NeuroImage: Clinical 23 (2019) 101814

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Talker normalization in typical Cantonese-speaking listeners and congenital amusics: Evidence from event-related potentials

Jing Shao[a,b], Caicai Zhang[a,b,*]

[a] The Hong Kong Polytechnic University, Department of Chinese and Bilingual Studies, Hong Kong, China
[b] Research Centre for Language, Cognition, and Neuroscience, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Keywords:
Talker normalization
Time course
Talker variability
Congenital amusia
Cantonese tone
ERPs

ABSTRACT

Despite the lack of invariance in the mapping between the acoustic signal and phonological representation, typical listeners are capable of using information of a talker's vocal characteristics to recognize phonemes, a process known as "talker normalization". The current study investigated the time course of talker normalization in typical listeners and individuals with congenital amusia, a neurodevelopmental disorder of refined pitch processing. We examined the event-related potentials (ERPs) underling lexical tone processing in 24 Cantonese-speaking amusics and 24 typical listeners (controls) in two conditions: blocked-talker and mixed-talker conditions. The results demonstrated that for typical listeners, effects of talker variability can be observed as early as in the N1 time-window (100–150 ms), with the N1 amplitude reduced in the mixed-talker condition. Significant effects were also found in later components: the N2b/c peaked significantly earlier and the P3a and P3b amplitude was enhanced in the blocked-talker condition relative to the mixed-talker condition, especially for the tone pair that is more difficult to discriminate. These results suggest that the blocked-talker mode of stimulus presentation probably facilitates auditory processing and requires less attentional effort with easier speech categorization than the mixed-talker condition, providing neural evidence for the "active control theory". On the other hand, amusics exhibited comparable N1 amplitude to controls in both conditions, but deviated from controls in later components. They demonstrated overall later N2b/c peak latency significantly reduced P3a amplitude in the blocked-talker condition and reduced P3b amplitude irrespective of talker conditions. These results suggest that the amusic brain was intact in the auditory processing of talker normalization processes, as reflected by the comparable N1 amplitude, but exhibited reduced automatic attentional switch to tone changes in the blocked-talker condition, as captured by the reduced P3a amplitude, which presumably underlies a previously reported perceptual "anchoring" deficit in amusics. Altogether, these findings revealed the time course of talker normalization processes in typical listeners and extended the finding that conscious pitch processing is impaired in the amusic brain.

1. Introduction

1.1. Talker normalization in speech perception

The mapping between acoustic patterns and phonological categories is not one-to-one, but many-to-many (Liberman et al., 1967). Speech production is easily influenced by the neighboring phonemes, speaking rate and talker voice characteristics, and the consequence is a lack of invariance between acoustic signals and phonemes. However, typical listeners are capable of resolving the complex mapping problem by using the acoustic information of a talker's vocal characteristics (Syrdal and Gopal, 1986) or information from external acoustic signals (Ladefoged and Broadbent, 1957), a process referred to as "talker normalization" (Wong et al., 2004).

One classic paradigm to examine the processes of talker normalization is to present the speech stimuli produced by multiple talkers in two conditions, i.e., in a blocked-talker (i.e., a single talker in a block) vs. mixed-talker (i.e., multiple talkers intermixed in one block) manner. Numerous studies have found that the accuracy of speech recognition was lowered and the response time was longer in the mixed-talker condition than the blocked-talker condition (Green et al., 1997; Lee, 2009; Lee et al., 2010; Mullennix and Pisoni, 1990; Nusbaum and Morin, 1992; Strange et al., 1976; Wong and Diehl, 2003). Similar findings were reported for lexical tone perception. For example, Cantonese speakers identified Cantonese tones more accurately when the tone stimuli were blocked by the talker than when the talkers were

* Corresponding author at: The Hong Kong Polytechnic University, Department of Chinese and Bilingual Studies, Hong Kong, China.
E-mail address: caicai.zhang@polyu.edu.hk (C. Zhang).

https://doi.org/10.1016/j.nicl.2019.101814
Received 29 October 2018; Received in revised form 20 March 2019; Accepted 2 April 2019
Available online 03 April 2019

mixed in a block (Wong and Diehl, 2003).

The active control theory has been put forward to account for the phenomena mentioned above. It has been proposed that listeners use an active control mechanism to resolve the multiple mapping problem caused by talker variability (Magnuson and Nusbaum, 2007; Nusbaum and Morin, 1992). When there is a constant talker, the source of variability is constrained, and listeners can make use of the consistent acoustic-phonetic relationships to form a phonological representation and to estimate the talker's acoustic-phonetic space (Nusbaum and Morin, 1992). The formed representation and space can be used to quickly process the subsequent stimuli and the cognitive effort and attention are reduced consequently. In contrast, when there are multiple talkers, with the increased variation in the acoustic-phonetic structure and talker characteristics, the computational mapping becomes more complex and demanding, thus slowing down the response time and lowering the accuracy. In an fMRI study, Wong et al. (2004) found that in the mixed-talker condition, the middle/superior temporal and superior parietal regions were activated to a larger extent compared to the blocked-taker condition, suggesting that selective attention and processing of spectral and spatial (e.g., spatial location of talkers) acoustic cues are especially required in the mixed-talker condition, providing some neural evidence for the active control theory.

Concerning the time course of talker normalization, a few studies have investigated the neural manifestations of interactions between phonetic processing and talker voice processing, but the results were inconsistent. Kaganovich et al. (2006) examined the interaction between vowel and talker processing using the Garner paradigm via the comparison of two tasks: a filtering task (vowel identification with multiple talkers; talker identification with multiple vowel categories) and a baseline task (vowel identification with a constant talker; talker identification with the vowel held constant). It was found that the interaction of vowel and talker processing could be detected at early time windows, in that the N1 amplitude (starting 100 ms after auditory onset) was more negative in the filtering task than in the baseline task. The authors explained that the more negative N1 amplitude may reflect increased cognitive effort to extract information in the filtering task. In another ERPs study, Zhang et al. (2016) examined the temporal loci of the interaction of phonetic (lexical tone) and talker processing via a modified Garner paradigm. The results revealed that the interactions of phonetic and talker processing mainly origin from a posterior P3b and frontal negativity, which were interpreted as suggesting increased difficulty to categorize the lexical tones when the unattended talker dimension changed in the phonemic change detection task. However, no early interference effects were found in the N1 time-window, which differs from the findings from Kaganovich et al. (2006). While the discrepancy between the two previous studies can be largely explained by differences in the study design, more studies are needed to illuminate the time course of talker normalization.

1.2. Congenital amusia and deficits of talker normalization

Congenital amusia (amusia, hereafter) is a lifelong deficit in the processing of musical pitch in the absence of brain injury, affecting around 1.5–4% of the population (Albouy et al., 2013; Hyde et al., 2006, 2007, 2011; Hyde and Peretz, 2003; Nan et al., 2010; Peretz et al., 2002, 2008; Peretz and Vuvan, 2017; Wang et al., 2017; Wong et al., 2012). Evidence has shown that amusia is not specific to music, but influences speech pitch processing, including speech intonation and emotional prosody processing (Hutchins et al., 2010; Jiang et al., 2010; Liu et al., 2013; Patel et al., 2008; Thompson et al., 2012). Previous studies have also revealed that amusics seemed to be more impaired in musical pitch processing than speech pitch processing (Tillmann et al., 2011), and that they can imitate pitch patterns which they cannot discriminate (Hutchins and Peretz, 2012).

The deficits of amusia also affect lexical tone perception negatively. Non-tonal language speakers with amusia were found to show inferior performance in discriminating non-native tone pairs, such as in Mandarin and Thai, while also demonstrating degraded performance on the corresponding musical analogs (Liversage, 2012; Nguyen et al., 2009). For tonal language speakers with amusia, accumulating evidence has demonstrated that lexical tone perception is impaired (Liu et al., 2012, Liu et al., 2016; Shao et al., 2016). Moreover, several studies have suggested that the deficit in amusia prevails to high-level phonological processing of lexical tones, for example, compromising categorical perception of native tones in Chinese speakers with amusia (Huang et al., 2015; Jiang et al., 2012; Zhang et al., 2017b).

In addition, a recent study has revealed that amusics may also be impaired in talker normalization in lexical tone perception (Zhang et al., 2018), and more importantly, showed a potential “anchoring” deficit in perceiving multi-talker lexical tone stimuli (Shao et al., 2019). The “anchoring” deficit hypothesis pinpoints a deficit in the dynamics that link perception with perceptual memory through the implicit formation of perceptual anchors, and was originally proposed to account for the core deficit of dyslexia (Ahissar, 2007; Ahissar et al., 2006). In the conditions where a perceptual reference was provided and repeated, typical listeners could form and tune to an implicit perceptual anchor to guide the perception of incoming stimuli, so that the perception in these conditions was fast and automatic in contrast to the conditions in which the perceptual reference was absent. On the contrary, individuals with dyslexia failed to benefit from the repetitive perceptual reference and acoustic constancy, exhibiting no difference between the conditions where the perceptual anchor was present and absent. Analogous to the results of an “anchoring” deficit, Cantonese-speaking amusics failed to show significant improvement in tone discrimination in the blocked-talker condition where a constant talker was presented as a perceptual anchor over the mixed-talker condition where there was no constant talker (Shao et al., 2019). In contrast, the controls demonstrated significantly better tone discrimination in the blocked-talker condition. Moreover, while the controls exhibited significantly higher accuracy than amusics in both conditions, the effect size was larger in the blocked-talker condition. Since tone perception in the mixed-talker condition involved a greater demand of talker normalization and amusics showed degraded performance than controls in this condition, it might suggest that amusics are impaired in talker normalization during lexical tone perception. More importantly, they also demonstrated a deficit in taking advantage of talker constancy in the acoustic stimuli and less efficiently sharpened their perception of tones when perceptual anchors in talker were provided in the blocked-talker condition, suggesting a possible “anchoring” deficit in amusia. Altogether, these behavioral findings revealed several deficits in amusia. Nonetheless, due to the lack of neuroimaging studies that directly investigated talker normalization in amusics, the neural underpinnings of such behavioral deficits remain elusive.

1.3. The current study

The motivations for the current study are three-fold.

First, although a few studies have examined the time course of talker normalization, inconsistent results are reported. As such, it remains controversial whether the effects of talker normalization processes can be observed in early, sensory-driven processes (Obleser and Kotz, 2011), as reflected in the early N1 component (Kaganovich et al., 2006), or whether they more strongly affect later perceptual and categorization processes, as reflected in the P3 and frontal negativity components (Zhang et al., 2016). The discrepancy in results could be largely explained by different experiment designs used in the two studies. In Kaganovich et al. (2006), four speech stimuli (two vowels produced by two male talkers) were presented within a block at the probability of 25% each in the filtering task, whereas in the baseline task, two speech stimuli (two vowels produced by the same talker or one vowel produced by two talkers) were presented at the probability of 50% each. The unmatched probability of the stimuli in the filtering and

baseline tasks may result in differential neural habituation effects, as more frequent presentation of two stimuli in a block could habituate the neural activities more, reducing the N1 amplitude in the baseline task (cf. Budd et al., 1998; Zhang et al., 2016). In Zhang et al. (2016), four speech stimuli (two Cantonese tones produced by a female and a male talker) were presented within a block, in which each of the four speech stimuli was presented as the standard (probability = 81.25%) and the other three stimuli were presented as three types of deviants (talker change, tone change, and talker + tone change) at equal probabilities (6.25% each). This design ensured identical probability of different types of deviants and can avoid the habituation effect on the N1, which may partly explain the lack of early interference effects in the N1 time-window. Nonetheless, the demand of talker normalization was relatively low in Zhang et al. (2016), for the reason that the stimuli only involved one talker' voice change, which occurred infrequently in a small number of deviants. The low demand of talker normalization might also contribute to some extent to the lack of early interference effects in the N1 time-window.

To reexamine the time course of talker normalization in typical listeners, we compared the neural activities in the mixed- vs. blocked-talker conditions in an active oddball paradigm. We improved the experiment design relative to previous studies in three aspects. Firstly, to avoid the possible habituation effect on N1 in Kaganovich et al. (2006), the overall possibilities of standards and deviants were matched between the blocked- and mixed-talker conditions. Secondly, the demand of talker variability was increased by including four talkers' voices, in order to better observe the effects of talker normalization and avoid the possible influence of low demand of talker variability as in Zhang et al. (2016). Lastly, we included speech stimuli that were of small vs. large acoustic differences, with an aim to probe the possible different neural manifestations of talker normalization in different stimulus conditions. It is possible that talker variability may exert a greater influence on stimulus pairs with small acoustic differences. We hypothesize that if the effects of talker normalization can be detected as early as in the time-window of auditory processing, as found in Kaganovich et al. (2006), the difference between blocked-talker and mixed-talker conditions shall be observed in early ERP components such as the N1. Alternatively, if the effect of talker normalization is primarily manifested in later perceptual and categorization processes, as found in Zhang et al. (2016), the difference between blocked-talker and mixed-talker conditions shall be mainly observed in later ERP components such as the P3.

The second aim of the current study is to examine the neural bases of the "anchoring" deficit in amusia. Although some evidence shows that Cantonese-speaking amusics may be impaired in talker normalization of lexical tones and exhibit a possible "anchoring" deficit (Shao et al., 2019), the electrophysiological bases of the "anchoring" deficit remain unknown. Previous ERPs studies suggested that the deficit in the amusic brain primarily lies in the conscious detection of pitch differences, as reflected by reduced P3 amplitude, while their pre-attentive pitch processing is largely intact, as reflected by comparable mismatch negativity (MMN) responses to controls in passive listening conditions (Hyde et al., 2011; Moreau et al., 2013; Peretz et al., 2005, 2009; Zendel et al., 2015; Zhang and Shao, 2018). Since the primary neural deficits of amusia were captured by the P3, the current study aims to adopt an active oddball paradigm, focusing on the P3, to investigate the electrophysiological bases of the "anchoring" deficit in the amusic brain. According to the active control theory, the operating mechanism in the blocked-talker and mixed-talker conditions differed in terms of the amount of attention shift to phonetic changes, with greater cognitive load in the mixed-talker condition (Magnuson and Nusbaum, 2007; Nusbaum and Morin, 1992). Based on the behavioral impairment of amusics in taking advantage of acoustic constancy in the blocked-talker condition (Shao et al., 2019), there is reason to speculate that amusics may be less automatic in attentional shift to lexical tone changes in the block-talker condition than controls. As the P3a has been associated with automatic orientating of attention (Kok, 2001; Polich, 2007; Zhang et al., 2016), we hypothesize that the P3a amplitude may be particularly reduced in amusics compared to the controls in the blocked-talker condition.

Third, the previous findings suggested that amusics showed largely intact auditory processing when detecting the musical pitch and lexical tone differences, as indexed by a N1 component comparable to that of controls, and that the deficit in the amusic brain primarily lies in the conscious detection of pitch differences (Moreau et al., 2013; Peretz et al., 2005; Zendel et al., 2015; Zhang and Shao, 2018). However, it is still unclear whether such normal performance in auditory processing is due to the relatively easy task. The contrast of mixed- vs. blocked-talker conditions in the current study provided an opportunity to examine whether an auditory processing deficit could be revealed in more challenging listening conditions with greater pitch and acoustic variation. If amusics are intact in auditory processing of lexical tones, their neural activities in the early N1 time window in both blocked-talker and mixed-talker conditions are expected to be comparable to the typical controls. Alternatively, if the lack of group difference in the N1 time-window is at least partially affected by the low task difficulty in previous studies, amusics are expected to show reduced N1 in the mixed-talker condition than controls.

To sum up, examining the electrophysiological bases of the "anchoring" deficit during talker normalization in amusics is a primary aim of the current study. Nonetheless, we also focus on investigating the time course of talker normalization in typical listeners, which remains controversial due to inconsistent results reported in earlier studies. Comparing the neural underpinnings of talker normalization in typical and amusic listeners allow us to address both questions in a single design. We examined the ERP correlates of lexical tone processing in 24 Cantonese-speaking amusics and 24 matched controls. Neural activities during the processing of Cantonese tone pairs were compared in the blocked-talker (standard and deviant comprising a lexical tone stimulus from one talker) vs. mixed-talker conditions (standards and deviants comprising lexical tone stimuli from four talkers) and between tone pairs with small vs. large acoustic differences, through an active oddball paradigm. Early ERP components, such as N1, as well as late components, such as N2b/c, P3a and P3b, were analyzed.

2. Methods

2.1. Participants

24 congenital amusics and 24 musically intact controls participated in this experiment. Control participants were matched with amusic participants one by one in age, gender, and years of education. All participants were native speakers of Hong Kong Cantonese, right-handed, with no hearing impairment or brain injury, and no reported history of formal musical training. All participants were university students with self-reported normal IQs. Amusics and controls were identified using the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz et al., 2003). The MBEA consists of six subtests: three of them are pitch-based tests (scale, contour, and interval), two of them are duration-based tests (rhythm and meter), and the last one is a memory test. All amusic participants scored below 71% (Nan et al., 2010) in the global score, which is the mean of all six subtests, whereas all control participants scored higher than 80%. Demographic characteristics of the participants are summarized in Table 1. The experimental procedures were approved by the Human Subjects Ethics Sub-committee of The Hong Kong Polytechnic University. Informed written consent was obtained from participants in compliance with the experiment protocols.

2.2. Stimuli and procedure

The stimuli were three words contrasting three Cantonese tones on the syllable /ji/: high level tone (T1) – /ji55/ (醫 "a doctor"), mid level

Table 1
Demographic characteristics of the amusic and control participants.

	Amusics	Controls
No. of participants	24 (12 M, 12 F)	24 (12 M, 12 F)
Age (range)	22.45 ± 2.5 years (18.5–28.8 years)	22.58 ± 2.8 years (18.9–28.9 years)
MBEA (SD)		
Scale	52.9 (16.3)	91.2 (5.7)
Contour	60.8 (19.2)	94.5 (4.6)
Interval	54.7 (19.5)	91.5 (4.4)
Rhythm	56.9 (15.7)	94.3 (7.8)
Meter	49.7 (14.8)	74.5 (14.9)
Memory	65.3 (22.9)	97.8 (3.1)
Global	56.7 (15.1)	90.6 (3.9)

Note: Amusics and controls were identified using the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz et al., 2003). Amusics scored lower than 71% in the global score, which is the mean of all six subtests, whereas controls scored higher than 80%. M = male; F = female.

tone (T3) – /ji33/ (意 "meaning"), and low level tone (T6) – /ji22/ (二 "two") (Bauer and Benedict, 1997; Matthews and Yip, 2013). Two female and two male native Cantonese speakers (M01, M02, F01 and F02) were recorded reading aloud the words in a carrier sentence, 呢個字係 /li55 ko33 tsi22 hɐi22/ ("This word is") for six times. For each word, one clearly produced token was selected and segmented out of the carrier sentence for each talker. All selected words were normalized in duration to 350 ms, and in mean intensity to 70 dB using Praat (Boersma and Weenink, 2014). The three tones were grouped into two pairs: T1-T6 (/ji55/ – /ji22/) and T3-T6 (/ji33/ – /ji22/). The T1-T6 pair had a large pitch difference (high level vs. low level tone), whereas the T3-T6 pair had a small pitch difference (mid level vs. low level tone). We selected these two pairs with the speculation that the tone pair with smaller acoustic difference (T3-T6) would be affected more by talker variability. Fig. 1 displays the F0 contours of the three tones produced by the four talkers.

The two tone pairs were always presented in separate blocks, with T6 being the deviant and either T1 or T3 being the standard in a block. The same set of stimuli was presented in two conditions, the blocked-talker and mixed-talker condition (Magnuson and Nusbaum, 2007; Sjerps et al., 2018; Wong and Diehl, 2003). In the blocked-talker condition, the stimuli produced by the four talkers were blocked by the talker and presented in four sub-blocks. In each sub-block, the standard (T1/T3) was presented frequently at a probability of 0.85, and the deviant (T6) was presented infrequently at a probability of 0.15. A total of 119 standards and 21 deviants were binaurally presented through earphones to subjects in each sub-block. The inter-stimulus interval (offset to onset) was 800 ms. In total, there were two blocks each containing four sub-blocks (2 pairs × 4 talkers) and each sub-block lasted about 2 min. In the mixed-talker condition, each tone pair produced by the four talkers was intermixed in a single block. Again, T6 was always the deviant and either T1 or T3 was the standard in a block. Four speech stimuli corresponding to the four talkers' voices in T1/T3 were presented as the standards, while another four speech stimuli corresponding to the four talkers' voices in T6 were presented as the deviants. In each block, the standards were presented frequently at a probability of 0.85, and the deviants were presented infrequently at a probability of 0.15. A total of 476 standards (119 × 4 talkers) and 84 deviants (21 × 4 talkers) were binaurally presented through earphones to subjects in each block. The inter-stimulus interval (offset to onset) was 800 ms. In total, there were two blocks (2 pairs) and each block lasted about 8 min.

In both blocked-talker and mixed-talker conditions, the standards and deviants were presented pseudo-randomly, such that the first eight stimuli of a block were always standards and any two adjacent deviants were separated by at least two standards. The subjects were instructed to press a button on the computer keyboard when they detected a lexical tone change from the repeated standards. Note that to avoid the order effect, the presentation order of the two conditions was counterbalanced. Within each condition, the order of blocks was also counterbalanced as much as possible and kept identical between matched amusic and control subjects.

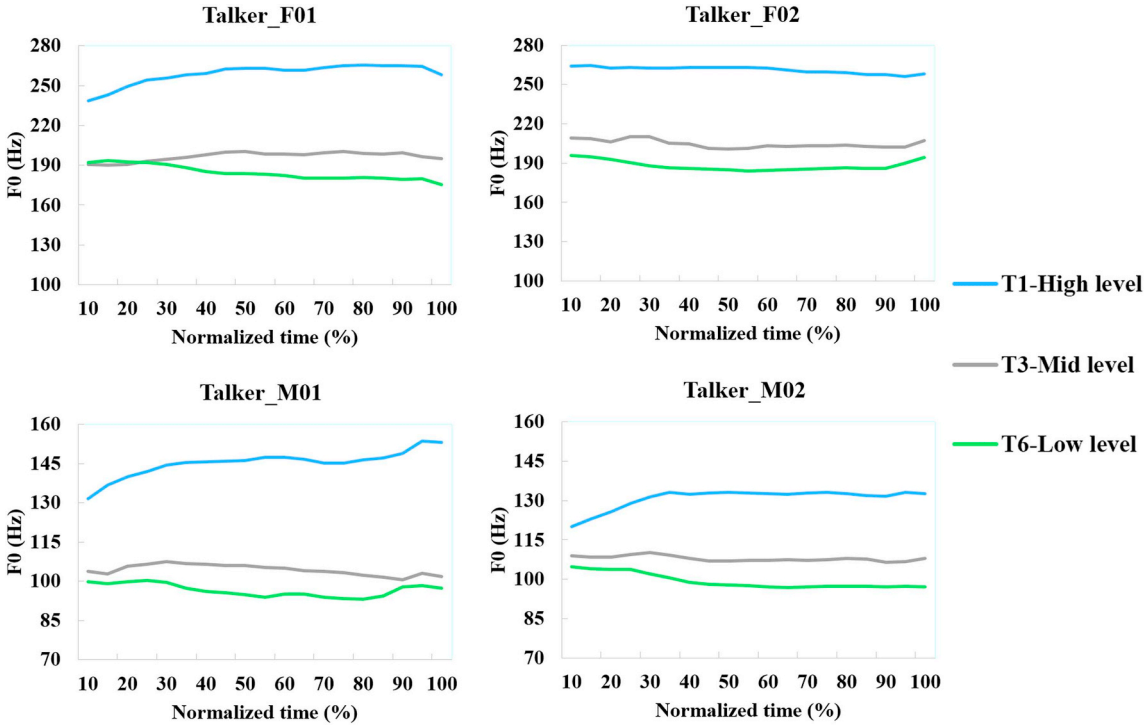

Fig. 1. F0 curve of the three Cantonese level tones (T1-T6, high level- low level tone; T3-T6, mid level-low level tone) produced by the four talkers used as stimuli in the experiment.

2.2.1. EEG data acquisition

EEG signals were recorded via a SynAmps 2 amplifier (NeuroScan, Charlotte, NC, U.S.A.) with a cap carrying 64 Ag/AgCI electrodes placed on the scalp at specific locations according to the extended international 10–20 system. Vertical electrooculography (EOG) was recorded using bipolar channels placed above and below the left eye, and horizontal EOG was recorded using bipolar channels placed lateral to the outer canthus of each eyes. Impedance between the reference electrode (located between Cz and CPz) and any recording electrode was kept below 5 kΩ. Alternating current signals (0.03–100 Hz) were continuously recorded and digitized with a 24-bit resolution at a sampling rate of 1000 Hz. Pre-processing of EEG signals was conducted using the BESA Version 7.1. The EEG recordings were re-filtered offline with a 0.01–30 Hz band-pass zero-phase shift digital filter (slope 12 dB/Oct in the low cutoff and slope 24 dB/Oct in the high cutoff).

2.2.2. EEG and behavioral data analysis

Epochs ranging from −100 to 800 ms after the onset of each deviant and the standard immediately preceding each deviant were analyzed. Epochs with amplitudes exceeding ± 75 μV at any channel were excluded from averaging. For the amusic group, the mean acceptance rate for deviants was 80.08% (N = 67.67/84, SD = 14.65%) for T1-T6 and 79.41% (N = 67.08/84, SD = 18.4%) for T3-T6 in the blocked-talker condition, and 80.90% (N = 68.5/84, SD = 14.2%) for T1-T6 and 80.04% (N = 67.45/84, SD = 12.09%) for T3-T6 in the mixed-talker condition. For the control group, the mean acceptance rate for deviants was 82.4% (N = 69.62/84, SD = 15.20%) for T1-T6 and 85.25% (N = 72/84, SD = 11.56%) for T3-T6 in the blocked-talker condition, and 82.79% (N = 69.95/84, SD = 18.93%) for T1-T6 and 85.58% (N = 72.25/84, SD = 13.48%) for T3-T6 in the mixed-talker condition. Independent-samples t-tests confirmed that the acceptance rate of the two groups was not significantly different for all conditions.

Four time-windows were determined from the global field power averaged from all deviants across all electrodes: N1 (100–150 ms), N2b/c (250–350 ms), P3a (350–500 ms) and P3b (500–800 ms). The selected time-windows largely coincided with those reported in previous P300 studies on Cantonese tone changes (Zhang et al., 2016; Zhang and Shao, 2018). Fig. 2 displays the ERP waveforms of the standards and deviants for each tone pair, each talker condition and each group at three midline electrodes, Fz, Cz and Pz.

Different sets of electrodes were selected for the analysis of the N1, N2b/c, P3a and P3b according to the topographic distributions (Fig. 3) and confirmed by the literature (Folstein and Petten, 2008; Polich, 2007; Zhang et al., 2016; Zhang and Shao, 2018). Three frontal electrodes (F3, Fz, and F4) where the N1 and N2b/c were expected to peak were selected for the N1 and N2b/c analysis, and three posterior electrodes (P3, Pz, and P4) where the P300 was expected to peak were selected for the P3a and P3b analysis. ERP waveforms were averaged across all selected electrodes for each condition. Analyses were conducted on the deviants only. The peak latency was determined from the time point with minimal (for N1, and N2b/c) or maximal deflection (for P3a and P3b) within the defined time-windows for each ERP component (Folstein and Petten, 2008; Polich, 2007; Zhang et al., 2016; Zhang and Shao, 2018) for each condition; the mean amplitude of each ERP component was obtained from the defined time-windows for each condition. Both peak latency and amplitude were analyzed, which can provide different information about the neural activities in different groups/conditions. Previous studies have primarily reported neural differences between amusics and controls in the ERP amplitude, such as reduced P3 amplitude in amusics (Moreau et al., 2013; Peretz et al., 2005, 2009; Zendel et al., 2015; Caicai Zhang and Shao, 2018), but it is possible that amusics might be impaired in the processing speed, which is better captured by the latency. *Group* × *condition* × *tone pair* repeated measures ANOVAs were conducted on the latency and amplitude of each ERP component respectively.

The behavioral accuracy and reaction time (RT) of the two groups in detecting deviants were also analyzed. The accuracy was the percentage of deviants correctly detected for each condition. RT, measured from the onset of the stimulus, was the mean reaction time of correctly detected deviants for each condition. *Group* × *condition* × *tone pair* repeated measures ANOVAs were conducted on the accuracy and RT respectively.

3. Results

3.1.1. Behavioral results

Fig. 4 displays the accuracy and RT of the two groups in detecting tonal deviants.

For the accuracy, *group* × *tone pair* × *condition* ANOVA found significant main effects of *group* ($F(1, 46) = 9.799$, $p = .003$, $\eta_p^2 = 0.176$), *tone pair* ($F(1, 46) = 70.989$, $p < .001$, $\eta_p^2 = 0.607$), *condition* ($F(1, 46) = 144.203$, $p < .001$, $\eta_p^2 = 0.758$) and a significant two-way interaction between *tone pair* and *condition* ($F(1, 46) = 73.509$, $p < .001$, $\eta_p^2 = 0.615$). Amusics exhibited overall lower accuracy than controls (M = 78.6%, SD = 23.1% vs. M = 85.2%, SD = 21.9%). Post hoc tests were conducted to analyze the *tone pair* × *condition* interaction. In the blocked-talker condition, the accuracy of T1-T6 and T3-T6 was both high and not significantly different (M = 95.4%, SD = 8% vs. M = 95.6%, SD = 9%; $t(94) = -0.095$, $p = .925$, $d = 0.023$); in the mixed-talker condition, the accuracy of T1-T6 was significantly higher than that of T3-T6 (M = 83.2%, SD = 17% vs. M = 53.2%, SD = 21%; $t(94) = 7.553$, $p < .001$, $d = 1.57$). For pair T1-T6, the accuracy elicited in the blocked-talker condition was significantly higher than that in the mixed-talker condition (M = 95.4%, SD = 8% vs. M = 83.2%, SD = 17%; $t(94) = -4.464$, $p < .001$, $d = 0.913$); for pair T3-T6, the pattern was similar, but the effect size (Rosnow and Rosenthal, 1996) was larger (M = 95.6%, SD = 9% vs. M = 53.2%, SD = 21%; $t(94) = -12.576$, $p < .001$, $d = 2.624$). These results confirmed that the accuracy of deviant tone detection was under greater influence of talker variability in the T3-T6 pair with smaller pitch differences.

For the RT, there were main effects of *group* ($F(1, 46) = 4.017$, $p = .051$, $\eta_p^2 = 0.08$), *tone pair* ($F(1, 46) = 43.025$, $p < .001$, $\eta_p^2 = 0.483$), *condition* ($F(1, 46) = 152.435$, $p < .001$, $\eta_p^2 = 0.768$), and significant two-way interaction between *tone pair* and *group* ($F(1, 46) = 5.788$, $p = .020$, $\eta_p^2 = 0.112$), and between *tone pair* and *condition* ($F(1, 46) = 6.069$, $p = .020$, $\eta_p^2 = 0.117$). The three-way interaction was also significant ($F(1, 46) = 5.402$, $p = .025$, $\eta_p^2 = 0.105$). To analyze the three-way interaction, *group* × *condition* repeated measures ANOVA was conducted within each tone pair, for the reason that the interest of investigation was whether and how amusics and controls were affected differently by talker variability (mixed- vs. blocked-talker). For pair T1-T6, there were significant main effects of *group* ($F(1, 46) = 9.402$, $p = .170$, $\eta_p^2 = 0.08$) and *condition* ($F(1, 46) = 77.873$, $p < .001$, $\eta_p^2 = 0.629$). Amusics used significantly longer RT than controls in detecting tonal changes (M = 546 ms, SD = 114 vs. M = 492 ms, SD = 90). Unsurprisingly, RT in the blocked-talker condition was significantly shorter than in the mixed-talker condition (M = 453 ms, SD = 79 vs. M = 585 ms, SD = 87). No other effects were significant. For pair T3-T6, there was a significant main effect of *condition* ($F(1, 46) = 148.552$, $p < .001$, $\eta_p^2 = 0.764$), and the two-way interaction was significant ($F(1, 46) = 8.646$, $p = .005$, $\eta_p^2 = 0.158$). Post hoc analyses were conducted to examine the two-way interaction. The controls showed significantly shorter RT in the blocked-talker condition than the mixed-talker condition (M = 465 ms, SD = 38 vs. M = 675 ms, SD = 97, $t(46) = 9.863$, $p < .001$, $d = 2.851$); amusics also demonstrated this pattern, but with a smaller effect size (M = 518 ms, SD = 79 vs. M = 646 ms, SD = 106; $t(46) = 4.824$, $p < .001$, $d = 1.394$). In the mixed-talker condition, the RT difference between amusics and controls was not significantly different (M = 675 ms, SD = 97 vs. M = 646 ms, SD = 106; $t(46) = -0.971$,

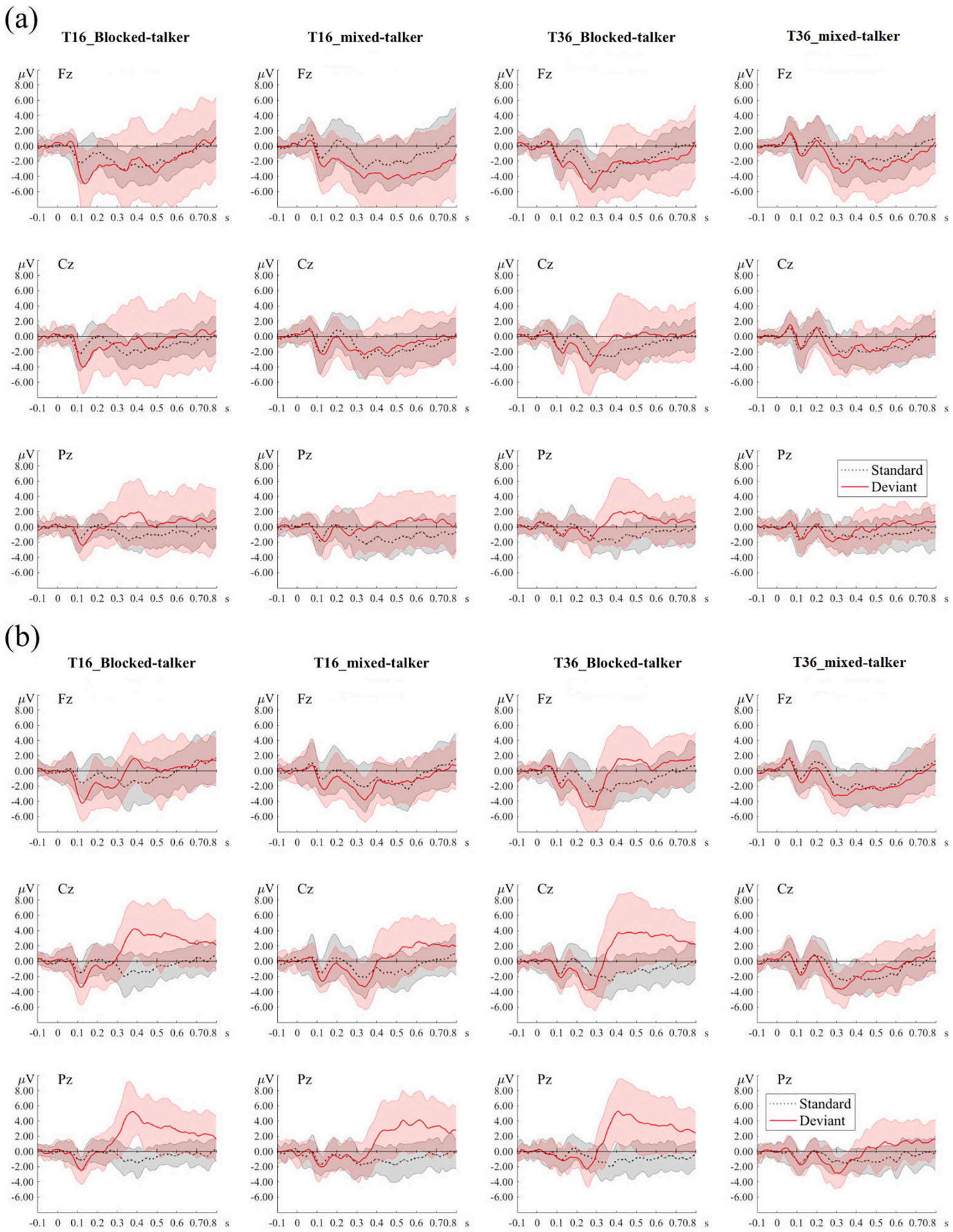

Fig. 2. Grand average ERP waveforms of the standards and deviants of the two tone-pairs (T1-T6 and T3-T6) in the two conditions (blocked- and mixed-talker). (a) The amusic group. (b) The control group.

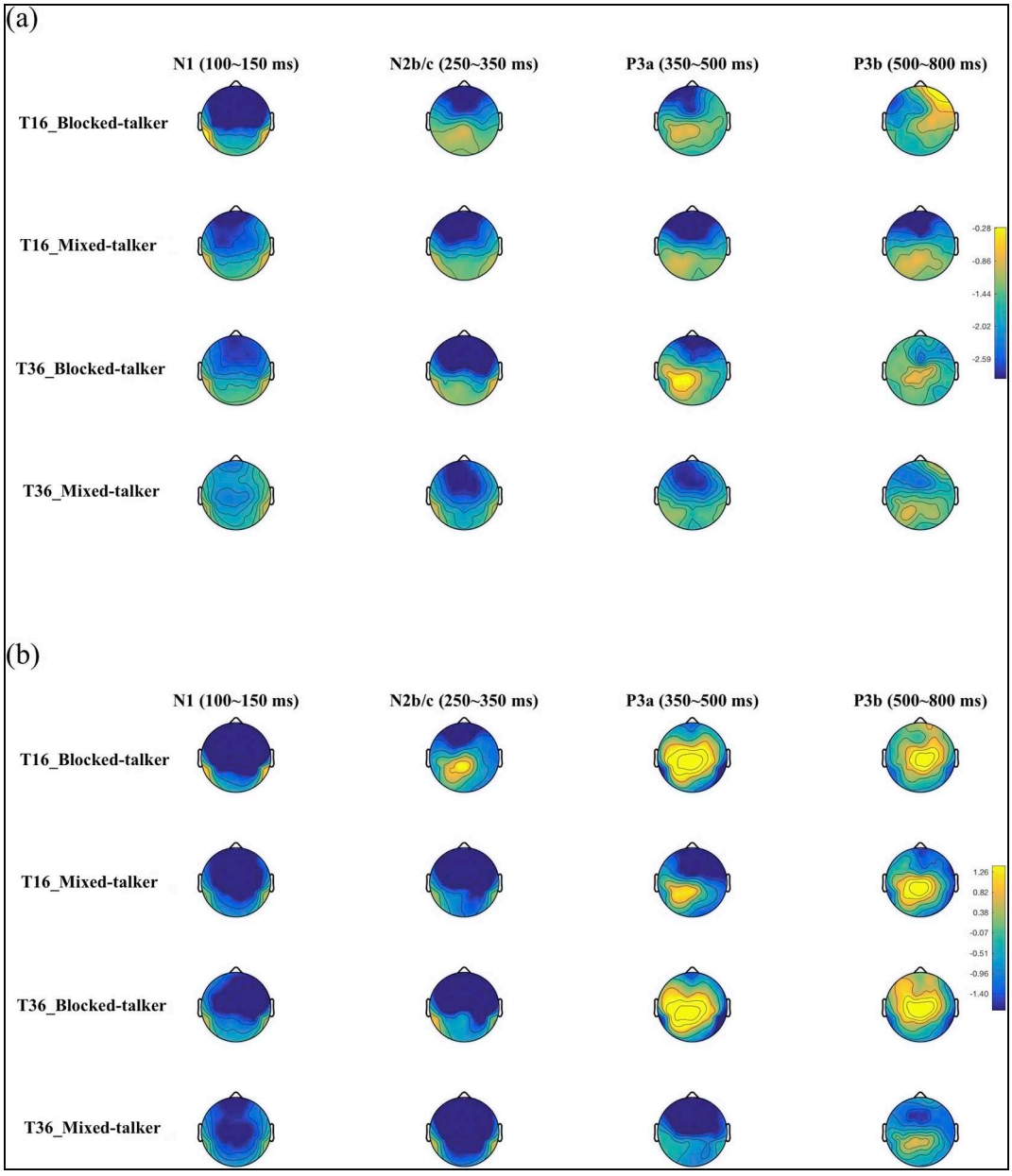

Fig. 3. Topographic distributions of the N1 (100–150 ms), N2b/c (250–350 ms), P3a (350–500 ms), and P3b (500–800 ms) of the two tone-pairs (T1-T6 and T3-T6) in the two conditions (blocked- and mixed-talker). (a) The amusic group. (b) The control group.

$p = .337$, $d = 0.285$), but in the blocked-talker condition, amusics showed significantly longer RT than controls ($M = 518$ ms, $SD = 75$vs. $M = 465$ ms, $SD = 38$; $t(46) = 3.082$, $p = .003$, $d = 0.891$).

3.1.2. ERP results

The peak latency and mean amplitude of the following four time-windows were analyzed using *group* × *condition* × *tone pair* repeated measures ANOVAs: N1 (100–150 ms), N2b/c (250–350 ms), P3a (350–500 ms) and P3b (500–800 ms). The peak latency and mean amplitude of these four ERP components for each condition are displayed in Fig. 5.

3.1.3. N1

For the N1 latency, no effects were significant.

For the N1 amplitude, there were significant main effects of *tone pair* ($F(1, 46) = 35.283$, $p < .001$, $\eta_p^2 = 0.434$) and *condition* ($F(1, 46) = 30.784$, $p < .001$, $\eta_p^2 = 0.401$). The blocked-talker condition elicited significantly larger (more negative) N1 amplitude than the mixed-talker condition ($M = -3.014\,\mu V$, $SD = 2.816$ vs. $M = -1.572\,\mu V$, $SD = 2.552$; $p < .001$). Tone pair T1-T6 elicited significantly larger N1 amplitude than T3-T6 ($M = -3.022\,\mu V$, $SD = 2.928$ vs. $M = -1.564\,\mu V$, $SD = 2.417$; $p < .001$). These results generally reflected more difficult auditory processing of tonal changes in the mixed-talker condition and in the tone pair with small pitch differences.

3.1.4. N2b/c

For the N2b/c latency, there were significant main effects of *group* ($F(1, 46) = 5.495$, $p = .023$, $\eta_p^2 = 0.107$), *condition* ($F(1, 46) = 51.915$, $p < .001$, $\eta_p^2 = 0.530$) and significant interaction between *condition* and *tone pair* ($F(1, 46) = 10.681$, $p = .002$, $\eta_p^2 = 0.188$). The N2b/c peaked significantly later in the amusic group than in the control group ($M = 0.309$ ms, $SD = 0.036$ vs. $M = 0.296$ ms, $SD = 0.034$). Post hoc pairwise comparisons were conducted to examine the interaction between *condition* and *tone pair*. For pair T3-T6, the N2b/c in the blocked-talker condition peaked significantly earlier than in the mixed-talker

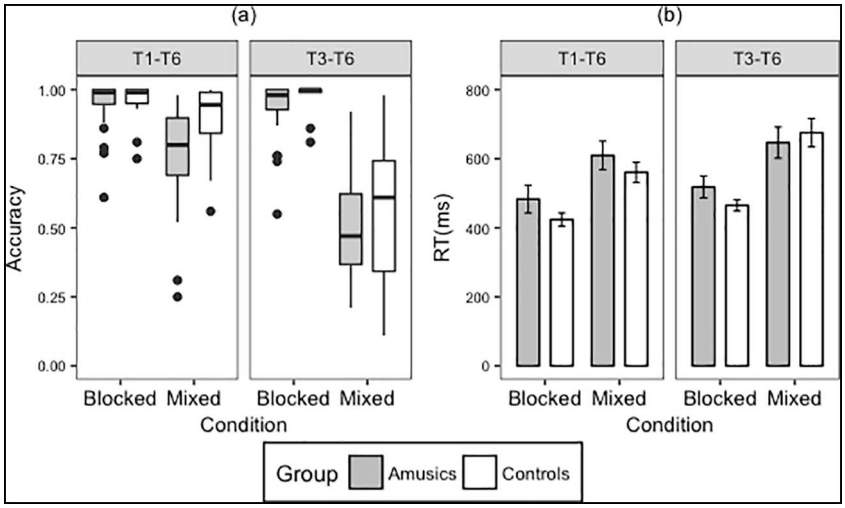

Fig. 4. Behavioral performance. (a) Accuracy of deviant detection. (B) Reaction time of accurately detected deviants.

condition (M = 0.278 ms, SD = 0.030 vs. M = 0.319 ms, SD = 0.026; t(94) = −7.164, p < .001, d = 1.461); for pair T1-T6, the results were similar but with a smaller effect size (M = 0.297 ms, SD = 0.039 vs. M = 0.314 ms, SD = 0.031; t(94) = −2.152, p = .034, d = 0.482). In the blocked-talker condition, pair T1-T6 peaked significantly earlier than T3-T6 (M = 0.297 ms, SD = 0.039 vs. M = 0.278 ms, SD = 0.03; t(94) = 2.759, p = .007, d = 0.546), but in the mixed-talker condition, the effect was not significant (M = 0.314 ms, SD = 0.031 vs. M = 0.319 ms, SD = 0.026; t(94) = −1.107, p = .312, d = 0.174).

In terms of the amplitude, there was a significant main effect of *tone pair* (F(1, 46) = 8.124, p = .007, η_p^2 = 0.150), and a significant interaction between *tone pair* and *condition* (F(1, 46) = 14.792, p < .001, η_p^2 = 0.243). The T3-T6 pair elicited significantly larger (more negative) N2b/c amplitude than the T1-T6 pair in the blocked-talker condition (M = −3.936 μV, SD = 3.328, vs. M = −1.992 μV, SD = 3.156; t(94) = 2.934, p = .004, d = 0.599), but not in the mixed-talker condition (M = −2.798 μV, SD = 2.572, vs. M = −2.956 μV, SD = 2.819; t(94) = −0.286, p = .775, d = 0.058). No other effects were significant.

3.1.5. P3a

For the P3a latency, there were significant main effects of *tone pair* (F(1, 46) = 12.632, p = .001, η_p^2 = 0.215) and *condition* (F(1, 46) = 9.662, p = .003, η_p^2 = 0.174). The T3-T6 pair peaked significantly later than the T1-T6 pair (M = 0.437 ms, SD = 0.043 vs. M = 0.418 ms, SD = 0.048; p = .001). The mixed-talker condition elicited a significantly later peaking P3a than the blocked-talker condition (M = 0.438 ms, SD = 0.047 vs. M = 0.417 ms, SD = 0.044; p = .003). No other effects were significant.

For the P3a amplitude, there were significant main effects of *group* (F(1, 46) = 8.060, p = .007) and *condition* (F(1, 46) = 42.217, p < .001), and significant interactions between *group* and *condition* (F(1, 46) = 6.329, p = .015) and between *condition* and *tone pairs* (F(1, 46) = 6.448, p = .015). Post hoc tests were first conducted to examine

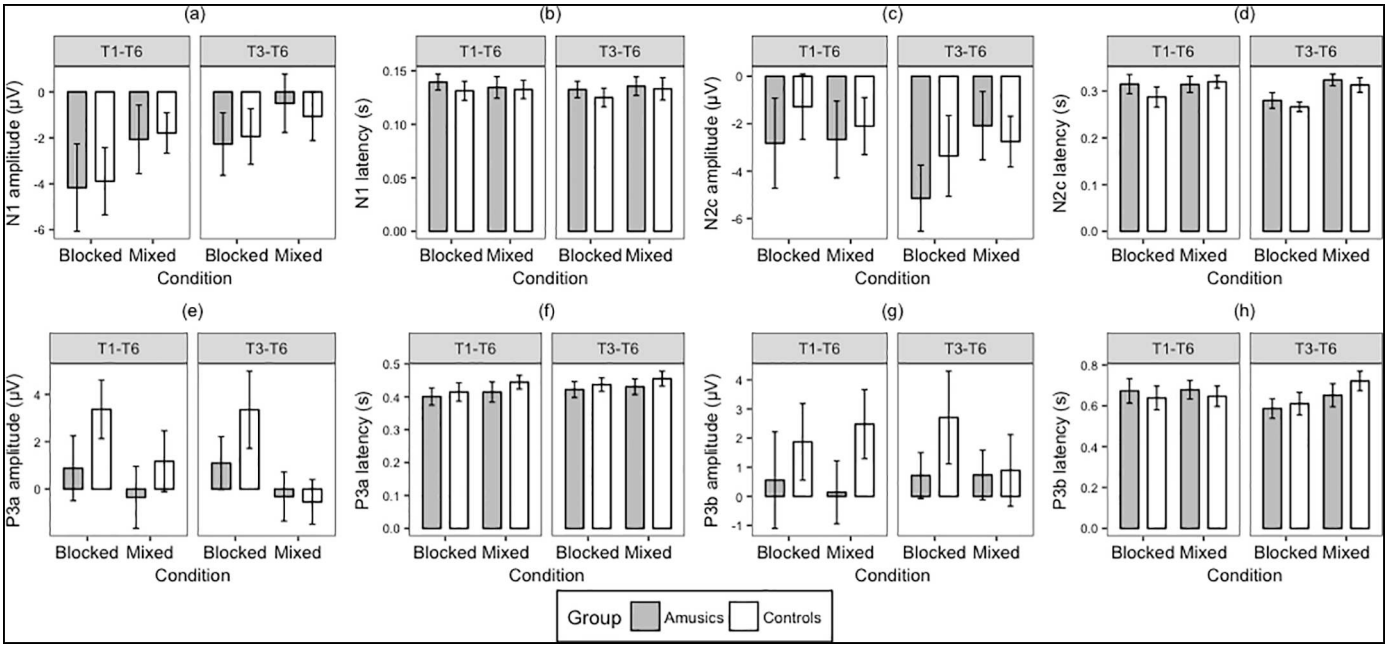

Fig. 5. ERP results. (a) N1 (100–150 ms) mean amplitude. (b) N1 (100–150 ms) peak latency. (c) N2b/c (250–350 ms) mean amplitude. (d) N2b/c (250–350 ms) peak latency. (e) P3a (350–500 ms) mean amplitude. (f) P3a (350–500 ms) peak latency. (g) P3b (500–800 ms) mean amplitude. (h) P3b (500–800 ms) peak latency.

the interaction between *group* and *condition*. In the control group, the P3a amplitude elicited in the blocked-talker condition was significantly larger than in the mixed-talker condition ($M = 3.669\,\mu V$, $SD = 2.838$ vs. $M = 0.574\,\mu V$, $SD = 2.422$; $t(94) = 5.747$, $p < .001$, $d = 1.173$). The amusic group demonstrated a similar pattern, but the effect size was much smaller ($M = 1.153\,\mu V$, $SD = 3.269$ vs. $M = -0.222\,\mu V$, $SD = 2.317$; $t(94) = 2.378$, $p = .019$, $d = 0.481$). In the blocked-talker condition, amusics demonstrated significantly smaller P3a amplitude than the control group ($M = 1.153\,\mu V$, $SD = 3.269$ vs. $M = 3.669\,\mu V$, $SD = 2.838$; $t(94) = -4.025$, $p < .001$, $d = 0.829$); in the mixed-talker condition, the group difference was not significant ($M = -0.222\,\mu V$, $SD = 2.317$ vs. $M = 0.574\,\mu V$, $SD = 2.422$; $t(94) = -1.645$, $p = .103$, $d = 0.335$). Post hoc tests were then conducted to explore the *tone pair* by *condition* interaction. In the blocked-talker condition, the P3a amplitude between tone pair T1-T6 and T3-T6 was not significantly different ($M = 2.232\,\mu V$, $SD = 3.264$ vs. $M = 2.589\,\mu V$, $SD = 3.355$; $t(94) = -0.528$, $p = .599$, $d = 0.107$), whereas in the mixed-talker condition, T1-T6 elicited significantly larger P3a amplitude than T3-T6 ($M = 0.799\,\mu V$, $SD = 2.788$ vs. $M = -0.449\,\mu V$, $SD = 1.729$; $t(94) = 2.673$, $p = .010$, $d = 0.538$). For tone pair T1-T6, the P3a amplitude in the blocked-talker condition was significantly larger than in the mixed-talker condition ($M = 2.232\,\mu V$, $SD = 3.264$ vs. $M = 0.799\,\mu V$, $SD = 2.788$; $t(94) = 2.312$, $p = .023$, $d = 0.472$); for tone pair T3-T6, the pattern was similar, but the effect size was larger ($M = 2.589\,\mu V$, $SD = 3.355$ vs. $M = -0.449\,\mu V$, $SD = 1.729$; $t(94) = 5.577$, $p < .001$, $d = 1.138$).

3.1.6. P3b

For the P3b latency, there was a significant main effect of *condition* ($F(1, 46) = 14.712$, $p < .001$, $\eta_p^2 = 0.242$), and a significant interaction between *tone pair* and *condition* ($F(1, 46) = 11.578$, $p = .001$, $\eta_p^2 = 0.201$). Post hoc analyses were conducted to examine the two-way interaction. For tone pair T1-T6, the peak latency in the mixed-talker and blocked-talker conditions was not significantly different ($M = 0.651$ ms, $SD = 0.097$ vs. $M = 0.649$ ms, $SD = 0.103$; $t(94) = -0.750$, $p = .940$, $d = 0.019$), whereas tone pair T3-T6 peaked significantly later in the mixed-talker condition than in the blocked-talker condition ($M = 0.675$ ms, $SD = 0.101$ vs. $M = 0.591$ ms, $SD = 0.094$; $t(94) = -4.237$, $p < .001$, $d = 0.861$). In the blocked-talker condition, T3-T6 peaked significantly earlier than T1-T6 ($M = 0.591$ ms, $SD = 0.094$ vs. $M = 0.649$ ms, $SD = 0.103$, $t(94) = 2.876$, $p = .005$, $d = 0.588$), whereas in the mixed-talker condition, the latency difference between T1-T6 and T3-T6 was not significant ($M = 0.651$ ms, $SD = 0.097$ vs. $M = 0.675$ ms, $SD = 0.101$; $t(94) = 2.876$, $p = .221$, $d = 0.242$).

For the P3b amplitude, there was a significant main effect of *group* ($F(1, 46) = 10.250$, $p = .002$, $\eta_p^2 = 0.182$), and a significant two-way interaction between *tone pair* and *condition* ($F(1, 46) = 4.352$, $p = .046$, $\eta_p^2 = 0.084$). The amusic group exhibited overall reduced P3b amplitude than the control group ($M = 0.583\,\mu V$, $SD = 3.259$. vs. $M = 2.193\,\mu V$, $SD = 2.489$). Post hoc analyses were conducted to examine the interaction between *tone pair* and *condition*. For tone pair T1-T6, the P3b amplitude in the blocked-talker and mixed-talker condition was not significantly different ($M = 1.303\,\mu V$, $SD = 3.145$ vs. $M = 1.660\,\mu V$, $SD = 2.823$; $t(94) = -0.586$, $p = .559$, $d = 0.119$), but for the tone pair T3-T6, the P3b amplitude in the block-talker condition was significantly larger than that in the mixed-talker condition ($M = 1.714\,\mu V$, $SD = 2.567$ vs. $M = 0.768\,\mu V$, $SD = 1.945$; $t(94) = 2.036$, $p = .045$, $d = 0.415$).

To summarize, behaviorally, amusics exhibited overall lower accuracy than controls in detecting tone changes. Amusics responded to tone changes significantly more slowly than controls in the blocked-talker condition, but the group difference disappeared in the most challenging condition, that is, tone pair T3-T6 presented in the mixed-talker condition. For early ERP components, among typical listeners, the N1 amplitude was significantly larger in the blocked-talker condition than in the mixed-talker condition. For later ERP components, N2b/c, P3a and P3b mainly yielded an interaction effect between *condition* and *tone pair*, and the general pattern was that the effect of talker variability was more prominent on the challenging tone pair (T3-T6). For amusics, their neural impairment was mainly manifested on later components, such as delayed N2b/c peak latency and reduced P3a and P3b amplitude, but not on the early component N1. The P3a amplitude mainly revealed a *group* by *condition* interaction, where the group difference was only significant in the blocked-talker condition, whereas for the P3b amplitude, there was a main effect of group, irrespective of tone pairs and conditions.

4. Discussion

The current study examined the time course of talker normalization in typical listeners and amusics through the comparison of two conditions differing in the amount of talker variability via an active oddball paradigm. The aims were to determine the time course of talker normalization in typical listeners and to reveal the neural correlates of the impairment in talker normalization and "anchoring" deficit in the amusic brain. We analyzed both early and later ERP components and compared ERP components in the blocked-talker and mixed-talker conditions to identify the effects of talker variability on speech perception processes.

4.1. Time course of talker normalization in typical listeners

The behavioral results confirmed that talker variability in the mixed-talker condition increased the difficulty of speech recognition. For both tone pairs, the accuracy was significantly lower in the mixed-talker condition than in the blocked-talker condition, but the effect of talker variability was greater on T3-T6 with small acoustic differences. Likewise, RT was significantly longer in the mixed-talker condition, with a larger effect of talker variability on the T3-T6 pair than on the T1-T6 pair, especially for the control group (see Fig. 4; also confirmed by post hoc analyses of the three-way interaction). These results largely echoed with previous findings (Lee, 2009; Strange et al., 1976; Wong and Diehl, 2003) that at the behavioral level, presenting the same speech stimuli to listeners in mixed- vs. block-talker manner may invoke the active control mechanism, which mediates talker normalization. Talker constancy in the blocked-talker condition allows the listeners to learn the talker's vocal characteristics and form a perceptual "anchor", such that the computation between acoustic signals and phonological representations becomes less effortful and more automatic. However, in the mixed-talker condition, the possible many-to-many mapping between acoustic signals and phonological representations may require more attention and the task is more demanding, especially for the tone pair with small acoustic differences, thus lowering the accuracy and lengthening the RT.

The ERP results revealed the time course of talker normalization in typical listeners. We found that early components such as the N1 already exhibited an effect of talker variability. The N1 amplitude was significantly larger in the blocked-talker condition than in the mixed-talker condition. These findings are different from Zhang et al. (2016), where no difference in the N1 was detected among conditions of interest. This discrepancy could be explained by the low demand of talker normalization in Zhang et al. (2016), where the design only involved infrequent change of one talker' voice in a small number of deviants. It is possible that when the demand of talker normalization was high, the effect could emerge in the early N1 time-window. Moreover, these results also somehow differed from the finding of Kaganovich et al. (2006) that the N1 amplitude was more negative in the filtering task than the baseline task. The authors argued that the more negative N1 amplitude might reflect sustained attentional processes operating during the filtering task. Indeed, N1 is considered to be an index of attentional early auditory processing (Naatanen et al., 1978; R

Näätänen and Alho, 2004; Risto Näätänen, 1982), and tends to be influenced by the stimulus predictability and tasks (Lange, 2013). The results of the current study showed the opposite pattern, in that the N1 amplitude was larger (more negative) in the blocked-talker condition (similar to the baseline task of Kaganovich et al. (2006)). First, the possibility cannot be ruled out that the more negative N1 amplitude observed in Kaganovich et al. (2006) was influenced by the differential neural habituation effect, since the probability of stimuli presented in the filtering task was lower than the baseline task. Second, Kaganovich et al. (2006) used a classification task, requiring the participants to identify the vowel or the talker, which was an attention demanding task. In contrast, in the current study participants were asked to detect infrequent tonal changes, which was similar to a discrimination task and relatively less attention demanding. It is possible that the task demand may lead to an earlier influence of attentional processes in Kaganovich et al. (2006). As a matter of fact, the N1 results of the current study are better explained by the difficulty of auditory processing, given the observation that the N1 amplitude was also significantly larger in the pair T1-T6, where auditory processing was easier due to the larger pitch difference than the pair T3-T6. This interpretation is also consistent with a previous study (Zhang and Shao, 2018) that examined lexical tone change detection in an oddball paradigm, which revealed that the N1 amplitude was reduced on tone pairs where the acoustic difference was more difficult to detect. It could be because of the combined effects of task demand and controlled stimulus probability in the current study that the influence of attentional processes was more notably exerted on later perceptual processes, such as the N2b/c and P300.

We found that the later components, such as the N2b/c, P3a and P3b, were also modulated by talker variability. But different from the N1, which showed a main effect of talker condition, the later ERP components exhibited an interaction between *condition* and *tone pair*. The N2b/c peaked significantly earlier in the blocked-talker condition than in the mixed-talker condition, more so for T3-T6, revealing a greater influence of talker variability on the T3-T6 pair. An interaction between *condition* and *tone pair* was also observed on the N2b/c amplitude, such that the more challenging tone pair (T3-T6) yielded larger N2b/c amplitude than the more acoustically distinct tone pair (T1-T6) in the blocked-talker condition, but not in the mixed-talker condition. It has been suggested that the N2 component often reflects attention, stimulus novelty or cognitive control (Folstein and Petten, 2008). A possible explanation for the N2b/c results is that more attentional resources or control effort were directed to detecting tonal changes in the T3-T6 pair with small pitch differences compared to the T1-T6 pair in the blocked-talker condition, eliciting enlarged N2b/c amplitude and later N2b/c latency. In contrast, in the mixed-talker condition, the task was more demanding, which may have compressed and leveled off the difference between the two tone-pairs.

It has been reported that the P300 amplitude is modulated by the overall arousal level in the experiment: when task conditions are less demanding, the P300 amplitude is relatively larger (Kok, 2001). The P300 is believed to include two subcomponents: an earlier fronto-central P3a and a later parietal P3b. The P3a has been associated with novelty detection and automatic orientating of attention (Polich, 2007; Zhang et al., 2016). The results of the current study showed that for both tone pairs, the P3a amplitude in the blocked-talker condition was larger than in the mixed-talker condition, but the effect of talker variability was more robust on T3-T6; the P3a also peaked significantly later in the mixed-talker condition than in the blocked-talker condition. Given that the P3a is an index of involuntary attentional switch (Kok, 2001; Polich, 2007; Zhang et al., 2016), these findings revealed that at the neural level, involuntary attentional switch to lexical tone changes was easier or more automatic in the blocked-talker condition than in the mixed-talker condition, especially for the T3-T6 pair with small pitch differences. These findings are consistent with Kaganovich et al. (2006), which found that the peak amplitude of the subsequent positive component, P3 (248–500 ms), was significantly reduced in the filtering task (similar to the mixed-talker condition of the current study) compared with the baseline task (similar to the blocked-talker condition of the current study).

On the other hand, the P3b is likely to be associated with stimulus categorization. It has been found that stimuli easier to categorize elicit larger P3b amplitude (McCarthy and Donchin, 1981; Mecklinger and Ullsperger, 1993; Polich, 2007; Zhang et al., 2016). Our results showed that for T3-T6, the P3b amplitude in the blocked-talker condition was larger than in the mixed-talker condition, suggesting that for the tone pair that was of smaller pitch differences, talker variability significantly interfered with speech categorization, increasing the difficulty of speech categorization in the mixed-talker condition. In contrast, for tone pair T1-T6 with larger pitch differences and thus being less susceptible to the influence of talker variability, stimulus categorization was relatively easy and the P3b amplitude did not differ significantly between the blocked- and mixed-talker conditions. Although obtained with a different experiment paradigm, these results are in line with the findings of Zhang et al. (2016), which showed interactions between lexical tone and talker processing on the P3b (500–800 ms after the stimulus onset).

Taken together, the ERP findings revealed the time course of talker normalization in lexical tone perception among typical listeners. We found that the influence of talker variability on tone perception was characterized by ERP components at both early and late time-windows, but with different effects revealed. The N1 amplitude was smaller in the mixed-talker condition than the blocked-talker condition and smaller in the T3-T6 pair than in the T1-T6 pair, which may indicate reduced auditory processing in the perceptually more difficult conditions. The later ERP components such as N2b/c, P3a and P3b yielded an interaction of *condition* and *tone pair*, which may suggest that at a higher perceptual level, acoustic variability might be tolerable within a certain range (e.g., for the T1-T6 pair presented in mixed- and blocked-talker conditions), and the influence of talker variability was generally more prominent on the T3-T6 pair. These findings provided neural evidence for the active control theory (Magnuson and Nusbaum, 2007; Nusbaum and Morin, 1992). They are also consistent with Wong et al. (2004), which reported neural evidence consistent with the hypothesis that talker changes engage controlled attentional processing during speech perception. The resource-demanding condition (mixed-talker condition) resulted in increased activity not only in cortical areas associated with speech processing (e.g., the posterior superior temporal gyrus), but also in cortical areas possibly sub-serving shift of attention to spectral information (e.g., the superior parietal cortex).

4.2. Neural correlates of the deficits in talker normalization and perceptual anchoring in amusia

Compared with typical listeners, behaviorally, amusics obtained overall lower accuracy of tone change detection than controls. Given that the controls' accuracy was approaching ceiling, RT results may be more meaningful. The RT results revealed a *group* by *condition* interaction for T3-T6. The controls showed significantly shorter RT in the blocked-talker condition than in the mixed-talker condition, whereas amusics demonstrated a similar pattern but with a smaller effect size; moreover, the RT difference between amusics and controls was not significant in the mixed-talker condition, but amusics showed significantly longer RT than controls in the blocked-talker condition. The RT results suggested that amusics exhibited a reduced gain of RT when the talker was fixed in the blocked-talker condition for the challenging tone pair T3-T6, which is partially consistent with the previous finding of an “anchoring” deficit that amusics are impaired in taking advantage of acoustic constancy in the blocked-talker condition (Shao et al., 2019).

With regard to the ERP results, for the early component N1, we hypothesized that if amusics are not impaired in the auditory

processing of lexical tones, their neural responses related to early auditory processing (N1) in both blocked- and mixed-talker conditions should be comparable to typical controls, irrespective of task difficulty. Alternatively, if amusics show reduced N1 in the mixed-talker condition than controls, it implies that the lack of group difference in N1 reported in previous studies (Peretz et al., 2005) might be at least partially affected by the low task difficulty. The results revealed that amusics did not show significantly different N1 amplitude from typical listeners in either blocked-talker or mixed-talker condition. As a matter of fact, amusics demonstrated larger N1 amplitude in the blocked-talker condition than in the mixed-talker condition, a pattern similar to typical listeners. These results confirmed the hypothesis that amusics are relatively intact in early auditory processing of musical and speech pitch stimuli (Hyde et al., 2011; Moreau et al., 2013; Peretz et al., 2005; Zendel et al., 2015; Zhang and Shao, 2018).

As for the N2b/c, although the group difference on the N2b/c amplitude was not significant, the amusic group did show an overall later N2b/c latency than controls. Given that the N2 component often reflects attention, stimulus novelty or cognitive control (Folstein and Petten, 2008), this result might suggest that the amusic brain is slower in detecting the stimulus novelty or slower in directing attention or cognitive control towards the stimulus change. Note that no group difference on the N2b/c was reported in Zhang and Shao (2018), which also compared amusics and controls on lexical tone change detection in an active oddball paradigm but did not manipulate talker variability. This difference may be due to the higher task demand in the current study, where greater talker variability was involved.

For the P3 components, such as P3a, we hypothesized that if amusics are less automatic in tone processing in the blocked-talker condition, the P3a amplitude, an index of involuntary orientation of attention, would be reduced compared to the controls in the blocked-talker condition, but less so in the mixed-talker condition. The results confirmed this hypothesis, revealing a *group* by *condition* interaction on the P3a amplitude. Note that the P3a is the only component that showed a *group* by *condition* interaction, while other components such as N2b/c and P3b revealed a main *group* effect. The controls showed larger P3a amplitude in the blocked-talker condition than in the mixed-talker condition, whereas the difference was reduced in the amusic group. Moreover, the group difference was only significant in the blocked-talker condition where the talker voice was constant and the acoustic variation was limited. The reduced P3a amplitude, which is associated with less automatic attention orientation, can account for the possible "anchoring" deficit manifested behaviorally in amusia (Shao et al., 2019). It showed that amusics were particularly impaired in the blocked-talker condition where a constant talker was presented, suggesting a possible "anchoring" deficit similar to dyslexics (Shao et al., 2019). With this deficit, the perceptual system of amusics may be less resilient to external variation, and the processing of speech stimuli requires more attentional effort even in the blocked condition (Zhang et al., 2017a). We argue that the reduced P3a amplitude observed in the blocked-talker condition in the amusic brain is an ERP signature of a deficiency in the allocation of attentional effort, which presumably underlies the "anchoring" deficit observed in amusics.

The less automatic attention shift as reflected by P3a in the amusic brain is also consistent with the poor attention skills reported in amusics (Jones et al., 2009). It has been found that about 40% of the amusic individuals suffer from an attention deficit. With the attention deficit, amusics may fail to benefit from the condition that is free from talker variation. In fact, Ahissar (2007) has suggested that the "anchoring" deficit in dyslexia can be viewed as a type of attention deficit. Amusics may also be under the influence of such attention deficits and therefore showed less automatic orientation of attention. Altogether, it is likely that the P3a is a potential neural landmark of the "anchoring" deficit and attention deficit previously observed in amusia.

As for the P3b, there was an overall group difference on the amplitude, suggesting that the ability to consciously categorize lexical tones may be impaired in the amusic brain, in both blocked- and mixed-talker conditions. These results are consistent with previous behavioral findings that Chinese amusics were impaired in the categorization and categorical perception of tones (Huang et al., 2015; Jiang et al., 2012; Shao et al., 2016; Zhang et al., 2017b). This finding is also consistent with a previous ERP study that reported reduced P3b amplitude when amusics detected lexical tone changes with small pitch differences (Zhang and Shao, 2018).

Altogether, the ERP results revealed that amusics exhibited comparable N1 activities to controls, but deviated from controls in later components, showing delayed N2b/c latency, reduced P3a amplitude in the blocked-talker condition, and overall reduced P3b amplitude. These results increased the understanding of the deficiency mechanism of amusics in pitch processing, by revealing that even in the more complex and demanding perceptual task, early auditory processing as indicated by the N1 amplitude appears to be intact in the amusic brain. On the other hand, the P3a amplitude revealed a *group* and *condition* interaction in directions consistent with the prediction of the "anchoring" deficit, indicating that the P3a is a potential neural signature of the "anchoring" deficit in amusia. These findings are largely consistent with the hypothesis that amusics are relatively intact in early auditory processing, and are primarily impaired in later, conscious perceptual evaluation or categorization of pitch stimuli (Hyde et al., 2011; Moreau et al., 2013; Peretz et al., 2005; Zendel et al., 2015; Zhang and Shao, 2018).

Lastly, combining the ERP results from typical and amusic listeners, there appears to be some connection between the active control mechanism and the "anchoring" deficit hypothesis. Among typical listeners, distinct patterns between early (main effect of talker condition) and late processes (interaction of talker and tone pair) of talker normalization were observed. We argued that the early component N1 primarily reflected difficulty of auditory processing, whereas the influence of attentional processes was more notably exerted on later perceptual processes, such as the N2b/c and P300. Among these later components that presumably reflect the active control mechanism, atypical neural activities in latency or amplitude were observed in the amusical brain. Most importantly, the P3a, an index of involuntary orientation of attention, is likely to be a neural signature of the "anchoring" deficit in amusia. These results imply that the perceptual anchoring ability is probably related to the active control mechanism, both of which are deficient in the amusical brain. Since both hypotheses essentially emphasize the critical role of attention and/or cognitive control in accommodating acoustic variation in perception, it is perhaps not surprising that the two hypotheses are related to some extent, like two sides of the same coin. Future studies can further illuminate the relationship between the active control mechanism and the "anchoring" hypothesis.

5. Conclusion

The current study revealed the time course of talker normalization in typical listeners and the neurophysiological bases of the talker normalization and "anchoring" deficits in amusia. For typical listeners, early components such as the N1 already exhibited a main effect of talker variability, while the later ERP components such as N2b/c, P3a and P3b mainly exhibited greater influence of talker variability on the T3-T6 pair with small pitch differences. Amusics demonstrated comparable N1 amplitude to controls, suggesting intact auditory processing even in the mixed-talker condition, and atypical neural activities in later components. Importantly, the P3a amplitude was especially reduced in the blocked-talker condition in amusics compared to controls, and we argued that it is presumably a neural signature of the "anchoring" deficit previously observed in amusics. Altogether, these findings shed some light on the time course of talker normalization and the neural underpinnings of perceptual impairment in amusia. The current study also has some limitations, in that explicit tests of attention

and working memory can be conducted as control tasks in the identification of amusics for a more precise characterization of amusics, and to better understand the potential effects of attention and memory deficits in amusia on talker normalization.

Acknowledgements

This work was supported by grants from the Research Grants Council of Hong Kong (ECS: 25603916), the National Natural Science Foundation of China (NSFC: 11504400), and the PolyU Start-up Fund for New Recruits. We thank Ms. Rebecca Yick Man Lau and Ms. Phyllis Oi Ching Tang for help with data collection. The first author is also affiliated with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences.

References

Ahissar, M., 2007. Dyslexia and the anchoring-deficit hypothesis. Trends Cogn. Sci. 11 (11), 458–465. https://doi.org/10.1016/j.tics.2007.08.015.

Ahissar, M., Lubin, Y., Putter-Katz, H., Banai, K., 2006. Dyslexia and the failure to form a perceptual anchor. Nat. Neurosci. 9 (12), 1558–1564. https://doi.org/10.1038/nn1800.

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.-E., ... Tillmann, B., 2013. Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. Brain 136 (5), 1639–1661. https://doi.org/10.1093/brain/awt082.

Bauer, R.S., Benedict, P.K., 1997. Modern Cantonese Phonology. Mouton de Gruyter, Berlin Retrieved from. https://books.google.com/books?id=QWNj5Yj6_CgC&pgis=1.

Boersma, P., Weenink, D., 2014. Praat: Doing Phonetics by Computer.

Budd, T.W., Barry, R.J., Gordon, E., Rennie, C., Michie, P.T., 1998. Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. Int. J. Psychophysiol. 31 (1), 51–68. https://doi.org/10.1016/S0167-8760(98)00040-3.

Folstein, J.R., Petten, C.V.A.N., 2008. Influence of Cognitive Control and Mismatch on the N2 Component of the ERP : A Review. vol. 45. pp. 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x.

Green, K.P., Tomiak, G.R., Kuhl, P.K., 1997. The encoding of rate and talker information during phonetic perception. Percept. Psychophys. 59 (5), 675–692. https://doi.org/10.3758/BF03206015.

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

Hutchins, S., Peretz, I., 2012. Amusics can imitate what they cannot discriminate. Brain Lang. 123 (3), 234–239. https://doi.org/10.1016/j.bandl.2012.09.011.

Hutchins, S., Gosselin, N., Peretz, I., 2010. Identification of changes along a continuum of speech intonation is impaired in congenital amusia. Front. Psychol. 1 (DEC), 1–8. https://doi.org/10.3389/fpsyg.2010.00236.

Hyde, K.L., Peretz, I., 2003. “Out-of-pitch” but still “in-time.”. Ann. N. Y. Acad. Sci. 999 (1), 173–176. https://doi.org/10.1196/annals.1284.023.

Hyde, K.L., Zatorre, R.J., Griffiths, T.D., Lerch, J.P., Peretz, I., 2006. Morphometry of the amusic brain: a two-site study. Brain 129 (10), 2562–2570. https://doi.org/10.1093/brain/awl204.

Hyde, K.L., Lerch, J.P., Zatorre, R.J., Griffiths, T.D., Evans, A.C., Peretz, I., 2007. Cortical thickness in congenital amusia: when less is better than more. J. Neurosci. 27 (47), 13028–13032. https://doi.org/10.1523/JNEUROSCI.3039-07.2007.

Hyde, K.L., Zatorre, R.J., Peretz, I., 2011. Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. Cereb. Cortex 21, 292–299.

Jiang, C., Hamm, J.P., Lim, V.K., Kirk, I.J., Yang, Y., 2010. Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. Neuropsychologia 48 (9), 2630–2639. https://doi.org/10.1016/j.neuropsychologia.2010.05.009.

Jiang, C., Hamm, J.P., Lim, V.K., Kirk, I.J., Yang, Y., 2012. Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. Mem. Cogn. 40 (7), 1109–1121. https://doi.org/10.3758/s13421-012-0208-2.

Jones, J.L., Zalewski, C., Brewer, C., Lucker, J., Drayna, D., 2009. Widespread auditory deficits in tune deafness. Ear Hear. 30 (1), 63–72. https://doi.org/10.1097/AUD.0b013e31818ff95e.

Kaganovich, N., Francis, A.L., Melara, R.D., 2006. Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. Brain Res. 1114 (1), 161–172. https://doi.org/10.1016/j.brainres.2006.07.049.

Kok, A., 2001. On the Utility of P3 Amplitude as a Measure of Processing Capacity. pp. 557–577.

Ladefoged, P., Broadbent, D.E., 1957. Information conveyed by vowels. J. Acoust. Soc. Am. 29 (1), 98–104. https://doi.org/10.1121/1.1908694.

Lange, K., 2013. The ups and downs of temporal orienting: a review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory N1 with opposite effects of attention and prediction. Front. Hum. Neurosci. 7 (June), 1–14. https://doi.org/10.3389/fnhum.2013.00263.

Lee, C., 2009. Identifying isolated, multispeaker Mandarin tones from brief acoustic input: a perceptual and acoustic study. J. Acoust. Soc. Am. 125 (2), 1125–1137. https://doi.org/10.1121/1.3050322.

Lee, C., Tao, L., Bond, Z.S., 2010. Identification of multi-speaker Mandarin tones in noise by native and non-native listeners. Speech Comm. 1–11. https://doi.org/10.1016/j.specom.2010.01.004.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. Psychol. Rev. 74 (6), 431.

Liu, F., Jiang, C., Thompson, W.F., Xu, Y., Yang, Y., Stewart, L., 2012. The Mechanism of Speech Processing in Congenital Amusia: Evidence from Mandarin Speakers. vol. 7(2)https://doi.org/10.1371/journal.pone.0030374.

Liu, F., Jiang, C., Pfordresher, P.Q., Mantell, J.T., Xu, Y., Yang, Y., Stewart, L., 2013. Individuals with congenital amusia imitate pitches more accurately in singing than in speaking: implications for music and language processing. Atten. Percept. Psychophys. 75 (8), 1783–1798. https://doi.org/10.3758/s13414-013-0506-1.

Liu, F., Chan, A.H.D., Ciocca, V., Roquet, C., Peretz, I., Wong, P.C.M., 2016. Pitch perception and production in congenital amusia: evidence from Cantonese speakers. J. Acoust. Soc. Am. 140 (1), 563–575. https://doi.org/10.1121/1.4955182.

Liversage, T., 2012. IVO ANDRI{Ć}–As Perceived In Denmark through his books. Forum Bosnae 2 (JUN). https://doi.org/10.3389/fpsyg.2011.00120.

Magnuson, J.S., Nusbaum, H.C., 2007. Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. J. Exp. Psychol. Hum. Percept. Perform. 33 (2), 391–409. https://doi.org/10.1037/0096-1523.33.2.391.

Matthews, S., Yip, V., 2013. Cantonese: A Comprehensive Grammar. Routledge.

McCarthy, G., Donchin, E., 1981. A metric for thought: a comparison of P300 latency and reaction time. Science 211 (4477), 77–80. https://doi.org/10.1126/science.7444452.

Mecklinger, A., Ullsperger, P., 1993. P3 varies with stimulus categorization rather than probability. Electroencephalogr. Clin. Neurophysiol. 86 (6), 395–407. https://doi.org/10.1016/0013-4694(93)90135-I.

Moreau, P., Jolicœur, P., Peretz, I., 2013. Pitch discrimination without awareness in congenital amusia: evidence from event-related potentials. Brain Cogn. 81 (3), 337–344. https://doi.org/10.1016/j.bandc.2013.01.004.

Mullennix, J.W., Pisoni, D.B., 1990. Stimulus variability and processing dependencies in speech perception. Percept. Psychophys. 47 (4), 379–390. https://doi.org/10.3758/BF03210878.

Näätänen, R., 1982. Processing negativity: an evoked-potential reflection. Psychol. Bull. 92 (3), 605.

Näätänen, R., Alho, K., 2004. Mechanisms of attention in audition as revealed by the event-related potentials of the brain. Cogn. Neurosci. Atten. 194–206.

Naatanen, R., Gaillard, A.W.K., Mantysalo, S., 1978. early selective-amention effect on evoked potential reinterpreted. Acta Psychol. 42 (4), 313–329 (Retrieved from All Papers/N/Naatanen et al. 1978 – EARLY SELECTIVE-AmENTION EFFECT ON EVOKED POTENTIAL REINTERPRETED.pdf).

Nan, Y., Sun, Y., Peretz, I., 2010. Congenital amusia in speakers of a tone language: association with lexical tone agnosia. Brain 133 (9), 2635–2642. https://doi.org/10.1093/brain/awq178.

Nguyen, S., Tillmann, B., Gosselin, N., Peretz, I., 2009. Tonal language processing in congenital amusia. In: Annals of the New York Academy of Sciences. 1169. pp. 490–493. https://doi.org/10.1111/j.1749-6632.2009.04855.x.

Nusbaum, H.C., Morin, T.M., 1992. Paying attention to differences among talkers. In: Tohkura Sagasaka, Y.Y., Vatikiotis-Bateson, E. (Eds.), Speech Perception, Speech Production, and Linguistic Structure. OHM, Tokyo, pp. 113–134.

Obleser, J., Kotz, S.A., 2011. Multiple brain signatures of integration in the comprehension of degraded speech. NeuroImage 55 (2), 713–723. https://doi.org/10.1016/j.neuroimage.2010.12.020.

Patel, A.D., Wong, M., Foxton, J., Lochy, A., Peretz, I., 2008. Speech intonation perception deficits in musical tone deafness (congenital amusia). Music. Percept. 25 (4), 357–368. https://doi.org/10.1525/mp.2008.25.4.357.

Peretz, I., Vuvan, D.T., 2017. Prevalence of congenital amusia. Eur. J. Hum. Genet. 25 (5), 625–630. https://doi.org/10.1038/ejhg.2017.15.

Peretz, I., Ayotte, J., Zatorre, R.J., Mehler, J., Ahad, P., Penhune, V.B., Jutras, B., 2002. Congenital amusia: a disorder of fine-grained pitch discrimination. Neuron 33 (2), 185–191. https://doi.org/10.1016/S0896-6273(01)00580-3.

Peretz, I., Champod, A.S., Hyde, K.L., 2003. Varieties of musical disorders. Ann. N. Y. Acad. Sci. 999 (1), 58–75. https://doi.org/10.1196/annals.1284.006.

Peretz, I., Brattico, E., Tervaniemi, M., 2005. Abnormal electrical brain responses to pitch in congenital amusia. Ann. Neurol. 58 (3), 478–482.

Peretz, I., Gosselin, N., Tillmann, B., Cuddy L., Gagnon, B., Trimmer G., C., … Bouchard, B. (2008). On-line identification of congenital amusia. Music. Percept., 25(4), 331–343. https://doi.org/10.1525/mp.2008.25.4.331

Peretz, I., Brattico, E., Jrvenp, M., Tervaniemi, M., 2009. The amusic brain: in tune, out of key, and unaware. Brain 132 (5), 1277–1286. https://doi.org/10.1093/brain/awp055.

Polich, J., 2007. Updating P300: An integrative theory of P3a and P3b. vol. 118. pp. 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019.

Rosnow, R.L., Rosenthal, R., 1996. Computing contrasts, effect sizes, and counternulls on other people's published data: general procedures for research consumers. Psychol. Methods 1 (4), 331–340. https://doi.org/10.1037/1082-989X.1.4.331.

Shao, J., Zhang, C., Peng, G., Yang, Y., Wang, W.S.-Y., 2016. Effect of noise on lexical tone perception in Cantonese-speaking amusics. In: Proceedings of the Interspeech. San Francisco, U.S.A.

Shao, J., Lau, R.Y.M., Tang, P.O.C., Zhang, C., 2019. The effects of acoustic variation on the perception of lexical tone in cantonese-speaking congenital amusics. J. Speech Language Hear. Res. 62 (1), 190–205. https://doi.org/10.1044/2018_JSLHR-H-17-0483.

Sjerps, M.J., Zhang, C., Peng, G., 2018. Lexical tone is perceived relative to locally

surrounding context, vowel quality to preceding context. J. Exp. Psychol. Hum. Percept. Perform. 44 (6), 914–924. https://doi.org/10.1037/xhp0000504.
Strange, W., Verbrugge, R.R., Shankweiler, D.P., Edman, T.R., 1976. Consonant environment specifies vowel identity. J. Acoust. Soc. Am. 60 (1), 213–224.
Syrdal, A.K., Gopal, H.S., 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. J. Acoust. Soc. Am. 79 (4), 1086–1100. https://doi.org/10.1121/1.393381.
Thompson, W.F., Marin, M.M., Stewart, L., 2012. Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. Proc. Natl. Acad. Sci. 109 (46), 19027–19032. https://doi.org/10.1073/pnas.1210344109.
Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umiltà, C., Peretz, I., 2011. Fine-grained pitch processing of music and speech in congenital amusia. J. Acoust. Soc. Am. 130 (6), 4089–4096. https://doi.org/10.1121/1.3658447.
Wang, J., Zhang, C., Wan, S., Peng, G., 2017. Is congenital amusia a disconnection syndrome? A study combining tract- and network-based analysis. Front. Hum. Neurosci. 11 (473). https://doi.org/10.3389/fnhum.2017.00473.
Wong, P.C.M., Diehl, R.L., 2003. Perceptual normalization for inter- and intratalker variation in Cantonese level tones. J. Speech Language Hear. Res. 46 (2), 413–421.
Wong, P.C.M., Nusbaum, H.C., Small, S.L., 2004. The neural basis of talker normalization. J. Cogn. Neurosci. 16, 1–13.
Wong, P.C.M., Ciocca, V., Chan, A.H.D., Ha, L.Y.Y., Tan, L.-H., Peretz, I., 2012. Effects of culture on musical pitch perception. PLoS One 7 (4), e33424.
Zendel, B.R., Lagrois, M.-E., Robitaille, N., Peretz, I., 2015. Attending to pitch information inhibits processing of pitch information: the curious case of amusia. J. Neurosci. 35 (9), 3815–3824. https://doi.org/10.1523/JNEUROSCI.3766-14.2015.
Zhang, C., Shao, J., 2018. Normal pre-attentive and impaired attentive processing of lexical tones in Cantonese-speaking congenital amusics. Sci. Rep. 8 (1), 8420. https://doi.org/10.1038/s41598-018-26368-7.
Zhang, C., Pugh, K.R., Mencl, W.E., Molfese, P.J., Frost, S.J., Magnuson, J.S., ... Wang, W.S.-Y., 2016. Functionally integrated neural processing of linguistic and talker information: an event-related fMRI and ERP study. NeuroImage 124. https://doi.org/10.1016/j.neuroimage.2015.08.064.
Zhang, C., Peng, G., Shao, J., Wang, W.S.-Y., 2017a. Neural bases of congenital amusia in tonal language speakers. Neuropsychologia 97, 18–28. https://doi.org/10.1016/j.neuropsychologia.2017.01.033.
Zhang, C., Shao, J., Huang, X., 2017b. Deficits of congenital amusia beyond pitch: evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics. PLoS One 12 (8), e0183151. https://doi.org/10.1371/journal.pone.0183151.
Zhang, C., Shao, J., Chen, S., 2018. Impaired perceptual normalization of lexical tones in Cantonese-speaking congenital amusics. J. Acoust. Soc. Am. 144 (2), 634–647. https://doi.org/10.1121/1.5049147.