cortical map specificity to sound features in A1 (Polley et al., 2006; Engineer et al., 2008). For example, rats trained to attend to frequency cues in sounds showed modifications of the cortical map in A1 corresponding to the trained frequency range only, while rats trained to attend to intensity cues showed modifications of the cortical map to the target intensity range only (Polley et al., 2006). Similarly, during musical training, trainees must focus on or direct attention to slight changes in pitch, intensity, and onsets/offsets of sounds, developing acoustical acuity in the temporal and spectral domains (Schneider et al., 2002; Marie et al., 2010). In turn, enhanced acoustical representations facilitate auditory object formation (acoustical feature binding) and thus allow for better selection and analysis of the acoustic scene (Shinn-Cunningham and Best, 2008). In other words, musicians' improved acuity to sound features, resulting in more veridical auditory object representations, may also shape their selective attention mechanisms, and thus improve their auditory scene analysis skills. Indeed, Strait et al. (2010) revealed that musicians perform better (have faster reaction times) than non-musicians in tasks requiring focused attention in the auditory but not in the visual modality. Additionally, attentive listening to music recruits general cognitive functions related to working memory, semantic processing, and target detection (Janata et al., 2002), which are all processes that influence auditory scene analysis. In sum, enhanced selective attention and other

top-down mechanisms that result from musical training may lead to superior concurrent sound segregation, which is an essential process during perception of speech in noisy environments, especially in individuals with HL.

CONCLUSION

While evidence of neuroplastic adaptations due to musical training is accumulating, the extent to which these neurophysiological changes transfer to speech perception remains inconclusive. Preliminary evidence suggests that musical training may influence a shared hierarchical auditory network underlying music and speech processing and thus can influence speech perception. Further studies are necessary to examine or clarify functional transfer due to musical training along different levels of the auditory pathway (e.g., MLR, MMN) and whether potential transfer effects are solely due to the overlap in acoustical features or also to the overlap in mechanisms that binds them. A possible limitation of this research is that much of the supporting evidence has come from the study of neuroplasticity in musicians, where genetically or ontogenetically determined acoustical abilities could, in principle, have influenced the decision to train musically. However, the dependence of enhanced brain responses in musicians on the duration of training and specific musical experience in addition to evidence provided by longitudinal studies, point to neuroplasticity as the crucial mechanism. Furthermore, future research addressing behavioral and neurophysiological influences of musical training on speech perception in adverse acoustical environments and in individuals with HL is necessary and timely. These topics go hand-in-hand, as speech in background noise creates a debilitating acoustical experience in individuals with HL. Initial evidence shows that musical training enhances pitch perception in individuals with HL. Moreover, musical training is correlated with enhanced behavioral and neurophysiological responses for speech in noise. These exciting preliminary findings provide a stepping stone toward future studies addressing the neurophysiological effects of musical training on speech perception in individuals with HL, which may influence our approach to devising targeted auditory training programs and thus prevention and intervention strategies for HL.

ACKNOWLEDGMENTS

The author thanks Kristina Backer, Dr. Larry Roberts, and Dr. Susan Nittrouer for their helpful comments.

REFERENCES

- Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature* 436, 1161–1165.
- Besson, M., and Ffytche, F. (1995). An event-related potential (ERP) study of musical expectancy: comparison of musicians with non-musicians. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 1278–1296.
- Bever, T. G., and Chiarello, R. J. (2009). Cerebral dominance in musicians and nonmusicians. 1974. *J. Neuropsychiatry Clin. Neurosci.* 21, 94–97.
- Bosnyak, D. J., Eaton, R. A., and Roberts, L. E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cereb. Cortex* 14, 1088–1099.
- Chan, A. S., Ho, Y. C., and Cheung, M. C. (1998). Music training improves verbal memory. *Nature* 396, 128.
- Chen, J. K., Chuang, A. Y., McMahon, C., Hsieh, J. C., Tung, T. H., and Li, L. P. (2011). Music training improves pitch perception in prelingually deafened children with cochlear implants. *Pediatrics*, 125: e793–800.
- Engineer, C. T., Perez, C. A., Chen, Y. H., Carraway, R. S., Reed, A. C., Shetake, J. A., Jakkamsetti, V., Chang, K. Q., and Kilgard, M. P. (2008). Cortical activity patterns predict speech discrimination ability. *Nat. Neurosci.* 11, 603–608.
- Freeman, J. A., and Nicholson, C. (1975). Experimental optimization of current source-density technique for anuran cerebellum. *J. Neurophysiol.* 38, 369–382.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci. (Regul. Ed.)* 6, 78–84.
- Fu, Q. J., and Galvin, J. J. III. (2008). Maximizing cochlear implant patients' performance with advanced speech training procedures. *Hear. Res.* 242, 198–208.
- Fujioka, T., Ross, B., Kakigi, R., Pantev, C., and Trainor, L. J. (2006). One year of musical training affects development

- of auditory cortical-evoked fields in young children. *Brain* 129, 2593–2608.
- Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., and Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *J. Cogn. Neurosci.* 16, 1010–1021.
- Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., and Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J. Cogn. Neurosci.* 17, 1578–1592.
- Gfeller, K., and Lansing, C. R. (1991). Melodic, rhythmic, and timbral perception of adult cochlear implant users. *J. Speech Hear. Res.* 34, 916–920.
- Hackett, T. A., Preuss, T. M., and Kaas, J. H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J. Comp. Neurol.* 441, 197–222.
- Hall, J. W. (2006). *New Handbook of Auditory Evoked Responses*. Pearson, Boston.
- Hamalainen, M. S., and Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32, 35–42.
- Hertrich, I., Mathiak, K., Lutzenberger, W., and Ackermann, H. (2000). Differential impact of periodic and aperiodic speech-like acoustic signals on magnetic M50/M100 fields. *Neuroreport* 11, 4017–4020.
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., and Schlaug, G. (2009). Musical training shapes structural brain development. *J. Neurosci.* 29, 3019–3025.
- Janata, P., Tillmann, B., and Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cogn. Affect. Behav. Neurosci.* 2, 121–140.
- Jancke, L. (2002). What is special about the brains of musicians? *Neuroreport* 13, 741–742.
- Jancke, L. (2008). Music, memory and emotion. *J. Biol.* 7, 21.
- Jentschke, S., and Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *Neuroimage* 47, 735–744.
- Kaiser, A. R., Kirk, K. I., Lachs, L., and Pisoni, D. B. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *J. Speech Lang Hear Res.* 46, 390–404.
- Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a “cocktail party”. *J. Neurosci.* 30, 620–628.
- Koelsch, S., Schroger, E., and Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport* 10, 1309–1313.
- Krishnan, A., Swaminathan, J., and Gandour, J. T. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J. Cogn. Neurosci.* 21, 1092–1105.
- Lee, K. M., Skoe, E., Kraus, N., and Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *J. Neurosci.* 29, 5832–5840.
- Lenz, D., Jeschke, M., Schadow, J., Naue, N., Ohl, F. W., and Herrmann, C. S. (2008). Human EEG very high frequency oscillations reflect the number of matches with a template in auditory short-term memory. *Brain Res.* 1220, 81–92.
- Lidji, P., Jolicoeur, P., Kolinsky, R., Moreau, P., Connolly, J. F., and Peretz, I. (2010). Early integration of vowel and pitch processing: a mismatch negativity study. *Clin. Neurophysiol.* 121, 533–541.
- Lidji, P., Jolicoeur, P., Moreau, P., Kolinsky, R., and Peretz, I. (2009). Integrated preattentive processing of vowel and pitch: a mismatch negativity study. *Ann. N. Y. Acad. Sci.* 1169, 481–484.
- Looi, V., and She, J. (2010). Music perception of cochlear implant users: a questionnaire, and its implications for a music training program. *Int. J. Audiol.* 49, 116–128.
- 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.
- Marie, C., Magne, C., and Besson, M. (2010). Musicians and the metric structure of words. *J. Cogn. Neurosci.* 23, 294–305.
- Menning, H., Roberts, L. E., and Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *Neuroreport* 11, 817–822.
- Meyer, M., Baumann, S., and Jancke, L. (2006). Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. *Neuroimage* 32, 1510–1523.
- 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.
- Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15894–15898.
- Musacchia, G., Strait, D., and Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear. Res.* 241, 34–42.
- Naatanen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590.
- Neuhaus, C., and Knösche, T. R. (2008). Processing of pitch and time sequences in music. *Neurosci. Lett.* 441, 11–15.
- Neuhaus, C., Knösche, T. R., and Friederici, A. D. (2006). Effects of musical expertise and boundary markers on phrase perception in music. *J. Cogn. Neurosci.* 18, 472–493.
- Nikjeh, D. A., Lister, J. J., and Frisch, S. A. (2009). Preattentive cortical-evoked responses to pure tones, harmonic tones, and speech: influence of music training. *Ear Hear.* 30, 432–446.
- Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., Katoh, A., and Imabayashi, E. (2001). Functional anatomy of musical perception in musicians. *Cereb. Cortex* 11, 754–760.
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature* 392, 811–814.
- Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., and Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport* 12, 169–174.
- Parbery-Clark, A., Skoe, E., and Kraus, N. (2009a). Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* 29, 14100–14107.
- Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009b). Musician enhancement for speech-in-noise. *Ear Hear.* 30, 653–661.
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36, 767–776.
- Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., and Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain* 117(Pt 6), 1283–1301.
- Polley, D. B., Steinberg, E. E., and Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *J. Neurosci.* 26, 4970–4982.
- Preisler, A. (1993). The influence of spectral composition of complex tones and of musical experience on the perceptibility of virtual pitch. *Percept. Psychophys.* 54, 589–603.
- Rauschecker, J. P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268, 111–114.
- Reinke, K. S., He, Y., Wang, C., and Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Brain Res. Cogn. Brain Res.* 17, 781–791.
- Riecke, L., Esposito, F., Bonte, M., and Formisano, E. (2009). Hearing illusory sounds in noise: the timing of sensory-perceptual transformations in auditory cortex. *Neuron* 64, 550–561.
- 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.
- Sammter, D., Koelsch, S., Ball, T., Brandt, A., Elger, C. E., Friederici, A. D., Grigutsch, M., Huppertz, H. J., Knosche, T. R., Wellmer, J., Widman, G., and Schulze-Bonhage, A. (2009). Overlap of musical and linguistic syntax processing: intracranial ERP evidence. *Ann. N. Y. Acad. Sci.* 1169, 494–498.
- Scherg, M. (1990). “Fundamentals of dipole source potential analysis,” in *Auditory Evoked Magnetic Fields and Electric Potentials*, Vol. 6, *Advances in Audiology*, eds F. Grandori, M. Hoke, and G. L. Romani (Basel: Karger), 40–69.
- Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* 5, 688–694.
- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., and Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in non-musicians and musicians. *Hum. Brain Mapp.* 32, 771–783.
- Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R., Seither, S., and Lutkenhoner, B. (2007). Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 743–751.
- Seither-Preisler, A., Krumbholz, K., and Lutkenhoner, B. (2003). Sensitivity of the neuromagnetic N100m deflection to spectral bandwidth: a function of the auditory periphery? *Audiol. Neurootol.* 8, 322–337.
- Shahin, A., Bosnyak, D. J., Trainor, L. J., and Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory

- evoked potentials in musicians. *J. Neurosci.* 23, 5545–5552.
- Shahin, A., Roberts, L. E., and Trainor, L. J. (2004). Enhancement of auditory cortical development by musical experience in children. *Neuroreport* 15, 1917–1921.
- Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., and Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport* 16, 1781–1785.
- Shahin, A. J., Bishop, C. W., and Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *Neuroimage* 44, 1133–1143.
- Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., and Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *Neuroimage* 41, 113–122.
- Shinn-Cunningham, B. G., and Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplif.* 12, 283–299.
- Snyder, J. S., and Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Brain Res. Cogn. Brain Res.* 24, 117–126.
- Song, J. H., Skoe, E., Wong, P. C., and Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *J. Cogn. Neurosci.* 20, 1892–1902.
- Sreenivasan, K. K., and Jha, A. P. (2007). Selective attention supports working memory maintenance by modulating perceptual processing of distractors. *J. Cogn. Neurosci.* 19, 32–41.
- Strait, D. L., Kraus, N., Parbery-Clark, A., and Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear. Res.* 261, 22–29.
- Tervaniemi, M., and Huottilainen, M. (2003). The promises of change-related brain potentials in cognitive neuroscience of music. *Ann. N. Y. Acad. Sci.* 999, 29–39.
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., and Schroger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Exp. Brain Res.* 161, 1–10.
- Tervaniemi, M., Kruck, S., De Baene, W., Schroger, E., Alter, K., and Friederici, A. D. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *Eur. J. Neurosci.* 30, 1636–1642.
- Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R. J., and Naatanen, R. (1999). Functional specialization of the human auditory cortex in processing phonetic and musical sounds: a magnetoencephalographic (MEG) study. *Neuroimage* 9, 330–336.
- Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., Van Zuijen, T. L., and Naatanen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Hum. Brain Mapp.* 10, 74–79.
- Tervaniemi, M., Rytönen, M., Schroger, E., Ilmoniemi, R. J., and Naatanen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learn. Mem.* 8, 295–300.
- Tervaniemi, M., Szameitat, A. J., Kruck, S., Schroger, E., Alter, K., De Baene, W., and Friederici, A. D. (2006). From air oscillations to music and speech: functional magnetic resonance imaging evidence for fine-tuned neural networks in audition. *J. Neurosci.* 26, 8647–8652.
- Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136.
- Tremblay, K., Kraus, N., McGee, T., Ponton, C., and Otis, B. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear Hear.* 22, 79–90.
- Tremblay, K. L., and Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *J. Speech Lang. Hear. Res.* 45, 564–572.
- Tremblay, K. L., Shahin, A. J., Picton, T., and Ross, B. (2009). Auditory training alters the physiological detection of stimulus-specific cues in humans. *Clin. Neurophysiol.* 120, 128–135.
- Wang, W., Zhou, N., and Xu, L. (2011). Musical pitch and lexical tone perception with cochlear implants. *Int. J. Audiol.* 50, 270–278.
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science* 167, 392–393.
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.
- Yucel, E., Sennaroglu, G., and Belgin, E. (2009). The family oriented musical training for children with cochlear implants: speech and musical perception results of two year follow-up. *Int. J. Pediatr. Otorhinolaryngol.* 73, 1043–1052.
- Zatorre, R. J., and Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11, 946–953.
- Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci. (Regul. Ed.)* 6, 37–46.
- Zendel, B. R., and Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *J. Cogn. Neurosci.* 21, 1488–1498.

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.

Received: 23 February 2011; accepted: 30 May 2011; published online: 13 June 2011.

Citation: Shahin AJ (2011) Neurophysiological influence of musical training on speech perception. *Front. Psychology* 2:126. doi: 10.3389/fpsyg.2011.00126

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

Copyright © 2011 Shahin. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.