The effects of acoustic variation on the perception of lexical tone in Cantonese-speaking congenital amusics

Jing Shao[a,b], Rebecca Yick Man Lau[a], Phyllis Oi Ching Tang[a], and Caicai Zhang[a,b,*]

[a] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China

[b] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

*Corresponding author:

Caicai Zhang: Room EF741, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China. Tel: (852) 34008465. Email address: caicai.zhang@polyu.edu.hk.

Funding Statement

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.

Conflict of Interest Statement

The authors have declared that no competing interests existed at the time of publication.

Running head: Effects of acoustic variation on amusia

Abstract

Purpose: Congenital amusia is an inborn neurogenetic disorder of fine-grained pitch processing. This study attempted to pinpoint the impairment mechanism of speech processing in tonal language speakers with amusia. We designed a series of perception tasks aiming at selectively probing low-level pitch processing and relatively high-level phonological processing of lexical tones, with an aim to illuminate the deficiency mechanism underlying tone perception in amusia.

Method: Sixteen Cantonese-speaking amusics and 16 matched controls were tested on the effects of acoustic (talker/syllable) variations on the identification and discrimination of Cantonese tones in two conditions. In the low variation condition, tones were always associated with the same talker or syllable; in the high variation condition, tones were associated with either different talkers (with the syllable controlled) or different syllables (with the talker controlled).

Results: Largely similar results were obtained in talker and syllable variation conditions. Amusics exhibited overall poorer performance than controls in tone identification. While amusics also demonstrated poorer performance in tone discrimination, the group difference was more obvious in low variation conditions, where more acoustic constancy was provided. Besides, controls exhibited a greater increase in discrimination sensitivity from high to low variation conditions, implying a stronger benefit of acoustic constancy.

Conclusions: The findings suggested that amusics' lexical tone perception abilities, in terms of both low-level pitch processing and high-level phonological processing, as measured in low and high variation conditions, are impaired. Importantly, amusics were more impaired in taking advantage of low acoustic variation contexts, and thus less efficiently sharpened their perception of tones when perceptual anchors in talker/syllable were provided, suggesting a possible 'anchoring deficit' in congenital amusia.

Keywords: congenital amusia, lexical tone perception, acoustic variation, Cantonese, anchoring deficit.

Introduction

Congenital amusia (amusia hereafter), also known as tone or tune deafness, is a lifelong disorder of musical pitch processing. It occurs without brain damage and affects about 1.5-4% of the population (Peretz & Hyde, 2003; Peretz & Vuvan, 2017). Individuals with amusia have difficulties detecting mistuned and out-of-key tones (Ayotte, Peretz, & Hyde, 2002). The primary deficit in congenital amusia lies in the processing of the pitch dimension (Foxton Dean, Jennifer L., Gee, Rosemary, Peretz, Isabelle, and Griffths, Timothy D., 2004; Hyde & Peretz, 2004; Peretz et al., 2002) and impaired short-term memory for pitch (Tillmann, Lévêque, Fornoni, Albouy, & Caclin, 2016; Tillmann, Schulze, & Foxton, 2009).

Empirical evidence has revealed that the pitch deficit in amusia is domain-general and influences speech processing negatively (Patel, Foxton, & Griffiths, 2005). This domain-general pitch-processing deficit hypothesis has been supported by several studies. For example, amusics showed inferior performance on intonation processing and identification of emotion status, where pitch is used extensively as a cue (Jiang et al., 2012; Jiang, Hamm, Lim, Kirk, & Yang, 2010; Liu, Patel, Fourcin, & Stewart, 2010; Lu, Ho, Liu, Wu, & Thompson, 2015; Thompson, Marin, & Stewart, 2012).

Deficits of lexical tone perception in amusia

In addition to intonation and emotion states, pitch is also used to systematically distinguish word meanings in tonal languages. Lexical tone is of both low-level pitch properties and high-level linguistic properties (Gu, Zhang, Hu, & Zhao, 2013; Zatorre & Gandour, 2008). Investigation on the perception of lexical tone in amusia may reveal different levels of processing and thus are of particular importance in understanding the speech processing

mechanism in amusia. Several studies have found that non-tonal language speakers with amusia showed degraded discrimination of non-native tone pairs in Mandarin and Thai, and also inferior performance on the corresponding musical analogs, suggesting a domain-general pitch-processing deficit (Nguyen, Tillmann, Gosselin, & Peretz, 2009; Tillmann et al., 2011). In a similar vein, in tonal languages, previous studies have found that the impairment in lexical tone perception in amusia might be attributed to a domain-general pitch-processing deficit (Liu et al., 2012, 2016; Shao, Zhang, Peng, Yang, & Wang, 2016). To intentionally reduce the demand for phonological processing of lexical tone, Liu et al. (2012) designed tone identification and discrimination tasks in which tones were always associated with the same segments, and words were auditorily presented with corresponding Chinese characters. In contrast to controls, Mandarin-speaking amusics showed impaired performance on word discrimination in natural speech and their gliding analogs, which were complex tones with the same pitch patterns as the speech stimuli. However, they performed as well as controls on word identification in natural speech. It thus indicates that the disorder in amusia is likely to be a domain-general pitch-processing deficit. Although it is possible that native Mandarin speakers may engage phonological processing of lexical tones in word discrimination, they can primarily rely on low-level pitch processing, especially when the tones were constantly carried by the same segments. Altogether, similar results from both non-tonal and tonal language speakers with amusia suggest that the impairment in lexical tone perception may be primarily due to a domain-general pitch-processing deficit in amusia, in that their pitch-processing deficit in the musical domain transfers to the low-level auditory pitch processing in the speech domain.

Different from the view that the deficit of amusia in lexical tone perception primarily lies at the low-level auditory pitch processing, some studies have suggested that high-level

phonological processing of lexical tones is impaired. For instance, several studies provided evidence that categorical perception of native tones in tonal language speakers with amusia is impaired (Huang, Liu, Dong, & Nan, 2015; Jiang, Hamm, Lim, Kirk, & Yang, 2012; Zhang, Shao, & Huang, 2017). Jiang et al. (2012) found that in contrast to the controls, Mandarin-speaking amusics showed no benefit for between-category tone discriminations, suggesting absence of categorical perception. Similar results were found in Huang et al. (2015), though there were individual differences in that only a subgroup of amusics with most severe tone identification and discrimination impairments showed reduced categorical perception. Zhang et al. (2017) examined the categorical perception by Cantonese-speaking amusics in four stimulus contexts: lexical tone, pure tone, vowel, and voice onset time (VOT). Amusics performed consistently worse than controls in the discrimination of frequency-based stimuli (lexical tone, pure tone and vowel), but preserved the ability of discriminating duration differences (VOT). Interestingly, there were differences in the impairment of frequency/spectral discrimination in speech and nonspeech contexts. Amusics exhibited less benefit in between-category discriminations than controls in speech contexts (lexical tone and vowel), suggesting reduced categorical perception; on the other hand, they performed inferiorly compared to controls regardless of between- and within-category discriminations in the nonspeech context (pure tone), suggesting impaired general auditory pitch processing. These differences imply that amusia may have a deficit of higher-level phonological processing in speech sounds, and also a deficit of lower-level auditory pitch processing in nonspeech sounds.

Further corroborating the phonological processing deficit hypothesis, several studies have reported reduced phonological awareness in amusics (Jones, Lucker, Zalewski, Brewer, & Drayna, 2009; Sun, Lu, Ho, & Thompson, 2017a). Jones et al. (2009) examined phonological

processing ability in amusics and controls in several tests of phonological processing, including auditory word discrimination, syllable segmentation, and the Comprehensive Test of Phonological Processing (CTOPP). The results showed that amusics demonstrated lower phonological and phonemic awareness abilities on all measures compared to controls. Sun et al. (2017) found that a subgroup of amusics with severe pitch impairment exhibited significantly worse performance than others in phonological awareness, which was further related to their pitch discrimination abilities.

Taken together, the above studies suggest that there are two main yet not mutually exclusive hypotheses on the deficit of amusia in speech processing. One hypothesis is that the deficit of amusia primarily lies in domain-general pitch processing (Liu et al., 2012, 2016; Nguyen et al., 2009; Tillmann et al., 2011), which is manifested as an impairment in low-level auditory pitch processing of lexical tones in non-tonal as well as tonal language speakers with amusia, especially when the segment variation was controlled. Another hypothesis is that the deficit already prevails to higher-level phonological processing, affecting categorical perception of lexical tones in native tonal language speakers (Huang et al., 2015; Jiang, Hamm, Lim, Kirk, & Yang, 2012; Zhang, Shao, et al., 2017) and phonological awareness (Jones, Lucker, et al., 2009; Sun et al., 2017a).

However, due to the scarcity of the studies that directly compare domain-general pitch processing and higher-level phonological processing of lexical tones, the impairment mechanism underlying the deficient tone perception performance in amusia is not well understood. The current study attempted to probe the impairment mechanism of tone perception in tonal language speakers with amusia. To this end, we designed a series of perception tasks aiming at selectively

tapping into relatively low- and high-level lexical tone perception, and by doing so aimed to shed light on the nature of the impairment mechanism in amusia.

Influence of acoustic variations on speech perception

Syllable and talker variations are two primary sources of variations in speech perception. It has been found that different degrees of syllable and talker variations employed in the stimulus set generate different degrees of task difficulty and can tap into relatively low- and high-level speech processing (Lee, Tao, & Bond, 2008, 2010; Mullennix & Pisoni, 1990; Shu, Peng, & McBride-Chang, 2008). Nan, Sun, & Peretz (2010) examined the performance of Mandarin-speaking amusics and controls in Mandarin tone discrimination, in which half of the tone pairs were associated with the same syllables, while the other half were associated with different syllables. Results revealed that amusics performed similarly to controls in discriminating tone pairs associated with the same syllables, whereas their performance was impaired in the different syllable condition. These results suggested that Mandarin-speaking amusics were particularly impaired in the different syllable condition, where the demand of phonological processing was presumably higher. Indeed, when the base segments were identical, tone discrimination can be conducted primarily through low-level pitch comparison, without necessarily engaging phonological processing of lexical tones. Thus lexical tone perception under this condition may primarily reflect low-level pitch processing. When the base segments were varied, it increased the demand for the extraction of tone categories independently from segmental processing, so that even typical Mandarin speakers found tone identification and discrimination tasks challenging (Lee et al., 2008; Shu et al., 2008). Lexical tone perception under this condition may mainly reveal the high-level phonological processing of lexical tones.

On the other hand, it is well known that talker variation influences word recognition and speech perception (Antoniou & Wong, 2015; Bradlow, Nygaard, & Pisoni, 1999; Peng, Zhang, Zheng, Minett, & Wang, 2012; Pisoni, 1993; Wong & Diehl, 2003; Zhang et al., 2016; Zhang, 2018; Zhang et al., 2016; Zhang, Peng, & Wang, 2012, 2013; Zhang & Chen, 2016). For example, Cantonese speakers identified Cantonese tones more accurately when the tone stimuli were blocked by the talker than when they were mixed across talkers (Wong & Diehl, 2003). When multi-talker speech stimuli were presented, it requires the normalization of talker variation, including the processing of talker-specific voice characteristics and the computing of mapping from talker-specific acoustic signals to phonological representations, thus increasing the demand for talker voice processing and phonological processing (Mullennix & Pisoni, 1990). Listeners were found to be slower and less accurate in speech recognition when exposed to multiple talkers as opposed to a single talker (Green, Tomiak, & Kuhl, 1997; Mullennix & Pisoni, 1990; Nusbaum & Morin, 1992). High talker variability also has an impact on the memory of words, such that listeners recalled fewer words from multi-talker lists than from single-talker lists (Martin, Mullennix, Pisoni, & Summers, 1989).

In sum, when tone stimuli were carried by identical segments or from a single talker, the acoustic variation is limited and lexical tone perception can primarily rely on low-level pitch processing. When the talker/syllable is varied, the task is more demanding, and lexical tone perception may require the extraction of abstract tone representations or the operation of talker normalization, and tap more into the high-level phonological processing of lexical tones. These hypotheses are in line with the theory underneath the high variability perceptual training (HVPT) paradigm in second language (L2) speech learning (Bradlow, 1999; Bradlow Pisoni, 1997; Pruitt, Jenkins, & Strange, 2006), which exposes L2 learners to speech stimuli embedded in multiple

phonetic contexts and produced by multiple talkers. HVPT is found to be more effective than low variability training in L2 speech learning, as the high-variation stimuli encourage and stimulate the leaners to form more abstract phonological representations (Bradlow, 2008).

In light of the aforementioned findings, low and high talker/syllable variation conditions offer an excellent scenario to examine the *low-level pitch processing* versus *higher-level phonological processing deficit* in tone perception in amusics. In the current study, we compared Cantonese-speaking amusics and musically intact controls on the effects of low versus high talker/syllable variation conditions, largely following the design of previous studies (Magnuson & Nusbaum, 2007; Nan et al., 2010; Sjerps, Zhang, & Peng, 2017; Wong & Diehl, 2003). In the high variation condition, tone stimuli produced by four different talkers (talker variation) or carried by four different base syllables (syllable variation) were inter-mixed in a single block, for the talker and syllable variation condition respectively. In the low variation condition, the same set of multi-talker/syllable stimuli were presented, but blocked by the talker for the talker variation condition and by the syllable for the syllable variation condition. We hypothesize that lexical tone perception in low and high talker/syllable variation contexts is supported by different processing mechanisms, with tone perception in the low talker/syllable variation context primarily mediated by low-level pitch processing, whereas that in the high talker/syllable variation context primarily driven by higher-level phonological processing. We further hypothesize that the amusics' performance would be impaired in both conditions, and that the high talker/syllable variation conditions, which require more phonological processing, would be particularly challenging for amusics. Since listeners have to extract abstract tonal representations from the highly variable speech stimuli through higher-level phonological operations for tone

identification and discrimination, amusics are expected to show greater impairment in the high talker/syllable variation conditions.

Method

Participants

16 congenital amusics and 16 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, and no reported history of formal musical training. Amusics and controls were identified using the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz, Champod, & Hyde, 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.

Stimuli

Talker variation condition

The stimuli were six words contrasting six Cantonese tones on the syllable "/ji/": high level tone (T1) – /ji55/ (醫 'a doctor'), high rising tone (T2) – /ji25/ (椅 'chair') , mid level tone (T3) –

/ji33/ (意 ‘meaning’), extra low level/low falling tone (T4) – /ji21/ (兒 ‘son’), low rising tone (T5) – /ji23/ (耳 ‘ear’), and low level tone (T6) – /ji22/ (二 ‘two’) (Bauer & Benedict, 1997; Matthews & Yip, 2013). Two female and two male native Cantonese speakers 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 650 ms, which was the mean duration of the selected tokens, and in mean intensity to 70 dB using Praat (Boersma & Weenink, 2014). Figure 1 displays the F0 contours of the six tones produced by the four talkers.

Syllable variation condition

The stimuli in the syllable variation condition were 24 meaningful words contrasting six Cantonese tones on syllables /ji/, /fɐn/, /fu/ and /wɐi/ (see Appendix). These four syllables were selected for the reasons that their syllable structures are relatively simple and that they occupy a large vowel space (e.g., /i/, /u/ and /ɐ/) that could maximize the pitch variation, given the well-established relationship between vowel height and pitch height (Fischer-Jorgensen, 1990; Lehiste, 1970). One female native Cantonese speaker was recorded reading aloud these 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. All selected words were normalized in duration to 650 ms and in mean intensity to 70 dB using Praat. Figure 2 shows the F0 trajectories of the six tones carried by the four syllables.

Procedure

Talker variation condition

The same set of stimuli was presented in two conditions, the low variation and high variation condition. The critical difference between the two conditions was that the stimuli from multiple talkers were presented in a blocked-talker manner in the low variation condition and in a mixed-talker manner in the high variation condition (Magnuson & Nusbaum, 2007; Sjerps et al., 2017; Wong & Diehl, 2003). In other words, there was no talker variation in the low variation condition. Each condition included an identification task and a discrimination task. The stimuli were presented using E-prime 2.0.

In the low variation condition, the stimuli produced by the four talkers (M01, M02, F01 and F02) were blocked by the talker and presented in four sub-blocks. In the *identification* task, each set of six words (/ji/) from one talker was presented in a sub-block, generating four sub-blocks. Six words in a talker set were repeated twice and presented randomly within the sub-block, generating a total of 48 trials in the identification task. In each trial, a fixation first occurred on the computer screen for 500 ms, followed by the presentation of a 650-ms speech stimulus via headphones. Subjects were instructed to identify the tone of the word by pressing buttons 1-6 on a keyboard. There was a break after each sub-block. Subjects were given a list of words contrasting the six tones beforehand to get familiar with the tone distinctions and to facilitate tone identification in the subsequent experiment. In the *discrimination* task, six words (/ji/) in each talker set were grouped into 15 different tone pairs and six same tone pairs. Each talker set was presented in a sub-block, generating four sub-blocks in total. Within a sub-block, 15 different tone pairs were presented twice in forward and reversed order and six same tone pairs were repeated five times, generating equal number of different and same tone pairs, which were intermixed and randomly presented. The total number of trials in the discrimination task was 240

(4 talker blocks × 60 tone pairs). In each trial, a fixation first occurred on the computer screen for 500 ms, followed by the presentation of two 650-ms stimuli separated by an inter-stimulus-interval of 500 ms via the headphones. Subjects were instructed to judge whether the two words carried the same tone or different tones by pressing "left arrow" (same) or "right arrow" (different) on a keyboard within 3 seconds. There was a break after every 20 trials within a sub-block and also a break after each sub-block.

In the high variation condition, the stimuli produced by the four talkers (M01, M02, F01 and F02) were intermixed in a block. In the *identification* task, the four talker sets (6 tones × 4 talkers = 24 trials) were repeated twice and presented randomly within a block, generating a total of 48 trials. The trial procedure was the same as that in the low variation condition. In the *discrimination* task, two words in each tone pair (same or different) were always from two different talkers. In order to keep the experiment duration short, the four talkers were grouped into two sets, with set A including six talker pairs (M01-M02, F01-M01, M01-F02, F01-F02, M02-F01, and F02-M02), and set B including the same talker pairs in reversed order (M02-M01, M01-F01, F02-M01, F02-F01, F01-M02, and M02-F02). Each talker pair carried 15 different and six same tone pairs. The 15 different tone pairs were repeated twice and six same tone pairs were repeated five times, generating 60 trials for each talker pair. Within each group, half of the participants were randomly assigned to the set A and the other half to the set B. The subjects were informed that in the high variation condition the talkers were always different, and they were asked to ignore the voice difference and pay close attention to the tones. The trial procedure was the same as that described above. There was a break after every 20 trials within a block. The total number of trials was 360 in the discrimination task (6 different talker pairs × 60 tone pairs). Note that the total number of trials in low and high variation discrimination tasks was not the

same. However, this difference is not expected to have much influence on the results, as the variation difference was a within-subjects factor, and all participants did the same tasks.

In both identification and discrimination tasks, the stimuli were presented binaurally through earphones to the participants in a soundproof room. The stimuli were presented at a comfortable listening level to the subjects and the volume level was kept constant within an experiment and across all subjects. The presentation order of the identification and discrimination tasks was counterbalanced across the participants. The block order was kept identical between each amusic participant and the accordingly matched control participant. Before each task, a practice block was given to the participants to familiarize them with the procedure. In both identification and discrimination tasks, accuracy and reaction time (RT) were recorded.

Syllable variation condition

Similar to the talker variation condition, the same set of stimuli with syllable variations was presented in low and high variation conditions. Note that there was no syllable variation in the low variation condition. Each condition included an identification task and a discrimination task. The stimuli were presented using E-prime 2.0.

In the low variation condition, tone stimuli carried by the four syllables (/ji/, /fɐn/, /fu/ and /wɐi/) were blocked by the syllable and presented in four sub-blocks. In the *identification* task, each syllable set was presented in a sub-block, in which six words were repeated twice and presented randomly. The trial procedure was the same as that described above. In the *discrimination* task, six words in each syllable set were grouped into 15 different tone pairs and six same tone pairs. Each syllable set was presented in a sub-block, in which 15 different tone pairs were presented twice in forward and reversed order and six same tone pairs were repeated

five times, generating equal number of different and same tone pairs, which were intermixed and randomly presented. The trial procedure was the same as that described above.

In the high variation condition, tone stimuli carried by the four syllables (/ji/, /fɐn/, /fu/ and /wɐi/) were intermixed in a block. In the *identification* task, the four syllable sets (6 tones × 4 syllables = 24 trials) were repeated twice and presented randomly within a block. In the discrimination task, each tone pair (same or different) was always associated with different syllables. In order to keep the experiment duration short, the four syllables were grouped into two sets, with set A including six syllable pairs (/fɐn/-/ji/, /ji/-/wɐi/, /fu/-/wɐi/, /fɐn/-/wai/, /fɐn/-/fu/ and /fu/-/ji/) and set B including the same syllable pairs in reversed order (/ji/-/fɐn, /wɐi/-/ji/, /wɐi/-/fu/, /wɐi/-/fan/, /fu/-/fɐn/ and /ji/-/fu/). Each syllable pair carried 15 different and six same tone pairs. The 15 different tone pairs were repeated twice and six same tone pairs were repeated five times, generating a total of 60 trials for each syllable pair (6 different syllable pairs × 60 tone pairs). Within each group, half of the participants were randomly assigned to the set A and the other half to the set B. The subjects were also informed that in the high variation condition the base syllables were always different, and they were asked to ignore the syllable difference and pay close attention to the lexical tones in tone discrimination.

In both the identification and discrimination tasks, the stimuli were presented binaurally through earphones to the participants in a soundproof room. The stimuli were presented at a comfortable listening level to the subjects and the volume level was kept constant within an experiment and across all subjects. The presentation order of the identification and discrimination tasks was counterbalanced across the participants, and kept identical between matched amusic and control participants. Before each task, a practice block was given to the participants to familiarize them with the procedure. In all tasks, accuracy and RT were recorded.

Data analysis

For the identification tasks, accuracy and RT were analyzed. Response to each trial was coded as 1 or 0 (correct or incorrect) for each participant. To compare the accuracy of amusics and controls, generalized mixed-effects models were fitted on the responses to each trial (1 or 0) with *group* (amusics and controls) and *variation* (low talker/syllable variation and high talker/syllable variation) as two fixed effects, and with by-subject random intercept and slope as random effects; two-way interaction was also included as a fixed effect in the models. In order to test the significance of fixed effects, a simple model (m_0) with only the intercept as a factor was first fitted, and the factors *group* and *variation* were added to the model consecutively. The model with a fixed effect (e.g., *group*) was compared with a baseline model without it. Models were compared by likelihood ratio tests and p-values were obtained from those tests.

RT was measured from the offset of the stimuli to the time that a response was made. RT was analysed in addition to accuracy, for the reason that it provides an extra index to measure the cognitive effort used in the speech perception task, by revealing subtle temporal differences between different listener groups or different listening conditions. When the accuracy among the listeners groups is comparable, shorter RT usually suggests more rapid, robust and automatic speech processing and implies less cognitive effort involved in the task (Strange, 2011). Thus, RT was calculated in the current study to provide another dimension for evaluating the perceptual ability of amusics and controls. All trials were included in the RT analysis, irrespective of whether the response was correct or not. Incorrect trials were not discarded for the reasons that the accuracy was quite low in the identification tasks (41%-65%) and not evenly distributed across different conditions. Excluding incorrect trials therefore leads to the removal

of a significant portion of the data and imbalanced number of trials across different conditions. Statistical analyses had been conducted on the RT of correct and incorrect trials separately, and incorrect trials yielded either similar effects as correct trials or no significant effect (perhaps due to small number of trials) (please see the supplemental files S1 and S2). The results of RT analyses combining correct and incorrect trials are reported here. Linear mixed-effects models were fitted on the log-transformed RT data with *group* and *variation* as two fixed effects, and with by-subject random intercept and slope as random effects; two-way interaction was also included as a fixed effect in the models. Models were compared by likelihood ratio tests and p-values were obtained from those tests. The above two sets of analyses were performed with R (R Core Team, 2014), using the *lme4* package (Bates, Maechler, & Bolker, 2012), the *lmtest* package (Zeileis & Hothorn, 2002) and the *lsmeans* package (Lenth, 2016).

For the discrimination tasks, the sensitivity index d' (Macmillan & Creelman, 2005) and RT were analyzed. The discrimination responses were analyzed in terms of d' rather than accuracy for the reason that d' takes into consideration the response bias. Following previous studies on speech discrimination (Zhang, 2018; Zhang, Shao, et al., 2017), the d' was computed as the z-score value of the hit rate ("different" responses to different tone pairs) minus that of the false alarm rate ("different" responses to same tone pairs) for each tone pair per subject. Note that this formula for calculating d' (z(H)-z(F)) follows the assumption of an independent-observation strategy in fixed designs (Macmillan & Creelman, 2005). Nonetheless, the tone pairs were intermixed and randomly presented in a block in the current study, which is not unlike a roving design. Macmillan & Creelman (2005) suggested that in roving designs, listeners might adopt a differencing strategy, applying a threshold of difference to decide if two stimuli are different enough to be judged as different. But since the stimuli were all speech materials with

recognizable categories in the current study, there is reason to assume that listeners adopted the independent-observation strategy, which justifies the calculation of d' as z(H)-z(F). Trials with null responses were disregarded (talker variation condition: 0.5%; syllable variation condition: 2.6%). *Group* × *variation* repeated measures ANOVAs were conducted on the d' scores, using the Statistical Package for the Social Sciences (SPSS) (SPSS, 2011). Corrections for violations of sphericity were made, where appropriate, using the Greenhouse-Geisser method. Tukey's HSD post-hoc test was applied to make pairwise comparisons when necessary. All effects were reported as significant at $p < .05$.

In the discrimination tasks, RT was measured from the offset of the second stimulus in a pair to the time that a response was made. All trials with a response made within the time limit were included in the RT analysis, irrespective of whether the response was correct or not. Linear mixed-effects models were fitted on the log-transformed RT data with *group* and *variation* as two fixed effects, and with by-subject random intercept and slope as random effects, following the procedure described above. The results of RT analyses separated by correct and incorrect trials are reported in the supplemental files S1 and S2.

Results

Talker variation condition

Figure 3(a) illustrates the identification accuracy in each condition for the amusic and control group. Generalized mixed-effects model revealed a significant main effect of *group* ($\chi^2(1) = 7.88$, $p = .004$). Amusics showed significantly lower accuracy compared to controls (amusics: $M = 0.41$, $SD = 0.22$; controls: $M = 0.61$, $SD = 0.18$). No other effects were significant. Although the two-way interaction was not significant ($\chi^2(1) = 0.60$, $p = .438$), there was a trend of larger

group differences in the low variation condition (amusics: M = 0.41, SD = 0.24; controls: M = 0.63, SD = 0.18) compared to the high variation condition (amusics: M = 0.41, SD = 0.20; controls: M = 0.60, SD = 0.18). Furthermore, controls exhibited a trend of benefiting more from low talker variations (low variation: M = 0.63, SD = 0.18; high variation: M = 0.60, SD = 0.18) compared to amusics (low variation: amusics: M = 0.41, SD = 0.24; high variation: amusics: M = 0.41, SD = 0.20) in tone identification, though the difference was small.

Figure 3(b) displays the RT in the identification task. No significant effects were found.

Figure 3(c) shows the discrimination sensitivity d'. *Group* × *variation* repeated-measures ANOVA found significant main effects of *group* ($F(1, 30) = 24.413$, $p < .001$, $\eta_p^2 = 0.449$) and *variation* ($F(1, 30) = 119.974$, $p < .001$, $\eta_p^2 = 0.800$) and a significant two-way interaction ($F(1, 30) = 10.453$, $p = .003$, $\eta_p^2 = 0.258$). Independent-samples t-tests showed that amusics exhibited significantly lower d' than controls in both low variation ($t(30) = -4.762$, $p < .001$, $d = 1.684$) and high variation conditions ($t(30) = -3.430$, $p = .002$, $d = 1.212$), but the group difference was larger in the low variation condition. Controls demonstrated significantly higher d' in the low variation condition than in the high variation condition ($t(30) = -6.505$, $p < .001$, $d = 2.299$), with a large effect size. In contrast, while amusics also exhibited larger d' scores in the low variation condition than in the high variation condition ($t(30) = -2.047$, $p = .05$, $d = 0.723$), this effect was only marginally significant. These results indicate that amusics showed inferior performance than controls under both low and high talker variation conditions, but the group difference was larger in the low variation condition. Besides, controls benefited more from low talker variations in tone discrimination.

Figure 3(d) shows the RT in the discrimination task. Linear mixed-effects model found a significant main effect of *variation* ($\chi^2(1) = 394$, $p < .001$), where RT in the high talker variation

condition was significantly longer than in the low talker variation condition. No other effects were significant.

Syllable variation condition

Figure 4(a) illustrates the identification accuracy. Generalized mixed-effects models found a significant main effect of *group* ($\chi^2(1) = 7.17$, $p < .001$), where amusics showed significantly lower accuracy compared to controls (amusics: $M = 0.44$, $SD = 0.21$; controls: $M = 0.62$, $SD = 0.16$). No other effects were significant. Although the *group* by *variation* interaction was not significant ($\chi^2(1) = 1.529$, $p = .216$), the group difference appeared to be larger in the low syllable variation condition (amusics: $M = 0.44$, $SD = 0.23$; controls: $M = 0.65$, $SD = 0.18$) than that in the high variation condition (amusics: $M = 0.44$, $SD = 0.20$; controls: $M = 0.59$, $SD = 0.13$). Furthermore, controls appeared to benefit more from low syllable variations (low variation: $M = 0.65$, $SD = 0.18$; high variation: $M = 0.59$, $SD = 0.13$) compared to amusics (low variation: $M = 0.44$, $SD = 0.23$; high variation: $M = 0.44$, $SD = 0.20$) in tone identification.

Figure 4(b) shows the RT in the identification task. Linear mixed-effects model found a significant main effect of *group* ($\chi^2(1) = 4.012$, $p = .04$), *variation* ($\chi^2(1) = 126.3$, $p < .001$), and a significant two-way interaction ($\chi^2(1) = 6.036$, $p = .014$). Pairwise comparisons revealed that in the control group, RT in the high variation condition was significantly longer than that in the low variation condition ($z = 6.198$, $p < .001$); such effect was also significant in the amusic group, but the difference was smaller ($z = 2.669$, $p = .01$). Under the high variation condition, controls exhibited significantly longer RT than amusics ($z = -2.255$, $p = .03$), but such difference cannot be found in the low variation condition ($z = -1.581$, $p = .124$). These results suggest that the

control group was under greater influences of syllable variations in terms of RT compared to the amusic group.

Figure 4(c) displays the discrimination sensitivity d'. *Group* × *condition* repeated-measures ANOVA found significant main effects of *group* ($F(1, 30) = 33.637$, $p < .001$, $\eta_p^2 = 0.529$) and *variation* ($F(1, 30) = 133.709$, $p < .001$, $\eta_p^2 = 0.817$) and a significant two-way interaction ($F(1, 30) = 4.926$, $p = .034$, $\eta_p^2 = 0.141$). Post hoc analyses showed that although both amusic and control groups exhibited significantly higher d' scores in the low variation condition than in the high variation condition (amusics: $t(30)= -5.462$, $p < .001$, $d = 1.931$; controls: $t(30)= -9.017$, $p < .001$, $d = 3.188$), the effect was larger in the control group. Within each variation condition, the amusic group demonstrated significantly lower d' than controls (high variation: $t(30)= -4.585$, $p < .001$, $d = 1.621$; low variation: $t(30)= -4.595$ $p < .001$, $d = 1.624$). These results demonstrate that controls benefited more from low syllable variation compared to amusics.

Figure 4(d) shows RT in the discrimination task. Linear mixed-effects model found a significant interaction between *group* and *condition* ($\chi^2(1) = 196.16$, $p < .001$). Pairwise comparisons revealed that in the high variation condition, RT in the control group was significantly longer than the amusic group ($z = -2.187$, $p = .03$); in the low variation condition, RT in the two groups was similar ($z = -0.142$, $p = .88$). Within each group, RT in the high variation condition was significantly longer than that in the low variation condition, but the effect was more evident in the control group ($z = 32.886$, $p < .001$) than in the amusic group ($z = 12.983$, $p < .001$). These RT results suggest that controls tended to show greater benefit from low syllable variation, where they employed significantly shorter RT in the low variation condition than in the high variation condition.

To summarize, the amusics performed worse than the controls in both tone identification and discrimination tasks. In the talker variation condition, there was a significant group effect on the identification accuracy, where controls outperformed than amusics. Although the two-way interaction was not significant, there were trends that controls exhibited a greater increase of accuracy from high to low variation conditions. In the discrimination task, there was a significant *group* by *variation* interaction effect on the d' score, where only the controls demonstrated significantly higher d' in the low variation condition than in the high variation condition, and the group difference was more evident in the low variation condition. The discrimination RT mainly revealed an effect of variation, where RT in the low variation condition was shorter than the high variation condition. In the syllable variation condition, the perceptual patterns were similar, except that the RT data also revealed significant *group* by *variation* interactions. In the identification task, controls exhibited significantly shorter RT in the low variation condition than in the high variation condition, but this effect was smaller in amusics. Similarly, in the discrimination task, controls again tended to show greater benefit from the low variation condition, by employing shorter RT in the low variation condition than in the high variation condition.

Discussion

While amusia has been consistently reported to influence tone perception negatively (Liu et al., 2016; Nan et al., 2010; Shao et al., 2016), the mechanism underlying the deficient tone perception is not well understood. In this study, we examined this issue through a comparison of two conditions, *low* talker/syllable variation versus *high* talker/syllable variation. In the low variation condition, only one talker/syllable was associated with the tones, and acoustic

variations among the stimuli were low. It was hypothesized that tone perception in this condition primarily tapped into relatively low-level pitch processing. In the high variation condition, tones were always associated with different talkers or syllables, meaning that listeners had to extract abstract tonal representations out of speech signals with large acoustic variations. When associated with different talkers, tone perception additionally involved talker voice processing. It was hypothesized that tone perception in this condition primarily tapped into relatively high-level phonological processing of lexical tones. It was predicted that amusics would show greater impairment in the high talker/syllable variation conditions given the hypothesized phonological processing deficit.

We found that the general patterns in talker and syllable variation conditions were largely similar. Amusics exhibited overall lower accuracy than controls in tone identification tasks. In tone discrimination tasks, there were significant *group* by *variation* interactions, in that the controls exhibited a greater increase of d' scores from high to low variation conditions, and the group difference was larger in the low variation conditions. The results deviated somewhat from the predicted patterns of greater group differences in the high variation conditions. In the text below, we first discussed the observed impairment of amusics in low and high variation conditions, respectively, in relation to the hypothesized low-level pitch-processing and high-level phonological processing deficits of amusia. We then tried to explain the larger group differences observed in the low variation conditions, together with an exploration of the difference between identification and discrimination tasks.

We found that amusics performed significantly worse than the controls in *low* talker/syllable variation conditions, confirming that amusics are generally impaired in tone perception, compatible with the *low-level* pitch-processing deficit. These results are consistent with previous

findings of impaired tone perception in tonal and non-tonal language speakers with amusia when the segments were fixed (Liu et al., 2016; Nguyen et al., 2009; Shao et al., 2016; Tillmann et al., 2011). In Liu et al. (2016), four Cantonese tone pairs were selected and presented to amusics and controls. The four tone pairs were always carried by the same syllables, which corresponds to the low syllable variation condition in the current study. Liu and colleagues found that Cantonese-speaking amusics demonstrated impaired tone perception compared to controls. The current results also echo with Shao et al. (2016), where Cantonese-speaking amusics showed deficient tone perception when the syllable variation was low.

However, the results that amusics were impaired in tone discrimination when the tones were carried by the same syllables are inconsistent with those reported in Nan et al. (2010). Nan et al. (2010) found that Mandarin-speaking amusics did not score significantly below the controls in tone discrimination when the carrying syllables were identical, but they did exhibit poorer performance when the syllables were different. This discrepancy between Cantonese- and Mandarin-speaking amusics may be due to the different tonal systems of the two languages. Each of the four tones in Mandarin has a unique F0 direction, which may facilitate successful tone discrimination (Gandour, 1983; Gandour & Harshman, 1978; Peng et al., 2012). In contrast, the tonal space of Cantonese is much denser (Bauer & Benedict, 1997; Matthews & Yip, 2013; Peng, 2006; Peng et al., 2012). Discrimination of the six tones in Cantonese relies on both F0 level and F0 direction, which increases difficulties in tone discrimination (Gandour, 1983; Gandour & Harshman, 1978; Peng et al., 2012). It is thus not surprising that Mandarin-speaking amusics showed comparable performance to controls when the same syllables were associated to the tones. In contrast, Cantonese-speaking amusics demonstrated poorer performance in the low variation condition.

Our results also revealed that Cantonese-speaking amusics showed significantly worse tone perception performance than controls in *high* talker/syllable variation conditions, suggesting a deficit in *high-level* phonological processing of lexical tones. Tones carried by different talkers/syllables were more acoustically different, increasing the difficulty of reliably extracting tone categories from acoustical signals and the cognitive load (Johnson, 2005). Perception of tones carried by different talkers/syllables involves phonological processing, i.e., extracting abstract phonological representations of tones from different talkers and syllables, in addition to talker voice processing (talker condition) and segmental processing (syllable condition). High talker/syllable variations could decrease the "quality" of extracted tone categories, affecting tone perception accuracy. The worse performance of amusics in the high variation condition therefore may suggest a deficit in high-level phonological processing of lexical tones. These findings are in line with previous studies which reported that categorical perception of native tones is impaired in Chinese amusics (Huang et al., 2015; Jiang, Hamm, Lim, Kirk, & Yang, 2012; Zhang et al., 2017). They also coincide partly with the findings of Wang & Peng (2014). Wang & Peng (2014) indicated that Mandarin-speaking controls discriminated non-native Cantonese level tones more accurately when the tones were carried by native Mandarin syllables than by Cantonese syllables, whereas amusics failed to show such benefit, suggesting a possible deficit of amusics in benefiting from native syllables when processing non-native tones. Altogether, these findings indicate that the deficit of amusics is not confined to auditory pitch discrimination, but prevails to high-level phonological processing of lexical tones.

The impairment of amusia in phonological processing is also somewhat compatible with the observation of poor phonological awareness in amusia. Several studies have indicated that amusics show reduced phonemic awareness (Jones, Lucker, et al., 2009; Sun et al., 2017a),

which may account for their impoverished performance in high variation conditions in the current study, where much phonological decoding was demanded. When breaking down the structure of the syllables to extract phonological representations of tones independently from representations of segments, the amusics may encounter greater difficulties compared to controls, leading to worse tone perception performance.

Note that the high talker variation conditions may not only involve phonological processing, but also talker voice processing (Johnson, 2005). Worse tone perception performance of amusics in these conditions may imply an additional talker-processing deficit in amusia. It has been found that encoding an acoustic signal into a phonological representation is closely related to the process of encoding a talker's voice. Talker and phonetic information can not be selectively ignored when participants are instructed to attend to a specific dimension of the speech stimuli, suggesting that talker and phonetic information is processed in an integral fashion (Kaganovich, Francis, & Melara, 2006; Mullennix & Pisoni, 1990; Perrachione, Pierrehumbert, & Wong, 2009; Zhang et al., 2016). The current findings that amusics were impaired in tone perception in high talker variation conditions may be partially affected by a deficit in talker processing, in addition to the high-level phonological processing deficit.

It should be noted that we hypothesized that the group difference would be more pronounced in the high variation conditions, which were phonologically more demanding tasks. However, we observed the opposite patterns in that the group difference was generally larger in the low variation conditions, where the talker/syllable was consistently repeated and the acoustic variation was limited. This pattern was most noticeable in the discrimination tasks. For the identification accuracy, there was a similar trend of larger group differences in the low variation conditions but it was not statistically significant. This is probably because in the identification

task, there was no talker/syllable variability within a trial, unlike in the discrimination task where listeners were confronted with two stimuli with different talkers or syllables when discriminating the tone differences. Although there was talker/syllable variability across trials in the identification task, this may somehow have limited the influence of talker/syllable variability. This may explain why there was only an insignificant trend in the identification task. Nonetheless, the identification RT did reveal a significant interaction of *group* and *variation* in the syllable condition, in that the control group tended to exhibit significantly shorter RT in the low variation condition than in the high variation condition, but the effect was reduced in amusics. RT thus provides an additional index to evaluate the perceptual performance of amusics, complementing the identification accuracy. Overall, these findings suggest that controls benefited more from low acoustic variations, exhibiting more accurate and rapid tone perception, especially when there was within-trial talker/syllable variability (i.e., in discrimination tasks).

The finding that amusics were more impaired in the low variation condition can be explained by the 'anchoring theory' which is originally proposed to account for the phonological deficit in dyslexia (Ahissar, 2007; Ahissar, Lubin, Putter-Katz, & Banai, 2006). This hypothesis states that the deficit of dyslexics lies in the dynamics that link perception with perceptual memory through the implicit formation of perceptual anchors. These anchors may guide the perception of subsequent stimuli, help retain and retrieve the incoming stimuli. Empirical evidence showed that typical listeners tune to, or 'anchor to', incoming stimuli quickly. When the stimuli are subsequently repeated, they can react more accurately with the 'perceptual anchors'. On the contrary, individuals with dyslexia failed to benefit from such repetitions. For example, Ahissar et al. (2006) investigated nonspeech tone discrimination by individuals with dyslexia. In one condition, a standard tone was constantly presented within each trial; in another

condition, no standard tone was presented. The results showed that the dyslexics performed poorly in the condition where the standard tone was presented, while their performance on the difficult 'no standard' condition did not differ from the controls. Moreover, the dyslexics' performance did not differ between the two conditions; in contrast, the controls' performance in the standard condition was considerably better than in the 'no standard' condition. It was proposed that in the standard condition, listeners could form a perceptual anchor based on the repeated standard tones, and thus could make judgment quickly and accurately. Individuals with dyslexia failed to form such an 'anchor' and showed degraded performance in the standard condition. Altogether, it suggests that dyslexic individuals have difficulties in dynamically constructing "perceptual anchors", deriving from deficient adaptation mechanisms.

Results of the current study can also be accounted for by the 'anchoring theory'. In the low variation condition, tone pairs were always associated with the same syllable and talker. Anchoring to the consistent syllable or talker identity in the speech stimuli probably facilitates tone perception. Controls benefited from the consistency in syllable and talker identity, exhibiting higher accuracy and shorter RT in tone perception in the low variation condition than in the high variation condition. In contrast to controls, the performance of amusics improved less when the same talker or syllable was presented across trials, indicating that their ability to build and tune to an internal talker/syllable anchor might be impoverished. Therefore, without a properly functioning anchoring mechanism, the perceptual system of amusics may be less resilient to external variation, similar to the case of dyslexics. This explains why under conditions with low talker/syllable variation, the perception of controls greatly sharpened compared with the high talker/syllable variation conditions, whereas the perception of amusics showed less improvement. The hypothesis of a potential 'anchoring deficit' in amusia is also

consistent with the result of a previous functional MRI study that the brain of amusics showed abnormally strong activations towards repeated pitch stimuli, while the activation in the controls' brain attenuated (Zhang, Peng, Shao, & Wang, 2017).

The fact that amusics might exhibit a similar 'anchoring deficit' to individuals with dyslexia could be partly tracked down to the attention deficit and impaired short-term memory. For example, Hari & Renvall (2001) indicated that the core deficit of dyslexia involves poor attentional skills. Ahissar (2007) suggested that the anchoring deficit hypothesis in dyslexia can be viewed as a type of attention deficit hypothesis. Furthermore, the short-term memory is found to be closely related to the 'anchoring' performance in dyslexics (Ahissar et al., 2006). Amusics, on the other hand, are also found to display impaired short-term memory (Albouy et al., 2013; Gosselin, Jolicœur, & Peretz, 2009; Tillmann et al., 2016; Williamson & Stewart, 2010). Approximately 40% of the amusic individuals are found to have an attention deficit (Jones, Zalewski, Brewer, Lucker, & Drayna, 2009). With poor working memory and attention deficits, amusics may have greater difficulty benefiting from the clean and variation-free conditions. Altogether, it is likely that the potential anchoring deficit in amusia contributes to the impoverished lexical tone perception performance in the low acoustic variation conditions, beyond their pitch-processing deficit.

It is worth noting that while the results were largely similar between the talker and syllable variation conditions, some differences can be observed, especially in the RT results. In general, significant group differences or *group* by *condition* interactions were found in the RT of the syllable variation condition, but not in the talker variation condition, implying a stronger impact of syllable variations on revealing the disparity between amusics and controls in response time during tone perception. It is possible that syllables carried more linguistic information, and when

the syllables varied it was more distracting than when the same syllable varied in different talkers' voices. As a result, syllable variations may place a greater demand on the anchoring system with a stronger influence on tone perception.

Lastly, the tone perception impairment in amusics appears to be superficially similar to that of cochlear implant (CI) users, who also exhibit tone recognition deficiencies (Chang, Chang, Lin, & Luo, 2016). It has been found that Mandarin-speaking CI users performed worse in the mixed-talker condition than in the blocked-talker condition in tone and vowel recognition (Chang et al., 2016). While our results suggested that amusics only showed marginally better performance in the blocked-talker condition than the mixed-talker condition in Cantonese tone perception, suggesting that they failed to benefit as much from acoustic constancy in the blocked-talker condition. This discrepancy suggests that the deficiency mechanism of amusia is likely to be different from that the CI users, who appeared to have a normal anchoring mechanism, and impaired low-level pitch processing that is similar to the amusics. It also implies that the deficit of amusia is more cortically originated (Liu, Maggu, Lau, & Wong, 2015). Future studies can further explore the potentially different deficiency mechanisms of amusia and CI users.

To conclude, we found that Cantonese-speaking amusics were impaired in both low-level pitch processing and higher-level phonological processing of lexical tones. Furthermore, amusics were more impaired than controls in low variation contexts, where acoustic constancy in syllables/talkers can be used to construct perceptual anchors. These findings shed some light on the mechanism of the deficient tone perception in amusia. We propose that it may reflect impaired dynamics of the stimulus anchoring mechanism, suggesting a possible 'anchoring deficit' in amusia. A few unsolved issues arise from the current study, which can be further

examined in future studies. First, to explicitly test the anchoring deficit hypothesis in amusics, future studies can examine the performance of amusics using study designs identical to those in previous studies of the "anchoring theory" (Ahissar, 2007; Ahissar et al., 2006), together with tests of attention and short-term memory. Second, while the current study focused on impaired speech processing (lexical tone perception) in tonal language speakers with amusia, it is possible that the potential anchoring deficit revealed in lexical tone perception (to be further verified) is domain-general. Analogous to the manipulation of syllable and talker variations in the current study, future studies may examine the effect of timbre variation (e.g., piano, violin, etc.) on pitch processing in amusical speakers of a non-tonal language. Third, it is worth examining whether the amusics' ability to develop a perceptual anchor would improve with more exposure to the speech stimuli. Fourth, it is possible the potential anchoring deficit in amusics was manifested most strongly when the task was demanding and the stimulus set was highly variable and complex. In current study, there were four base syllables and four talkers' voices; in the high variation discrimination tasks, tones carried by different base syllables or different talkers' voices were always paired together to maximize the variability. The task difficulty under the current experiment settings was considerably high. It is possible that in less demanding experiment settings, the group difference might be compressed and the anchoring deficit might be less clearly demonstrated. Future studies may examine the effect of task demand and stimulus variability on the perception of lexical tones in amusia. Fifth, future studies with a larger sample size in each participant group may want to include tone/tone pair as a third factor in statistical analyses, to further examine whether certain tone or tone pairs may suffer more from high talker/syllable variations. Sixth, some aspects of stimulus presentation in the high variation conditions can be improved. For example, stimulus presentation may be better controlled by not

presenting stimuli with the same talker/syllable in consecutive trials, in order to maximize the variability across trials. Furthermore, the high variation conditions may be broken down into four sub-blocks, to be more comparable to the low variation conditions. Last, the number of trials in the high variation discrimination tasks was larger than that in the low variation discrimination tasks, which may have increased fatigue. Although this presumably did not affect the comparison between amusics and controls, future studies may try to equalize the number of trials by randomly selecting a subset of four talker/syllable pairs in the high variation discrimination tasks.

Acknowledgements

We thank the anonymous reviewers for constructive comments. We thank Mr. Yubin Zhang for help with data analysis. The data of the syllable variation condition were presented at the Speech Prosody 2018 in Poznań, Poland, during 13-16 June 2018.

References

Ahissar, M. (2007). Dyslexia and the anchoring-deficit hypothesis. *Trends in Cognitive Sciences*, *11*(11), 458–465. http://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. *Nature Neuroscience*, *9*(12), 1558–1564. http://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. http://doi.org/10.1093/brain/awt082

Antoniou, M., & Wong, P. C. M. (2015). Poor phonetic perceivers are affected by cognitive load when resolving talker variability; Poor phonetic perceivers are affected by cognitive load

when resolving talker variability. *Journal of Acoustical Society of America*, *138*(2), 571–574. http://doi.org/10.1121/1.4923362

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, *125*(2), 238–251. http://doi.org/10.1093/brain/awf028

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: linear mixed-effects models using S4 classes. R package version 0.999375-42. 2011.

Bauer, R., & Benedict, P. K. (1997). *Modern cantonese phonology*. Berlin: Mouton de Gruyter.

Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer.

Bradlow, A. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*.

Bradlow, A. (2008). Training non-native language sound patterns:Lessons from training Japanese adults on the English/p/–/l/ contrast.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, *61*(2), 206–219. http://doi.org/10.3758/BF03206883

Bradlow Pisoni, A. (1997). Training Japanese listeners to identify English /r/ and /l/: IV.ome effects of perceptual learning on speech production. *JASA*.

Chang, Y., Chang, R. Y., Lin, C.-Y., & Luo, X. (2016). Mandarin tone and vowel recognition in cochlear implant users: Effects of talker variability and bimodal hearing. *Ear and Hearing*, *37*(3), 271.

Fischer-Jorgensen, E. (1990). Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica*, *47*(3–4), 99–140.

Foxton Dean, Jennifer L., Gee, Rosemary, Peretz, Isabelle, and Griffths, Timothy D., J. M.

(2004). Characterization of deficits in pitch perception underlying “tone deafness.” *Brain*, *127*, 801–810. Journal Article.

Gandour, J. (1983). Tone perception in Far Eastern languages. *J Phon*, *11*, 49–175.

Gandour, J., & Harshman, R. (1978). Cross-language difference in tone perception: A multidimensional scaling investigation. *Lang Speech*, *21*, 1–33.

Gosselin, N., Jolicœur, P., & Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*(1), 270–272. JOUR. http://doi.org/10.1111/j.1749-6632.2009.04762.x

Green, K. P., Tomiak, G. R., & Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Attention, Perception, & Psychophysics*, *59*(5), 675–692.

Gu, F., Zhang, C., Hu, A., & Zhao, G. (2013). Left hemisphere lateralization for lexical and acoustic pitch processing in Cantonese speakers as revealed by mismatch negativity. *NeuroImage*, *83*(0), 637–645. Journal Article. http://doi.org/http://dx.doi.org/10.1016/j.neuroimage.2013.02.080

Hari, R., & Renvall, H. (2001). Impaired processing of rapid stimulus sequences in dyslexia. *Trends in Cognitive Sciences*, *5*(12), 525–532. http://doi.org/10.1016/S1364-6613(00)01801-5

Huang, W. T., Liu, C., Dong, Q., & Nan, Y. (2015). *Categorical perception of lexical tones in Mandarin-speaking congenital amusics*. *Frontiers in Psychology*. http://doi.org/10.3389/fpsyg.2015.00829

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, *15*(5), 356–360. http://doi.org/10.1111/j.0956-7976.2004.00683.x

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., Chen, X., & Yang, Y. (2012). Amusia results in

abnormal brain activity following inappropriate intonation during speech comprehension. *PLoS ONE*, *7*(7), e41411. http://doi.org/10.1371/journal.pone.0041411

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. http://doi.org/http://dx.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. *Memory & Cognition*, *40*(7), 1109–21. http://doi.org/10.3758/s13421-012-0208-2

Johnson, K. (2005). Speaker normalization in speech perception. In D. B. Pisoni and Remez, Robert E. (Ed.), *The handbook of speech perception* (pp. 363–389). Blackwell Publishing.

Jones, J. L., Lucker, J., Zalewski, C., Brewer, C., & Drayna, D. (2009). Phonological processing in adults with deficits in musical pitch recognition. *Journal of Communication Disorders*, *42*(3), 226–234. http://doi.org/10.1016/j.jcomdis.2009.01.001

Jones, J. L., Zalewski, C., Brewer, C., Lucker, J., & Drayna, D. (2009). Widespread auditory deficits in tune deafness. *Ear and Hearing*, *30*(1), 63.

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

Lee, C.-Y., Tao, L., & Bond, Z. S. (2008). Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics*, *36*(4), 537–563. http://doi.org/10.1016/j.wocn.2008.01.002

Lee, C.-Y., Tao, L., & Bond, Z. S. (2010). Identification of multi-speaker Mandarin tones in noise by native and non-native listeners. *Speech Communication*, *52*(11鈥?2), 900–910.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, M.A.: MIT Press.

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. http://doi.org/10.18637/jss.v069.i01

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. *The Journal of the Acoustical Society of America*, *140*(1), 563–575. http://doi.org/10.1121/1.4955182

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. *PLoS ONE*, *7*(2), e30374. http://doi.org/10.1371/journal.pone.0030374

Liu, F., Maggu, A. R., Lau, J. C. Y., & Wong, P. C. M. (2015). Brainstem encoding of speech and musical stimuli in congenital amusia: Evidence from Cantonese speakers. *Frontiers in Human Neuroscience*, *8*(1029). article. http://doi.org/10.3389/fnhum.2014.01029

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain*, *133*(6), 1682–1693. http://doi.org/10.1093/brain/awq089

Lu, X., Ho, H. T., Liu, F., Wu, D., & Thompson, W. F. (2015). Intonation processing deficits of emotional words among Mandarin Chinese speakers with congenital amusia: an ERP study. *Frontiers in Psychology*, *6*, 385. http://doi.org/10.3389/fpsyg.2015.00385

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology:*

Human Perception and Performance, *33*(2), 391–409.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 676.

Matthews, S., & Yip, V. (2013). *Cantonese: A Comprehensive Grammar*. Routledge.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, *47*(4), 379–390.

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. http://doi.org/10.1093/brain/awq178

Nguyen, S., Tillmann, B., Gosselin, N., & Peretz, I. (2009). Tonal language processing in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*(1), 490–493. JOUR. http://doi.org/10.1111/j.1749-6632.2009.04855.x

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

Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition*, *59*(3), 310–313. JOUR. http://doi.org/http://dx.doi.org/10.1016/j.bandc.2004.10.003

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics*, *34*(1), 135–154.

Peng, G., Zhang, C., Zheng, H.-Y., Minett, J. W., & Wang, W. S.-Y. (2012). The effect of inter-talker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems.

J Speech Lang Hear Res, *55*(2), 579–595. http://doi.org/10.1044/1092-4388(2011/11-0025)

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. http://doi.org/http://dx.doi.org/10.1016/S0896-6273(01)00580-3

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. *Annals of the New York Academy of Sciences*, *999*(1), 58–75.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, *7*(8), 362–367. http://doi.org/http://dx.doi.org/10.1016/S1364-6613(03)00150-5

Peretz, I., & Vuvan, D. T. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics*.

Perrachione Pierrehumbert, Janet B., and Wong, Patrick C. M., T. K. (2009). Differential neural contributions to native- and foreign-language talker identification. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1950–1960.

Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication*, *13*(1–2), 109–125. http://doi.org/10.1016/0167-6393(93)90063-Q

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America*, *119*, 1684.

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.

Shu, H., Peng, H., & McBride-Chang, C. (2008). Phonological awareness in young Chinese children. *Developmental Science*, *11*(1), 171–181. http://doi.org/10.1111/j.1467-7687.2007.00654.x

Sjerps, M. J., Zhang, C., & Peng, G. (2017). Lexical tone is perceived relative to locally surrounding context, vowel quality to preceding context. *Journal of Experimental Psychology: Human Perception and Performance*, No Pagination Specified-No Pagination Specified. http://doi.org/10.1037/xhp0000504

SPSS, I. B. M. (2011). IBM SPSS statistics for Windows, version 20.0. *New York: IBM Corp*.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466. http://doi.org/10.1016/j.wocn.2010.09.001

Sun, Y., Lu, X., Ho, H. T., & Thompson, W. F. (2017a). Pitch discrimination associated with phonological awareness: Evidence from congenital amusia. *Scientific Reports*, *7*(March), 44285. http://doi.org/10.1038/srep44285

Sun, Y., Lu, X., Ho, H. T., & Thompson, W. F. (2017b). Pitch discrimination associated with phonological awareness: Evidence from congenital amusia. *Scientific Reports*, *7*(March), 44285. http://doi.org/10.1038/srep44285

Team, R. C. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(46), 19027–19032. http://doi.org/10.1073/pnas.1210344109

Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., & Peretz, I. (2011). Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. *Frontiers in Psychology*, *2*(JUN). http://doi.org/10.3389/fpsyg.2011.00120

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

Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition*, *71*(3), 259–264. JOUR. http://doi.org/http://dx.doi.org/10.1016/j.bandc.2009.08.003

Wang, X., & Peng, G. (2014). Phonological processing in Mandarin speakers with congenital amusia. *Journal of the Acoustical Society of America*, *136*(6), 3360–3370.

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory*, *18*(6), 657–669. JOUR. http://doi.org/10.1080/09658211.2010.501339

Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research*, *46*(2), 413–421.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1087–1104.

Zeileis, A., & Hothorn, O. (2002). Diagnostic checking in regression relationships. *R News*, *2*(3), 7–10.

Zhang, C. (2018). Online adjustment of phonetic expectation of lexical tones to accommodate

speaker variation: a combined behavioural and ERP study. *Language, Cognition and Neuroscience*, *33*(2), 175–195. http://doi.org/10.1080/23273798.2017.1376752

Zhang, C., & Chen, S. (2016). Toward an integrative model of talker normalization. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(8), 1252–1268.

Zhang, C., Peng, G., Shao, J., & Wang, W. S. Y. (2017). Neural bases of congenital amusia in tonal language speakers. *Neuropsychologia*, *97*(July 2016), 18–28. http://doi.org/10.1016/j.neuropsychologia.2017.01.033

Zhang, C., Peng, G., & Wang, W. S.-Y. (2012). Unequal effects of speech and nonspeech contexts on the perceptual normalization of Cantonese level tones. *Journal of the Acoustical Society of America*, *132*(2), 1088–1099.

Zhang, C., Peng, G., & Wang, W. S.-Y. (2013). Achieving constancy in spoken word identification: time course of talker normalization. *Brain and Language*, *126*(2), 193–202. http://doi.org/10.1016/j.bandl.2013.05.010

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*, 536–549. http://doi.org/http://dx.doi.org/10.1016/j.neuroimage.2015.08.064

Zhang, C., Shao, J., & Huang, X. (2017). Deficits of congenital amusia beyond pitch: Evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics. *PloS One*, *12*(8), e0183151.

Supplemental file 1 (S1) shows the percentage of correct and incorrect trials in each condition.

Supplemental file 2 (S2) reports the RT results separated by correct and incorrect trials.

Figure 1. F0 curves of the six Cantonese tones used as stimuli in the talker variation condition.

Figure 2. F0 curves of the six Cantonese tones used as stimuli in the syllable variation condition.

Figure 3. Results in the talker variation condition for the amusic and control groups in low and high variation conditions. (a) Identification accuracy. The dotted line indicates the chance level accuracy (1/6 = 0.167; there were six tone choices in the identification task); (b) Identification RT combining correct and incorrect trials; (c) Identification RT of correct trials only; (d) Identification RT of incorrect trials only; (e) Discrimination sensitivity index d'; (f) Discrimination RT combining correct and incorrect trials; (g) Discrimination RT of correct trials only; (h) Discrimination RT of incorrect trials only. HV = high variation; LV = low variation.

Figure 4. Results in the syllable variation condition for the amusic and control groups in low and high variation conditions. (a) Identification accuracy. The dotted line indicates the chance level accuracy (1/6 = 0.167; there were six tone choices in the identification task); (b) Identification RT combining correct and incorrect trials; (c) Identification RT of correct trials only; (d) Identification RT of incorrect trials only; (e) Discrimination sensitivity index d'; (f) Discrimination RT combining correct and incorrect trials; (g) Discrimination RT of correct trials only; (h) Discrimination RT of incorrect trials only. HV = high variation; LV = low variation.

Table 1. Demographic characteristics of the amusic and control participants.

Subject information	Amusics	Controls
No. of participants	16 (8 M, 8 F)	16 (8 M, 8 F)
Age (range)	22.35 ± 2.8 years (19.1-27.5 years)	22.5 ± 3.1 years (18.7-28.5 years)
MBEA (SD)		
Scale	50.8 (17.7)	90.4 (5.6)
Contour	58.4 (19.6)	93.5 (4.9)
Interval	54.3 (18.1)	90.8 (4.3)
Rhythm	55.6 (15.0)	95.3 (3.6)
Meter	45.5 (10.4)	74.7 (14.3)
Memory	63.5 (23.3)	98.1 (2.9)
Global	54.7 (14.7)	90.5 (2.7)

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.

S1. Percentage of correct and incorrect trials in each condition.

Amusics	High variation		Low variation	
	Incorrect	Correct	Incorrect	Correct
Identification (talker)	59%	41%	59%	41%
Discrimination (talker)	27%	73%	20%	80%
Identification (syllable)	56%	44%	56%	44%
Discrimination (syllable)	46%	54%	24%	76%
Controls	**High variation**		**Low variation**	
	Incorrect	Correct	Incorrect	Correct
Identification (talker)	40%	60%	37%	63%
Discrimination (talker)	15%	85%	6%	94%
Identification (syllable)	40%	60%	35%	65%
Discrimination (syllable)	29%	71%	7%	93%

S2. RT results separated by correct and incorrect trials.

Talker variation condition

Identification task

For the identification task, no effects were significant for either correct or incorrect trials.

Discrimination task

Correct trials

There was only a significant main effect of *condition* ($\chi^2(1) = 8.702$, $p = 0.003$), where RT in the high variation condition was longer than that in the low variation condition.

Incorrect trials

No effects were statistically significant.

Syllable variation condition

Identification task

Correct trials

There were a main effect of *condition* ($\chi^2(1) = 14.437$, $p < 0.001$), and significant two-way interaction between *group* and *condition* ($\chi^2(1) = 8.769$, $p = 0.003$). Pairwise comparisons revealed that the effect of group was not significant in either condition. The RT in the high variation condition was longer than in the low variation condition in the control group ($z = 5.952$, $p < 0.001$), but not in the amusic group ($z = 1.238$, $p = 0.224$)

Incorrect trials

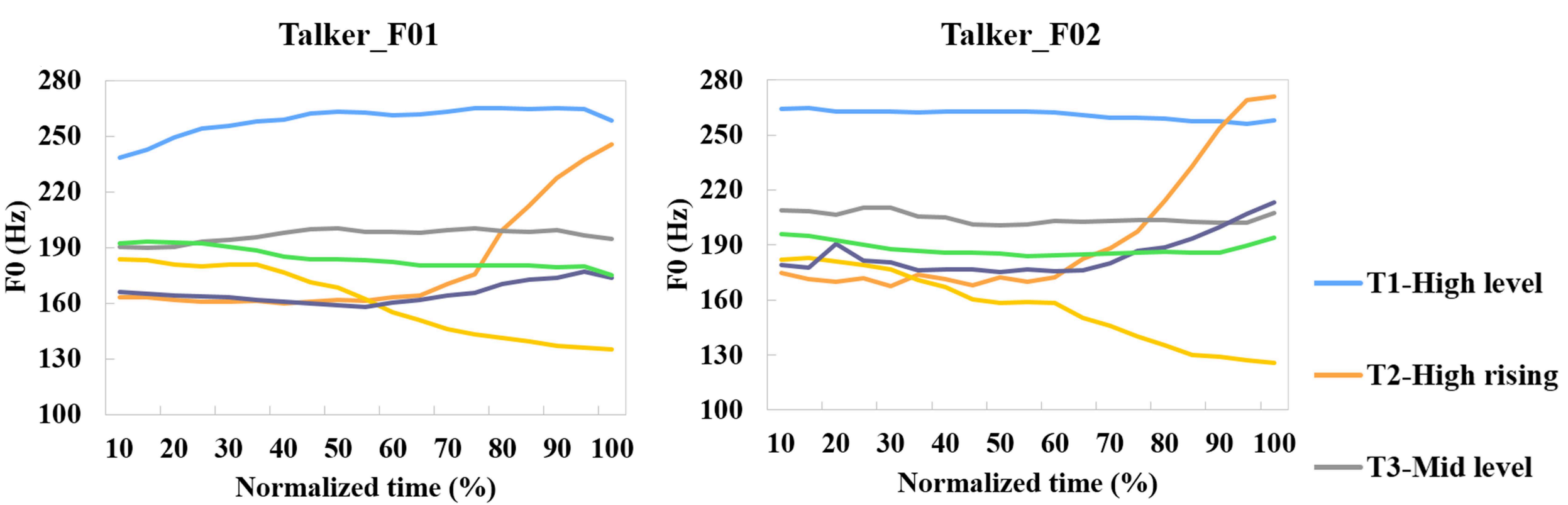
There was a significant main effect of *group* ($\chi^2(1) = 5.371$, $p < 0.001$), where the RT of the control group was longer than the amusic group. There was also a main effect of *condition* ($\chi^2(1) = 19.571$, $p < 0.001$), where the RT in the high variation condition was longer than that in the low variation condition.

Discrimination task

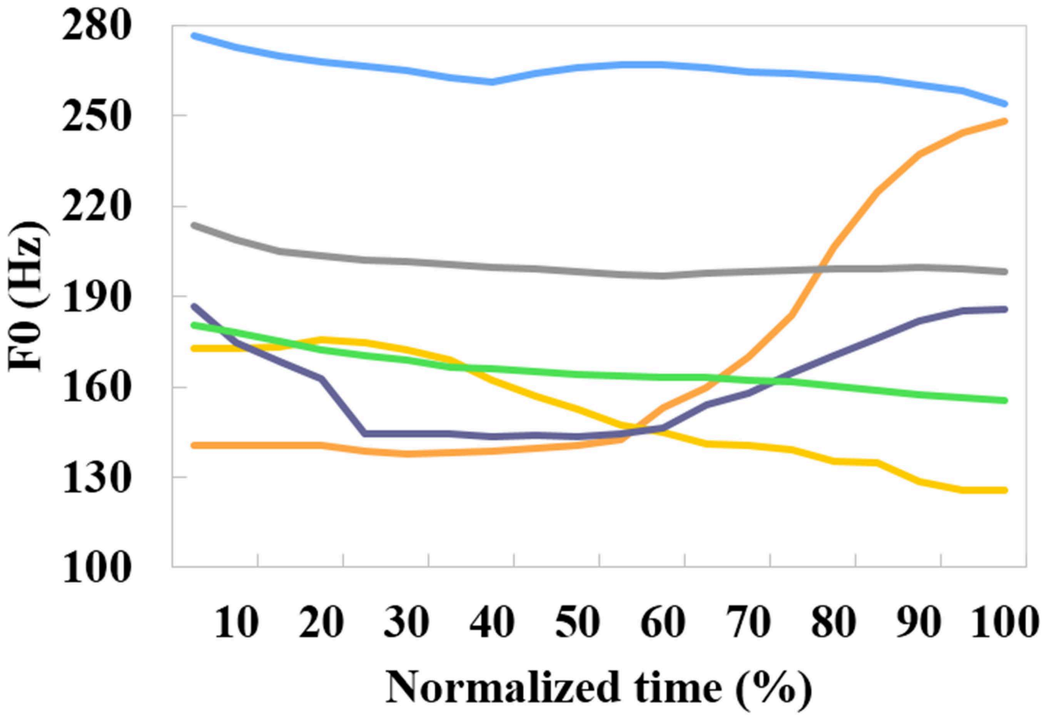
In the discrimination task, the RT of both correct and incorrect trials showed significant main effects of *condition* (correct trials: ($\chi^2(1) = 13.614$, $p < 0.001$); incorrect trials ($\chi^2(1) = 7.395$, $p < 0.001$), where the RT in the high variation condition was longer than that in the low variation condition. No other effects were significant.

Appendix. The four sets of syllables used in the syllable variation condition.

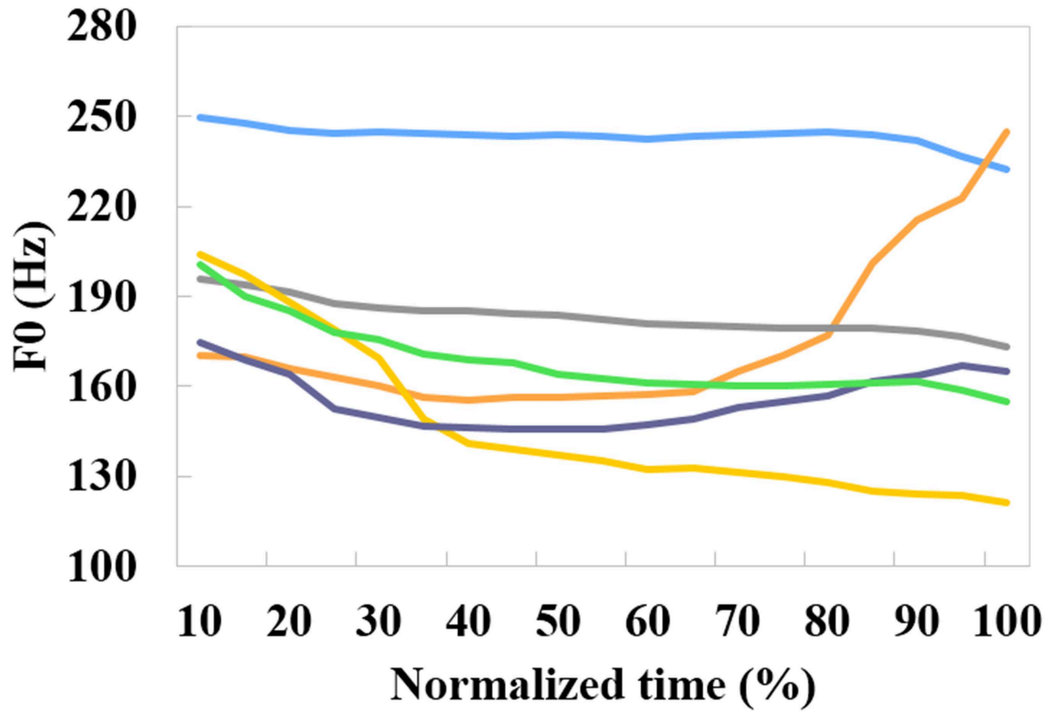
Tone	/ji/	/fɐn/	/fu/	/wɐi/
T1 high level /55/	醫 “doctor”	婚 “marriage”	夫 “husband”	威 “power”
T2 high rising /25/	椅 “chair”	粉 “pink”	父 “father”	委 “council”
T3 mid level /33/	意 “meaning”	訓 “train”	褲 “trousers”	餵 “feed”
T4 low falling /21/	兒 “son”	焚 “burn”	符 “symbol”	圍 “surround”
T5 low rising /23/	耳 “ear”	奮 “strive”	婦 “woman”	偉 “grand”
T6 low level /22/	二 “two”	份 “part”	負 “negative”	彗 “comet”


Talker_F01
Talker_F02
Talker_M01
Talker_M02
F0 (Hz)
Normalized time (%)
280
250
220
190
160
130
100
160
145
130
115
100
85
70
10
20
30
40
50
60
70
80
90
100
T1-High level
T2-High rising
T3-Mid level
T4-Low falling
T5-Low rising
T6-Low level

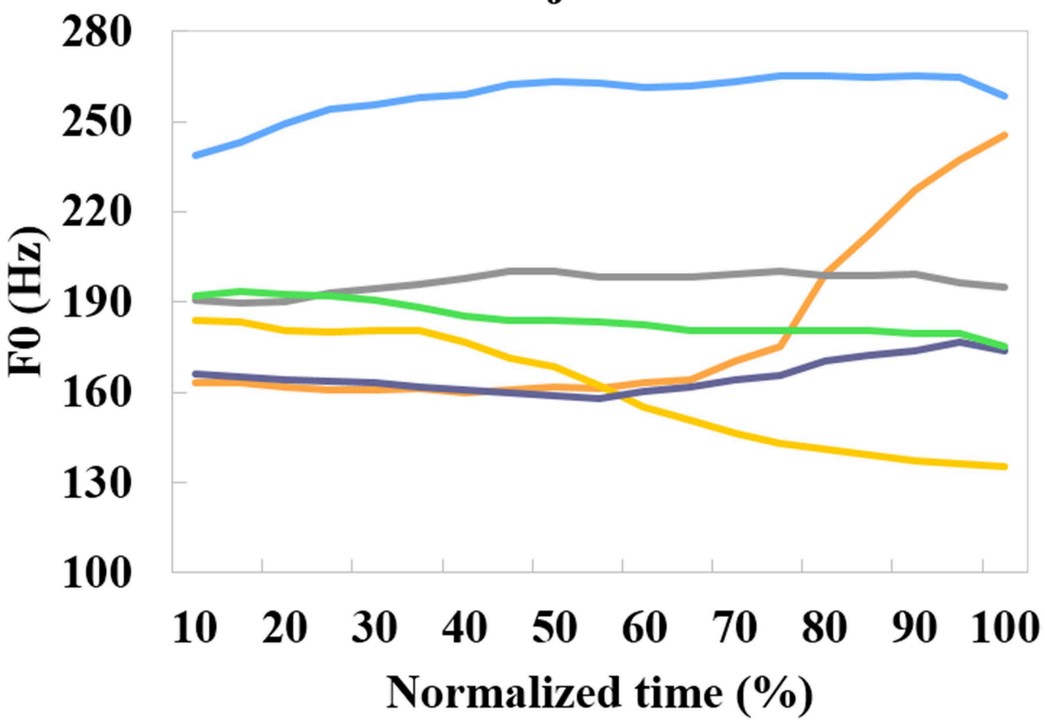
/fɐn/


280
250
220
190
160
130
100
F0 (Hz)
10 20 30 40 50 60 70 80 90 100
Normalized time (%)

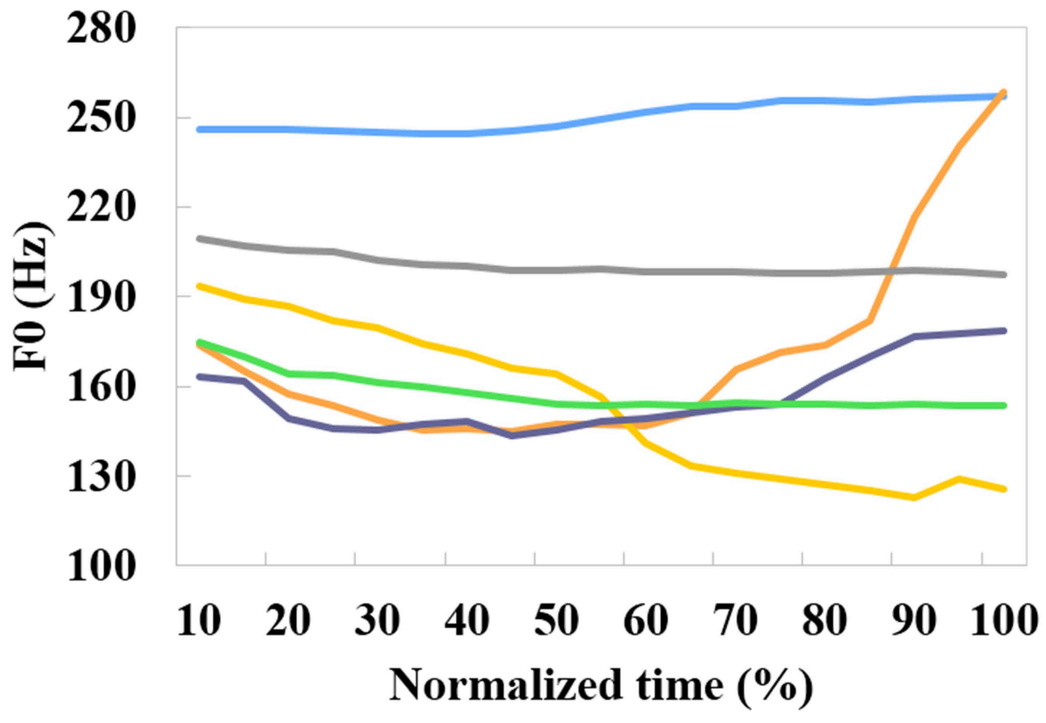
/fu/

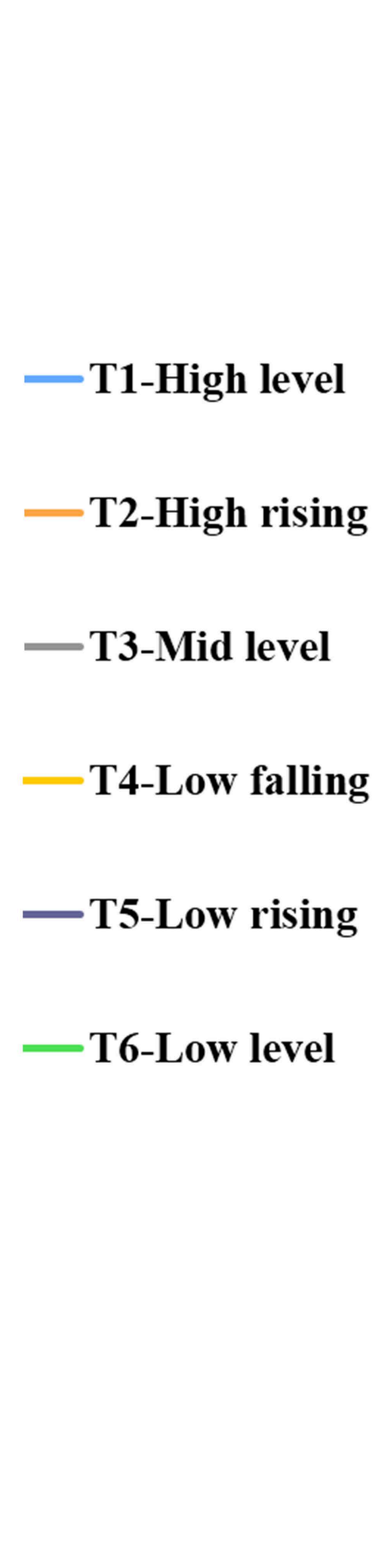

280
250
220
190
160
130
100
F0 (Hz)
10 20 30 40 50 60 70 80 90 100
Normalized time (%)

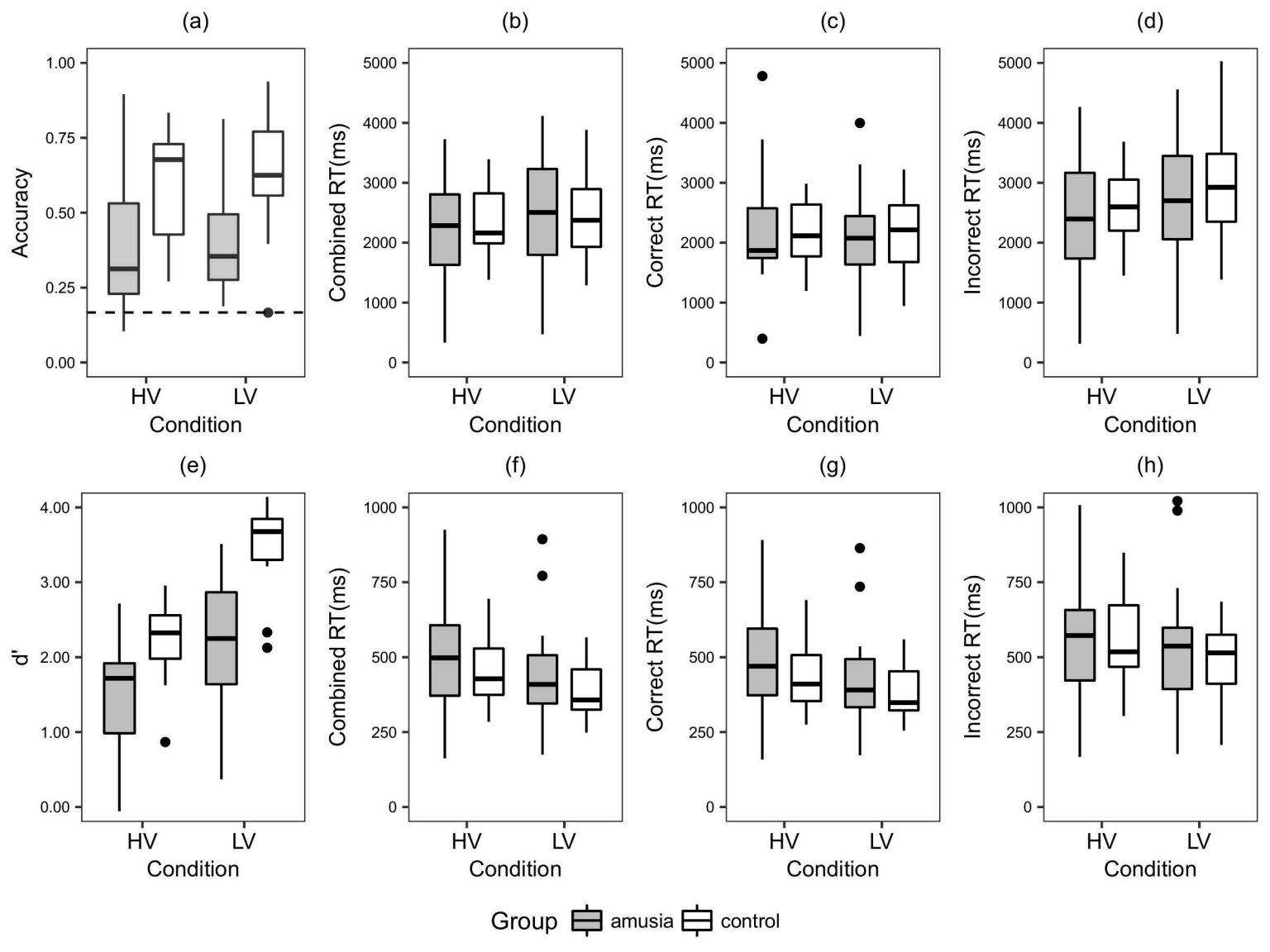
/ji/

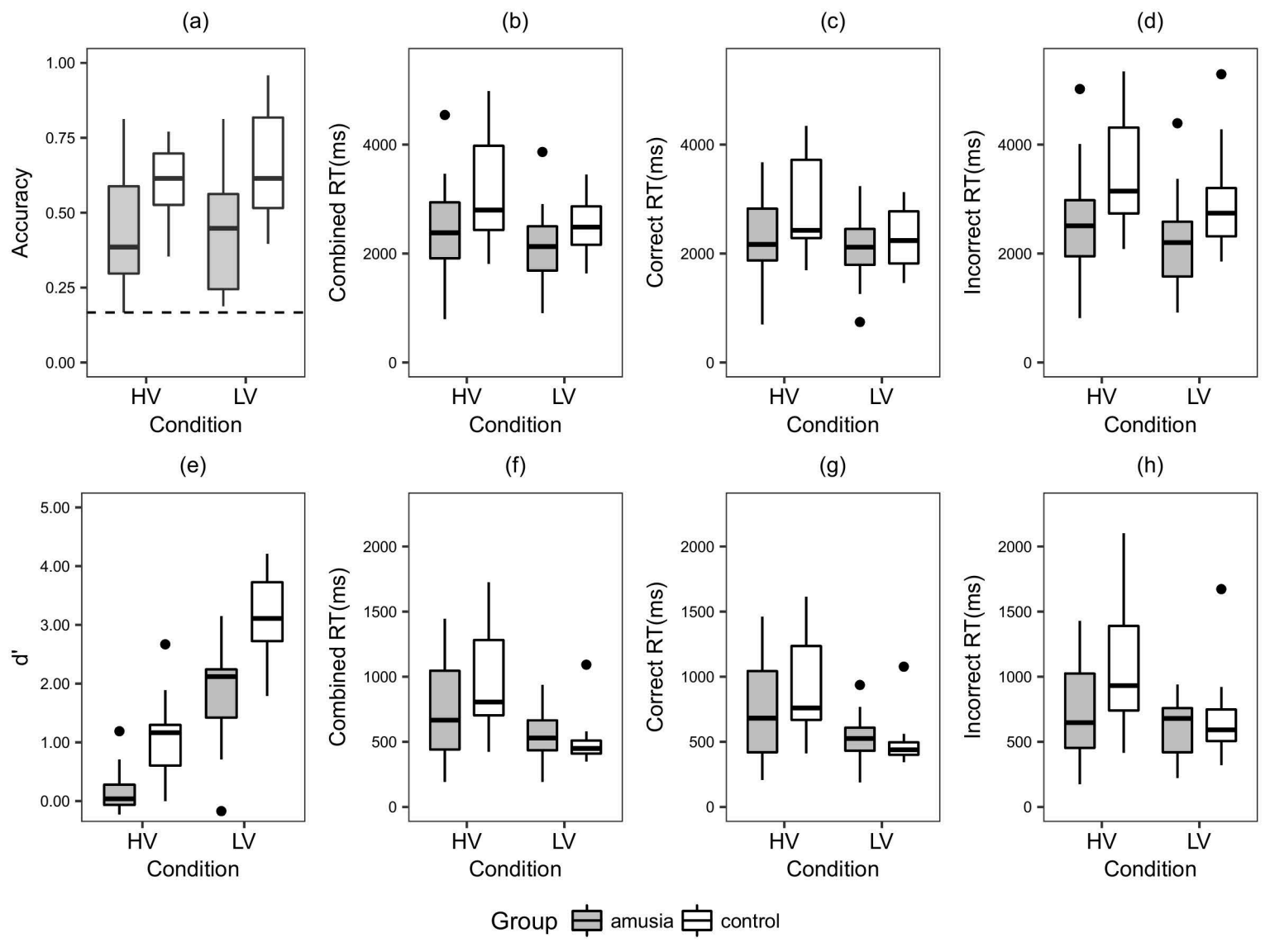

280
250
220
190
160
130
100
F0 (Hz)
10 20 30 40 50 60 70 80 90 100
Normalized time (%)

/wɐi/


280
250
220
190
160
130
100
F0 (Hz)
10 20 30 40 50 60 70 80 90 100
Normalized time (%)


T1-High level
T2-High rising
T3-Mid level
T4-Low falling
T5-Low rising
T6-Low level


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Accuracy
Combined RT(ms)
Correct RT(ms)
Incorrect RT(ms)
d'
HV
LV
Condition
Group
amusia
control


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Accuracy
Combined RT(ms)
Correct RT(ms)
Incorrect RT(ms)
d'
HV
LV
Condition
Group
amusia
control