Cross-domain effect of pitch processing in lexical and melodic contexts reflected by non-musicians: An MEG study

Zhaowei Liu¹, Yun Nan², Lingxi Lu³, Wei Cui⁴, Jianqiao Ge^{4*}, Jia-Hong Gao^{4,5,6*}

Pitch processing in language and music is generally considered to engage overlapped neural correlates. Previous studies on musicians showed that the ability of pitch processing in music could be transferred to language. It is known that music training can facilitate neural processing of speech. However, the underlying neural mechanisms of pitch processing in language and music are not fully understood, especially in non-musicians. Using magnetoencephalography (MEG), we presented a pitch anomaly paradigm which consists of language/music phrases ending in either congruous or incongruous tone/pitch to non-musicians. We found distinctive brain activity patterns between two groups of participants with varying (high vs. low) musical pitch perceptual abilities. The brain-behavior results showed a positive correlation between performance of musical pitch tasks and activation of the left frontotemporal cortical regions elicited by lexical tones. Our results suggested that the cross-domain effect of language and music could be generalized to people without formal music training.

Keywords: pitch processing, tonal language, music, non-musicians, MEG

INTRODUCTION

Pitch is a basic acoustic attribute shared by both language and music. Whereas different tones indicate different lexical semantics in tonal languages such as Mandarin Chinese, introducing changes in pitch across time produces melody in music. For decades, the relationship between neural mechanisms underlying pitch processing in the domains of language and music has drawn much research attention. Most of these investigations were based on people with extensive experience in musical training, such as the professional musicians (1, 2). However, the neural mechanism of the cross-domain effect between language and music in the general population still remains unclear.

Decades of neuroimaging studies on musicians have indicated both shared and distinct neural mechanisms underlie the processing of information in language and music (3–5). A previous electroencephalography (EEG) study on Mandarin-speaking musicians suggested that pitch processing is left lateralized for tonal language and right lateralized for music, revealed by the late positive component (6). A follow-up fMRI study found that the right superior temporal gyrus (STG) and the left inferior frontal gyrus (IFG) are engaged in the processing of pitch incongruities across both tonal language and music domains (7). Specifically, it was shown that the right frontotemporal network is engaged in the processing of pitch contour in tonal language (i.e., lexical tone) (8, 9). In the music domain, melody with varying pitch activates the Heschl's gyrus (HG) and planum temporale (PT) bilaterally (10). Nevertheless, how pitch in lexical and melodic contexts is processed in people without formal music training has not been well studied.

Interestingly, the neuroplastic accounts of associations between music training and speech perception were challenged. Recent research found that individuals without formal music training but with higher music perception ability showed stronger frequency-following responses to noise-degraded speech than those with

weaker music perception ability while listening to clear and noise-degraded speech sounds, which is consistent with the increased fidelity of speech encoding and speech-in-noise benefits observed previously in highly trained musicians (11). It has been suggested that the auditory neuroplasticity of music engagement likely involves a layering of both preexisting (nature) and experience-driven (nurture) factors in speech perception.

With the aim to investigate the influence of the music pitch perceptual ability on the processing of tonal speech in non-musicians, we manipulated the tone (or pitch) of the last word (or note) of language (or music) phrases and used an event-related field (ERF) paradigm design to study the interactional neural mechanisms of the pitch processing in the domains of language and music. The Profile of Music Perception Skills (PROMS) test was used to classify non-musicians in terms of their ability in pitch processing into high-scoring and low-scoring groups (12).

We hypothesized that for non-musicians, their preexisting ability on musical pitch processing may also have an impact on the processing of lexical tone in their brain and showed differentiated brain activation patterns in cortical regions that play an essential role in pitch processing across domains, such as the Broca's area and the left temporal lobe. To test this hypothesis, we used MEG to record cortical responses of non-musicians while listening to congruous and incongruous phrases of language and music. Source localization and brain-behavior analysis were further performed to investigate the brain activation patterns of people with different receptive music perception abilities on pitch.

RESULTS

Behavioral results

This study mainly focused on the musical ability of pitch processing. We selected the subtests of melody, tuning, and pitch from the PROMS test. Further, the scores of these three subtests were used to evaluate the musical pitch perception ability. A median (12) split of the total score divided non-musicians into low-scoring ($N = 14$) and high-scoring ($N = 13$) groups. The PROMS pitch processing score was 9.18 ± 1.16 for the low-scoring group, 15.80 ± 1.81 for the high-scoring group (mean $\pm$ SE). In addition, the PROMS total score was 17.70 ± 4.07 for the low-scoring group, and 28.72 ± 3.22 for the high-scoring group. There was a significant difference in terms of both the PROMS pitch processing scores and the PROMS total scores between the high-scoring and low-scoring groups (two-

¹Center for Excellence in Brain Science and Intelligence Technology (Institute of Neuroscience), Chinese Academy of Science, 200031, Shanghai, China. ²State Key Laboratory of Cognitive Neuroscience and Learning, International Data Group (IDG), McGovern Institute for Brain Research, Beijing Normal University, 100875, Beijing, China. ³Center for the Cognitive Science of Language, Beijing Language and Cultural University, 100873, Beijing, China. ⁴Center for MRI Research, Academy for Advanced Interdisciplinary Studies, Peking University, 100871, Beijing, China. ⁵PKU-IDG/McGovern Institute for Brain Research, Peking University, 100871, Beijing, China. ⁶Beijing City Key Laboratory for Medical Physics and Engineering, Institute of Heavy Ion Physics, School of Physics, Peking University, 100871, Beijing, China. *Corresponding authors. Email: gejq@pku.edu.cn (J.G.), jgao@pku.edu.cn (J.G.)

tailed two-sample equal variance t -test, $p_{Pitch} < 0.001$, $p_{PROMS} < 0.001$) (see Fig. 1).

Fig. 1. The distribution of the PROMS pitch processing score ($N = 27$). The median score of 12 was used to divide the low-scoring group (blue bars) and the high-scoring group (red bars). The PROMS pitch processing score of the low-scoring group was significantly lower than that of the high-scoring group ($* p < 0.001$).

MEG: Source-level signals

The incongruous vs. congruous contrast maps showed different patterns of brain activation in the language and music tasks across the low-scoring and high-scoring groups, indicating that the musical ability of pitch processing would exert different influences on the two domains. We also found distinct brain activation patterns for the language and music tasks across the low-scoring and high-scoring groups, indicating different networks for language and music pitch processing.

In the language congruity task, the incongruous condition (compared to the congruous condition) elicited stronger activation in the left IFG, left middle frontal gyrus (L MFG), left superior frontal gyrus (L SFG), left sensory motor cortex (L SMC), left supramarginal gyrus (L SMG), and left temporal cortex (L TC) of the low-scoring group within 200-600 ms relative to the onset of the fourth word. As for the high-scoring group, the incongruous condition (compared to the congruous condition) elicited stronger activation in the L IFG, L SMC, L STG, and left anterior portion of the superior temporal gyrus (L aSTG) within 300-400 ms and 700-1000 ms relative to the onset of the fourth word ($p < 0.01$ cluster-based permutation tests) (Fig. 2A).

In the music congruity task, the incongruous condition (compared to the congruous condition) elicited stronger activation in the right IFG (rIFG), right MFG (rMFG), right SMG (rSMG), and right TC (rTC) in the high-scoring group within 100-300 ms relative to the onset of the fourth note ($p < 0.01$ cluster-based permutation tests). However, no significant activation in the low-scoring group was observed in the incongruous vs. congruous contrast (Fig. 2B).

Furthermore, the source amplitude within 100-300 ms relative to the onset of the fourth word/note for each ROI in the right hemisphere was extracted from the source localization results of the incongruous and congruous conditions for both language and music tasks. In the language task, there was no significant difference in source amplitude between the incongruous and congruous conditions within the selected ROIs. In the music task, the source amplitudes under the incongruous condition of rSMC, rMFG, rTC, and rIFG were significantly stronger than those under the congruous condition (Fig. 3).

Fig. 2. Statistical maps for the source localization result of language and music tasks. (A) Cortical maps illustrate t -value for the language incongruous vs. congruous contrast of the low-scoring and the high-scoring groups ($p < 0.01$ cluster-based permutation tests). (B) Cortical maps illustrate t -value for the music incongruous vs. congruous contrast of the low-scoring and the high-scoring groups ($p < 0.01$ cluster-based permutation tests). The time window was 0-1000 ms from the onset of the fourth word. L and R indicated left and right hemisphere respectively.

Fig. 3. The comparison of source amplitude in the language and music tasks. The selected ROIs were rSMC, rMFG, rTC, and rIFG. In the language task, there was no significant difference in source amplitude between the incongruous and congruous conditions within the selected ROIs. In the music task, the source amplitudes under the incongruous condition of rSMC, rMFG, rTC, and rIFG were significantly higher than those under the congruous condition ($* p < 0.05$; $** p < 0.001$).

Brain-behavior results

The results of the Pearson correlation analysis between the averaged source amplitude of the ROIs and the PROMS pitch processing score showed the bi-directional influence between the processing of lexical tone and musical pitch (Fig. 4).

We found a significant positive correlation between the averaged source amplitude of the left pars opercularis of the IFG (L IFGop) within 200-600 ms relative to the onset of the fourth word under the language congruous condition and the PROMS pitch processing score ($R = 0.43$, $p = 0.02$). Besides, under the language incongruous condition, the averaged source amplitudes of L aSTG and left superior temporal pole (L STP) within 700-1000 ms relative to the onset of the fourth word were also significantly positively correlated with the PROMS pitch processing score ($R_{aSTG} = 0.42$, $p_{aSTG} = 0.03$; $R_{STP} = 0.42$, $p_{STP} = 0.02$).

Fig. 4. Brain-Behavior correlation results. During the shown time window of interest, the mean source amplitudes of IFGop in the language congruous condition, as well as aSTG and STP in the language incongruous condition were positively correlated with individual's PROMS pitch processing score.

DISCUSSION

The current study investigated the influence of an individual's receptive musical pitch perception ability on language pitch processing in their brain. Our MEG study recruited non-musicians as participants, who were then divided into low-scoring and high-scoring groups based on their scores in the behavioral PROMS test. Using source localization of the MEG recordings, we found that both groups showed similar significant activations of extensive left frontotemporal brain regions for processing pitch incongruity in the language task. However, the pitch incongruity in the music task elicited activations of the right frontotemporal cortex only in the high-scoring group. Importantly, the results also showed that receptive musical pitch perception ability could significantly predict brain activation of the left IFGop under the language congruous condition, as well as that of the aSTG and STP under the language incongruous condition, which indicated the pivotal function of these brain regions in pitch processing across both music and language domains.

In this study, the MEG source localization results of the language pitch congruity task reflected the cross-domain pitch processing effect from music to language – a directional inference made on

the basis of participant classification using receptive musical pitch perception ability. Our results demonstrated that incongruous tones (compared to congruous tones) elicited stronger activations in the left frontotemporal brain regions, as well as the SMC and SMG within the time windows of the N400 and 700-1000 ms relative to the onset of the fourth word in the language task, corroborating the N400 effect yielded by tonal inconsistency from a previous EEG study (13). With regard to spatial information, the activated brain regions under the incongruous vs. congruous contrast within 200-600 ms in our study consistently suggested lexical-semantic violation induced by incongruous tones (14). Furthermore, the activations of the frontotemporal brain regions within 700-1000 ms relative to the onset of the fourth word under the incongruous vs. congruous contrast in language task was found only in the high-scoring group, reflecting a musician-like auditory function when processing lexical tones (6).

From the perspective of musical pitch processing, our results showed differentiated brain activity patterns under incongruous vs. congruous contrast in the music task. There was no significant activation in the low-scoring group, which indicated their lower receptive musical pitch discrimination ability and provided MEG evidence for the behavioral PROMS test (12, 15). As for the high-scoring group, the incongruous notes elicited stronger activation in the right temporal cortex, right IFG, right SMG, and right SMC within 100-300 ms relative to the onset of the fourth note, which indicated their higher sensitivity in discriminating incongruous musical pitch (16, 17).

Accumulating neuroimaging evidences have indicated that language and music share domain-general subcortical and cortical network for pitch processing (1, 3, 5). In our study, the significant positive correlation between the activation of frontotemporal brain regions for language processing and the PROMS pitch processing scores suggested that the better an individual is at musical pitch processing, the stronger the activation is in the frontotemporal regions elicited by incongruity in language processing, providing consistent evidence for the bi-directional influence between tonal language and music. In addition, the incongruous musical pitch elicited stronger activation of rSMC, rMFG, rTC, and rIFG compared to congruous musical pitch, indicating acoustic analysis of pitch information (18, 19).

The present MEG study revealed the spatiotemporal features of brain activity on pitch processing across the tonal language and music domains in individuals without formal music training. Specifically, pitch incongruity in language task elicited stronger activation of the frontotemporal brain regions for individuals with better musical pitch perception ability, and these individuals were also found to be more sensitive towards pitch congruity of music. These results highlight the importance of left frontotemporal brain regions in modulating the pitch processing between music and tonal language, and suggested that the cross-domain effect of language and music may be related with individual's preexisting ability on music perception and thus could be generalized into people without professional music training.

MATERIALS AND METHODS

PROMS Musicality Test

We assessed receptive musicality objectively using the PROMS (12). The test battery consists of four subtests tapping listening skills in the domains of melody, tuning, accent, as well as tempo and pitch perception. Each subtest contains 14-18 trials.

Listeners heard two identical sound clips and then were asked if a third probe clip was the same as or different from the previous two. Scores for each subtest were calculated based on accuracy and confidence ratings (i.e., two points were given for correctly reporting “definitely the same” or “definitely different,” one point was given for “probably the same” or “probably different,” and no point was given for “I don’t know” or an incorrect answer).

Participants

Thirty-one non-musicians (17 males, aged between 18 and 30 years, with a mean age of 23.7 years) participated in this study. All participants were healthy with no history of mental disorder or language impairment. Non-musicians were required to have had <3 years of musical training, which must have occurred >7 years before their participation in the study (11, 20). Handedness was tested using the Edinburgh Handedness Questionnaire (21), and one female participant was excluded because of left-handedness. In addition, one female and two male participants were excluded from data analysis because of excessive electrooculogram (EOG) artifacts and head movements. Thus, the subsequent behavioral and MEG data analyses were based on the remaining 27 participants (15 males, aged between 18 and 28 years). The mean age was 23.31 ± 3.45 for the high-scoring group, and 24.71 ± 3.26 for the low-scoring group. There was no significant difference in age between the two groups ($p = 0.304$). Besides, the gender across each group was balanced (low-scoring group: 7 males and 7 females; high-scoring groups: 6 males and 7 females). Written informed consent was provided by all participants, and they were paid a modest honorarium for participation. The procedure for this study was approved by the Institutional Review Board of Peking University.

Stimuli

The following types of language and music stimuli were used in this experiment (Fig. 5): (1) language congruous phrases (LCs), (2) language incongruous phrases (LIs), (3) music congruous phrases (MCs), and (4) music incongruous phrases (MIs). The auditory stimuli were adopted from previous studies (6, 7).

The language congruous condition consisted of 60 Chinese four-character phrases, which were read out by a Chinese-native woman at a normal speaking rate. Each phrase lasted for 2000 ms and included four equal-length syllables. We shifted the tone of the last word of the language congruous phrases to create the language incongruous stimuli. The manipulation of tone shift led to the semantic anomaly effect during speech processing. The tone shifts of the last syllables between congruous and incongruous language phrases were evenly distributed among the four-pitched tone categories of Mandarin Chinese, with 16 (out of 60) phrases shifted to the neighboring pitch category, 31 (out of 60) to the next-nearest-neighboring pitch entry, and 13 (out of 60) to the next-next-nearest-neighboring entry (the longest pitch distance in the current study) (Fig. 5A).

A set of 60 four-note musical phrases with a timbre of grand piano were created using a YAMAHA DGX-620 keyboard. For each of the musical phrases, there were modified counterparts with the last note changed, which differed from the original phrases in that they built up dissonant intervals with the third note in each of the phrases (Fig. 5B). To ensure that the changed last notes also served as incongruous endings with respect to the whole phrases, only those intervals of augmented fourth (36 pieces out of 60), diminished fifth (20 out of 60), and major seventh (four out of 60) were selected. As a result, the pitch shifts of the last notes

	Congruous	Incongruous
A	铿锵有力(li4)	厘(li2)
	控制实验(yan4)	言(yan2)
	良师益友(you3)	邮(you2)
B

Fig. 5. Stimulus illustration. (A) The illustration of language phrases. (B) The illustration of music phrases. The stimuli were divided into language task and music task. The first three words/notes constituted preparation, and the fourth word/note formed a congruous or incongruous relationship with the context on the pitch level. The experiment was divided into four conditions: language congruous condition, language incongruous condition, music congruous condition and music incongruous condition. The duration of each word/note was 500 ms.

between congruous and incongruous musical phrases ranged from one to six semitones, with 15 one-semitone, 11 two-semitone, four three-semitone, 14 four-semitone, four five-semitone, and 12 six-semitone shifts respectively. Musical phrases were also 2000 ms in duration with four 500 ms notes, matching the language phrases.

All the language and musical stimuli were digitized at a 16 bit/44.1 kHz sampling rate. We used the Adobe Audition (San Jose, California, Adobe Systems, Inc.) to edit the amplitude of the sound to ensure no significant difference between the language and the music tasks ($p < 0.05$).

Procedures

Before the MEG experiment, the participants completed the PROMS Musicality Test. Then, the participants were seated in a dimly lit and magnetically shielded room (VACUUMSCHMELZE GmbH & Co. KG, Hanau, Germany). The experimental stimuli were presented using the Presentation software (Neurobehavioral Systems, Albany CA, USA) and delivered binaurally through MEG-compatible insert earphones. Visual instructions were projected to the screen positioned 1 m in front of the center of the participant’s visual field.

The congruous and incongruous conditions in the language and music tasks contained 240 stimuli, which were divided into 4 conditions on average. The order of the sessions and the trials in each session was pseudorandomized for each participant. In each trial, the auditory stimulus was first presented with a visual cue of a black fixation displayed on the screen. The auditory stimulus was terminated after 2000 ms. Then, a black fixation cue was presented for 500 ms, followed by a visual cue for congruity judgment, which lasted for 1500 ms, during which the participants were instructed to judge the congruity of the stimuli by pressing the corresponding buttons using the middle/index finger at the presentation of a visual cue for congruity judgment. The arrangement of the buttons for congruity judgment was counterbalanced across participants. The feedback visual cue (“correct”, “wrong”, or “missed”) would

present subsequently for 1000 ms. Finally, a fixation cross was presented for 1600-1895 ms in the center of the screen. Before the formal experiment, participants received a practice session to ensure that they could perform the task correctly.

MEG and structural MRI

Continuous neuromagnetic signals were recorded using a 306-channel, whole-head MEG system (Elekta-Neuromag TRIUX, Helsinki, Finland) at Peking University. Before each session started, the head position of each participant was determined by four head-position indicator (HPI) coils. The EOG signal was simultaneously captured by two electrodes placed around the eyes: one above the left eye and one below the right eye. The continuous MEG data were online bandpass filtered at 0.1 - 330 Hz, and the sampling rate was 1000 Hz.

The participant's structural MRI images were obtained with a 3T GE Discovery MR750 MRI scanner (GE Healthcare, Milwaukee, WI, USA). A three-dimensional (3D) fast gradient-recalled acquisition in the steady-state with radiofrequency spoiling (FSPGR) sequence was used to obtain $1 \times 1 \times 1 \text{ mm}^3$ -resolution T1-weighted anatomical images. We coregistered the MEG data with the MRI data based on the location of three fiducial marks (the nasion and two preauricular points) and approximately 150 digitalization points on each participant's scalp.

MEG data preprocessing and source localization

MEG data preprocessing and source localization were conducted according to a well-accepted pipeline (22). Raw MEG data were first processed using the temporal extension of the signal-space separation (tSSS) technique with the default setting of parameters (length of raw data buffer: 30 seconds; subspace correlation limit: 0.98) to minimize the confounding contribution of magnetic sources from outside the head and reduce within-sensor artifacts (23). Head movements between each session were corrected based on the positions of HPI coils, implemented in MaxFilter 2.2 software (Elekta Neuromag). Artifacts due to eye blinks and eye movements were removed using independent component analysis (ICA), which was performed with MNE-Python (24). Further, a 40-Hz low-pass filter was applied. T1-weighted images were processed using the Freesurfer software (<http://surfer.nmr.mgh.harvard.edu/>) for cortical reconstruction and volumetric segmentation.

After preprocessing, source localization was performed with Brainstorm (25) (<http://neuroimage.usc.edu/brainstorm/>). MEG data were first visually inspected to exclude trials that contained bad segments caused by head movements and eye blinks. There was no difference across conditions among rejected trials. The bad MEG channels were also marked and excluded from event-related analyses (performed -200 to 1000 ms relative to the onset of the fourth word/note) and source estimation.

Minimum-norm estimates (26) with overlapping spheres (27) were used for source construction with a source space consisting of 15000 vertices. The same source reconstruction analysis was conducted based on the unfiltered data for Granger causality analysis. The source localization results were then normalized with a z-score based on the baseline from -200 to 0 ms relative to the onset of stimuli. In addition, the root mean square (RMS) of each vertex in the source space was calculated over a 100-ms time window from 0 to 1000 ms following the onset of the fourth word/note. The RMS result of each time window per participant was spatially smoothed using an isotropic Gaussian kernel with a full-width at maximum (FWHM) of 5 mm. The source localization

results were projected to the default anatomy of Colin27 (28) with a 1-mm resolution in the Montreal Neurological Institute (MNI) standard space for further group analysis. The RMS results of the sequential 100-ms time windows of each participant in the source space were used for statistical analysis. The differences in cortical activation among the congruous and incongruous conditions of the language and music tasks were assessed using paired *t*-tests (two-tailed). The cluster-based nonparametric permutation test (29), implemented in FieldTrip (<http://fieldtriptoolbox.org/>) was applied for multiple-comparison correction (1000 times Monte-Carlo simulations, $\alpha = 0.01$ for cluster inclusion and $\alpha = 0.01$ for permutation significance).

ROI selection

We selected the brain regions of interest (ROI) that showed significant activations under the congruous condition compared with the incongruous condition in the language and music congruity tasks within high-scoring and low-scoring groups (Table 1). Specifically, we found no significant activation under the contrast of music incongruous condition vs. music congruous condition in the low-scoring group. The selected ROIs were L SFG, L MFG, L IFG, L SMC, L SMG, L aSTG, left posterior portion of the superior temporal gyrus (L pSTG), L STP, left middle TC, rSMC, rMFG, rIFG, rTC. The time courses of significant activated brain regions among the contrasts (low-scoring group: LI > LC, 200-600 ms; High-scoring group: LI > LC, 300-400 ms and 700-900 ms; High-scoring group: MI > MC, 100-300 ms) were extracted for further brain-behavior analysis.

Table 1. ROI selection

Region	MNI coordinates x, y, z; (mm)	Size (Vertices)	Peak <i>t</i> value
Low-scoring group			
LI > LC			
L SFG	-22.5, 1.4, 72.2	214	5.72
L MFG	-35.8, 24.2, 34.5	301	4.79
L IFG	-59.9, 24.6, 11.0	240	5.02
L SMC	-62.9, -17.6, 44.9	221	8.59
L SMG	-63.5, -35.8, 38.8	305	6.99
L aSTG	-56.8, 11.2, -21.6	109	5.86
L pSTG	-70.6, -23.7, 3.7	143	5.86
High-scoring group			
LI > LC			
L IFG	-56.4, 35.6, 3.2	101	3.01
L SMC	-65.6, -17.0, 35.4	75	4.94
L SMG	-67.6, -33.6, 36.5	118	5.48
L STP	-50.2, 14.3, -34.9	190	8.40
L aSTG	-62.8, -6.3, -3.5	151	5.80
L MTC	-70.6, -20.5, -18.5	155	5.88
High-scoring group			
MI > MC			
R SMC	52.5, -6.7, 31.2	171	6.83
R MFG	48.2, 35.7, 34.4	290	5.54
R IFG	59.6, 28.3, 10.7	266	6.83
R TC	50.9, -9.4, -12.5	311	3.84

Note: left superior frontal gyrus (L SFG), left middle frontal gyrus (L MFG), left inferior frontal gyrus (L IFG), left sensory motor cortex (L SMC), left supramarginal gyrus (L SMG), left anterior portion of the superior temporal gyrus (L aSTG), left posterior portion of the superior temporal gyrus (L pSTG), left superior temporal pole (L STP), left middle temporal cortex (L MTC), right sensory motor cortex (R SMC), right middle frontal gyrus (R MFG), right inferior frontal gyrus (R IFG), right temporal cortex (R TC).

Brain-behavior analysis

According to the selection of the ROI, the similar ROIs (left inferior frontal gyrus, left superior temporal pole, left anterior

temporal gyrus, and left middle temporal cortex) under the contrast of LI > LC between low-scoring and high-scoring group were selected for brain-behavior correlation analysis. Based on the source localization results, the source amplitude for each of these ROIs under LI and LC conditions within 200-600 and 700-1000 ms relative to the onset of the fourth word was extracted. Subsequently, we separately calculated the Pearson correlation coefficient of the average source amplitude of these brain regions and the PROMS pitch processing score to investigate the function of the left inferior frontal gyrus and the left temporal cortex in music pitch processing.

REFERENCES

- G. M. Bidelman, J. T. Gandour, A. Krishnan, Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* **23**, 425-434 (2011).
- P. C. M. Wong, E. Skoe, N. M. Russo, T. Dees, N. Kraus, Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* **10**, 420-422 (2007).
- D. Schön, C. Magne, M. Besson, The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology* **41**, 341-349 (2004).
- S. Moreno, C. Marques, A. Santos, M. Santos, S. L. Castro, M. Besson, Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cereb. Cortex* **19**, 712-723 (2009).
- S. S. Asaridou, J. M. McQueen, Speech and music shape the listening brain: Evidence for shared domain-general mechanisms. *Front. Psychol.* **4**, 321 (2013).
- Y. Nan, A. D. Friederici, H. Shu, Y. Jia Luo, Dissociable pitch processing mechanisms in lexical and melodic contexts revealed by ERPs. *Brain Res.* **1263**, 104-113 (2009).
- Y. Nan, A. D. Friederici, Differential roles of right temporal cortex and broca's area in pitch processing: Evidence from music and Mandarin. *Hum. Brain Mapp.* **34**, 2045-2054 (2013).
- H. Luo, J. T. Ni, Z. H. Li, X. O. Li, D. R. Zhang, F. G. Zeng, L. Chen, Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19558-19563 (2006).
- J. Ge, G. Peng, B. Lyu, Y. Wang, Y. Zhuo, Z. Niu, L. H. Tan, A. P. Leff, J.-H. Gao, Cross-language differences in the brain network subserving intelligible speech. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2972-2977 (2015).
- R. D. Patterson, S. Uppenkamp, I. S. Johnsrude, T. D. Griffiths, The processing of temporal pitch and melody information in auditory cortex. *Neuron* **36**, 767-776 (2002).
- K. Mankel, G. M. Bidelman, Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 13129-13134 (2018).
- L. N. C. Law, M. Zentner, Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLoS ONE* **7**, e52508 (2012).
- X. Li, Y. Yang, P. Hagoort, Pitch accent and lexical tone processing in Chinese discourse comprehension: An ERP study. *Brain Res.* **1222**, 192-200 (2008).
- Z. Liu, S. Shu, L. Lu, J. Ge, J. H. Gao, Spatiotemporal dynamics of predictive brain mechanisms during speech processing: An MEG study. *Brain Lang.* **203**, 104755 (2020).
- A. M. Proverbio, A. Orlandi, F. Pisanu, Brain processing of consonance/dissonance in musicians and controls: A hemispheric asymmetry revisited. *Eur. J. Neurosci.* **44**, 2340-2356 (2016).
- K. L. Hyde, J. P. Lerch, R. J. Zatorre, T. D. Griffiths, A. C. Evans, I. Peretz, Cortical thickness in congenital amusia: When less is better than more. *J. Neurosci.* **27**, 13028-13032 (2007).
- K. L. Hyde, R. J. Zatorre, I. Peretz, Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cereb. Cortex* **21**, 292-299 (2011).
- G. Hickok, D. Poeppel, The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393-402 (2007).
- J. Obleser, R. J. S. Wise, M. A. Dresner, S. K. Scott, Functional integration across brain regions improves speech perception under adverse listening conditions. *J. Neurosci.* **27**, 2283-2289 (2007).
- A. Parbery-Clark, E. Skoe, N. Kraus, Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* **29**, 14100-14107 (2009).
- R. C. Oldfield, The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* **9**, 97-113 (1971).
- J. Gross, S. Baillet, G. R. Barnes, R. N. Henson, A. Hillebrand, O. Jensen, K. Jerbi, V. Litvak, B. Maess, R. Oostenveld, L. Parkkonen, J. R. Taylor, V. van Wassenhove, M. Wibral, J.-M. Schoffelen, Good practice for conducting and reporting MEG research. *Neuroimage* **65**, 349-363 (2013).
- S. Taulu, M. Kajola, Presentation of electromagnetic multichannel data: The signal space separation method. *J. Appl. Phys.* **97**, 124905 (2005).
- A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, M. Hamalainen, MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, 267 (2013).
- F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, R. M. Leahy, Brainstorm: A user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* **2011**, 879716 (2011).
- M. S. Hämäläinen, R. J. Ilmoniemi, Interpreting magnetic fields of the brain: Minimum norm estimates. *Med. Biol. Eng. Comput.* **32**, 35-42 (1994).
- M. X. Huang, J. C. Mosher, R. M. Leahy, A sensor-weighted overlapping-sphere head model and exhaustive head model comparison for MEG. *Phys. Med. Biol.* **44**, 423-440 (1999).
- C. J. Holmes, R. Hoge, L. Collins, R. Woods, A. W. Toga, A. C. Evans, Enhancement of MR images using registration for signal averaging. *J. Comput. Assist. Tomogr.* **22**, 324-333 (1998).
- E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177-190 (2007).

Acknowledgments: **Funding:** This work was supported by the National Natural Science Foundation of China (81790650, 81790651, 81790654, 81727808, 31771253 and 31421003), the National Key Research and Development Program of China (2017YFC0108900) and the Beijing Brain Initiative of Beijing Municipal Science & Technology Commission (Z181100001518003). We thank the National Center for Protein Sciences at Peking University in Beijing, China, for assistance with the MEG and MRI data acquisition and data analyses. **Author contributions:** Z.W. Liu, Y. Nan and J.H. Gao designed the study; Z.W. Liu, L.X. Lu and W. Cui performed the experiments; Z.W. Liu and J.Q. Ge analyzed the data; Z.W. Liu, J.Q. Ge and J.H. Gao wrote the manuscript; all authors have read and approved the final version of the manuscript. **Competing Interests:** The authors declared no conflict of interest. **Data and materials availability:** The data and materials are available upon request to the corresponding authors.