Reduced Sensitivity to Between-Category Information but Preserved Categorical Perception of Lexical Tones in Tone Language Speakers With Congenital Amusia

Fei Chen[1,2] and Gang Peng[2,3]*

[1]School of Foreign Languages, Hunan University, Changsha, China, [2]Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, China, [3]Chinese Academy of Sciences (CAS) Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Previous studies have shown that for congenital amusics, long-term tone language experience cannot compensate for lexical tone processing difficulties. However, it is still unknown whether such difficulties are merely caused by domain-transferred insensitivity in lower-level acoustic processing and/or by higher-level phonological processing of linguistic pitch as well. The current P300 study links and extends previous studies by uncovering the neurophysiological mechanisms underpinning lexical tone perception difficulties in Mandarin-speaking amusics. Both the behavioral index (d') and P300 amplitude showed reduced within-category as well as between-category sensitivity among the Mandarin-speaking amusics regardless of the linguistic status of the signal. The results suggest that acoustic pitch processing difficulties in amusics are manifested profoundly and further persist into the higher-level phonological processing that involves the neural processing of different lexical tone categories. Our findings indicate that long-term tone language experience may not compensate for the reduced acoustic pitch processing in tone language speakers with amusia but rather may extend to the neural processing of the phonological information of lexical tones during the attentive stage. However, from both the behavioral and neural evidence, the peakedness scores of the d' and P300 amplitude were comparable between amusics and controls. It seems that the basic categorical perception (CP) pattern of native lexical tones is preserved in Mandarin-speaking amusics, indicating that they may have normal or near normal long-term categorical memory.

Keywords: amusia, P300, acoustic processing, phonological processing, categorical perception

INTRODUCTION

As two old and unique products of the human brain, speech and music are present across all cultures (Patel, 2008). Although they belong to different domains with different representations, an increasing number of neuroimaging studies point to a large neural overlap in responses to speech and musical stimuli (see Peretz et al., 2015 for a review), which implies a close

relationship between musicality and speech-processing capacity. Pitch, an important psycho-acoustical attribute in both music and speech, is often at the center of researchers' attention in studies on the mutual influence between speech and music. Accumulating evidence suggests that experience in one domain could be transferred to facilitate processing in the other: for example, a tone language background is often associated with better performance in musical pitch processing (e.g., Pfordresher and Brown, 2009; Bidelman et al., 2013; Peng et al., 2013). On the other hand, compared with non-musicians, musicians tend to be more skilled at detecting non-native lexical tone contrasts (e.g., Wong et al., 2007; Lee and Hung, 2008; Chandrasekaran et al., 2009) and even more sensitive to native linguistic tone categories (Tang et al., 2016). These studies imply a two-way transferability of pitch expertise across the domains of music and speech.

From another perspective, the existence of congenital amusia (simply amusia hereinafter), a lifelong neuro-genetic disorder primarily affecting fine-grained melodic pitch processing in the absence of brain injury (Peretz, 2001; Peretz et al., 2002), offers an ideal window to investigate the influence of impaired musical ability on lexical tone processing in speech. As indicated in various behavioral studies, impaired or reduced lexical tone perception skills are consistently reported among non-tonal language speaking amusics (Nguyen et al., 2009; Tillmann et al., 2011) and in tone language speaking amusics at the whole group level (Jiang et al., 2012; Wang and Peng, 2014; Zhang et al., 2017a) or in a subgroup called "tone agnosics" (Nan et al., 2010; Huang et al., 2015). On the basis of these behavioral findings, we can conclude that the pitch processing deficit in amusia tends to be domain-general regardless of language background. The processing deficit in musical melody may be transferred to pitch-based speech processing.

Although there is ample behavioral evidence of pitch impairment in amusia, the evidence regarding the neurophysiological bases that could account for the nature of amusics' pitch processing deficits is not completely conclusive. The structural abnormalities in the right inferior frontal gyrus (IFG) of amusics and its reduced connectivity with the auditory cortex (Hyde et al., 2006, 2007; Albouy et al., 2013) were corroborated in a functional MRI (fMRI) study (Hyde et al., 2011), which demonstrated a functional deactivation in these brain networks of amusics. However, the auditory cortex (superior temporal gyrus, STG) of amusic individuals seems to function normally to pitch change (Hyde et al., 2011), although brain structural abnormalities have been detected in this core auditory area (Hyde et al., 2007; Mandell et al., 2007). Nevertheless, there is conflicting evidence implying an anomalous functioning of the auditory cortex of amusics (Albouy et al., 2013; Zhang et al., 2017b). The discrepancy regarding the functioning of the auditory cortex could be partly explained by the experimental tasks utilized in these functional neuroimaging studies. Specifically, the normal functional activation of auditory cortices was found when subjects were passively listening to pitch deviations without focal attention (Hyde et al., 2011). However, abnormal functional activation of auditory cortices was shown in both non-tonal language speakers (Albouy et al., 2013) and

tone language speakers with amusia (Zhang et al., 2017b) when the focal attention was directed to the auditory stimuli. Consequently, attentional requirements are likely to play a key role in the neurophysiological manifestation of amusia.

Consistent with this claim, accumulating evidence from event-related potential (ERP) studies converges with the notion that pre-attentive implicit processing of pitch information might remain normal in amusics, whereas their explicit judgment or awareness of pitch perception might be compromised (Peretz et al., 2005, 2009; Moreau et al., 2013; Omigie et al., 2013; Zhang and Shao, 2018; but see Nan et al., 2016 for evidence showing impaired pre-attentive tone processing in a small proportion of Mandarin-speaking amusics called tone agnosics). For instance, the mismatch negativity (MMN), which is an early component reflecting an automatic cortical response to auditory changes without subjects' focal attention (Näätänen, 2000), was found to remain intact in amusics even for small 25-cent musical pitch differences; instead, unlike the controls who also showed a prominent P3b in the attentive condition, no P3b response was shown in the amusics when they were asked to actively detect such small pitch deviations in music (Moreau et al., 2013). In the speech domain, for tone language speakers with amusia, a recently conducted ERP study (Zhang and Shao, 2018) examined how typical Cantonese tone stimuli were processed in both pre-attentive (MMN) and attentive (P300) conditions and showed that Cantonese-speaking amusics could perceive lexical tone changes normally pre-attentively but showed a deficiency in consciously detecting the same lexical tone differences at the later attentive level (Zhang and Shao, 2018). In short, the findings above suggest a neuro-dynamic mechanism of intact pre-attentive pitch processing but impaired active/attentive pitch processing among amusics in the domains of both music and speech.

A plethora of studies have demonstrated that long-term access to a tonal language environment may shape the neural architectures of pitch processing in native speakers (e.g., Gandour et al., 2004; Bidelman et al., 2011; Gu et al., 2013). So far, the neurophysiological mechanisms underlying amusics' behavioral linguistic tone deficits remain unclear. This topic is also related to the question of whether and how long-term experience of a tonal language affects the neural bases of amusia. It is well known that in a tonal language, the pitch is exploited to distinguish phonological contrasts and word meanings at the syllable level. In stark contrast to non-tonal language speakers, Cantonese-speaking amusics tend to exhibit no functional abnormality in the right IFG. Instead, they mainly exhibit neural deficits in the right STG when processing lexical tones, which partly overlaps with neural activation of lexical tone processing (Zhang et al., 2017b), implying that language background might modulate the neural correlates of amusics' brains. Similar to non-tonal language speakers with amusia, at different processing stages along the auditory pathway, the lexical processing deficits in tonal language speakers with amusia have also been demonstrated to occur at a later time window when consciously detecting the lexical tone changes at the attentive stage (Zhang and Shao, 2018). However, it is still unknown whether, during this time window (P300), such an

attentive lexical tone processing deficit in tone language speakers with amusia is restricted to the domain-transferred insensitivity to the acoustic pitch differences or, alternatively, further extends to a reduced phonological processing of lexical tone categories.

As suggested by the functional hypothesis (van Lancker, 1980), the accurate and complete perception of lexical tones in native tone language speakers involves the processing of acoustic information as well as the phonological information carried by the tonal signal. Specifically, the acoustic information contains pitch features, such as pitch height and contour variations, while the phonological information differentiates different lexical semantics on the basis of different tonal categories (Gandour et al., 2000). Several neuroimaging studies have indicated that the acoustic information and the phonological information of lexical tones in native speakers are essentially two distinct auditory inputs, which exhibit different cognitive and physiological characteristics (Xi et al., 2010; Zhang et al., 2011, 2012; Yu et al., 2014, 2017). The phonological processing, which develops early in native speakers, is driven by an accumulation of perceptual development from the ambient native sound inputs (Kuhl, 2004). Thus, listeners' speech processing performance can be modulated and reshaped by their experience with acoustic and phonological information (Chandrasekaran et al., 2009).

As for lexical tone perception, it has been reported that amusics perform worse than musically intact controls in their perception of non-native lexical tone categories (Nguyen et al., 2009; Tillmann et al., 2011; Wang and Peng, 2014), which reflects a domain-transferred acoustic processing deficit across the domains of music and speech. However, it remains unclear whether the native lexical tone perception deficits in tone language speakers with amusia (Nan et al., 2010; Jiang et al., 2012; Huang et al., 2015; Zhang et al., 2017a; Zhang and Shao, 2018) are merely due to the similar domain-transferred deficit in acoustic processing or additionally caused by reduced higher-level phonological processing. Answers to this question can help uncover whether and how reduced lower-level acoustic processing could negatively influence higher-level phonological processing. In order to fill the aforementioned research gap, the current P300 study examined the contrast between acoustic information and phonological information as the contrast between within-category and across-category variations by adopting a modified categorical perception (CP) paradigm (see Yu et al., 2014, 2017), which directly dissociates the acoustic and phonological processing of lexical tones for native amusic and control participants. According to model of Fujisaki and Kawashima (1971), lower-level acoustic processing is required to discriminate within-category discrimination pairs by the physical F0 difference, while higher-level phonological processing is mainly employed to conduct between-category discrimination tasks based on the information of phonemic categories. Moreover, the CP paradigm is more suitable because it reflects a fine-grained processing mode along a pitch continuum, with the perceptual pitch differences being relatively small.

To this end, the present P300 study compared Mandarin-speaking amusic and control participants' behavioral and ERP responses to within-category and between-category deviants in both speech and nonspeech contexts. For native speakers, there appears to be some carry-over influence of the long-term phonological processing of lexical tone on the perception of the pitch counterparts from the speech to the nonspeech domain (Xu et al., 2006). Consequently, incorporating nonspeech stimuli enabled us to investigate whether the potential acoustic/phonological processing deficit in amusia is domain-specific or domain-general. All the stimuli were presented in an active oddball paradigm, randomly interspersing standards having the rising pitch contour of a high rising tone (Tone 2 in Mandarin) with equally spaced deviants of two types: a within-category deviant – an exemplar of the same linguistic category as the standard (Tone 2) but with a different pitch contour slope – and a between-category deviant – an exemplar of a distinct tone category (Tone 1 in Mandarin) from the standard, also with a different pitch contour slope. As consistently indicated in previous ERP studies of tone language speakers with amusia, the early-occurring components of P1, N1, and N2 were not significantly different between the amusic and control groups (Nan et al., 2016; Zhang and Shao, 2018), and the deficits occurred at the later attentive stage (P300 component) when consciously detecting lexical tone changes (Zhang and Shao, 2018). In this study, the amusic and control participants' sensitivity index (d' score), reaction time (RT), mean amplitude, and peak latency of the P300 component were specifically analyzed to determine which types of deviant – within-category or between-category – in which context – speech or nonspeech – were different between the amusics and the controls.

In particular, we considered both behavioral and P300 evidence in favor of the following hypotheses: the amusics would show less sensitivity to a within-category deviant due to a domain-transferred inferior skill in acoustic processing. If the higher-level phonological processing – developed early in native speakers – is robust enough to resist the influence of reduced acoustic sensitivity, it is possible that the behavioral and P300 indices in responses to a between-category deviant would be similar between the two subject groups. Alternatively, if the acoustic processing deficits further extend to a reduced phonological processing of lexical tone categories, it is possible that Mandarin-speaking amusics would show compromised behavioral and neural processing of between-category variations compared to musically intact controls.

MATERIALS AND METHODS

Participants

Fifteen amusic participants (seven females) and 15 musically intact control participants (seven females) from universities in Mainland China matched for chronological age, gender, and level of education participated in the current study. The sample size of the amusic and control participants in the current study was largely comparable to that usually reported in previous behavioral or ERP studies on lexical tone perception in amusics (cf. Jiang et al., 2012; Huang et al., 2015; Nan et al., 2016; Zhang et al., 2017a; Zhang and Shao, 2018). All the participants were right-handed and native Mandarin speakers from northern

China with no reported history of hearing impairment, neurological illness, or long-term formal musical training. To identify the presence or absence of amusia, the online Montreal Battery of Evaluation of Amusia[1] was utilized at the screening stage, with the cutoff average score set to 71% (Peretz et al., 2008). This online version includes two pitch-based subtests (out-of-key and mistuned subtests) and one rhythm/duration-based subtest (offbeat subtest), and it has been widely adopted as a diagnostic tool for tone language speakers with amusia (e.g., Wong et al., 2012; Wang and Peng, 2014; Zhang et al., 2017a; Zhang and Shao, 2018). The mean accuracy of the three subtests was 60.67% for the amusics and 83.73% for the controls. The results of an independent-samples *t*-test confirmed that both the global score and the scores of the three subtests for the amusic participants were much lower than those of the control participants (all *p*s < 0.001). The demographic characteristics of all the participants are presented and summarized in **Table 1**. Approval for this study was obtained from the Human Research Ethics Committee of the Shenzhen Institutes of Advanced Technology, the Chinese Academy of Sciences.

Stimuli and Experimental Design

The Mandarin monosyllabic morpheme /i/ with the high-level tone (Tone 1, around 290 Hz) was recorded by a native female speaker (22,050 Hz sampling rate, 16-bit resolution). On the basis of the natural speech template with Mandarin Tone 1, a tone continuum containing 11 tone stimuli was resynthesized using the overlap-add re-synthesis function in Praat (Boersma and Weenink, 2009), with an equal step size of 6 Hz. The pitch contours are illustrated in **Figure 1A**; the procedures for synthesizing these stimuli followed those described in Peng et al. (2010). These pitch contours formed bilinear approximations (Wang, 1976) of high-rising (Tone 2, meaning "aunt") and high-level (Tone 1, meaning "clothes") tones in Mandarin. Eleven additional nonspeech counterparts were synthesized using the triangle wave (cf. Chen and Peng, 2016) with the same pitch contours as the speech stimuli. All the speech and nonspeech stimuli had a duration of 300 ms, with the pitch direction in the first 60 ms kept unchanged (see **Figure 1**). The nonspeech stimuli sound much lower perceptually when

[1]http://www.brams.org/amusia-public/

TABLE 1 | Control and amusic participants' sample size, chronological ages, and Online Identification Test of Congenital Amusia scores. The results (*p*) of independent-samples *t*-tests comparing the two groups are also reported. M, male; F, female; SD, standard deviation.

	Controls	Amusics	*p*
No. of participants	15 (8 M, 7 F)	15 (8 M, 7 F)	N/A
Age (SD)	22.87 years (2.85)	21.73 years (2.43)	*p* = 0.47
Online identification test of congenital amusia			
Out-of-key (SD)	88.67% (4.92%)	63.67% (11.30%)	*p* < 0.001
Offbeat (SD)	82.80% (7.73%)	62.00% (9.86%)	*p* < 0.001
Mistuned (SD)	78.73% (10.12%)	55.80% (10.94%)	*p* < 0.001
Mean score (SD)	83.73% (4.25%)	60.67% (5.49%)	*p* < 0.001

compared to the speech stimuli of the same intensity. This is because, compared with speech sound, the nonspeech sound (triangle wave) is much simpler in terms of spectro-temporal complexity. For the purpose of matching the loudness level, the speech and nonspeech stimuli were presented at 65 and 80 dB, respectively, which were based on the perceptual judgments of three native speakers (Zhang et al., 2013). Moreover, the intensity envelopes were closely matched and were kept constant across each speech and nonspeech continuum.
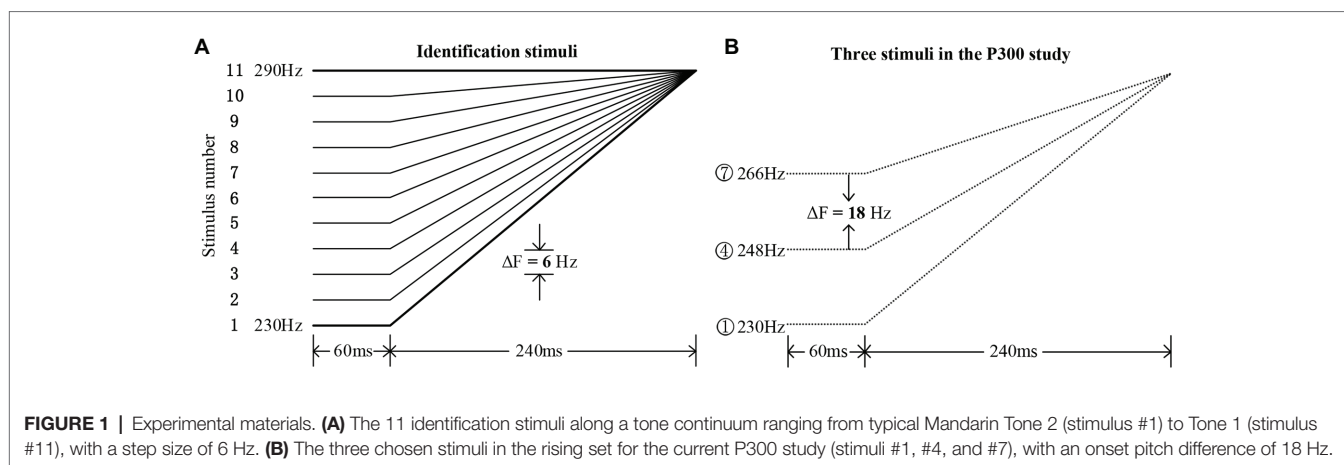
Based on a pilot identification test among several native Mandarin speakers who did not participate in the current P300 study, the obtained average boundary position (Finney, 1971) straddling the two categories was around stimulus #5 in both the speech and nonspeech continua. Then, in this P300 study, a rising set with three rising pitch contours, containing stimuli #1, #4, and #7 for both speech and nonspeech stimuli, was chosen (see **Figure 1B**). Based on the identification boundary, a within-category stimulus pair (#1 and #4) and a between-category stimulus pair (#4 and #7) were chosen for the ERP experiment using the oddball paradigm. Specifically, stimulus #4 was presented as the standard stimulus, stimulus #1 as the within-category deviant, and stimulus #7 as the between-category deviant. The onset acoustic distance between each type of deviant and standard was set to be 18 Hz, indicating that the between- or within-category deviants were equally spaced when compared with the standard stimuli. All stimuli were presented binaurally to both the control and amusic participants *via* a pair of E·A·R Tone 3A Insert Earphones manufactured by Etymotic Research Inc. Participants were seated in front of a monitor at a distance of approximately 60 cm in a sound-attenuated and electrically shielded chamber. All the sound stimuli were run by the E-Prime 2.0 program (Psychology Software Tools Inc., United States), which was also used to collect the behavioral RT and accuracy data.

Two oddball sessions (speech and nonspeech) were presented to each participant with a counterbalanced sequence. Each experimental session began with six standard trials, followed by 600 test trials in total, including 80% standards (480 trails), 10% within-category deviants (60 trials), and 10% between-category deviants (60 trials). The order of the test stimuli within each session was pseudo-randomized with at least three standard stimuli between any adjacent deviant stimuli. To fully control for the sequence effect, those three consecutive standard trials directly following deviants were disregarded in further ERP analyses. The stimulus-onset-asynchrony (SOA) jittered in the range from 1,100 to 1,500 ms. In each session, the stimuli were presented among five blocks, with 1-min breaks between blocks. Given that native speakers may find it harder to detect within-category deviants than between-category deviants (Zhang et al., 2012; Zheng et al., 2012), of the 15 participants in each subject group, eight were instructed to press the left button of the mouse after hearing the within-category deviants and the right button for the between-category deviants using their two thumbs, while the other seven were instructed to press the two buttons in the opposite manner. Moreover, all subjects were instructed not to press any mouse buttons after hearing the standards. The participants were asked to make

FIGURE 1 | Experimental materials. **(A)** The 11 identification stimuli along a tone continuum ranging from typical Mandarin Tone 2 (stimulus #1) to Tone 1 (stimulus #11), with a step size of 6 Hz. **(B)** The three chosen stimuli in the rising set for the current P300 study (stimuli #1, #4, and #7), with an onset pitch difference of 18 Hz.

their responses as fast as possible based on their first judgment. During the training stage, the three types of speech or nonspeech stimuli were played to each participant five, and only five, times and a practice run was presented before recording to familiarize participants with the task and stimuli. The participants were instructed to minimize their physical movements and eye blinks throughout the ERP experiment to reduce artifacts.

After the electrophysiological (EEG) recordings, all the participants carried out a behavioral identification posttest to the determine the identification boundary and further justify how they categorized the three chosen stimuli in this P300 study (i.e., stimuli #1, #4, and #7) along the speech or nonspeech continuum. For both the speech and nonspeech stimuli, each participant gave their responses by clicking the corresponding button on the keyboard labeled "Tone 1" or "Tone 2." The 11 stimuli of each speech or nonspeech continuum (see **Figure 1A**) were repeated nine times in one block. Two blocks of tone identification, speech and nonspeech, were presented to each subject, with the order of the two blocks randomized.

EEG Recording and Data Analyses

The EEG data were recorded in NetStation (V5.1.2) using an EGI (Electrical Geodesics, Inc.) GES 410 system with 64-channel HydroCel GSN electrode nets. Vertical and horizontal eye movements were monitored by electrodes placed on the supra- and infra-orbital ridges of each eye and electrodes near the external canthi of each eye, respectively. The continuous online recordings were referenced to the vertex electrode (Cz), digitized at a sampling rate of 1 kHz, and amplified with an analog band-pass filter of 0.1–75 Hz. Electrode impedances were generally kept below 20 kΩ, considerably below the 50 kΩ threshold recommended by the manufacturer (Electrical Geodesics, 2006).

For the analysis of the posttest behavioral data, the identification score was calculated as the percentage of responses in which the participants identified a particular sound as resembling either tone category (Tone 1 or Tone 2). The boundary position, defined as the corresponding 50% crossover point in a continuum, was obtained using Probit analysis (Finney, 1971). For the analysis of the behavioral data in this P300 study, both the sensitivity index (d' score) and mean

RT of each participant's behavioral responses were computed. For the analysis of the behavioral data in P300 paradigm, the sensitivity index d' (Macmillan and Creelman, 2008), was calculated as the z-score of the hit rate (corresponding button-press responses to deviant) minus that of the false alarm rate (any button-press responses to the standards). For instance, for one specific subject, the hit rate (H) of between-category or within-category deviants was calculated as the ratio of pressing the corresponding mouse button to the total number of each type of the deviant, and the false alarm rate (F) of standard stimuli was calculated as the ratio of pressing any mouse button to the total number of standard stimuli (since the participants were instructed not to press any button when the standard was presented). Then, the d' score of between-category or within-category deviants was calculated using the following formula: $d' = z(H) – z(F)$. Furthermore, the d' score of between-category deviants minus that of within-category deviants is referred to as the "peakedness d'" (Jiang et al., 2012). The RTs were analyzed for the correct responses to within- and between-category deviants, respectively, and those that fell outside of two SDs from the participant's mean RT were not included.

For the offline analysis of the P300 data, the continuous EEG data were digitally refiltered with a 0.5–30 Hz band-pass filter and epoched with 100 ms of pre-stimulus intervals and 800 ms of post-stimulus intervals, with the pre-stimulus interval (−100–0 ms) used for baseline correction. Two electrodes attached to the left and right mastoids were used as offline re-references. Trials with ocular artifacts were excluded from averaging to ERP. Moreover, if more than 20% of the experimental trials for a particular participant were contaminated by ocular artifacts or body movements, the entire EEG data for that participant were excluded. Under this criterion, although we initially recruited 18 controls and 17 amusics to conduct this study, the EEG recordings obtained from five participants (three controls; two amusics) were excluded from further analyses due to excessive artifacts. The current study focused on the mean amplitude and peak latency of the P300 component at the later attentive stage. The time window of the P300 component (320–600 ms) was determined from global field
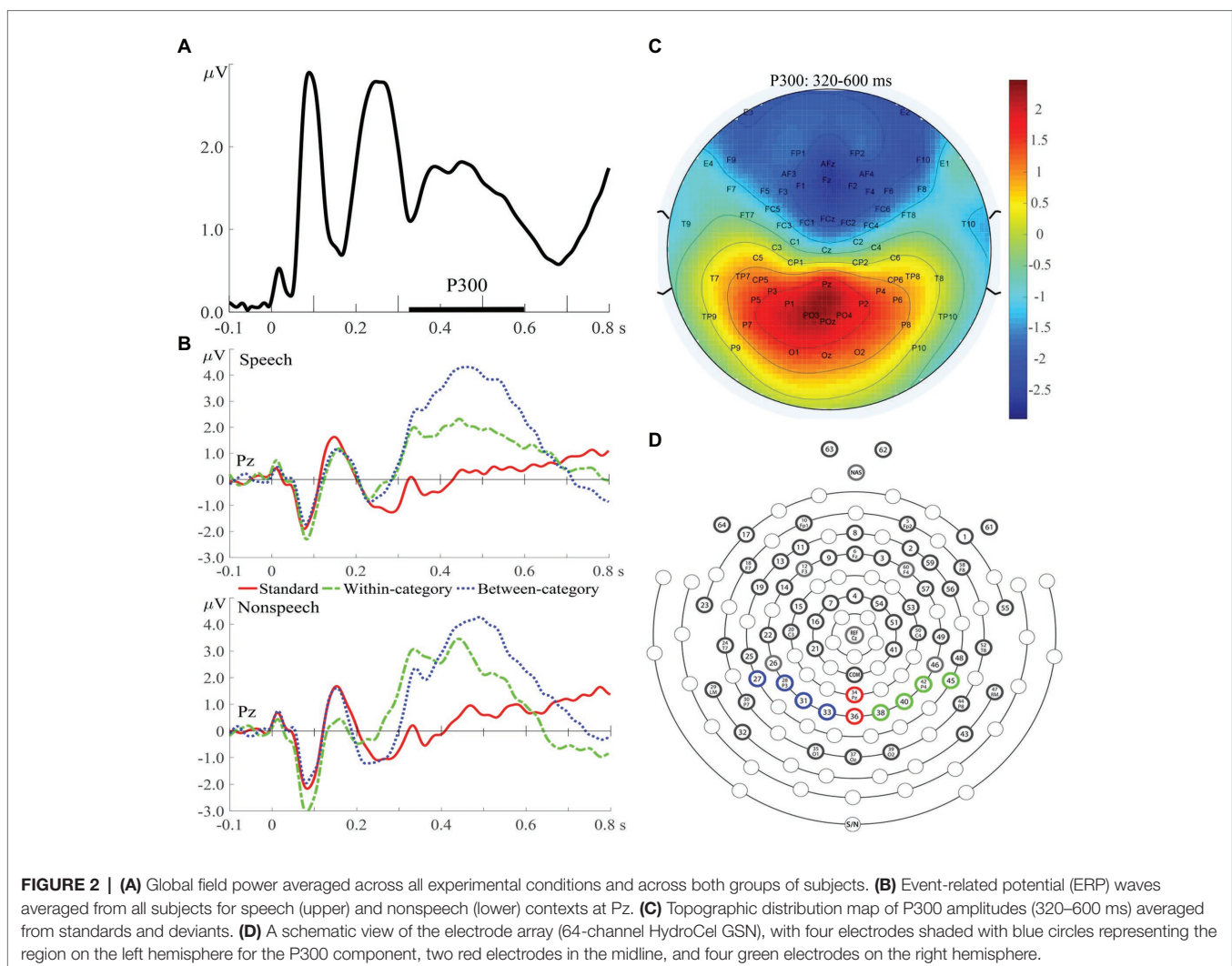
power (Zheng et al., 2012; Zhang et al., 2013), calculated by taking the square root of the mean square ERP values averaged across all electrodes, all experimental conditions, and all subjects (**Figure 2A**), and further confirmed from the ERP waveforms (**Figure 2B**). The electrode locations of P300 were constrained to the region of interest on the basis of previous P300 studies and confirmed by the topographic distribution map (**Figure 2C**) in this study. In total, 10 parieto-occipital electrodes where the P300 amplitude was expected to peak (Polich, 2007) were chosen in terms of different brain hemispheres (**Figure 2D**): left (P1, P3, P5, and PO3), middle (Pz, POz), and right (P2, P4, P6, and PO4). Specifically, difference waves were obtained by subtracting the standard ERPs from each type of deviant ERPs and further submitted to statistical analyses. The amplitude of the P300 difference wave was determined by calculating the adaptive mean amplitude (Scerif et al., 2006) over electrodes within the specified time window (320–600 ms). The peak latency of the P300 difference wave was defined as the timing point corresponding to the maximal point of the second-order polynomial fitted curve to the difference wave.
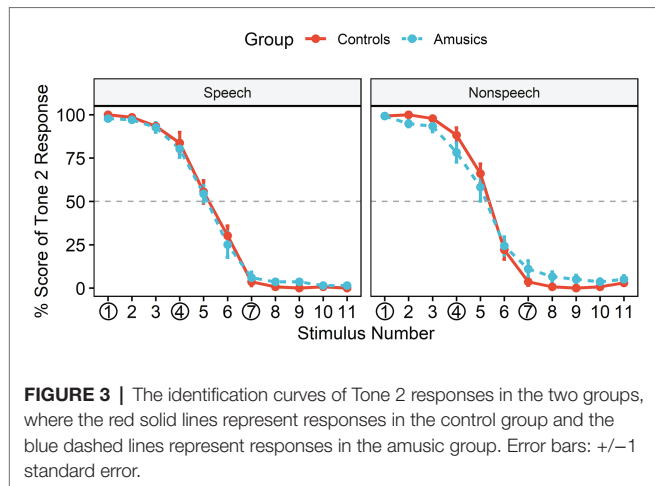
RESULTS

Posttest Identification Results

Figure 3 shows the posttest identification curves in the speech and nonspeech contexts among the controls and amusics. The typical stimuli at the rising end (stimulus #1) were consistently perceived as Mandarin Tone 2, with an identification rate close to 100% in both groups. All participants in most cases perceived stimulus #4 as belonging to the same category as stimulus #1 (i.e., "Tone 2") for both the speech stimuli (Controls: 83.73%; Amusics: 80.27%) and the nonspeech stimuli (Controls: 88.27%; Amusics: 78.27%), while they perceived stimulus #7 as a different sound category (i.e., "Tone 1") both for the speech stimuli (Controls: 96.33%; Amusics: 93.93%) and the nonspeech stimuli (Controls: 96.33%; Amusics: 88.93%).

The obtained mean boundary positions (SDs) were 5.14 (0.67) and 5.15 (0.59) in speech and 5.34 (0.57) and 5.31 (0.73) in nonspeech for the controls and the amusics, respectively. The boundary position was submitted to a two-way ANOVA with *group* (controls vs. amusics) as a between-subjects factor


FIGURE 2 | (A) Global field power averaged across all experimental conditions and across both groups of subjects. **(B)** Event-related potential (ERP) waves averaged from all subjects for speech (upper) and nonspeech (lower) contexts at Pz. **(C)** Topographic distribution map of P300 amplitudes (320–600 ms) averaged from standards and deviants. **(D)** A schematic view of the electrode array (64-channel HydroCel GSN), with four electrodes shaded with blue circles representing the region on the left hemisphere for the P300 component, two red electrodes in the midline, and four green electrodes on the right hemisphere.

FIGURE 3 | The identification curves of Tone 2 responses in the two groups, where the red solid lines represent responses in the control group and the blue dashed lines represent responses in the amusic group. Error bars: +/−1 standard error.

FIGURE 4 | Sensitivity index (*d'* scores) of within-category and between-category deviants in speech **(A)** and nonspeech **(B)** stimuli for the control (dark bars) and amusic (light bars) participants. Error bars: +/−1 standard error.

and *stimulus type* (speech vs. nonspeech) as a within-subjects factor. Greenhouse-Geisser corrections were conducted when appropriate. Neither the main effects nor the interaction effect reached significance (all *p*s > 0.05), indicating that the boundary position straddling the two categories was similar in the two subject groups and for different stimulus types. Therefore, the posttest identification results confirmed that both the control and amusic participants in this study indeed perceived #4 vs. #7 as a between-category contrast and #1 vs. #4 as a within-category contrast.

Behavioral Results: *d'* Score and Reaction Time

Figure 4 displays both the control and amusic participants *d'* scores for detecting within-category and between-category deviants in speech and nonspeech contexts. A three-way repeated measures ANOVA was conducted on the sensitivity index *d'* by indicating *stimulus type* (speech vs. nonspeech) and *category type* (within-category vs. between-category deviant) as two within-subjects factors and *group* (controls vs. amusics) as a between-subjects factor. A significant main effect of *group*, $F(1, 28) = 12.91$, $p = 0.001$, $\eta_p^2 = 0.32$, was observed on *d'* scores, while none of its two-way and three-way interactions with other factors reached significance (all *p*s > 0.05), indicating that the amusics were overall less accurate than the controls in detecting both within-category and between-category deviants behaviorally for both the speech and nonspeech stimuli.

Moreover, there were a significant main effect of *category type* on *d'* values [$F(1, 28) = 26.56$, $p < 0.001$, $\eta_p^2 = 0.49$] and a significant interaction of *category type* by *stimulus type* [$F(1, 28) = 16.77$, $p < 0.001$, $\eta_p^2 = 0.38$]. No other effects were significant. Next, a simple main effect analysis of the *category type × stimulus type* interaction was performed with Bonferroni adjustment. First, both groups of participants showed a much higher sensitivity to the between-category deviants (mean = 3.14) as compared with the within-category deviants (mean = 2.22) in the speech context [$F(1, 28) = 39.20$, $p < 0.001$, $\eta_p^2 = 0.58$], while the *d'* values of the two types of deviants
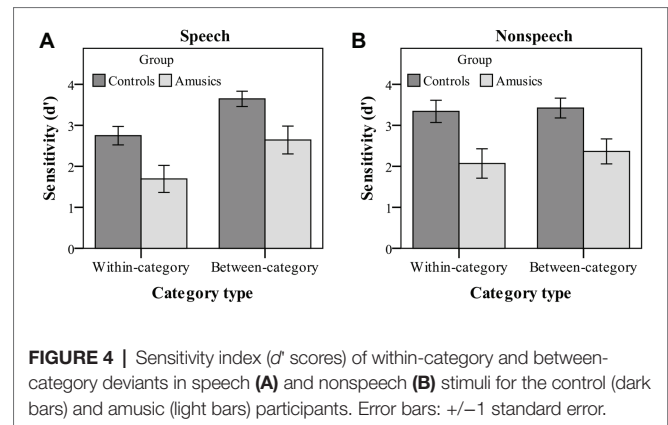
were not different from each other in the nonspeech context [$F(1, 28) = 2.00$, $p = 0.17$, $\eta_p^2 = 0.07$]. Then, a follow-up analysis of peakedness *d'* (the *d'* score of between-category deviants minus that of within-category deviants) was further conducted in the speech context, with no group differences being detected between the control and amusic groups [$t(28) = -0.18$, $p = 0.86$]. Second, the *d'* value of between-category deviants in the speech stimuli was comparable to that in the nonspeech stimuli, $F(1, 28) = 1.07$, $p = 0.31$, $\eta_p^2 = 0.04$, while nonspeech (mean = 2.71) yielded a marginally higher *d'* value of within-category deviants than speech (mean = 2.22) for both the control and amusic groups, $F(1, 28) = 3.72$, $p = 0.06$, $\eta_p^2 = 0.12$.

Figure 5 shows the RTs for detecting within-category and between-category deviants, in which participants made a behavioral response in speech and nonspeech contexts. RT was analyzed using three-way repeated measures ANOVA, with *group* as the between-subjects factor and *stimulus type* and *category type* as two within-subjects factors. The only significant effect was the interaction of *category type* by *stimulus type*, $F(1, 28) = 13.53$, $p < 0.001$, $\eta_p^2 = 0.33$. No other main effects or interactions reached significance (all *p*s > 0.05). These results indicated that the amusics and controls generally required similar amounts of RT to respond to the deviants. After this, simple main effect analyses were performed with Bonferroni adjustment. In the speech stimuli (see **Figure 5A**), the participants showed similar RTs to respond to within-category deviants (mean = 676.25 ms) and between-category deviants (mean = 652.55 ms), $F(1, 28) = 2.40$, $p = 0.13$, $\eta_p^2 = 0.08$. However, in the nonspeech stimuli (see **Figure 5B**), both groups of participants (controls and amusics) made a button press more quickly in response to the within-category deviants of stimulus #1 (mean = 663.17 ms) than in response to the between-category deviants of stimulus #7 (mean = 699.28 ms), $F(1, 28) = 6.99$, $p < 0.05$, $\eta_p^2 = 0.20$.

Electrophysiological Results

Difference waves were obtained by subtracting the standard ERPs from each of the deviant ERPs. **Figures 6, 7**, respectively, show the P300 difference waves of the within- and
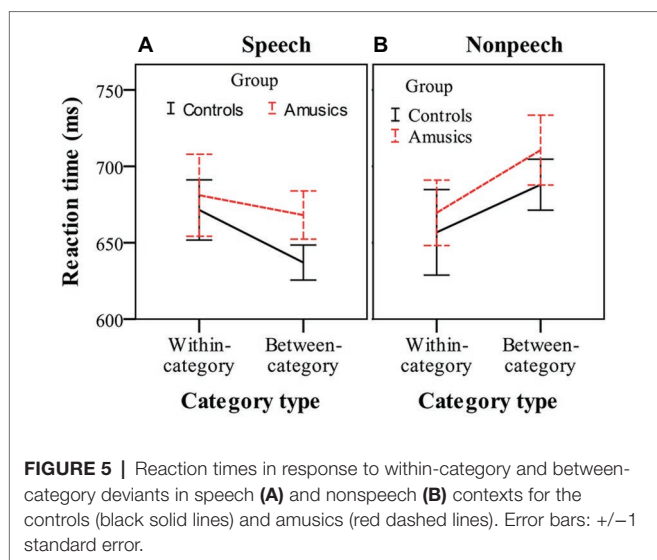
FIGURE 5 | Reaction times in response to within-category and between-category deviants in speech **(A)** and nonspeech **(B)** contexts for the controls (black solid lines) and amusics (red dashed lines). Error bars: +/−1 standard error.

between-category deviants and their topographic voltage maps elicited from both the control and amusic participants in the speech and nonspeech stimuli. Four-way repeated measures ANOVAs were conducted on both the peak latency and mean amplitude of the P300 difference wave (320–600 ms), with three within-subjects factors – *stimulus type* (speech vs. nonspeech), *category type* (within-category vs. between-category deviant), and *hemisphere* (left, middle, and right) – and one between-subjects factor, *group* (controls vs. amusics). Corrections for violations of sphericity were made using the Greenhouse-Geisser method whenever necessary.

P300 Amplitude

Figure 8 plots the mean P300 amplitudes of within-category and between-category difference waves among the two groups of subjects for both speech and nonspeech stimuli. With regard to the P300 amplitude, significant main effects of *group* [$F(1, 28) = 8.26$, $p < 0.05$, $\eta_p^2 = 0.21$], *category type* [$F(1, 28) = 14.11$, $p = 0.001$, $\eta_p^2 = 0.34$], and *hemisphere* [$F(2, 56) = 29.28$, $p < 0.001$, $\eta_p^2 = 0.51$] were observed. Moreover, there was a significant two-way interaction of *category type × hemisphere* [$F(2, 56) = 10.31$, $p = 0.001$, $\eta_p^2 = 0.27$] and a significant three-way interaction of *category type × hemisphere × stimulus type* [$F(2, 56) = 4.02$, $p < 0.05$, $\eta_p^2 = 0.13$]. Firstly, these results indicated that a greater P300 amplitude was elicited from the controls than from the amusics for the neural processing of both within- and between-category stimuli in both the speech and nonspeech stimuli (see **Figure 8**).

Next, the *category type × hemisphere × stimulus type* interaction was further analyzed in terms of different stimulus types. First, in the speech context, there were significant main effects of *category type* [$F(1, 28) = 26.08$, $p < 0.001$, $\eta_p^2 = 0.47$] and *hemisphere* [$F(2, 58) = 23.27$, $p < 0.001$, $\eta_p^2 = 0.45$] and a significant interaction between *category type* and *hemisphere* [$F(2, 58) = 13.01$, $p < 0.001$, $\eta_p^2 = 0.31$]. Further analyses showed that in the speech stimuli, at all

the electrodes on the midline and the left and right hemispheres, the processing of between-category deviants elicited a greater P300 amplitude than the processing of within-category deviants (all ps < 0.001) for both groups. Then, a follow-up analysis of peakedness amplitude (the amplitude of between-category deviants minus that of within-category deviants) was performed in the speech context (see **Figure 9**), and the results revealed no group differences between the controls and amusics in any of the three recording sites (all ps > 0.05). Furthermore, electrodes on the midline elicited a larger P300 amplitude than electrodes in the left and right hemispheres for the processing of both types of deviants in speech (all ps < 0.05). The left and right hemispheres were not significantly different from each other ($p = 0.46$) for the processing of within-category deviants in speech, while the P300 amplitude of the response to between-category deviants in the speech stimuli was larger over the left hemispheric recording sites (mean = 2.68 μV) than that in the right hemisphere (mean = 1.98 μV; $p < 0.01$).

Second, in the nonspeech context, only the main effect of *hemisphere* was detected, $F(2, 58) = 19.61$, $p < 0.001$, $\eta_p^2 = 0.40$. These results indicated that for all subjects, there was no significant difference in P300 amplitude for the processing of within-category and between-category deviants in nonspeech stimuli (see **Figure 8B**). Pair-wise comparison with Bonferroni adjustment showed that the P300 amplitude of the response to nonspeech stimuli at electrodes on the midline was significantly greater than that in the left and right hemispheres (all ps < 0.001), while the left and right hemispheres were not significantly different from each other ($p = 0.11$) for the neural processing of both types of deviants in nonspeech stimuli.

P300 Latency

For the peak latency of the P300 difference wave, there was a significant main effect of *category type* [$F(1, 28) = 5.21$, $p < 0.05$, $\eta_p^2 = 0.16$] and a significant four-way interaction of *category type × stimulus type × hemisphere × group* [$F(2, 56) = 3.61$, $p < 0.05$, $\eta_p^2 = 0.11$]. No other effects reached significance. Next, simple main effect analysis indicated that in the speech context, no effects reached significance (all ps > 0.05), indicating that in the speech stimuli, the peak latency elicited from within-category deviants was similar to that elicited from between-category deviants. In the nonspeech context, only the main effects of *category type* [$F(1, 28) = 9.13$, $p < 0.01$, $\eta_p^2 = 0.25$] and *hemisphere* [$F(2, 56) = 3.86$, $p < 0.05$, $\eta_p^2 = 0.12$] reached significance. The significant main effect of *category type* indicated that for both the controls and amusics, the mean peak latency of within-category deviants (mean = 424.22 ms) shifted earlier compared to that of between-category deviants (mean = 465.64 ms) in the nonspeech stimuli. Regarding *hemisphere*, however, pair-wise comparison with Bonferroni adjustment showed that none of the peak latencies between two hemispheres differed from the others (all ps > 0.05). Furthermore, these statistical results also implied that the group difference was not significant for the P300 latency in either the speech or nonspeech contexts.
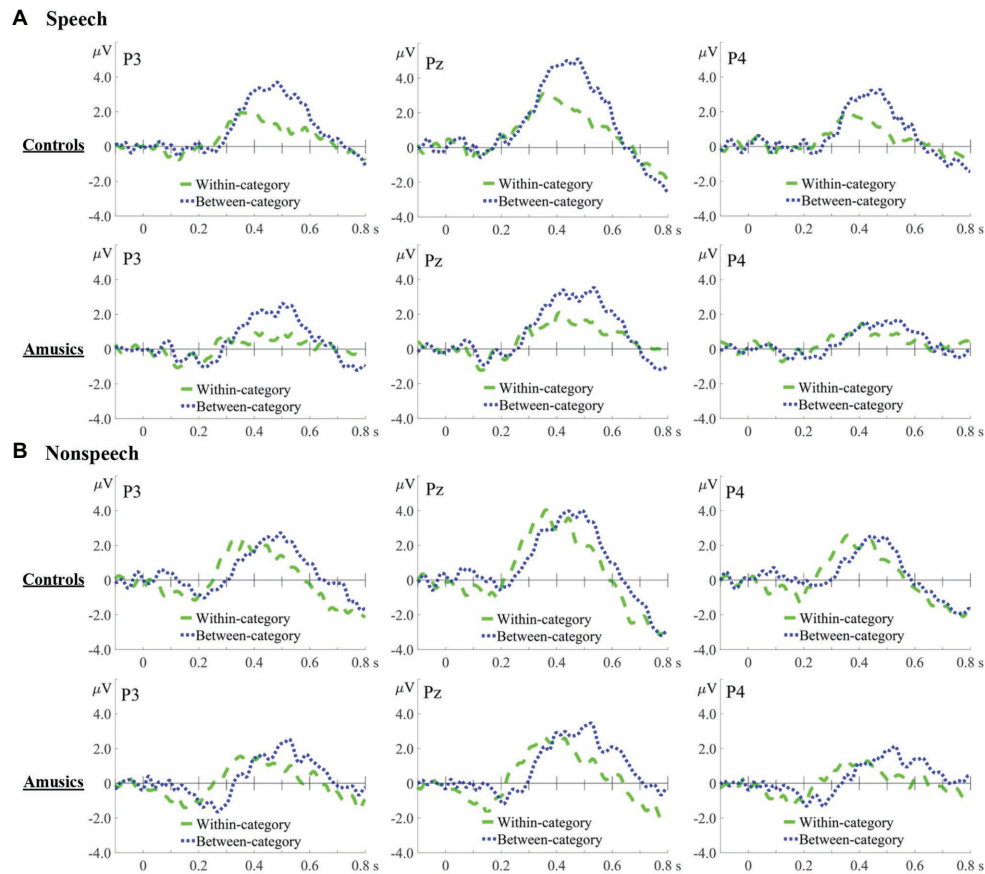
FIGURE 6 | Difference waves in speech **(A)** and nonspeech **(B)** contexts for within-category deviants minus standards (green dashed lines) and between-category deviants minus standards (blue dotted lines) in the controls and amusics, which are displayed at three representative electrodes: P3 (left), Pz (middle), and P4 (right).
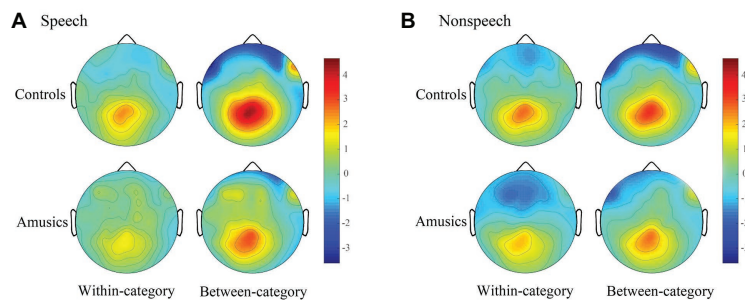

FIGURE 7 | Topographic maps in speech **(A)** and nonspeech **(B)** contexts which demonstrate the scalp distributions of the P300 amplitudes (320–600 ms) of within-category and between-category difference waves (deviants minus standards) across the controls and amusics.

DISCUSSION

Reduced Attentive Processing of Between-Category Information of Lexical Tones in Mandarin-Speaking Amusics

In order to illuminate the level of lexical tone processing difficulties in Mandarin-speaking amusics, this study manipulated tonal category (within-category/across-category) along a pitch continuum to directly dissociate acoustic processing from the phonological processing of lexical tones for native speakers, an approach that was widely adopted in previous ERP studies (Xi et al., 2010; Zhang et al., 2012; Yu et al., 2014, 2017). A posttest identification test after the current P300 experiment confirmed that both the controls and amusics perceived (1) stimulus #1 (within-category deviant) as the same tone category as stimulus #4 (standard) and (2) the equally spaced stimulus #7
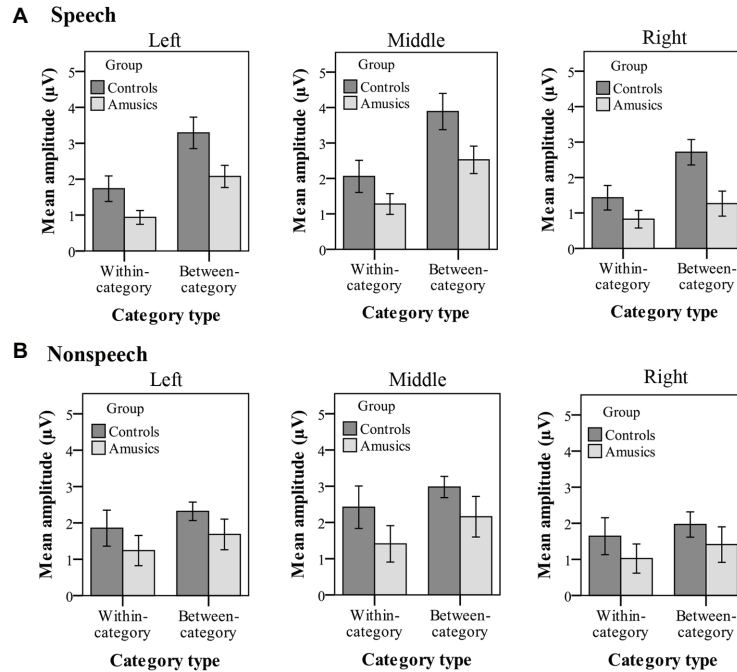
FIGURE 8 | Mean P300 amplitude of within-category and between-category difference waves elicited in **(A)** speech stimuli and **(B)** nonspeech stimuli for both groups. The P300 amplitude was calculated from 320 to 600 ms by averaging the voltage measured at electrodes on the left hemisphere (P1, P3, P5, and PO3), in the midline (Pz, POz), and on the right hemisphere (P2, P4, P6, and PO4), respectively. Error bars: +/−1 standard error.
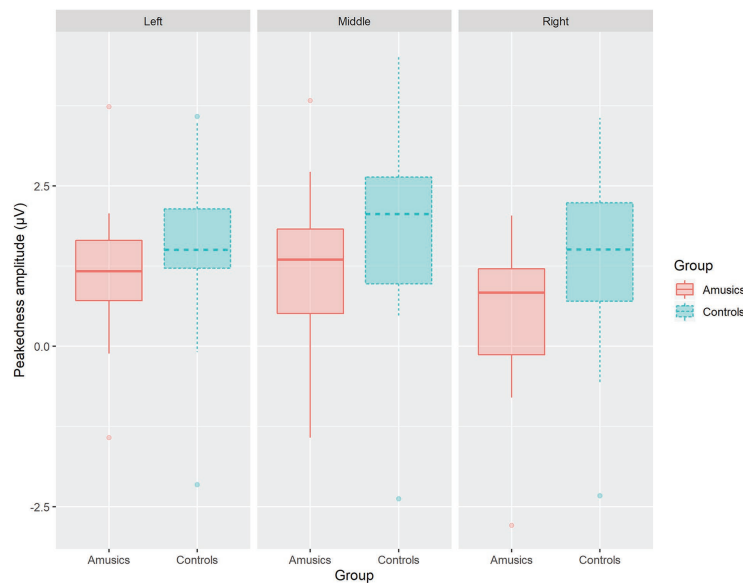

FIGURE 9 | The boxplot of peakedness amplitude (the amplitude of between-category deviants minus that of within-category deviants) between the controls and amusics in the speech context.

(between-category deviant) as a distinct tone category in both speech and nonspeech contexts (see **Figure 3**). As reflected by both the behavioral and electrophysiological measures, the d' scores (see **Figure 4**) and P300 amplitudes (see **Figure 8**)

in both speech and nonspeech contexts were systematically lower in the amusics than the controls when responding to both within- and between-category deviants, indicating that both lower-level acoustic processing and higher-level phonological

processing were compromised among the tone language speakers with amusia, regardless of the linguistic status of the signal.

At the acoustic processing level, the result that the amusics exhibited reduced neural and behavioral processing of within-category deviants is in accordance with well-established findings that have consistently demonstrated that amusics are typically less sensitive to various fine-grained acoustic properties, such as the pitch information of music, nonspeech, and speech sounds (e.g., Peretz, 2001; Peretz et al., 2002; Nan et al., 2010; Jiang et al., 2012), and even to other acoustic correlates beyond pitch (Whiteford and Oxenham, 2017; Zhang et al., 2017a). Especially in regard to lexical tone perception, non-tonal language speakers with amusia demonstrate inferior performance of lexical tone discrimination in Mandarin Chinese (Nguyen et al., 2009; Tillmann et al., 2011) and Thai (Tillmann et al., 2011). Moreover, the native Mandarin (a tone language) speakers with amusia also show an impoverished perceptual skill in discriminating non-native Cantonese level tones (Wang and Peng, 2014). In brief, the aforementioned compromised processing of non-native lexical tones in amusics from different language backgrounds mainly reflects a domain-general acoustic processing deficit in the speech domain transferred from the music domain. Similarly, our findings further corroborated that the attentive processing of acoustic variations when detecting the within-category changes along native tonal stimuli is also compromised in both speech and nonspeech contexts. Thus, long-term tone language experience cannot compensate for the acoustic processing deficits in amusics even when the acoustic pitch is embedded in native tone categories, suggesting that the acoustic processing of pitch across different domains may share overlapping cognitive resources and/or processes (Patel, 2008, 2012).

At the phonological processing level, it has been proved that the developmental trajectory of native phonological categories starts early, from around 6 months in infants (Kuhl, 2004). The development of higher-level phonological perception in native speakers is thus driven by an accumulation of perceptual development from the tonal input of the ambient native sound environment (Chen et al., 2017). The main focus of this study was to figure out whether the higher-level phonological processing of native tone categories in amusics is more or less intact due to long-term exposure to the native tonal language or, alternatively, compromised at the attentive stage. The ERP component of P300 (P3b), which is associated with controlled processing in the attentive condition (Sutton et al., 1965), can act as an indicator of phonological processing capacity in the speech context (Maiste et al., 1995; Frenck-Mestre et al., 2005), and this feature may extend to nonspeech contexts (Chandrasekaran et al., 2009). The elicited P300 is modulated by the subject's overall arousal level as resources are allocated to process different types of stimuli: the more demanding the task, the smaller the amplitude of the P300, and vice versa (Polich, 2007). In this study, the significantly greater amplitude elicited from the control group in response to between-category speech deviants implies the greater ease with which the controls engaged in phonological processing to distinguish between the F0 contours of distinct lexical tone categories. This difference also implies that the attentive

phonological processing of lexical tone differences in Mandarin-speaking amusics might be compromised, which is in line with the reduced d' values of between-category variations among the amusic cohort found in the behavioral measurements. To conclude, the results of the present experiment support the notion that the reduced acoustic processing in amusia is manifested broadly and profoundly and further persists into the higher-level phonological processing that involves the attentive processing of different lexical tone categories.

The Influence of Lower-Level Acoustic Processing on the Higher-Level Phonological Processing of Lexical Tones

To accurately understand speech signals, listeners should process the basic acoustic characters and assign these highly variable speech sounds efficiently to various phonemic categories. These two types of information, acoustic information and phonological information are also dealt with when a native speaker processes lexical tones in Mandarin Chinese. Previous studies suggest that lower-level acoustic processing and higher-level phonological processing are represented differently in our human brain. One line of evidence reveals that speech is processed hierarchically along the auditory pathways, with the upstream areas (e.g., the dorsal STG areas) performing acoustic processing and the downstream regions (e.g., the ventral superior temporal sulcus, STS, and middle temporal gyrus, MTG regions) responsible for phonological processing (Wessinger et al., 2001; Okada et al., 2010; Zhang et al., 2011). The other line of evidence regarding brain lateralization indicates that although lexical tonal processing engages both hemispheres, pure acoustic processing tends to be processed in the right hemisphere, while phonological and semantic processing mainly occurs in the left hemisphere (Gandour et al., 2004; Gandour, 2006). Especially for Mandarin tone processing, Xi et al. (2010) found that between-category tonal deviants (i.e., phonological processing) elicited larger MMN than within-category tonal deviants (i.e., acoustic processing) in the left hemisphere, while the within-category deviants elicited larger MMN in the right hemisphere. At the attentive stage, Zhang et al. (2012) found that that for the P300 component, the amplitudes elicited by the within-category deviants were similar between the left and the right recording sites. However, the across-category deviants elicited greater P300 amplitudes in the left recording sites compared with the right sites. This hemispheric pattern found in Zhang et al. (2012) is the same as our current findings in both Mandarin-speaking controls and amusics when processing speech sounds, confirming that the phonological processing of lexical tones in tone language speakers happens in both hemispheres, although there is a bias toward the left side of our brain.

Although the aforementioned neuroimaging studies have implied that a dissociation exists between lower-level acoustic and higher-level phonological processing at different functional hierarchy levels (dorsal STG vs. ventral STS/MTG) and in different hemispheres (right vs. left), the way these two levels of processing interact with each other in lexical tone perception is not well understood. In one direction, phonological processing may influence acoustic processing. Zhang et al. (2011) found

that compared to within-category variations, between-category stimuli elicit stronger activation in the middle MTG, reflecting the higher-level phonological representations. Meanwhile, activation on the dorsal plane of STG decreases significantly, suggesting that lower-level acoustic processing is modulated and dominated by higher-level phonological processing through "feedback mechanisms" (Hickok and Poeppel, 2007). Moreover, Zhao and Kuhl (2015) conducted a behavioral study by investigating the relative contribution of higher- and lower-level influences when both are present. They also concluded that higher-level linguistic categories dominate lower-level acoustics in lexical tone processing.

In the other direction, the "feed-forward mechanisms" (Binder, 2000; Scott and Wise, 2004) suggest that speech processing starts from the core auditory areas (STG) to downstream brain areas and then to more lateral and anterior regions, implying that initial bottom-up acoustic processing lays the foundation of phonological processing. The current findings provide direct evidence on the feed-forward mechanisms by showing that lower-level acoustics underlie higher-level phonological processing in lexical tone perception since the reduced acoustic processing skill in amusics extends to phonological processing in native perceivers. As congenital amusia is regarded as being inborn (Peretz, 2008), it is very likely that the impoverished acoustic pitch processing skill in this cohort interferes with the formation of native phonological categories from the very beginning. Taken together, the roles of both feed-forward and feedback mechanisms imply the existence of cortical dynamics (Zhang et al., 2011) for a bidirectional interaction between bottom-up acoustic analysis and top-down phonological processing in speech perception. The cortical dynamics support the dual-stream model of speech processing (Hickok and Poeppel, 2004, 2007) that claims reciprocal interactions across different brain regions. With more neuroimaging studies in the future, more will be revealed about the neurophysiological nature of the relative influence of acoustic and phonological processing in lexical tone perception.

Neurophysiological Evidence for Preserved CP of Lexical Tones in Tone Language Speakers With Amusia

Categorical perception refers to a tendency for native listeners of a particular language to classify the sounds used in their language as discrete, categorical, and linguistic representations. The CP pattern of speech sounds is characterized by a sharp identification curve and an enhanced discrimination accuracy around the boundary position (Liberman et al., 1957). Thus, the cross-boundary benefit is likely to be the defining feature of CP (Massaro, 1987; Zhang, 2016). In other words, there is no support for claiming that CP occurs if there is no benefit for between-category discrimination compared to within-category discrimination. In terms of lexical tone perception, there is ample behavioral evidence that native perception of the Mandarin tonal continuum, due to the necessary involvement of contour tones, tends to be categorical, with higher sensitivity to between-category discrimination pairs relative to within-category pairs

(e.g., Wang, 1976; Xu et al., 2006; Peng et al., 2010; Shen and Froud, 2016; Chen et al., 2017). Moreover, several ERP studies have provided electrophysiological evidence for the CP of lexical tones in both the pre-attentive stage (Xi et al., 2010; Yu et al., 2014, 2017; Shen and Froud, 2019) and the attentive stage (Zheng et al., 2012; Shen and Froud, 2019) by exhibiting more negative amplitudes (MMN) and more positive amplitudes (P300), respectively, elicited from between-category deviants relative to within-category deviants in native tone language speakers.

The previous studies investigating the CP of lexical tones in tone language speakers with amusia were all behavioral studies and drew different conclusions (Jiang et al., 2012; Huang et al., 2015; Zhang et al., 2017a). At the whole group level, Jiang et al. (2012) found an impaired CP of Mandarin tones in amusics, with no improvement for discrimination pairs that crossed the classification boundary. However, Zhang et al. (2017a) found that Cantonese-speaking amusics did perceive the lexical tones categorically, as indicated by the higher d' for between-category pairs than for within-category pairs. Consistent with Zhang et al. (2017a), our behavioral results also indicate a higher sensitivity to between-category variations in Mandarin-speaking amusics. Besides, the current P300 results further offer electrophysiological evidence for claiming a preserved CP of lexical tones in Mandarin-speaking amusics. On the one hand, for both groups, controls and amusics, the across-category deviants did elicit a significantly higher amplitude than the within-category deviants in the speech context. On the other hand, the peakedness measure represents the magnitude of the benefit for discriminations that cross the identification boundary. In the current study, both the peakedness d' of the behavioral index and the peakedness amplitude of the ERP index consistently showed no group differences.

The discrepancy regarding the CP pattern in the literature may largely be attributed to acoustic differences in the step size utilized in various studies. It is important to note that the step size of the discrimination pairs used in Jiang et al. (2012) was only 6 Hz, which falls in the range of just-noticeable differences (JNDs) for F0 discrimination (4–8 Hz) even among healthy Mandarin-speaking individuals (Liu, 2013). Given that the amusics generally showed a fine-grained pitch perception deficit, the 6-Hz acoustic difference used in the discrimination task might be too small to reveal the categorical nature in amusics. As indicated in Jiang et al. (2012), both the between- and within-category accuracies were close to the chance level due to a floor effect in the amusic group. The essential characteristic of the CP pattern reemerged in the tone language speakers with amusia when the acoustic distance of step size was enlarged (around 9–10 Hz on average in Zhang et al., 2017a, and 18 Hz in the current study). Furthermore, Huang et al. (2015) also confirmed that most Mandarin-speaking amusics (called "pure amusics" in their study) show a cross-boundary benefit when perceiving lexical tones behaviorally, in support of a preserved CP pattern, while a small portion of amusics (called "tone agnosics," who showed very severe tone perception difficulties behaviorally, with an accuracy of 3 SDs below the mean of controls) still show a lack of CP of native tones. The view of subgroup differences in the CP

of lexical tones in amusics needs to be corroborated with neurobiological investigation in future studies. At the very least, the present findings, from both behavioral and neural evidence, demonstrate a preserved CP pattern in Mandarin-speaking amusics at the whole group level.

Furthermore, as indicated by the indices of d' value and P300 amplitude, the CP pattern of cross-boundary benefit in the speech context was not transferred to nonspeech counterparts for either group in this study, whereas such a cross-boundary benefit was often observed in the nonspeech context when it exhibited some of the critical features of speech (i.e., pitch; e.g., Xu et al., 2006; Xi et al., 2010; Zhang et al., 2017a), reflecting a carry-over influence of long-term phonological processing from the speech to nonspeech domain. In this study, for both the controls and amusics, the d' values of between-category de*via*nts in speech were indeed comparable to those in nonspeech, whereas nonspeech yielded a higher d' value of within-category deviants than speech (see **Figure 4**). As mentioned, CP is characterized as enhanced between-category discrimination as well as "dulled" within-category discrimination in the speech context, while within-category discriminations can be dulled less in the nonspeech context (Xu et al., 2006). The pattern that nonspeech yields higher within-category sensitivity than speech seems to be generalizable to all native tone-language speakers, be they controls (Xu et al., 2006; Zheng et al., 2012), musicians (Wu et al., 2015), or amusics (this study). Since the current study utilized a much larger step size (18 Hz) compared to other related studies (Xu et al., 2006; Zheng et al., 2012; Wu et al., 2015), this may lead to un-dulled within-category sensitivity in nonspeech that is comparable to the between-category sensitivity. Moreover, the within-category stimuli (#1) utilized in this study showed a much higher rising contour slope compared with the between-category stimuli (#7), which may partly explain the shorter RT (see **Figure 5B**) and earlier peak latency (see **Figure 6B**) in response to the within-category deviants in the nonspeech context. Nevertheless, no such differences between the two types of deviants were observed in the speech counterparts, probably due to the much easier detection of between-category speech deviants, which might make their processing time course shift earlier as well.

Another issue that arises with the current study is the underlying mechanisms to explain the reduced acoustic and phonological processing skills but preserved CP pattern in tone language speakers with amusia. This perceptual pattern in Mandarin-speaking amusics is essentially different from that found in non-tonal language speaking individuals who show both phonological processing deficit and an impaired CP in their perception of non-native lexical tones (e.g., Wang, 1976; Xu et al., 2006; Peng et al., 2010; Shen and Froud, 2016, 2019). These non-tonal language speakers show a phonological processing deficit due to a lack of exposure to lexical tones, but this could be improved with increased exposure and practice (Shen and Froud, 2016). However, for tone language speakers with amusia, although their phonological processing capacity was reduced to some extent, but not totally impaired due to the long-term native language exposure. According to the multistore model (Xu et al., 2006), three

forms of memory – sensory memory and the short- and long-term forms of categorical memory – are involved in CP. Among these forms of memory, long-term categorical memory facilitates the categorization of speech sounds only in native speakers, and short-term categorical memory may become permanently preserved in long-term memory *via* long-term language exposure. Different from non-tonal language speakers, the presence of CP pattern in Mandarin-speaking amusics indicates that they might have normal or near normal long-term categorical memory (Jiang et al., 2012; Zhang et al., 2017a). The reduced acoustic and phonological processing skills in Mandarin amusics could possibly be attributed to difficulties in sensory memory and short-term categorical memory, respectively, which is in line with the proposed short-term pitch memory deficit in non-tonal language speakers with amusia (Gosselin et al., 2009; Tillmann et al., 2009, 2016; Albouy et al., 2015).

CONCLUSION

In light of the positive transfer effects between music and language, it has been proposed that the music processing deficit in amusics would affect speech perception, such as lexical tone perception. This study aimed to uncover the neural mechanisms underpinning the lexical tone processing deficits in Mandarin-speaking amusics during the attentive stage (P300): whether such deficits in tone language speakers with amusia are merely caused by a domain-transferred deficit in lower-level acoustic processing or additionally influenced by reduced higher-level processing of phonological categories. Both the behavioral and P300 findings indicate that compared to the controls, both lower-level acoustic processing and higher-level phonological processing were reduced among the Mandarin-speaking amusics regardless of the linguistic status of the signal. Moreover, the current results imply an intact CP pattern in Mandarin-speaking amusics at the whole group level by exhibiting a cross-boundary perceptual benefit and a comparable peakedness in lexical tone perception, indicating that they may have normal or near normal long-term categorical memory due to long-term exposure to native tone categories. Considering the previous findings and the present results together, we suggest that long-term tone language experience may not compensate for the acoustic pitch processing deficit in tone language speakers with amusia but rather may be extended to the neural processing of the phonological information of lexical tones during the attentive stage.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Human Research Ethics Committee of the Shenzhen

Institutes of Advanced Technology, the Chinese Academy of Sciences. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FC and GP conceived and designed the experiment. FC implemented the experiment, collected, and analyzed the data. FC and GP interpreted the data and wrote the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P. E., et al. (2013). Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain* 136, 1639–1661. doi: 10.1093/brain/awt082

Albouy, P., Mattout, J., Sanchez, G., Tillmann, B., and Caclin, A. (2015). Altered retrieval of melodic information in congenital amusia: insights from dynamic causal modeling of MEG data. *Front. Hum. Neurosci.* 9:20. doi: 10.3389/fnhum.2015.00020

Bidelman, G. M., Gandour, J. T., and Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* 23, 425–434. doi: 10.1162/jocn.2009.21362

Bidelman, G. M., Hutka, S., and Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PLoS One* 8:e60676. doi: 10.1371/journal.pone.0060676

Binder, J. (2000). The new neuroanatomy of speech perception. *Brain* 123, 2371–2372. doi: 10.1093/brain/123.12.2371

Boersma, P., and Weenink, D. (2009). Praat: Doing phonetics by computer [Computer program]. Available at: http://www.praat.org/ (Accessed November 14, 2018).

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. doi: 10.1016/j.bandl.2008.02.001

Chen, F., and Peng, G. (2016). Context effect in the categorical perception of mandarin tones. *J. Signal Process. Sys.* 82, 253–261. doi: 10.1007/s11265-015-1008-2

Chen, F., Peng, G., Yan, N., and Wang, L. (2017). The development of categorical perception of mandarin tones in four-to seven-year-old children. *J. Child Lang.* 44, 1413–1434. doi: 10.1017/S0305000916000581

Electrical Geodesics (2006). Net station viewer and waveform tools tutorial, S-MAN-200-TVWR-001; December 21, 2006. Eugene, OR: Author.

Finney, D. J. (1971). *Probit analysis. 3rd Edn.* Cambridge: Cambridge University Press.

Frenck-Mestre, C., Meunier, C., Espesser, R., Daffner, K., and Holcomb, P. (2005). Perceiving nonnative vowels: the effect of context on perception as evidenced by event-related brain potentials. *J. Speech Lang. Hear. Res.* 48, 1496–1510. doi: 10.1044/1092-4388(2005/104)

Fujisaki, H., and Kawashima, T. (1971). A model of the mechanisms for speech perception-quantitative analysis of categorical effects in discrimination. Annual report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo; 30, 59–68.

Gandour, J. (2006). "Brain mapping of Chinese speech prosody" in *Handbook of east Asian psycholinguistics. Vol. 1.* eds. P. Li, L. H. Tan, E. Bates and O. J. L. Tzeng (New York, NY: Cambridge University Press), 308–319.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., et al. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage* 23, 344–357. doi: 10.1016/j.neuroimage.2004.06.004

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Lancker, D. V., and Hutchins, G. D. (2000). A cross-linguistic PET study of tone perception. *J. Cogn. Neurosci.* 12, 207–222. doi: 10.1162/089892900561841

Gosselin, N., Jolicoeur, P., and Peretz, I. (2009). "Impaired memory for pitch in congenital amusia" in *Neurosciences and music III: Disorders and plasticity. Vol. 1169.* eds. S. DallaBella, N. Kraus, K. Overy, C. Pantev, J. S. Snyder, M. Tervaniemi et al. (New York, NY: New York Academy of Sciences), 270–272.

Gu, F., Zhang, C., Hu, A., and Zhao, G. (2013). Left hemisphere lateralization for lexical and acoustic pitch processing in Cantonese speakers as revealed by mismatch negativity. *Neuroimage* 83, 637–645. doi: 10.1016/j.neuroimage.2013.02.080

Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi: 10.1016/j.cognition.2003.10.011

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113

Huang, W. T., Liu, C., Dong, Q., and Nan, Y. (2015). Categorical perception of lexical tones in mandarin-speaking congenital amusics. *Front. Psychol.* 6:829. doi: 10.3389/fpsyg.2015.00829

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. doi: 10.1523/JNEUROSCI.3039-07.2007

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., and Peretz, I. (2006). Morphometry of the amusic brain: a two-site study. *Brain* 129, 2562–2570. doi: 10.1093/brain/awl204

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. doi: 10.1093/cercor/bhq094

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., and Yang, Y. (2012). Impaired categorical perception of lexical tones in mandarin-speaking congenital amusics. *Mem. Cognit.* 40, 1109–1121. doi: 10.3758/s13421-012-0208-2

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843. doi: 10.1038/nrn1533

Lee, C. Y., and Hung, T. H. (2008). Identification of mandarin tones by English-speaking musicians and nonmusicians. *J. Acoust. Soc. Am.* 124, 3235–3248. doi: 10.1121/1.2990713

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

Liu, C. (2013). Just noticeable difference of tone pitch contour change for English- and Chinese-native listeners. *J. Acoust. Soc. Am.* 134, 3011–3020. doi: 10.1121/1.4820887

Macmillan, N. A., and Creelman, C. D. (2008). *Detection theory: A user's guide.* New York: Psychology Press.

Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., and Picton, T. W. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear Hear.* 16, 68–90. doi: 10.1097/00003446-199502000-00006

Mandell, J., Schulze, K., and Schlaug, G. (2007). Congenital amusia: an auditory-motor feedback disorder? *Restor. Neurol. Neurosci.* 25, 323–334.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Erlbaum.

Moreau, P., Jolicœur, P., and Peretz, I. (2013). Pitch discrimination without awareness in congenital amusia: evidence from event-related potentials. *Brain Cogn.* 81, 337–344. doi: 10.1016/j.bandc.2013.01.004

Näätänen, R. (2000). Mismatch negativity (MMN): perspectives for application. *Int. J. Psychophysiol.* 37, 3–10. doi: 10.1016/s0167-8760(00)00091-x

Nan, Y., Huang, W. T., Wang, W. J., Liu, C., and Dong, Q. (2016). Subgroup differences in the lexical tone mismatch negativity (MMN) among mandarin speakers with congenital amusia. *Biol. Psychol.* 113, 59–67. doi: 10.1016/j.biopsycho.2015.11.010

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. doi: 10.1093/brain/awq178

Nguyen, S., Tillmann, B., Gosselin, N., and Peretz, I. (2009). Tonal language processing in congenital amusia. *Ann. N. Y. Acad. Sci.* 1169, 490–493. doi: 10.1111/j.1749-6632.2009.04855.x

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

Omigie, D., Pearce, M. T., Williamson, V. J., and Stewart, L. (2013). Electrophysiological correlates of melodic processing in congenital amusia. *Neuropsychologia* 51, 1749–1762. doi: 10.1016/j.neuropsychologia.2013.05.010

Patel, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.

Patel, A. D. (2012). "Language, music, and the brain: a resource-sharing framework" in *Language and music as cognitive systems*. eds. P. Rebuschat, M. Rohrmeier, J. Hawkins and I. Cross (Oxford, U.K.: Oxford University Press), 204–223.

Peng, G., Deutsch, D., Henthorn, T., Su, D., and Wang, W. S. (2013). Language experience influences nonlinguistic pitch perception. *J. Chinese Linguist.* 41, 447–467.

Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., and Wang, W. S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *J. Phon.* 38, 616–624. doi: 10.1016/j.wocn.2010.09.003

Peretz, I. (2001). Brain specialization for music. New evidence from congenital amusia. *Ann. N. Y. Acad. Sci.* 930, 153–165. doi: 10.1111/j.1749-6632.2001.tb05731.x

Peretz, I. (2008). Musical disorders: from behavior to genes. *Curr. Dir. Psychol. Sci.* 17, 329–333. doi: 10.1111/j.1467-8721.2008.00600.x

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., et al. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron* 33, 185–191. doi: 10.1016/s0896-6273(01)00580-3

Peretz, I., Brattico, E., Järvenpää, M., and Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain* 132, 1277–1286. doi: 10.1093/brain/awp055

Peretz, I., Brattico, E., and Tervaniemi, M. (2005). Abnormal electrical brain responses to pitch in congenital amusia. *Ann. Neurol.* 58, 478–482. doi: 10.1002/ana.20606

Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., et al. (2008). On-line identification of congenital amusia. *Music Percept.* 25, 331–343. doi: 10.1525/mp.2008.25.4.331

Peretz, I., Vuvan, D., Lagrois, M. É., and Armony, J. L. (2015). Neural overlap in processing music and speech. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140090. doi: 10.1098/rstb.2014.0090

Pfordresher, P. Q., and Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Atten. Percept. Psychophysiol.* 71, 1385–1398. doi: 10.3758/APP.71.6.1385

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Scerif, G., Worden, M. S., Davidson, M., Seiger, L., and Casey, B. J. (2006). Context modulates early stimulus processing when resolving stimulus-response conflict. *J. Cogn. Neurosci.* 18, 781–792. doi: 10.1162/jocn.2006.18.5.781

Scott, S. K., and Wise, R. J. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition* 92, 13–45. doi: 10.1016/j.cognition.2002.12.002

Shen, G., and Froud, K. (2016). Categorical perception of lexical tones by English learners of mandarin Chinese. *J. Acoust. Soc. Am.* 140, 4396–4403. doi: 10.1121/1.4971765

Shen, G., and Froud, K. (2019). Electrophysiological correlates of categorical perception of lexical tones by English learners of mandarin Chinese: an ERP study. *Biling. Lang. Cogn.* 22, 253–265. doi: 10.1017/S136672891800038X

Sutton, S., Braren, M., Zubin, J., and John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science* 150, 1187–1188. doi: 10.1126/science.150.3700.1187

Tang, W., Xiong, W., Zhang, Y. X., Dong, Q., and Nan, Y. (2016). Musical experience facilitates lexical tone processing among mandarin speakers: behavioral and neural evidence. *Neuropsychologia* 91, 247–253. doi: 10.1016/j.neuropsychologia.2016.08.003

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

Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., and Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Res.* 1640, 251–263. doi: 10.1016/j.brainres.2015.10.035

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. doi: 10.1016/j.bandc.2009.08.003

van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Paper. Linguist. Int. J. Hum. Commun.* 13, 201–277. doi: 10.1080/08351818009370498

Wang, W. S. -Y. (1976). "Language change" in *Origins and evolution of language and speech*. *Vol. 280.* eds. S. R. Harnad, H. D. Steklis and J. Lancaster (New York: New York Academy of Sciences), 61–72.

Wang, X., and Peng, G. (2014). Phonological processing in mandarin speakers with congenital amusia. *J. Acoust. Soc. Am.* 136, 3360–3370. doi: 10.1121/1.4900559

Wessinger, C. M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., and Rauschecker, J. P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J. Cogn. Neurosci.* 13, 1–7. doi: 10.1162/089892901564108

Whiteford, K. L., and Oxenham, A. J. (2017). Auditory deficits in amusia extend beyond poor pitch perception. *Neuropsychologia* 99, 213–224. doi: 10.1016/j.neuropsychologia.2017.03.018

Wong, P. C., Ciocca, V., Chan, A. H., Ha, L. Y., Tan, L. H., and Peretz, I. (2012). Effects of culture on musical pitch perception. *PLoS One* 7:e33424. doi: 10.1371/journal.pone.0033424

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

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

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

Yu, K., Wang, R., Li, L., and Li, P. (2014). Processing of acoustic and phonological information of lexical tones in mandarin Chinese revealed by mismatch negativity. *Front. Hum. Neurosci.* 8:729. doi: 10.3389/fnhum.2014.00729

Yu, K., Zhou, Y., Li, L., Su, J. A., Wang, R., and Li, P. (2017). The interaction between phonological information and pitch type at pre-attentive stage: an ERP study of lexical tones. *Lang. Cogn. Neurosci.* 32, 1164–1175. doi: 10.1080/23273798.2017.1310909

Zhang, Y. (2016). "Categorical perception" in *Encyclopedia of Chinese language and linguistics*. eds. R. Sybesma, W. Behr, Y. Gu, Z. Handel, J. Huang and J. Myers (Leiden, Netherlands: Brill).

Zhang, C., Peng, G., Shao, J., and Wang, W. S. Y. (2017b). Neural bases of congenital amusia in tonal language speakers. *Neuropsychologia* 97, 18–28. doi: 10.1016/j.neuropsychologia.2017.01.033

Zhang, C., Peng, G., and Wang, W. S. Y. (2013). Achieving constancy in spoken word identification: time course of talker normalization. *Brain Lang.* 126, 193–202. doi: 10.1016/j.bandl.2013.05.010

Zhang, C., and Shao, J. (2018). Normal pre-attentive and impaired attentive processing of lexical tones in Cantonese-speaking congenital amusics. *Sci. Rep.* 8:8420. doi: 10.1038/s41598-018-26368-7

Zhang, C., Shao, J., and Huang, X. (2017a). Deficits of congenital amusia beyond pitch: evidence from impaired categorical perception of vowels in

Cantonese-speaking congenital amusics. *PLoS One* 12:e0183151. doi: 10.1371/journal.pone.0183151

Zhang, L., Xi, J., Wu, H., Shu, H., and Li, P. (2012). Electrophysiological evidence of categorical perception of Chinese lexical tones in attentive condition. *Neuroreport* 23, 35–39. doi: 10.1097/WNR.0b013e32834e4842

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

Zhao, T. C., and Kuhl, P. K. (2015). Higher-level linguistic categories dominate lower-level acoustics in lexical tone processing. *J. Acoust. Soc. Am.* 138, EL133–EL137. doi: 10.1121/1.4927632

Zheng, H. Y., Minett, J. W., Peng, G., and Wang, W. S. (2012). The impact of tone systems on the categorical perception of lexical tones: an event-related potentials study. *Lang. Cognit. Process.* 27, 184–209. doi: 10.1080/01690965.2010.520493