

Musical experience modulates categorical perception of lexical tones in native Chinese speakers

Han Wu¹, Xiaohui Ma¹, Linjun Zhang^{2*}, Youyi Liu¹, Yang Zhang³ and Hua Shu^{1*}

¹ State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China, ² Faculty of Linguistic Sciences, Beijing Language and Culture University, Beijing, China, ³ Department of Speech-Language-Hearing Sciences and Center for Neurobehavioral Development, University of Minnesota, Minneapolis, MN, USA

OPEN ACCESS

Edited by:

Isabelle Peretz,
Université de Montréal, Canada

Reviewed by:

Peter Pfordresher,
University at Buffalo State University
of New York, USA
Patrick Wong,
The Chinese University of Hong Kong,
Hong Kong

*Correspondence:

Hua Shu,
State Key Laboratory of Cognitive
Neuroscience and Learning, Beijing
Normal University, No. 19 Xijiekouwai
Street, Beijing 100875, China
shuh@bnu.edu.cn;

Linjun Zhang,
Faculty of Linguistic Sciences, Beijing
Language and Culture University,
No. 15 Xueyuan Road, Beijing
100083, China
zhanglinjun75@gmail.com

Specialty section:

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

Received: 03 January 2015

Accepted: 27 March 2015

Published: 13 April 2015

Citation:

Wu H, Ma X, Zhang L, Liu Y, Zhang Y
and Shu H (2015) Musical experience
modulates categorical perception
of lexical tones in native Chinese
speakers.
Front. Psychol. 6:436.
doi: 10.3389/fpsyg.2015.00436

Although musical training has been shown to facilitate both native and non-native phonetic perception, it remains unclear whether and how musical experience affects native speakers' categorical perception (CP) of speech at the suprasegmental level. Using both identification and discrimination tasks, this study compared Chinese-speaking musicians and non-musicians in their CP of a lexical tone continuum (from the high level tone, Tone1 to the high falling tone, Tone4). While the identification functions showed similar steepness and boundary location between the two subject groups, the discrimination results revealed superior performance in the musicians for discriminating within-category stimuli pairs but not for between-category stimuli. These findings suggest that musical training can enhance sensitivity to subtle pitch differences between within-category sounds in the presence of robust mental representations in service of CP of lexical tonal contrasts.

Keywords: musical training, categorical perception, Chinese lexical tone, within-category discrimination, between-category discrimination

Introduction

Categorical perception (CP), which refers to the phenomenon that gradually morphed sounds in a stimulus continuum tend to be perceived as discrete representations, has been studied for more than 50 years. While most of the early CP studies focused on segmental features (consonants and vowels; Liberman et al., 1957, 1961; Fry et al., 1962), there has been a recent surge of interests in suprasegmental features such as vowel duration contrasts of quantity languages (Nenonen et al., 2003; Ylinen et al., 2005) and lexical tone contrasts of tonal languages (Francis et al., 2003; Hallé et al., 2004; Xu et al., 2006; Xi et al., 2010). Evidence shows that the categorical nature of lexical tone perception depends on the pitch trajectory. For a continuum of flat tones ranging from one level to another, it is not categorically perceived (Francis et al., 2003), whereas a continuum involving contour tones is perceived categorically (Wang, 1976; Xu et al., 2006). In Mandarin Chinese there are four lexical tones, only one of which is a level tone (Howie, 1976). Thus, native perception of any Mandarin Chinese tonal continuum, due to the necessary involvement of contour tones, tends to be categorical with better sensitivity to between-category distinction relative to within-category differences. In recent years, much research interest has been oriented toward understanding the relationship between musical training and speech perception because music and speech share many acoustic commonalities as well as cognitive mechanisms. In fact, both signals convey information by means of timing, pitch, and timbre cues. Of these cues, pitch is of special interest because of its important roles in both domains. In music, two types of pitch information, i.e., contour

and interval, are necessary to create melodies. Early studies have established that musicianship influences CP performance for pitch-related and duration-related auditory stimuli. For instance, the judgment of tonal intervals is shown to be more categorical in professionally trained musicians than non-musicians (Locke and Kellar, 1973; Siegel and Siegel, 1977; Burns and Ward, 1978). The evidence of categorical processing of musical pitch suggests that CP extends to signals other than speech and that it may be acquired from a special learning experience even when there are no natural sensory cues or physiologically built-in auditory discontinuities available to the listener (See Harnad, 1987 for discussions on theoretical explanations for CP). In speech, variations in pitch constitute essential prosodic patterns as associated with stress and intonational structure. Furthermore, in a tonal language like Chinese, pitch is also exploited to distinguish phonological contrasts at the syllable level. The advantage of musicianship in pitch processing has been confirmed by many previous studies. For example, compared to non-musicians, musicians are more skilled at learning to use non-native tonal contrasts to distinguish word meanings (Wong and Perrachione, 2007), and they are also more sensitive to pitch rises on final words of utterances irrespective of whether such changes occurred in their native language (Magne et al., 2006) or in a foreign language (Marques et al., 2007). Moreover, musicians are superior to non-musicians in discriminating subtle differences (Marie et al., 2012) and tracking accuracy (Wong et al., 2007) of pitch trajectories of both speech and non-speech sounds. In a recent study, it was reported that musicians showed greater accuracy and reaction time consistency than non-musicians for all types of stimuli in a discrimination task involving lexical tones, low-pass filtered speech tones, and violin sounds that carry the pitch contrasts of lexical tones (Burnham et al., 2014).

While the previous studies provide evidence for the transfer effects of long-term musical training on pitch processing of speech, it is theoretically important to investigate the extent of overlapping or domain specificity regarding the enhancement effects to better understand the perceptual and cognitive mechanisms involved in language and music processing. Auditory cognitive neuroscience models for language processing all posit that speech perception involves multiple neural representations along the auditory pathway with dedicated low- and high-level neural structures performing acoustic analysis and phonological processing, respectively, that can be shaped and reshaped by learning experience (Hickok and Poeppel, 2007; Friederici, 2011). The study of lexical tone processing in the context of presence or absence of musical training for tonal-language speakers provides a great opportunity to explore the relationship between processing of speech and music. It remains unclear whether and how musical training affects CP of lexical tones by native tonal language speakers because the previous studies focused on pitch perception of musicians and non-musicians who are non-tonal language speakers. As a matter of fact, it is hotly debated whether the facilitatory effects of musical training on native language speech perception are attributed to musicians' higher sensitivity to specific acoustic features or enhanced internal representation of phonological categories (Patel, 2011; Sadakata and Sekiyama, 2011; Marie et al., 2012; Kühnis et al., 2013; Bidelman et al.,

2014). Previous research has provided robust evidence indicating that compared to non-musicians, musicians are more sensitive to various acoustic properties other than pitch (Milovanov et al., 2009; Marie et al., 2011, 2012; Sadakata and Sekiyama, 2011; Kühnis et al., 2013). However, whether the facilitatory effects extend to the additional higher-order phonological operation is not well known. CP of speech sounds arguably provides an optimal window for investigating whether and how musical training affects acoustic and/or linguistic processing of native language speech because both acoustic and phonological processing levels are involved in CP processes that perceptually evaluate carefully controlled within-category and between-category differences.

In the present study, identification and discrimination tasks were adopted to compare the performance of Chinese-speaking musicians and non-musicians in their CP of the lexical tone continuum from Tone1 to Tone4. Considering previous results reporting that musicians are superior in acoustic processing of pitch information, we predicted that musicians would be more accurate at discriminating within-category pairs. In terms of linguistic processing of between-category stimuli, we considered it an open-ended question whether musicians would have greater accuracies than non-musicians. Given the fact that CP of pitch direction by native Chinese speakers is also influenced by stimulus complexity (speech vs. non-speech; Xu et al., 2006), sine-wave tones re-synthesized from the Tone1 to Tone4 continuum were also included in order to explore domain specificity of speech processing and examine whether the stimulus complexity effect could also be modulated by musical experience.

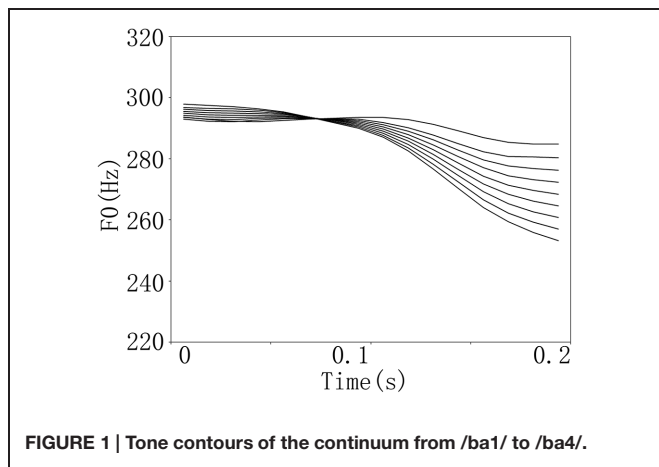
Materials and Methods

Subjects

Sixty-four native speakers of Mandarin Chinese participated in the experiment: Thirty-two musicians (18 female, 14 male; mean age = 19.2, range 17–23) and 32 non-musicians (20 female, 12 male; mean age = 19.7, range 18–25). The musicians had undergone at least seven years of continuous formal western instrumental music training, and they had regular practice and current opportunities to play an instrument. The non-musicians were selected with the inclusion criterion that they had never received formal musical training within the last five years and less than 2 years of musical experience prior to that (Wong et al., 2007; Wayland et al., 2010). The two groups of subjects were, respectively, recruited from China Conservatory of Music and Beijing Normal University. They reported having no history of hearing impairment, neurological, psychiatric, or neuropsychological problems. All signed an informed consent in compliance with a protocol approved by the research ethics committee of Beijing Normal University and were paid for their participation.

Stimuli

There were two sets of continua for the speech and non-speech stimuli (**Figure 1**). The speech stimuli were the Chinese monosyllables /pa/ that differed in their lexical tones (i.e., the high level tone, Tone1, and the high falling tone, Tone4). The original



stimuli were first recorded at a sampling rate of 44.1 kHz from a female native Chinese speaker. The syllables were then digitally edited to have a duration of 200 ms using Sound-Forge (SoundForge9, Sony Corporation, Japan). In order to isolate the lexical tones and keep the rest of the acoustic features equivalent, pitch tier transfer was performed using the Praat software (<http://www.fon.hum.uva.nl/praat/>). This procedure generated two stimuli, /pa1/ and /pa4/, which were identical with each other except for the pitch contour difference. The /pa1/ and /pa4/ stimuli were then taken as the endpoint stimuli to create a lexical tone continuum. A morphing technique was performed in Matlab (Mathworks Corporation, USA) using STRAIGHT (Kawahara et al., 1999) in eight equal intervals. The non-speech stimuli were sine-wave tones with the same pitch contours as the speech stimuli. All stimuli were normalized in RMS intensity.

Procedures

Categorical perception of the speech and non-speech continua was examined in discrimination and identification tasks. In order to avoid the possible effects that perception of one stimulus type might have on the other due to familiarity with the experimental procedure, we included a 6-month interval between speech and non-speech tests. Furthermore, the use of discrimination/identification tasks and speech/non-speech stimuli were counterbalanced among the participants in each subject group to control effects due to stimulus presentation sequence.

In the identification task, participants listened to stimuli from the speech/non-speech continuum presented in isolation. They were instructed to press the “F” key upon hearing a “level” pitch, or the “J” key on a computer keyboard upon hearing a “falling” pitch. There were six occurrences of each of the nine stimuli (54 trials) presented in random order. The rate of presentation was self-paced with 1-s pause after response.

In the discrimination task, two stimuli were presented consecutively with a 500-ms interstimulus interval (ISI) and participants were instructed to judge whether the two sounds were identical or not by pressing the “1” (same) or 2 (different) key on the keyboard. There were three occurrences of each of the fourteen 2-step pairs (42 trials) randomly presented, either in forward (1–3, 2–4, 3–5, 4–6, 5–7, 6–8, 7–9) or reverse order (3–1, 4–2, 5–3,

6–4, 7–5, 8–6, 9–7). Foil trials involving identical stimuli in the pairwise presentation were also included. The rate of presentation was self-paced as in the identification task.

Data Analysis

To investigate the effects of group (musician vs. non-musician) and stimulus type (speech vs. non-speech) on identification and discrimination performance, five essential characteristics of CP, i.e., sharpness and location of category boundary, between- and within-category discrimination accuracy, and peakedness of discrimination function were calculated (Xu et al., 2006).

Logistic regression between identification score and step number was used to obtain the identification function. The Generalized Estimating Equations estimated regression coefficient b_1 was used to evaluate the slope of the fitted logistic curve which is an indication of the sharpness of the categorical boundary (Liang and Zeger, 1986). The categorical boundary location was derived from the value of step number (X_{cb}) corresponding to the 50% identification score.

The obtained discrimination data were examined by three different measures: between-category discrimination accuracy (P_{bc}), which was measured from the comparison unit corresponding to the categorical boundary (X_{cb}) determined from the subgroup identification functions ($P_{bc} = P_{46}$); within-category discrimination accuracy (P_{wc}), which was the average of the two comparison units at the ends of the continuum (P_{13} and P_{79}); and peakedness of the discrimination function (P_{pk}), estimated by P_{bc} minus P_{wc} . All the discrimination measures were obtained by computing the proportion of “different” responses.

Results

Table 1 shows the estimated regression coefficients for the mean logistic response functions. **Table 2** shows the d -prime measures for discrimination of each subject group and stimulus type. The identification and discrimination curves are depicted in **Figure 2**.

TABLE 1 | GEE estimates of regression coefficients (b_0 , b_1) and the derived categorical boundary (X_{cb}) for each subgroup (Musician and Non-musician Groups; Speech and Non-speech conditions).

Subgroup	b_0	b_1	$X_{cb} = -b_0/b_1$
MusicianSP	6.8277	-1.4527	4.7216
MusicianNonSP	6.2251	-1.2193	5.0771
Non-musicianSP	7.3825	-1.7135	4.3619
Non-musicianNonSP	7.3710	-1.3778	5.2870

TABLE 2 | d -prime measures for discrimination of each subgroup (Musician and Non-musician Groups; Speech and Non-speech conditions).

Subgroup	P_{bc} d -prime	P_{wc} d -prime
MusicianSP	1.60 (0.27)	0.50 (0.14)
MusicianNonSP	1.59 (0.22)	0.61 (0.18)
Non-musicianSP	1.68 (0.24)	0.09 (0.18)
Non-musicianNonSP	1.77 (0.19)	0.67 (0.14)

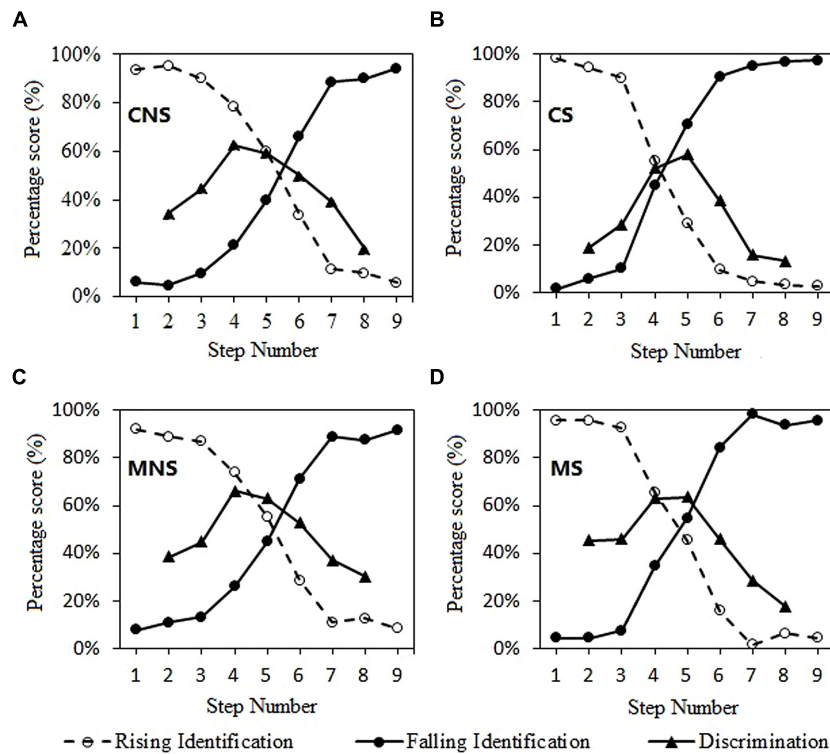


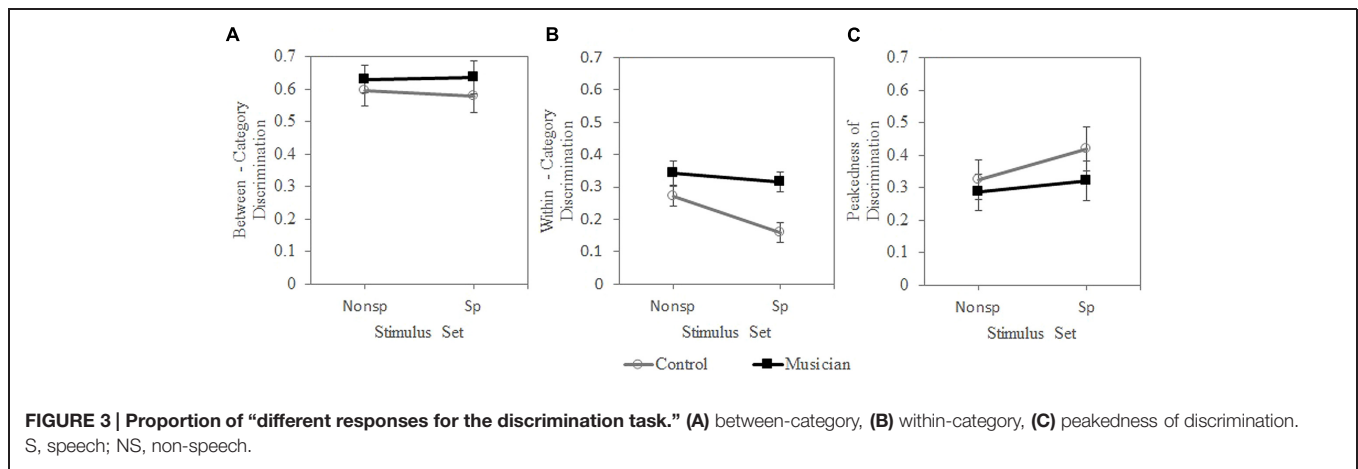
FIGURE 2 | Logistic identification functions and two-step discrimination curves. The “level” logistic response functions were plotted by reflecting the “falling” logistic response functions ($P_{\text{level}} = 1 - P_{\text{falling}}$). The discrimination curves were obtained from proportion of “different responses.” (A) non-musician, non-speech; (B) non-musician, speech; (C) musician, non-speech; (D) musician, speech.

Each of the five measures was analyzed by a two-way ANOVA model for group (musician vs. non-musician) and stimulus type (speech vs. non-speech) effects. The steepness of the category boundary (b1) had a significant main effect of stimulus type [$F(1,62) = 6.575, p = 0.013$], indicating that CP of the speech continuum is stronger than non-speech. There was a significant main effect of stimulus type [$F(1,62) = 25.264, p < 0.001$] and group \times stimulus type interaction effect [$F(1,62) = 4.997, p = 0.029$] for the location of category boundary (X_{cb}). The category boundary of the speech continuum shifted to the high level end compared with the non-speech continuum for both groups. This boundary shift phenomenon was more obvious in the non-musician group.

The within-category discrimination accuracy (P_{wc}) had significant main effects of subject group [$F(1,62) = 11.596, p = 0.001$] and stimulus type [$F(1,62) = 5.227, p = 0.026$]. The group \times stimulus type interaction effect was not significant, indicating that musicians were superior to non-musicians in discriminating the within-category pitch difference in both the speech and non-speech continua. For both groups, non-speech yielded better within-category discrimination accuracy than speech. For both the between-category discrimination accuracy (P_{bc}) and peakedness of the discrimination function (P_{pk}), there was no significant main effect or interaction (Figure 3).

Discussion

While previous work has shown that musical expertise enhances the ability to categorize linguistically relevant sounds of a second language (Wong and Perrachione, 2007; Sadakata and Sekiyama, 2011; Mok and Zuo, 2012), and to discriminate some important acoustic cues of both speech and non-speech sounds (Magne et al., 2006; Marques et al., 2007; Marie et al., 2012; Kühnis et al., 2013), the current results link and extend these results by demonstrating similarities as well as differences in how Chinese musicians and non-musicians perform in the CP of native tonal contrasts. Our identification results showed that it was more obvious in the non-musician group that category boundary of the speech continuum shifted to the high level end compared with the non-speech continuum. Nevertheless, at the subject group level, musicians and non-musicians did not differ in either the locations or sharpness of the category boundaries of both continua. The identification results were further confirmed by the discrimination findings showing that the two groups had similar between-category discrimination accuracy and peakedness of the discrimination function. However, the musicians had higher within-category discrimination accuracy irrespective of the stimulus type. Taken together, these results suggest that musical training affects native Chinese speakers' CP of lexical tones resulting from differential perceptual weighting for the acoustic and linguistic dimensions of processing.



Positive transfer effects of musical training to various cognitive abilities have been confirmed by many previous studies. For example, musical training is associated with an increase in general IQ (Schellenberg, 2006) as well as enhanced working memory (Tierney et al., 2008). Various linguistic abilities such as verbal memory (Ho et al., 2003) and reading (Anvari et al., 2002) have also been shown to correlate with musical skills. However, at the more fundamental level of speech perception, whether musical training enhances the ability to categorize the speech sounds of native language is not well known. To our knowledge, this issue is directly examined only by two studies but the results are inconsistent with each other. Bidelman et al. (2014) found that musicians had a steeper category boundary than non-musicians in a vowel identification task, indicating that musicians have heightened internal representations of native phonological categories, whereas musicians' advantage in identifying native vowels is not found in another study (Sadakata and Sekiyama, 2011). Using both identification and discrimination tasks, the current study showed that musicians and non-musicians did not differ in terms of linguistic operations for between-category identification and discrimination of native speech sounds. We argue that the phonetic inventories along with the phonetic boundaries in the speech continua for the native language were acquired and refined in early development (Zhang et al., 2005), i.e., long before the onset of music lessons, and thus more resistant to change brought about by musical experience. The result that musical training enhanced within-category discrimination accuracy is in accordance with a large number of previous findings that have consistently demonstrated musicians are typically more sensitive to various acoustic properties, in particular, pitch information of pure tones, music, and speech sounds (Wong et al., 2007; Milovanov et al., 2009; Marie et al., 2011, 2012; Sadakata and Sekiyama, 2011; Kühnis et al., 2013). Together with previous findings, our identification and discrimination results seem to indicate that musical training strongly enhances the sensitivity to subtle acoustic difference between within-category sounds, while the perceptual space between phonological contrasts of native language is more robust and less likely to be affected.

Neuroimaging studies have revealed hierarchical brain processing that operates on the flow of information from acoustic

to phonological facets of the speech network, with the upstream areas (e.g., Heschl's gyrus and the superior temporal gyrus) performing initial acoustic analysis and the downstream regions (e.g., the superior temporal sulcus and middle temporal gyrus) performing higher-level phonological processing (Wessinger et al., 2001; Joanisse et al., 2007; Okada et al., 2010; Zhang et al., 2011). The impact of intensive musical training on auditory processing has been well documented, however, functional and structural changes associated with musical training mainly take place in the upstream brain areas, especially Heschl's gyrus and the planum temporale rather than the downstream regions (Pantev et al., 2001; Hyde et al., 2009; Moreno et al., 2009; Elmer et al., 2012). In this regard, our behavioral results are also consistent with the neuroimaging findings.

Finally, our study revealed an effect of stimulus type (speech vs. non-speech) in both identification and discrimination tasks. Compared with speech, perception of the non-speech continuum had a shallower category boundary, higher within-category discrimination accuracy and shifted the boundary location toward the high falling end of the continuum. These results are consistent with an earlier study (Xu et al., 2006) adopting the same tasks and similar stimuli, indicating that stimulus complexity affects the CP of pitch direction of native Chinese speakers. More importantly, only the boundary location showed an interaction effect between stimulus type and group, i.e., different boundary locations between the speech and non-speech continua were only observed in the non-musician group. Such results indicate that musical experience contributes to the approximation of the category boundaries of the speech/non-speech continua, likely by improved fine-grained auditory skills in the professional musicians.

Acknowledgments

This research was supported by Program for New Century Excellent Talents in University (NCET-13-0691) to LZ, the Natural Science Foundation of China (31271082) and the Natural Science Foundation of Beijing (7132119) to HS.

References

- Anvari, S. H., Trainor, L. J., Woodside, J., and Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *J. Exp. Child Psychol.* 83, 111–130. doi: 10.1016/S0022-0965(02)00124-8
- Bidelman, G. M., Weiss, M. W., Moreno, S., and Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* 40, 2662–2673. doi: 10.1111/ejn.12627
- Burnham, D., Brooker, R., and Reid, A. (2014). The effects of absolute pitch ability and musical training on lexical tone perception. *Psychol. Music* doi: 10.1177/0305735614546359
- Burns, E. M., and Ward, W. D. (1978). Categorical perception phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *J. Acoust. Soc. Am.* 63, 456–468. doi: 10.1121/1.381737
- Elmer, S., Meyer, M., and Jäncke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cereb. Cortex* 22, 650–658. doi: 10.1093/cercor/bhr142
- Francis, A. L., Ciocca, V., and Ng, B. K. (2003). On the (non) categorical perception of lexical tones. *Percept. Psychophys.* 65, 1029–1044. doi: 10.3758/BF03194832
- Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392. doi: 10.1152/physrev.00006.2011
- Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Lang. Speech* 5, 171–189. doi: 10.1177/002383096200500401
- Hallé, P. A., Chang, Y. C., and Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *J. Phonetics* 32, 395–421. doi: 10.1016/S0095-4470(03)00016-0
- Harnad, S. (1987). “Psychophysical and cognitive aspects of categorical perception: A critical overview,” in *Categorical Perception: The Groundwork of Cognition*, ed. S. Harnad (New York, NY: Cambridge University Press), 1–52.
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Ho, Y. C., Cheung, M. C., and Chan, A. S. (2003). Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology* 17, 439–450. doi: 10.1037/0894-4105.17.3.439
- Howie, J. M. (1976). *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press.
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., et al. (2009). The effects of musical training on structural brain development. *Ann. N. Y. Acad. Sci.* 1169, 182–186. doi: 10.1111/j.1749-6632.2009.04852.x
- Joanisse, M. F., Zevin, J. D., and McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cereb. Cortex* 17, 2084–2093. doi: 10.1093/cercor/bhl124
- Kawahara, H., Masuda-Katsuse, I., and de Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds. *Speech Commun.* 27, 187–207. doi: 10.1016/S0167-6393(98)00085-5
- Kühnis, J., Elmer, S., Meyer, M., and Jäncke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: an EEG study. *Neuropsychologia* 51, 1608–1618. doi: 10.1016/j.neuropsychologia.2013.04.007
- Liang, K. Y., and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13–22. doi: 10.1093/biomet/73.1.13
- Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368. doi: 10.1037/h0044417
- Liberman, A. M., Harris, K. S., Kinney, J. A., and Lane, H. (1961). The discrimination of relative onset-time of components of certain speech and nonspeech patterns. *J. Exp. Psychol.* 61, 379–388. doi: 10.1037/h0049038
- Locke, S., and Kellar, L. (1973). Categorical perception in a non-linguistic mode. *Cortex* 9, 355–369. doi: 10.1016/S0010-9452(73)80035-8
- Magne, C., Schön, D., and Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *J. Cogn. Neurosci.* 18, 199–211. doi: 10.1162/jocn.2006.18.2.199
- Marie, C., Kujala, T., and Besson, M. (2012). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. *Cortex* 48, 447–457. 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. doi: 10.1162/jocn.2010.21413
- Marques, C., Moreno, S., Castro, S. L., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463. doi: 10.1162/jocn.2007.19.9.1453
- 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. doi: 10.1016/j.neulet.2009.05.063
- Mok, P. K., and Zuo, D. (2012). The separation between music and speech: evidence from the perception of Cantonese tones. *J. Acoust. Soc. Am.* 132, 2711–2720. doi: 10.1121/1.4747010
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex* 19, 712–723. doi: 10.1093/cercor/bhn120
- Neenonen, S., Shestakova, A., Huottilainen, M., and Näätänen, R. (2003). Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Brain Res. Cogn. Brain* 16, 492–495. doi: 10.1016/S0926-6410(03)00055-7
- Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., et al. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cereb. Cortex* 20, 2486–2495. doi: 10.1093/cercor/bhp318
- Pantev, C., Roberts, L. E., Schulz, M., Engelen, A., and Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport* 12, 169–174. doi: 10.1097/00001756-200101220-00041
- 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
- Sadakata, M., and Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: a cross-linguistic study. *Acta Psychol.* 138, 1–10. doi: 10.1016/j.actpsy.2011.03.007
- Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *J. Educ. Psychol.* 98, 457–468. doi: 10.1037/0022-0663.98.2.457
- Siegel, J. A., and Siegel, W. (1977). Categorical perception of tonal intervals: musicians can't tell sharp from flat. *Percept. Psychophys.* 21, 399–407. doi: 10.3758/BF03199493
- Tierney, A. T., Bergeson-Dana, T. R., and Pisoni, D. B. (2008). Effects of early musical experience on auditory sequence memory. *Empir. Musicol. Rev.* 3, 178–186.
- Wang, W. S. Y. (1976). Language change. *Ann. N.Y. Acad. Sci.* 280, 61–72. doi: 10.1111/j.1749-6632.1976.tb25472.x
- Wayland, R., Herrera, E., and Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *J. Phonetics* 38, 654–662. doi: 10.1016/j.wocn.2010.10.001
- Wessinger, C., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., and Rauschecker, J. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J. Cogn. Neurosci.* 13, 1–7. doi: 10.1162/089892901564108
- Wong, P., and Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Appl. Psycholinguist.* 28, 565–585. doi: 10.1017/S0142716407070312
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422. doi: 10.1038/nn1872
- Xi, J., Zhang, L., Shu, H., Zhang, Y., and Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience* 170, 223–231. doi: 10.1016/j.neuroscience.2010.06.077
- Xu, Y., Gandour, J. T., and Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J. Acoust. Soc. Am.* 120, 1063–1074. doi: 10.1121/1.2213572

- Ylinen, S., Shestakova, A., Alku, P., and Huotilainen, M. (2005). The perception of phonological quantity based on durational cues by native speakers, second-language users and nonspeakers of Finnish. *Lang. Speech* 48, 313–338. doi: 10.1177/00238309050480030401
- Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., and Li, P. (2011). Cortical dynamics of acoustic and phonological processing in speech perception. *PLoS ONE* 6:e20963. doi: 10.1371/journal.pone.0020963
- Zhang, Y., Kuhl, P., Imada, T., Kotani, M., and Tohkura, Y. (2005). Effects of language experience: neural commitment to language-specific auditory patterns. *Neuroimage* 26, 703–720. doi: 10.1016/j.neuroimage.2005.02.040

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

Copyright © 2015 Wu, Ma, Zhang, Liu, Zhang and Shu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.