ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Individuals With Congenital Amusia Show Degraded Speech Perception but Preserved Statistical Learning for Tone Languages

Jiaqiang Zhu,[a] Xiaoxiang Chen,[a] Fei Chen,[a] and Seth Wiener[b]

[a] College of Foreign Languages, Hunan University, Changsha, China [b] Department of Modern Languages, Carnegie Mellon University, Pittsburgh, PA

ARTICLE INFO

Article History:
Received July 13, 2021
Revision received August 30, 2021
Accepted September 8, 2021

Editor-in-Chief: Bharath Chandrasekaran
Editor: Stephanie Borrie

https://doi.org/10.1044/2021_JSLHR-21-00383

ABSTRACT

Purpose: Individuals with congenital amusia exhibit degraded speech perception. This study examined whether adult Chinese Mandarin listeners with amusia were still able to extract the statistical regularities of Mandarin speech sounds, despite their degraded speech perception.
Method: Using the gating paradigm with monosyllabic syllable–tone words, we tested 19 Mandarin-speaking amusics and 19 musically intact controls. Listeners heard increasingly longer fragments of the acoustic signal across eight duration-blocked gates. The stimuli varied in syllable token frequency and syllable–tone co-occurrence probability. The correct syllable–tone word, correct syllable-only, correct tone-only, and correct syllable–incorrect tone responses were compared respectively between the two groups using mixed-effects models.
Results: Amusics were less accurate than controls in terms of the correct word, correct syllable-only, and correct tone-only responses. Amusics, however, showed consistent patterns of top-down processing, as indicated by more accurate responses to high-frequency syllables, high-probability tones, and tone errors all in manners similar to those of the control listeners.
Conclusions: Amusics are able to learn syllable and tone statistical regularities from the language input. This extends previous work by showing that amusics can track phonological segment and pitch cues despite their degraded speech perception. The observed speech deficits in amusics are therefore not due to an abnormal statistical learning mechanism. These results support rehabilitation programs aimed at improving amusics' sensitivity to pitch.

Congenital amusia (amusia hereafter) is the condition in which individuals lack musical abilities common to most individuals, such as singing and recognizing music (Peretz, 2001). Amusia is an innate lifelong neurogenetic disorder present in an estimated 1.5%–4% of the general population (Kalmus & Fry, 1980; Peretz & Vuvan, 2017). Individuals with amusia (amusics hereafter) have difficulties perceiving fine-grained musical pitch differences, detecting mistuned melodies, and memorizing familiar tunes (Peretz et al., 2002, 2008). For many amusics, listening to a musical performance is akin to listening to foreign speech (Allen, 1878). In this study, we build on previous research related to amusics' perception of Mandarin speech—a language that requires pitch perception for lexical meaning—in order to examine to what degree amusics are able to track and make use of the statistical patterns of Mandarin speech sounds.

Correspondence to Xiaoxiang Chen: xiaoxiangchensophy@hotmail.com, and Fei Chen: chenfeianthony@gmail.com. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Amusia Affects Both Music Processing and Speech Processing

Early studies argued that amusia was domain specific and interfered with music processing alone (e.g.,

Ayotte et al., 2002; Hyde & Peretz, 2003; Peretz & Hyde, 2003). Ayotte et al. (2002) modified melodies by changing the pitch of one note by one semitone lower or higher and presented these melodies to amusics and typical listeners with normal hearing. Participants were asked whether a "wrong note" was contained in the melody. Amusics performed close to chance and well below typical listeners in terms of percentage of hits minus false alarms. The authors proposed that the observed impairments were not attributed to amusics' hearing loss, lack of musical exposure, or general cognitive slowing because of the matched sample demographics between amusics and typical listeners. The deficient music perception has been replicated by an array of studies (Loui et al., 2009; Peretz et al., 2008, 2009), with further results demonstrating that amusics are characterized as having elevated thresholds for pitch changes and directions (Foxton et al., 2004; Hyde & Peretz, 2004; F. Liu et al., 2015) and impaired short-term memory of pitch (Tillmann et al., 2015, 2016; Williamson & Stewart, 2010).

In contrast with these studies that argued amusia is domain specific, a growing body of recent studies has shown that amusia can affect other domains, particularly speech perception (e.g., Nan et al., 2010; Patel et al., 2005; Vuvan et al., 2015; C. Zhang, Peng, et al., 2017; C. Zhang, Shao, et al. 2017). For example, Nguyen et al. (2009) tested French-speaking amusics and typical listeners and found that amusics were less accurate in Chinese lexical tone discrimination relative to typical listeners. In the study by Jiang et al. (2012), native Mandarin-speaking amusics and controls were tested in their categorical perception of lexical tones. Interestingly, although amusics barely reported that they encountered difficulties in daily communication using lexical tones, they showed less accurate identification and discrimination of lexical tones as compared to the musically intact counterparts. Chen and Peng (2018) further tested categorical perception of tone by enlarging the pitch interval between stimulus steps for Mandarin-speaking amusics and typical listeners. Results demonstrated that amusics had a higher sensitivity (d') for between-category comparisons than for within-category comparisons, indicating that amusics largely preserved their capacities to perceive lexical tones categorically; nevertheless, amusics' performance was overall poorer than that of controls (i.e., a broader boundary width and a reduced between-category sensitivity). These results imply that amusics can use lexical tones as necessary to access their mental lexicon; however, their overall perception of tone is poorer than that of typical Mandarin listeners (Chen & Peng, 2020).

Amusics' reduced behavioral performance in laboratory-based studies stems from their degraded speech perception (F. Liu et al., 2021). Recent studies have shown that the effects of amusia are present not only at the suprasegmental level but also at the segmental level in speech. C. Zhang, Shao, et al. (2017) tested Cantonese-speaking amusics using various stimuli, including lexical tones, vowels, and stop consonants in order to explore whether amusics show deficits in the processing of segments in a manner similar to that of lexical tones. The results revealed that, in comparison to typical listeners, amusics exhibited poorer performance when listening to lexical tones and vowels but comparable performance when listening to stop consonants—which primarily involve processing the cue of voice onset time. Likewise, the degraded vowel perception in amusics was observed in two subsequent studies (Li et al., 2019; W. Tang et al., 2018). Besides, F. Liu et al. (2015) showed that, in spoken sentences with both natural pitch information and neutralized/flattened pitch information, amusics had a lower accuracy in comprehension tasks as compared to typical listeners. In summary, amusia affects speech processing of both suprasegments and segments, which require the perception of frequency or spectral information within the acoustic signal (C. Zhang, Shao, et al., 2017).

Possible Origins of the Impairments Across Domains in Amusia

There have been two prevailing theories to explain the deficits of amusia in the extant literature. First, adopting a meta-analytic approach, Vuvan et al. (2015) developed an acoustic account. In their meta-analysis, the authors analyzed the effect size from the amusia literature involving the performance gap between amusics and controls across music and language domains. The performance gap was indeed moderated by the size of the pitch shifts in the stimuli. Huang, Nan, et al. (2015) found that Chinese-speaking amusics showed significantly larger just-noticeable differences (JNDs) than the matched counterparts in lexical tone perception (see C. Liu, 2013, for discussion of JNDs of lexical tones). The elevated pitch thresholds are thus connected to amusics' poorer processing of pitch-related materials in both music and speech.

Researchers also consider amusia to involve short-term memory of pitch. Due to the sequential nature of stimulus presentation, most psychophysical paradigms require participants to memorize the stimulus of one presentation and compare it with another presentation (Tillmann et al., 2016). Consequently, amusics' deficits in pitch processing can be attributable to a weakened or faulty memory trace. For example, Gosselin et al. (2009) found that amusics' performance decreased when the to-be-remembered musical tone sequence was increased from three to five tones and the to-be-compared musical tones were inserted with irrelevant tones. Control listeners showed no such performance decrease. In a related study, Tillmann et al. (2014) made use of the gating paradigm (Grosjean, 1980, 1996)—the paradigm we use in this study—to examine

amusics' memory of music. In this paradigm, listeners are presented with increasingly longer fragments of a stimulus and asked to make a behavioral response. Tillmann et al. played amusics fragments of familiar and unfamiliar instrumental musical pieces and asked listeners to rate their familiarity with the perceived music. This allowed for an examination of whether amusics had encoded music in memory, in spite of their limited musical pitch processing. The results revealed that, for familiar musical pieces, amusics responded slower for longer fragments of musical pieces than the matched controls did, which indicated that amusics needed more time to access their long-term memory as compared to the matched controls. Nevertheless, the amusics performed consistently with the control listeners, suggesting that amusics store musical pieces in long-term memory.

In addition to the acoustic account and memory account, there is one more potential factor that may be related to amusics' learning capacities: the inability to track the statistics of a language's speech sounds. In other words, the reduced performance shown by amusics raises the question as to whether their innate statistical learning mechanism functions well given that they are afflicted with amusia (Peretz et al., 2012). This issue lies at the heart of this study. A large body of work by Saffran et al. (e.g., Saffran, 2003; Saffran et al., 1996) lends support for an innate statistical learning mechanism used to help acquire music and language. This account states that the statistical regularities present in the environment can be extracted by learners via mere exposure. Evidence for this claim has been found in infants acquiring their first language (L1) and even adults acquiring their second language (L2) (Saffran et al., 1999; T. Wang et al., 2020).

This line of research was examined by Peretz et al. (2012), who tested whether amusics could learn the statistical regularities involved in music and language. French-speaking adults with and without amusia were played three-syllable nonsense words or three-tone motifs with the identical statistical structure (i.e., the same transitional probability) and then tested on their learning. The results showed that, as compared to typical listeners, amusics could learn nonsense words but systematically failed in learning musical motifs. This suggests that amusics are unable to track the statistical information in the acoustic input of music but can track the statistical information in language. In a related study, Loui and Schlaug (2012) created a novel musical system for participants, known as the Bohlen–Pierce scale. Musical systems, in general, are based on the octave with a 2:1 frequency ratio, whereas the Bohlen–Pierce scale is based on the "tritave" with a 3:1 frequency ratio. The stimuli from the Bohlen–Pierce scale were presented auditorily to the listeners. The results showed that unlike typical listeners, amusics did not learn the frequency structure stemming from the new musical scale. Taken together, Peretz et al. and Loui and Schlaug indicate that amusics demonstrate a certain insensitivity to the statistical learning of musical pitch but may be able to learn the statistical patterns of speech.

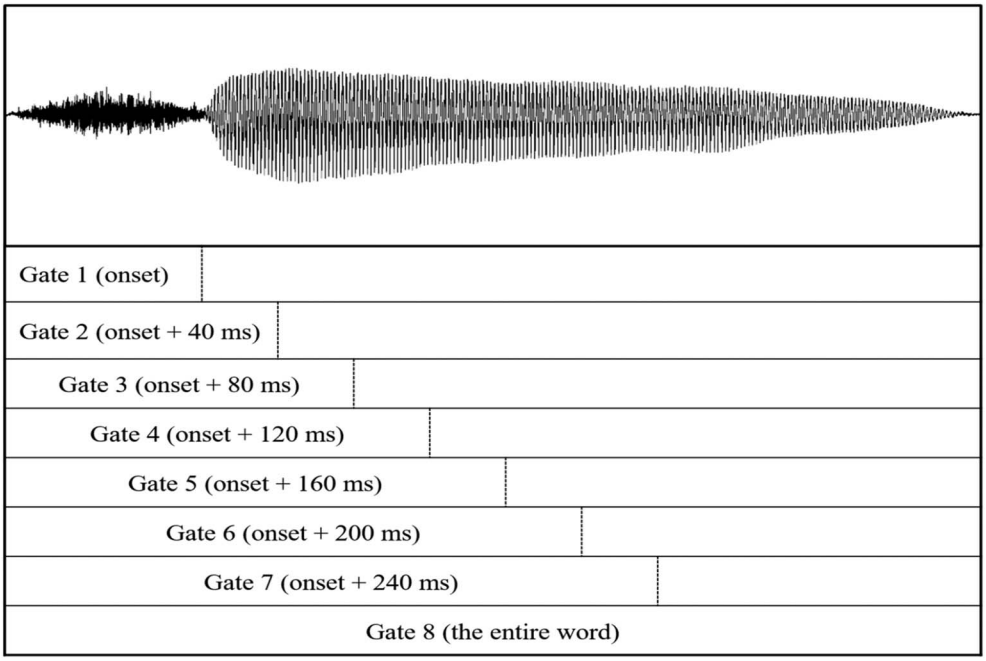
Notably, since the pitch information of the words in Peretz et al. (2012) and Omigie and Stewart (2011) did not involve listeners' phonological processing abilities, it remains unclear as to whether amusics' statistical learning mechanism involving lexical tones is preserved given amusics' impaired lexical tone perception. This study further explores amusics' abilities to compute the statistics of speech involving the co-occurrence of syllables with lexical tones. A growing body of research has shown that adult native and even nonnative L2 Mandarin listeners are sensitive to a variety of statistical information in the speech signal including syllable token frequency and syllable–tone co-occurrence probabilities (e.g., Lee & Wiener, 2020; Wiener & Ito, 2015, 2016; Wiener et al., 2018, 2021; Wiener & Turnbull, 2016). This type of statistical learning could be challenging for amusics because they need to simultaneously process frequency information (i.e., lexical tones) and probability information (i.e., syllable–tone combinations) as they track the statistical properties in the input of a tone language.

This Study

The stimuli tested in Omigie and Stewart (2011) or Peretz et al. (2012) were artificial words that did not contain phonologically or lexically meaningful pitch information. This study uses the tone language Mandarin Chinese in which pitch information carries a lexical role (Gandour, 1983; P. Tang et al., 2019). Standard Mandarin has four lexical tones that differ in pitch contour and pitch height: Tone 1 (level), Tone 2 (rising), Tone 3 (falling–rising), and Tone 4 (falling). Lexical tones are phonemically contrastive with the primary acoustic cue being fundamental frequency (F0; the physical correlate of pitch; Chao, 1965), with tonal variations representing different words (H. Zhang et al., 2020). For example, the monosyllable "ma" with Tone 1 represents "mother," but with Tone 4, it means "to scold." The occurrence of lexical tones in our auditory stimuli thus made it possible to examine the functioning of amusics' statistical learning when they processed segments with lexical tones, both of which have been found to be reduced in the perception of Mandarin speech relative to healthy adult listeners (Chen & Peng, 2018; Jiang et al., 2012; C. Zhang, Shao, et al., 2017).

This study involves presenting increasingly longer speech fragments or "gates" of a spoken Mandarin syllable–tone word and asking the listener to report the perceived word. Figure 1 shows how at each successive gate more acoustic information is presented to the listener. Gating, therefore, allows for an estimation of the amount of acoustic

Figure 1. The schematic illustration of the gating experiment as instantiated by the monosyllabic word "qi1." The first gate is the syllable onset, and the last gate contains the entire word, with the intermediate gates gradually increasing at the incremental size of 40 ms herein.



information needed to correctly identify segments, tones, and their combination as words. Importantly, in early gates when the acoustic information is limited or insufficient for correct identification, gating forces listeners to draw on their prior experience with the language and predict likely syllables, tones, and their combination, that is, draw on their knowledge of the statistical distribution of Mandarin speech. Wiener and Ito (2016) employed the gating paradigm to test native Mandarin listeners using monosyllabic words that consisted of high and low syllable token frequencies paired with a tone that was either most or least likely to co-occur with the syllable based on a spoken word corpus (SUBTLEX-CH: Cai & Brysbaert, 2010). Participants heard eight gates of increasing length beginning with the onset only (Gate 1), followed by 40-ms increments (Gates 2–7) until the full word (Gate 8). After each stimulus was played, participants were asked to type the perceived word using the *Pinyin* romanization system, which specified the syllable and tone number (e.g., "ma3"). Results indicated that participants identified high token–frequency syllables more accurately than low token–frequency syllables. Importantly, the authors analyzed responses that contained the correct syllable but incorrect tone, that is, enough acoustic information for segmental identification but not suprasegmental identification. In these errors, listeners could use either limited F0 to report acoustically similar tones (e.g., Tone 1 and Tone 4 both start with relatively high F0 onsets) or their knowledge of the most likely tone given the syllable (e.g., the syllable "kao" is most probable as "kao3" and least probable as "kao1"). The authors found that listeners responded with more probable tones on low-frequency syllables, given that these syllables tend to almost always occur with a high-probability tone. The authors argued that healthy adult listeners immediately made use of the statistical regularities (i.e., syllable frequency information and tonal probability information) at an early stage of word recognition in order to overcome fragmented speech.

To summarize, we designed the current gating study to examine whether amusics are able to extract the statistical regularities of monosyllabic syllable–tone words given their degraded speech perception as compared to that of typical individuals whose statistical learning mechanism remains intact. It was hypothesized that amusics would perform similarly to the matched controls as a result of their possibly spared implicit knowledge of the native language (Chen & Peng, 2018; Peretz et al., 2012; Tillmann et al., 2014), specifically with respect to syllable frequency information, but with a reduced accuracy in overall spoken word recognition due to amusics' degraded perception of lexical tones and vowels (W. Tang et al., 2018; Tillmann et al., 2011; C. Zhang, Shao, et al., 2017). It may also be the case that, in early gates, amusics may rely on syllable–tone co-occurrence probabilities to a larger degree than matched controls given their degraded pitch perception. That is, amusics may turn to greater knowledge-based processing of likely syllable–tone combinations to overcome their reduced acoustic-based processing of tone.

Method

Participants

Nineteen Mandarin-speaking amusic participants and 19 musically intact listeners were recruited. Initially, each group encompassed 20 participants; however, two were dropped from the experiment as a result of their unavailability. Because amusics are relatively rare in the general population (Peretz & Vuvan, 2017), over 300 individuals from the student body in universities of Mainland China were screened based on their own reports experiencing music in their daily routines (Hyde & Peretz, 2004). These individuals were then tested using the online Montréal Battery of Evaluation of Amusia without repeat under the experimenter's supervision, with amusics identified using the cutoff average score set to 71% (Peretz et al., 2008). The online diagnostic protocols have been widely adopted among tone language speakers in the extant literature (Chen & Peng, 2018, 2020; Shao & Zhang, 2018; X. Wang & Peng, 2014; Wong et al., 2012; C. Zhang, Peng, et al., 2017; C. Zhang, Shao, et al. 2017), which consist of three subtests measuring listeners' abilities of pitch processing (out-of-key and mistuned subtests) and rhythm processing (offbeat subtest). The mean global accuracy was 65.26 (3.21) for amusics and 88.21 (3.38) for typical listeners, with standard deviation (*SD*) presenting in parentheses. The results of the independent-samples *t* tests further confirmed that both the global score and the scores of the composite subtests were significantly lower for amusic listeners than those for normal participants (*p*s < .001). Although our amusics were unaware that they had amusia, they did report having difficulties in singing instead of in speech, which is in line with previous studies showing mild and severe impairments in speech and music, respectively (F. Liu et al., 2021; Peretz et al., 2008; Tillmann et al., 2015).

The amusic group (M_{age} = 19.47 ± 0.96 years, range: 18–21, 12 women) and the control group (M_{age} = 19.42 ± 0.90 years, range: 18–22, 11 women) were all nonmusicians and right-handed according to a handedness questionnaire adapted from a modified Chinese version of the Edinburgh Handedness Inventory (Oldfield, 1971). In addition, the digit span tests in either forward or reverse order were used to index participants' working memory, which were derived from Wechsler Adult Intelligence Scale–Revised by China (Gong, 1992). The amusic group did not differ from the nonamusic group in working memory or age based on the independent-samples *t* tests (both *p*s > .05), confirming that the two groups had comparable general memory (Tillmann et al., 2016; Williamson & Stewart, 2010; Yang et al., 2014). Because listeners needed to type their responses for the auditory stimuli, a short interview with each participant was carried out prior to testing. All participants were confirmed to be native speakers of Mandarin Chinese, had learned the *Pinyin* system and the four lexical tones very early in life, and thus had the relevant knowledge of Mandarin's syllable–tone combinations. None of the participants reported having hearing loss, brain injuries, or neurological disabilities among themselves or their family members. The two groups in this study were matched in sex, age, education, handedness, musical training background, and working memory (see Table 1). The experiment was approved by the ethics review board at the College of Foreign Languages of Hunan University. Each participant gave the written consent and was paid for their participation.

Stimuli

The auditory stimuli in this study were adapted from previous studies by Wiener et al. (Wiener & Ito, 2015, 2016; Wiener et al., 2019). All experimental items were legal Chinese monosyllabic words in line with W. S. Y. Wang's (1973) analysis of Mandarin phonology. As Figure 2 shows, a Mandarin monosyllabic word requires both the syllable and lexical tone; responses contradicting the structure were judged to be illegal. Two statistical characteristics of the stimuli were manipulated: syllable token frequency and syllable–tone co-occurrence probability. Following Wiener and Ito (2015, 2016), all token occurrences of a particular syllable (irrespective of tone) in SUBTLEX-CH (Cai & Brysbaert, 2010) were computed with the median common log frequency being 4.4; next, syllables with a log frequency above 4.8 were considered high token frequency (F+), whereas those below 4.0 were considered low token frequency (F−). In total, 24 syllables were chosen: 12 F+ with a mean log frequency at 5.08 and 12 F− with a mean log frequency at 3.84.

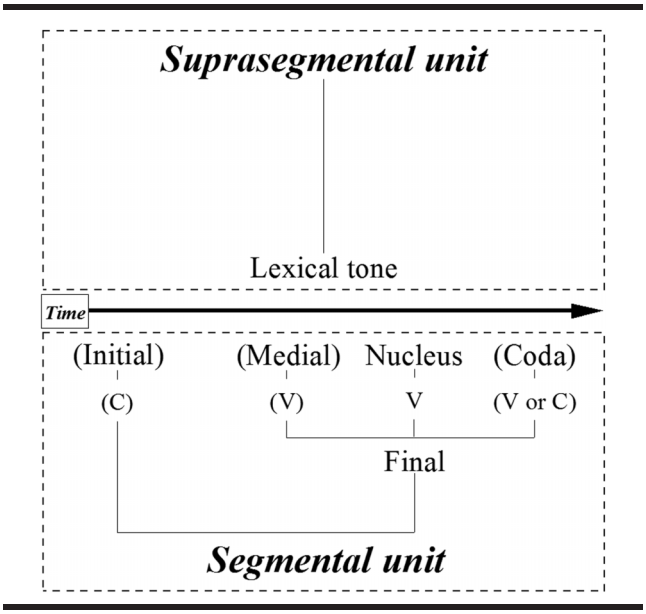
Tonal probability was calculated by dividing the token count for a given syllable–tone word by the token count of that particular syllable irrespective of tone; subsequently, the most probable (P+) tone was identified if it emerged in more

Table 1. Demographic characteristics of amusics and controls.

Subject information	Amusics	Controls
No. of participants (M/F)	19 (7/12)	19 (8/11)
Age in years, *M* (*SD*)	19.47 (0.96)	19.42 (0.90)
Working memory, *M* (*SD*)	13.95 (0.97)	14.53 (0.96)
Online identification test of congenital amusia, *M* (*SD*)		
Out of key	68.84 (6.17)	91.68 (6.47)
Offbeat	71.37 (10.75)	83.42 (8.09)
Mistuned	54.58 (6.51)	89.11 (7.79)
Global score	65.26 (3.21)	88.21 (3.38)

Note. M = male; F = female.

Figure 2. The structure of the monosyllabic word in Mandarin Chinese with optional elements contained in parentheses (Chen et al., 2017; Gandour, 1983; W. S. Y. Wang, 1973). V = vowel; C = consonant.



than 50% of the syllables' utterances in the SUBTLEX-CH corpus, whereas the least probable (P−) tone occurred in less than 20% of the syllables' utterances in the corpus. Thus, for each of the 24 syllables, the P+ and P− tones were respectively assigned to each syllable resulting in 48 words.

The 48 items were recorded by a female native speaker from Beijing, China, at 44.1 kHz in a soundproof booth (see the Appendix). Five additional native speakers who did not take part in the current experiment confirmed the naturalness of the words and their constituent segments and lexical tones with 100% agreement. Each of these words was fragmented into eight gates using Praat (Boersma & Weenink, 2018). The first gate (Gate 1) encompassed the consonant onset up to the beginning of the first regular periodicity of the vowel. The last gate (Gate 8) contained the full word. Gates 2–7 were developed with six 40-ms gradual increments on the rime (see Figure 1). In summary, there were 384 auditory stimuli in this study (48 words × 8 gates).

Procedure

Participants were instructed that they would listen to Chinese words, which could be either fragments or the full syllable–tone combinations. This was done to avoid the possible floor effect especially in early gates. Although all participants were interviewed prior to testing, an additional pretest training took place. This training involved clarifying that the four lexical tones were represented as the numbers 1, 2, 3, and 4 (hence, listeners felt comfortable typing the tones using the keyboard), along with the initials and finals separately, which could be combined to form Chinese monosyllabic words (see Figure 2). The training materials involved paper handouts that listed the Mandarin speech elements with no syllable–tone combinations given. The listeners were told that they could move to the next phase after they reviewed these speech elements at their own pace. All participants finished their training within 15 min.

In the testing phase, each participant was seated in a quiet room and wore headphones to listen to the stimuli, which were presented randomly at each duration-blocked gate using E-Prime 2.0 (Psychology Software Tools, n.d.). The order of gates was fixed from the shortest gate to the longest one (i.e., from Gate 1 to Gate 8). Participants were asked to type Chinese *Pinyin* and the tone number such as "wu2." Chinese characters were avoided because there are polyphonic characters with multiple pronunciations, including different tones (e.g., "奇" has two different syllable–tone combinations of "ji1" and "qi2"), which could contaminate the final results (see Wiener & Ito, 2016, for discussion). Participants could familiarize themselves with the procedure by doing practice trials prior to the experiment. The practice trials did not contain the words presented in the main experiment. In total, the experiment lasted approximately 40 min.

Data Analysis

Responses containing illegal syllable–tone combinations (nonwords, e.g., "sho3" and "gong2") were excluded (2%, cf. 4% in Wiener & Ito, 2016). Given the relatively limited number of participants tested, we followed the approach by Wiener et al. (Wiener & Ito, 2015, 2016; Wiener & Lee, 2020; Wiener et al., 2019) to increase statistical power by combining responses from two consecutive gates into one "window" (i.e., Window 1 by Gates 1 and 2, Window 2 by Gates 3 and 4, Window 3 by Gates 5 and 6, and Window 4 by Gates 7 and 8). In other words, both patterns of accuracy and correct syllable–incorrect tone errors were analyzed across four windows, with each window containing data from two consecutive gates.

First, we examined the accuracy of the syllable–tone word (e.g., the response "wu2" to the stimulus "wu2"), syllable-only (e.g., "da3" to "da2"), and tone-only (e.g., "gong4" to "hong4") responses. Response to each trial was coded as 0 or 1 (incorrect or correct, respectively) for each participant. Second, we examined correct syllable–incorrect tone errors. Following Wiener and Ito (2016), these errors were classified as either acoustic-based errors or probability-based errors. Acoustic-based errors stemmed from reporting an acoustically similar tone given the two tones' similar F0 onsets. For example, Tones 2 and 3 start in the low register and confuse listeners due to their initial acoustic ambiguity (Moore & Jongman, 1997). As an example, the response "shou2" to

the stimulus "shou3" was coded as an acoustic-based error. Probability-based errors stemmed from reporting the statistically more probable tone given the perceived syllable. These errors occurred despite the two tones' acoustic dissimilarity. For instance, the response "tie3" to the stimulus "tie1" was regarded as a probability-based error given that these tones started in opposite registers and were not acoustically similar. Notably, there existed responses that did not match these two types of error (e.g., "zhu3" to "zhu4") or happened to be the case of belonging to these two types of error simultaneously (e.g., "bin1" to "bin4"). These responses were excluded (3%). The empirical log of the error ratio was calculated with the formula (log[(probability error + 0.5)/(acoustic error + 0.5)]) for each participant at each gate as a function of syllable frequency (i.e., F+ and F−). A positive log ratio indicated that an individual made more probability-based errors than acoustic-based errors, whereas a negative value suggested that a listener made more acoustic-based errors than probability-based errors.

All statistical analyses were conducted using generalized (accuracy) or linear (error log ratio) mixed-effects models with the lme4 package (Bates et al., 2015) in R (R Core Team, 2014). Specifically, for the analysis of accuracy, the models based on the four windows were built with group (amusics and controls), syllable frequency (F+ and F−), and tonal probability (P+ and P−) acting as fixed factors; for the analysis of error log ratio, the models were built with group (amusics and controls), tonal probability (P+ and P−), and window (Windows 1, 2, 3, and 4) acting as fixed factors. Two-way and three-way interaction terms were also included as fixed effects in the models. When fitting the models, working memory was always regarded as the controlled covariate. By-subject and by-item random intercepts and slopes for all possible fixed factors were included in the initial model (Barr et al., 2013). For both accuracy and error analyses, a simplified model that excluded a specific fixed factor was compared to the initial model using the analysis of variance function in lmerTest package (Kuznetsova et al., 2017). Pairwise comparisons were computed with Tukey adjustment using the lsmeans package (Lenth, 2016).

Results

Accuracy Results

Figure 3 depicts the mean syllable–tone word accuracy, syllable accuracy, and tone accuracy by amusics and typical listeners at different gates, with error bars showing 1 standard error (*SE*). Both groups responded more accurately as more acoustic information was available. Tones were most accurately identified, followed by syllables and complete syllable–tone combinations.

Correct Syllable–Tone Combinations

Figure 4 plots the mean correct syllable–tone responses among amusics and controls faceted by syllable frequency and tonal probability at different windows. Both groups exhibited a trend of rising accuracy when exposed to more of the acoustic signal. The mean accuracy by

Figure 3. Mean correct tone, syllable, and syllable–tone responses by amusics and typical listeners (error bar = ±1 *SE*). ST Combinations = syllable–tone combinations.

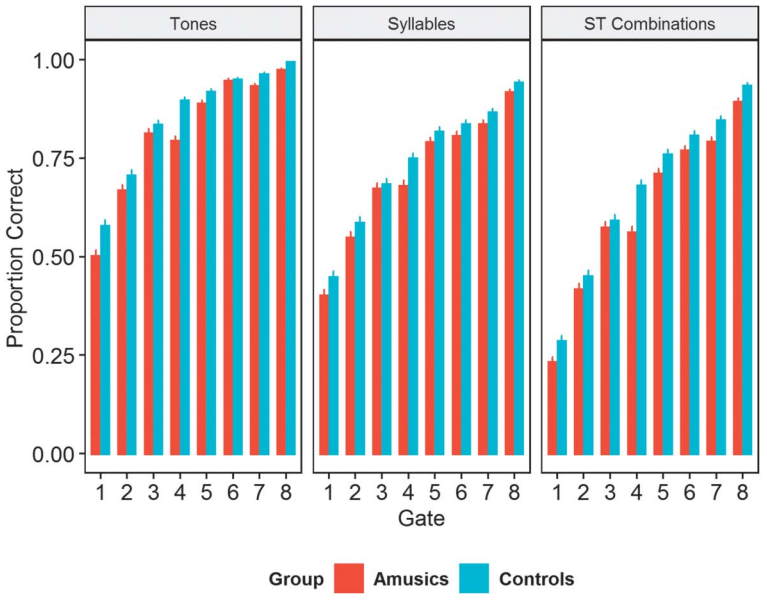
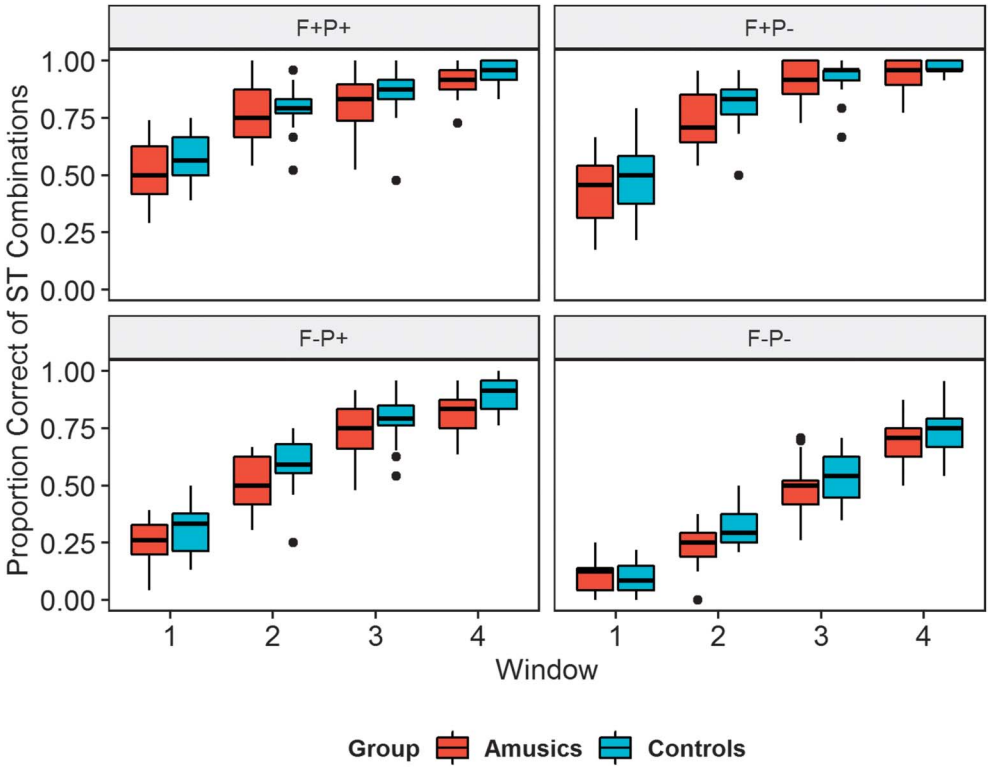


Figure 4. Mean correct syllable–tone responses by syllable frequency and tonal probability by amusics and typical listeners at each window. ST Combinations = syllable–tone combinations.



amusics was slightly lower than that of the matched counterparts. The mixed-effects models revealed that, at Window 1, there was a significant effect of syllable frequency, $\chi^2(1) = 21.78$, $p < .001$, indicating that both groups identified high token–frequency (F+) syllables more accurately than low token–frequency (F−) syllables regardless of tonal probability. Additionally, a significant effect of tonal probability, $\chi^2(1) = 5.16$, $p < .05$, and a significant two-way interaction between group and tonal probability, $\chi^2(1) = 3.96$, $p < .05$, were found at Window 1. Further analysis of this interaction revealed that, in comparison to controls ($M = 0.45$, $SD = 0.50$), amusics ($M = 0.38$, $SD = 0.49$) had marginally lower accuracy ($\beta = -0.39$, $SE = 0.21$, $t = -1.88$, $p = .06$) when the syllables carried low-probability (P−) tones. Controls more accurately identified syllables with high-probability (P+) tones than syllables with P− tones ($\beta = 1.20$, $SE = 0.44$, $t = 2.73$, $p < .01$). Amusics exhibited a similar trend for tonal probability ($\beta = 0.82$, $SE = 0.44$, $t = 1.87$, $p = .06$).

At Window 2, there were significant main effects of group, $\chi^2(1) = 4.49$, $p < .05$, and syllable frequency, $\chi^2(1) = 19.77$, $p < .001$, which revealed that, independent of tonal probability, amusics ($M = 0.57$, $SD = 0.50$) had a lower accuracy than controls ($M = 0.63$, $SD = 0.48$), and F+ syllables were more accurately recognized than F− syllables for both groups. At Window 3, significant main effects of group, $\chi^2(1) = 3.88$, $p < .05$ (amusics: $M = 0.74$, $SD = 0.44$; controls: $M = 0.78$, $SD = 0.41$), and syllable frequency, $\chi^2(1) = 18.33$, $p < .001$, were found in line with those results from the first two windows. Additionally, a significant two-way interaction between syllable frequency and tonal probability, $\chi^2(1) = 8.79$, $p < .01$, was found. Further analysis of this interaction revealed that, as compared to P− tones, P+ tones were more accurately identified when they co-occurred with F− syllables ($\beta = 1.63$, $SE = 0.60$, $t = 2.71$, $p < .01$). F+ syllables were more accurately identified than F− syllables by both groups when they co-occurred with P− tones ($\beta = 3.46$, $SE = 0.64$, $t = 5.37$, $p < .001$). At Window 4, significant main effects of group, $\chi^2(1) = 12.29$, $p < .001$ (amusics: $M = 0.84$, $SD = 0.37$; controls: $M = 0.89$, $SD = 0.31$), and syllable frequency, $\chi^2(1) = 18.49$, $p < .001$, and a significant two-way interaction between syllable frequency and tonal probability, $\chi^2(1) = 5.58$, $p < .05$, were found. Further analysis of this interaction revealed that, as compared to P− tones, P+ tones were more accurately recognized on F− syllables ($\beta = 0.97$, $SE = 0.47$, $t = 2.09$, $p < .05$). The F+ syllables were more accurately recognized with P− tones than F− syllables ($\beta = 2.48$, $SE = 0.50$, $t = 4.96$, $p < .001$).

In summary, the correct syllable–tone word results revealed that, at all four windows of analysis, amusics were less accurate than the control listeners and that both groups were more accurate in identifying F+ syllables than F− syllables. At Window 1, the control listeners were more accurate at P+ tones than P− tones, irrespective of syllable frequency. The amusics only showed a marginal trend for tonal probability. In Windows 3 and 4, both groups were more accurate at identifying P+ tones (as compared to P− tones) on F− syllables and P− tones on F+ syllables (as compared to F− syllables). We next examine the correct tone-only responses.

Correct Tone-Only Responses

Figure 5 plots the mean correct tone responses irrespective of syllables by amusic and musically intact participants faceted by syllable frequency and tonal probability at different windows. Similar to correct syllable–tone combinations, both groups were more accurate given more of the acoustic signal. The identification function of amusics generally remained lower than that of the controls. The mixed-effects models revealed a significant effect of group, $\chi^2(1) = 4.86$, $p < .05$, at Window 1, with amusics ($M = 0.59$, $SD = 0.49$) performing less accurately than typical listeners ($M = 0.64$, $SD = 0.48$). Similarly, an effect of group, $\chi^2(1) = 4.89$, $p < .05$, was found at Window 2 with typical listeners ($M = 0.86$, $SD = 0.34$) outperforming amusics ($M = 0.80$, $SD = 0.40$). There was no group difference at Window 3, $\chi^2(1) = 0.50$, $p = .48$, and only a trend for poorer performance by amusics ($M = 0.95$, $SD = 0.22$) than controls ($M = 0.98$, $SD = 0.15$) at Window 4, $\chi^2(1) = 3.59$, $p = .06$. No other effects were significant. In summary, the typical listeners were more accurate at tone identification than the amusics at Windows 1 and 2. We next examine correct syllable-only responses.

Correct Syllable-Only Responses

Figure 6 plots the syllable accuracy regardless of tones between amusics and musically intact listeners faceted by syllable frequency and tonal probability at different windows. Once again, both groups were more accurate as more acoustic information was available in the signal; moreover, amusics showed slightly lower accuracy than controls. The mixed-effects models revealed that amusics did not differ from controls at Window 1, $\chi^2(1) = 1.99$, $p = .159$; however, a significant main effect of syllable frequency, $\chi^2(1) = 28.04$, $p < .001$, was found, indicating that both groups perceived F+ syllables more accurately than F− syllables. At Window 2, significant main effects of group, $\chi^2(1) = 4.39$, $p < .05$, and syllable frequency, $\chi^2(1) = 29.92$, $p < .001$, were found, which indicated that amusics ($M = 0.67$, $SD = 0.47$) had poorer

Figure 5. Mean correct tone-only responses by syllable frequency and tonal probability by amusics and typical listeners at each window.

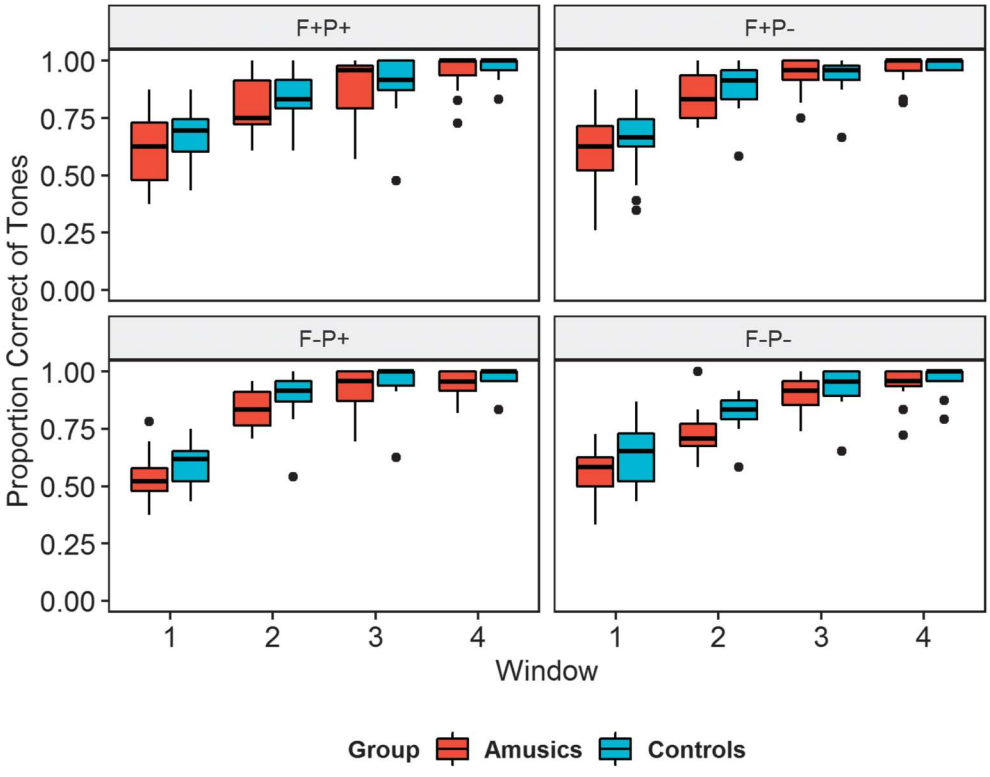
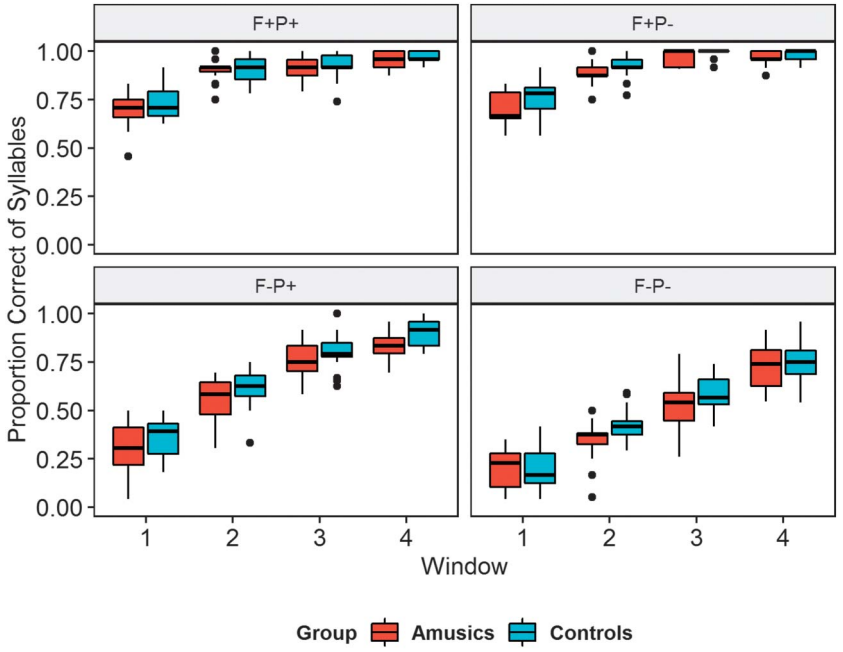


Figure 6. Mean correct syllable-only responses by syllable frequency and tonal probability by amusics and typical listeners at each window.



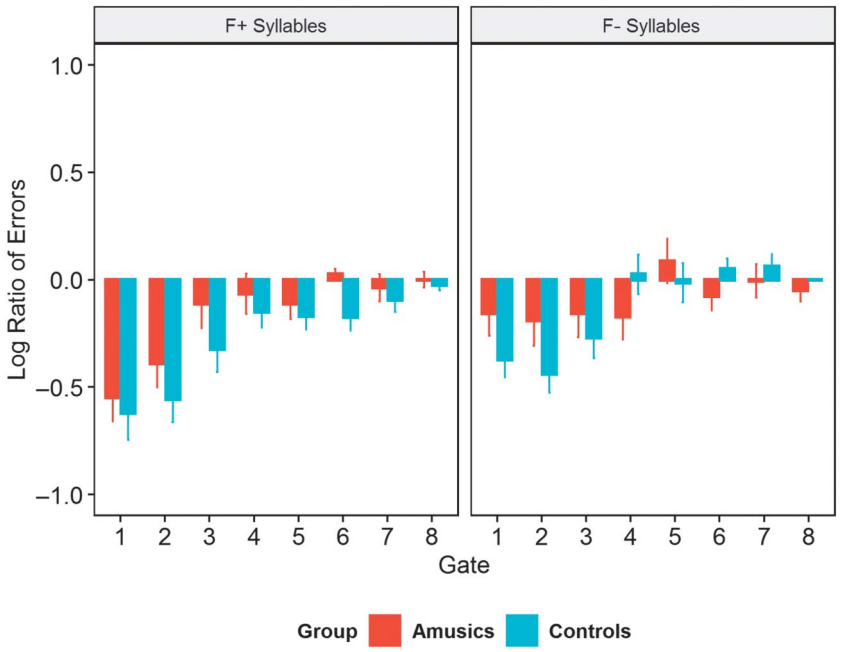
performance than controls (M = 0.71, SD = 0.45), but both groups identified F+ syllables more accurately than F− syllables independent of tonal probability. At Window 3, significant main effects of group, $\chi^2(1) = 9.53$, $p < .01$, and syllable frequency, $\chi^2(1) = 32.60$, $p < .001$, and a significant interaction between group and syllable frequency, $\chi^2(1) = 4.21$, $p < .01$, were found. Further analysis of this interaction showed that amusics (M = 0.65, SD = 0.48) were less accurate than controls (M = 0.70, SD = 0.46) at identifying F− syllables ($\beta = -0.49$, $SE = 0.24$, $t = -2.01$, $p < .05$). At Window 4, significant main effects of group, $\chi^2(1) = 6.47$, $p < .05$, and syllable frequency, $\chi^2(1) = 22.43$, $p < .001$, were found. Amusics (M = 0.87, SD = 0.33) identified syllables less accurately than typical listeners (M = 0.90, SD = 0.30) regardless of syllable frequency. Both groups identified F+ syllables more accurately than F− syllables. In summary, both groups identified F+ syllables more accurately than F− syllables across all four windows. In Windows 2, 3, and 4, typical listeners were more accurate than amusics overall. In Window 3, typical listeners were more accurate at identifying F− syllables than amusics. We next examine incorrect responses in which the correct syllable–incorrect tone cases were reported.

Correct Syllable–Incorrect Tone Errors

Following Wiener et al. (Wiener & Ito, 2016; Wiener & Lee, 2020; Wiener et al., 2019), we examine responses in which participants heard sufficient acoustic information to correctly identify the syllable, but not necessarily enough acoustic information to accurately identify meaningful F0 for tone categorization. In these instances, participants were forced to either use the limited F0, which resulted in an acoustic-based error, or rely on their knowledge of the most probable tone given the perceived syllable, resulting in a probability-based error.

Figure 7 displays the mean log ratio of tone errors faceted by syllable frequency for amusic and nonamusic participants. Listeners of both groups made fewer acoustic-based errors as more of the acoustic signal was available. The mixed-effects models revealed no effect of group, $\chi^2(1) = 1.17$, $p = .28$; however, significant effects of window, $\chi^2(1) = 4.63$, $p < .05$, and syllable frequency, $\chi^2(1) = 4.63$, $p < .05$, as well as their two-way interaction, $\chi^2(1) = 4.63$, $p < .05$, were found. Further analysis revealed that both groups showed a significantly more negative log ratio for F+ syllables compared to F− syllables ($\beta = -0.24$, $SE = 0.06$, $t = -3.87$, $p < .001$) at Window 1, indicative of participants' more acoustic-based errors when tones were combined with F+ syllables than F− syllables. In addition, both groups made fewer acoustic-based errors irrespective of syllable frequency between Window 1 and other later Windows 2, 3, and 4 (respectively, ps < .05), suggesting that the increased acoustic input resulted in fewer acoustic-based errors. Taken together, the results demonstrated that amusics and typical listeners showed a comparable pattern of errors in that both groups primarily made acoustic-based errors on F+ syllables in the

Figure 7. Mean log ratio of tone errors by syllable frequency by amusics and typical listeners at each window (error bar = ±1 *SE*).



early gates and that these errors became less common with more acoustic input.

Discussion

This study used the gating paradigm to test typical Mandarin listeners and Mandarin listeners with amusia in their perception of truncated Mandarin speech. The gating stimuli were based on the spoken word corpus SUBTLEX-CH (Cai & Brysbaert, 2010) and varied in syllable token frequency and syllable–tone co-occurrence probability. We specifically examined whether amusics, despite their degraded speech perception, would make use of the statistical structure of Mandarin speech sounds and report more frequent syllables and more probable tones similar to typical Mandarin listeners (Wiener & Ito, 2016; Wiener & Lee, 2020; Wiener et al., 2019). We analyzed syllable–tone word accuracy, syllable-only accuracy, tone-only accuracy, and correct syllable–incorrect tone errors. It was noted that the separation of syllable accuracy and tone accuracy did not mean that segmental and suprasegmental information were processed independently of each other in spoken word recognition; it instead clarified whether amusics' language deficits existed in both levels of speech given that Mandarin Chinese and other tone languages consist of both meaningful tones and syllables (Gandour, 1983). We report two key findings.

First, we corroborated the well-established findings that amusics' perception of both segments and suprasegmentals was degraded relative to typical, musically intact listeners (Jiang et al., 2012; Shao et al., 2019; C. Zhang, Peng, et al., 2017). The amusics were less accurate than the typical listeners in all of our statistical analyses. This finding held for the first two windows of the tone-only analysis, indicating that amusics particularly struggled to identify the tone when the F0 information was limited (onset and up to 120 ms of the vowel). For the syllable-only analysis, this finding held for Windows 2, 3, and 4 (onset and 80 ms or more of the vowel). This suggests that even when the full syllable was available to listeners, amusics still struggled to process spectral cues (Li et al., 2019; W. Tang et al., 2018). For the syllable–tone word analysis, amusics were found to be less accurate than the control group in all four windows. These findings underscore the difficulty amusics have with speech perception relative to typical listeners not only in lexical tones (Chen & Peng, 2018, 2020; Nguyen et al., 2009; C. Zhang et al., 2021) but also in vowels and their combination of spectral and F0 cues (Li et al., 2019; C. Zhang, Shao, et al., 2017).

Second, we found that in our syllable-only and syllable–tone word accuracy analyses, amusics, like typical listeners, made use of statistical information to correctly identify high-frequency (F+) syllables more accurately than low-frequency (F−) syllables. This pattern was observed in all four windows for both groups and in both analyses involving the syllable. This corroborates previous findings by Wiener et al. (Wiener & Ito, 2016; Wiener & Lee, 2020; Wiener et al., 2019) and strengthens the notion that the syllable plays an important role in Mandarin speech perception (and production) for

Mandarin listeners, including amusics (Chen et al., 2002; You et al., 2012). Importantly, this finding demonstrates that even though amusics struggle to accurately perceive the vowels within Mandarin syllables, they are still able to track syllables' frequency distribution. This finding is in line with previous studies by Omigie and Stewart (2011) and Peretz et al. (2012), which showed that amusics were able to track transitional probabilities between syllables spanning word boundaries. Crucially, whereas Omigie and Stewart removed all pitch information from the syllables used and Peretz et al. did not use F0 at a phonological level, we used stimuli containing phonological F0 information. Thus, we advance previous work by showing that although amusics' perception of vowels is degraded relative to typical listeners, their statistical learning of vowels and syllables appears to be preserved even when meaningful F0 information is present in the stimuli. It is therefore possible to posit that, in speech or language processes, the degraded perception of one cue or dimension does not necessarily lead to the degraded statistical computation for that cue or dimension. Further investigations, however, are needed to fully support this tentative claim.

More importantly, we found novel evidence that amusics track syllable–tone co-occurrence probabilities similarly to typical listeners. Amusics showed a marginal trend in the first window by identifying words with more probable (P+) tones more accurately than words with less probable (P−) tones. In Windows 3 and 4, amusics, like typical listeners, identified P+ tones more accurately than P− tones on low-frequency (F−) syllables and P− tones on high-frequency (F+) syllables more accurately than P− tones on F− syllables. This again corroborates previous findings by Wiener et al. (Wiener & Ito, 2016; Wiener & Lee, 2020; Wiener et al., 2019) and demonstrates that amusics can track not only syllable token distributions but also which tone is most likely given the particular consonant–vowel structure (i.e., dimension-based statistical learning; see Idemaru & Holt, 2011; Wiener et al., 2018, 2021). We also found that both groups made similar acoustic-based errors in their correct syllable–incorrect tone responses and that these errors were primarily on F+ syllables given that these syllables were easier to predict than F− syllables.

Taken together, the statistical learning mechanism involved in language processing (and presumably acquisition, see Saffran et al., 1996, 1999) appears to be spared in amusics given that amusics managed to track the statistical regularities (i.e., syllable frequency and tonal probability) embedded in Mandarin speech. This extends previous studies by Loui and Schlaug (2012) and Tillmann et al. (2014) by demonstrating that this is the case even when pitch plays a phonological role as it does in a tone language such as Mandarin. Amusics performed the task by drawing on their implicit knowledge derived from a lifetime of exposure to the language (Tillmann et al., 2014, 2015, 2016). Specifically, amusics relied on their knowledge of Mandarin syllable frequency and tonal probability in order to overcome fragmented speech and report the correct syllable–tone words. This implicit knowledge was combined with their ability to use categorical perception to bridge lower level acoustics with higher level phonological categories involving segments and tone (e.g., Francis et al., 2003; Yu et al., 2019; Zhu et al., 2021). Our results are in line with Xu et al.'s (2006) model about categorical perception and categorical memory (Ma et al., 2021; C. Zhang, Shao, et al., 2017). Even though prior studies broadly reported that amusics were impaired in their categorical perception (e.g., Huang, Liu, et al., 2015; Jiang et al., 2012), none of the studies specified that amusics were deprived of this capacity for categorical perception (Chen & Peng, 2018, 2020; F. Liu et al., 2021). Our findings strongly support the claim that amusics can perceive tones in a categorical manner, similar to that of musically intact controls.

The mechanisms involved in statistical learning of speech and categorical perception appear to be preserved in congenital amusics. Amusics' degraded speech perception appears to be primarily ascribed to the central deficit of impaired pitch processing (Peretz, 2016; Vuvan et al., 2015; C. Zhang, Shao, et al., 2017). Lexical tones and vowels have their respective primary acoustic cues of F0 and formants, but both are frequency-based sounds in the context of speech; hence, the possible deficit in the processing of spectral information results in amusics' poorer performance when listening to lexical tones and vowels as compared to controls (C. Zhang, Shao, et al., 2017). Because amusia is a congenital and lifelong disorder (Peretz et al., 2002, 2008), amusics are likely to struggle more than normal infants in the process of shaping the language boundaries between 6 and 12 months of age (Aslin & Pisoni, 1980), which in turn can persist into adulthood with inaccurate lexical representations (Kuhl, 2004). The finding that amusics performed worse than controls even in the last window containing the full acoustic signal was not entirely unexpected but, rather, supports previous studies that adopted auditory stimuli with sufficient and complete acoustic information (e.g., Jiang et al., 2012; Nan et al., 2010; Tillmann et al., 2011).

We note that there is an alternative account for the observed results: the potential anchoring deficit in amusia. The anchoring deficit, associated with deficient adaptation mechanisms, was initially proposed to explain the phonological deficit in the developmental disorder of dyslexia (Ahissar, 2007; Ahissar et al., 2006). Dyslexic patients tend to have difficulties in constructing perceptual anchors dynamically (Shao et al., 2019). Previous studies have shown that amusics demonstrated overall lower accuracy in experiments involving their perceptual adaptation of speech, including lexical tone integration with a meaningful (Shao & Zhang, 2018) or anomalous carrier sentence (C. Zhang et al., 2018), lexical tone adaptation in a high-

or low-pitched context (F. Liu et al., 2021), and lexical tone identification and discrimination with either different talkers or different syllables (Shao et al., 2019). In contrast to typical listeners, amusics show reduced connectivity between frontal and temporal regions in the brain, which is a disconnection syndrome (Hyde et al., 2011) and interferes with top-down processing from higher cortical levels onto lower level auditory processes. In this study, the auditory stimuli were presented in isolation with each token involving different syllables and tones and successively longer auditory fragments/gates. This could be considered as a dynamic environment, which demanded perceptual adaptation for participants (Shao et al., 2019). Future research will need to examine whether amusics' behavior was the result of their deficient perceptual adaptation, their impaired pitch processing, or the combination of these two factors.

With respect to practical applications, a straightforward finding from our results is that amusics can acquire an L2 even if the L2 is a tone language. Given that amusics are able to track both syllable and tone regularities in their native language (L1), amusics should be able to acquire L2 syllable–tone words. Although this study is not directly comparable to the gating study by Wiener et al. (2019), which tested adult L1 English–L2 Mandarin classroom learners, this study showed a similar trend in that amusics and L2 learners reached similar levels of accuracy by the last few gates. Future work on amusics' acquisition of tonal (and nontonal) L2s is desperately needed to better understand the acquisition process.

Moreover, our finding of amusics' preserved statistical learning mechanism is relevant for training programs aimed at rehabilitating congenital amusia. Importantly, combining our results with a prior gating study of music by Tillmann et al. (2014), it is suggested that regimens with pitch-related materials in either speech or music can be part of a treatment to lessen amusics' insensitivity to pitch. Given that amusics can attend to cues other than pitch in context (Nan et al., 2010) and the use of pitch in speech is generally not as rigorous and accurate as it is in music (Patel, 2014), amusics did not report that they had difficulties in communication and any improved communication skills thereof may not be observed. However, the lifelong deficit of impaired pitch-processing abilities can be potentially softened. Based on our findings, amusics are able to identify, discriminate, or acquire pitch information via learning and exercises in the training phase, through which their pitch thresholds are likely to be narrowed down. There have been a handful of studies concerning amusics' treatment in the extant literature. For example, from an educational perspective, Anderson et al. (2012) conducted a training program lasting 7 weeks to remediate amusia. The authors found improvements in both perception and production skills even though this was reported in three of five amusics. Although there can be genetic factors contributing to amusia (Peretz & Vuvan, 2017), our study explicitly manifests the possibility for amusics to learn and acquire pitch information despite their degraded pitch-processing abilities. Future studies are needed to conduct more translational research to soften this disorder, which is meanwhile significant to advance the understanding of the nature of amusia.

We conclude by noting the limitations of the study. First, our stimuli were recorded by only one speaker, which followed Wiener and Ito's (2016) approach. A design with more talkers may better clarify amusics' statistical learning mechanism in the speech domain and clarify to what degree perceptual adaptation (or the lack thereof) affected the results. Second, our participants were tested in Mainland China and were all largely monolingual Mandarin speakers. This stands in contrast with Wiener and Ito's participants who were tested in the United States and all functionally bilingual Mandarin–English speakers. We assume language dominance, language mode, and L2 immersion all had an effect on Wiener and Ito's participants' cognition and behavior (Grosjean, 1988, 1996; Kroll et al., 2015; Qin et al., 2021). At this point, the differences between the typical listeners in the two studies were potentially due to the population, testing environment, or the combination of the two. Third, it remains unclear to what degree amusics' performance may be influenced by noise typical to most speech communication. Presenting the stimuli in noise may yield dramatically different results. Fourth, our study only examined the statistical learning of speech. Saffran (2003) proposed that music and language share the same statistical learning mechanism. Future studies are needed to further clarify the domain generality or domain specificity of statistical learning involving both music and language in a single study (Peretz, 2006; Peretz et al., 2012).

In conclusion, individuals with amusia showed degraded speech perception of Mandarin syllables, tones, and their combinations as syllable–tone words in line with well-established amusia findings. However, despite this deficit, amusics showed a reliable pattern of statistical learning similar to that of typical, musically intact Mandarin listeners. This serves as novel evidence of amusics' preserved statistical learning mechanism, particularly when pitch plays a phonological role in a tone language.

Acknowledgments

This work was, in part, supported by the Fundamental Research Funds for the Central Universities of China (531118010660 awarded to Fei Chen), the Social Science Foundation of Ministry of Education of China (20YJC740041),

and Hunan Provincial Social Science Foundation of China (20ZDB003 awarded to Fei Chen, and 19YBQ112).

References

Ahissar, M. (2007). Dyslexia and the anchoring-deficit hypothesis. *Trends in Cognitive Sciences, 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. *Nature Neuroscience, 9*(12), 1558–1564. https://doi.org/10.1038/nn1800

Allen, G. (1878). I.—Note-deafness. *Mind, os-3*(10), 157–167. https://doi.org/10.1093/mind/os-3.10.157

Anderson, S., Himonides, E., Wise, K., Welch, G., & Stewart, L. (2012). Is there potential for learning in amusia? A study of the effect of singing intervention in congenital amusia. *Annals of the New York Academy of Sciences, 1252*(1), 345–353. https://doi.org/10.1111/j.1749-6632.2011.06404.x

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Volume 2, perception* (pp. 67–96). Academic Press.

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. https://doi.org/10.1093/brain/125.2.238

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models Usinglme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer*. http://www.praat.org

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLOS ONE, 5*(6), e10729. https://doi.org/10.1371/journal.pone.0010729

Chao, Y. R. (1965). *A grammar of spoken Chinese*. University of Berkeley Press.

Chen, F., & Peng, G. (2018). Lower-level acoustics underlie higher-level phonological categories in lexical tone perception. *The Journal of the Acoustical Society of America, 144*(3), EL158–EL164. https://doi.org/10.1121/1.5052205

Chen, F., & Peng, G. (2020). Reduced sensitivity to between-category information but preserved categorical perception of lexical tones in tone language speakers with congenital amusia. *Frontiers in Psychology, 11,* 581410. https://doi.org/10.3389/fpsyg.2020.581410

Chen, F., Peng, G., Yan, N., & Wang, L. (2017). The development of categorical perception of Mandarin tones in four- to seven-year-old children. *Journal of Child Language, 44*(6), 1413–1434. https://doi.org/10.1017/s0305000916000581

Chen, J. Y., Chen, T. M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*(4), 751–781. https://doi.org/10.1006/jmla.2001.2825

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffths, T. D. (2004). Characterization of deficits in pitch perception underlying "tone deafness." *Brain, 127*(4), 801–810. https://doi.org/10.1093/brain/awh105

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non) categorical perception of lexical tones. *Perception & Psychophysics, 65*(7), 1029–1044. https://doi.org/10.3758/bf03194832

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics, 11*(2), 149–175. https://doi.org/10.1016/s0095-4470(19)30813-7

Gong, Y. X. (1992). *Wechsler Adult Intelligence Scale–Revised in China Version*. Hunan Medical College.

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

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*(4), 267–283. https://doi.org/10.3758/BF03204386

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Processes, 3*(3), 233–274. https://doi.org/10.1080/01690968808402089

Grosjean, F. (1996). Gating. *Language and Cognitive Processes, 11*(6), 597–604. https://doi.org/10.1080/016909696386999

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

Huang, W. T., Nan, Y., Dong, Q., & Liu, C. (2015). Just-noticeable difference of tone pitch contour change for Mandarin congenital amusics. *The Journal of the Acoustical Society of America, 138*(1), EL99–EL104. https://doi.org/10.1121/1.4923268

Hyde, K. L., & Peretz, I. (2003). "Out-of-pitch" but still "in-time." *Annals of the New York Academy of Sciences, 999*(1), 173–176. https://doi.org/10.1196/annals.1284.023

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

Hyde, K. L., Zatorre, R. J., & Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex, 21*(2), 292–299. https://doi.org/10.1093/cercor/bhq094

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance, 37*(6), 1939–1956. https://doi.org/10.1037/a0025641

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), 1109–1121. https://doi.org/10.3758/s13421-012-0208-2

Kalmus, H., & Fry, D. (1980). On tune deafness (dysmelodia): Frequency, development, genetics and musical background. *Annals of Human Genetics, 43*(4), 369–382. https://doi.org/10.1111/j.1469-1809.1980.tb01571.x

Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, mind, and brain. *Annual Review of Linguistics, 1*(1), 377–394. https://doi.org/10.1146/annurev-linguist-030514-124937

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lee, C. Y., & Wiener, S. (2020). Acoustic-based and knowledge-based processing of Mandarin tones by native and non-native speakers. In H. M. Liu, F. M. Tsao, & P. Li (Eds.), *Speech perception, production and acquisition: Multidisciplinary approaches in Chinese languages* (Vol. 11, pp. 37–57). Springer Nature. https://doi.org/10.3389/fpsyg.2020.00214

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

Li, M., Tang, W., Liu, C., Nan, Y., Wang, W., & Dong, Q. (2019). Vowel and tone identification for Mandarin congenital amusics: Effects of vowel type and semantic content. *Journal of Speech, Language, and Hearing Research, 62*(12), 4300–4308. https://doi.org/10.1044/2019_JSLHR-S-18-0440

Liu, C. (2013). Just noticeable difference of tone pitch contour change for English- and Chinese-native listeners. *The Journal of the Acoustical Society of America, 134*(4), 3011–3020. https://doi.org/10.1121/1.4820887

Liu, F., Jiang, C., Wang, B., Xu, Y., & Patel, A. D. (2015). A music perception disorder (congenital amusia) influences speech comprehension. *Neuropsychologia, 66,* 111–118. https://doi.org/10.1016/j.neuropsychologia.2014.11.001

Liu, F., Yin, Y., Chan, A., Yip, V., & Wong, P. C. M. (2021). Individuals with congenital amusia do not show context-dependent perception of tonal categories. *Brain and Language, 215*(1), 104908. https://doi.org/10.1016/j.bandl.2021.104908

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *The Journal of Neuroscience, 29*(33), 10215–10220. https://doi.org/10.1523/JNEUROSCI.1701-09.2009

Loui, P., & Schlaug, G. (2012). Impaired learning of event frequencies in tone deafness. *Annals of the New York Academy of Sciences, 1252*(1), 354–360. https://doi.org/10.1111/j.1749-6632.2011.06401.x

Ma, J., Zhu, J., Yang, Y., & Chen, F. (2021). The development of categorical perception of segments and suprasegments in Mandarin-speaking preschoolers. *Frontiers in Psychology, 12,* 693366. https://doi.org/10.3389/fpsyg.2021.693366

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America, 102*(3), 1864–1877. https://doi.org/10.1121/1.420092

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. *Annals of the New York Academy of Sciences, 1169*(1), 490–493. https://doi.org/10.1111/j.1749-6632.2009.04855.x

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Omigie, D., & Stewart, L. (2011). Preserved statistical learning of tonal and linguistic material in congenital amusia. *Frontiers in Psychology, 2,* 109. https://doi.org/10.3389/fpsyg.2011.00109

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research, 308,* 98–108. https://doi.org/10.1016/j.heares.2013.08.011

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. https://doi.org/10.1016/j.bandc.2004.10.003

Peretz, I. (2001). Brain specialization for music. New evidence from congenital amusia. *Annals of the New York Academy of Sciences, 930*(1), 153–165. https://doi.org/10.1111/j.1749-6632.2001.tb05731.x

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition, 100*(1), 1–32. https://doi.org/10.1016/j.cognition.2005.11.004

Peretz, I. (2016). Neurobiology of congenital amusia. *Trends in Cognitive Sciences, 20*(11), 857–867. https://doi.org/10.1016/j.tics.2016.09.002

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., Brattico, E., Järvenpää, 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

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

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

Peretz, I., Saffran, J., Schon, D., & Gosselin, N. (2012). Statistical learning of speech, not music, in congenital amusia. *Annals of the New York Academy of Sciences, 1252*(1), 361–366. https://doi.org/10.1111/j.1749-6632.2011.06429.x

Peretz, I., & Vuvan, D. T. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics, 25*(5), 625–630. https://doi.org/10.1038/ejhg.2017.15

Psychology Software Tools. (n.d.). *E-Prime* (Version 2.0) [Computer software]. Author.

Qin, Z., Zhang, C., & Wang, W. S. Y. (2021). The effect of Mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones. *The Journal of the Acoustical Society of America, 149*(1), 435–446. https://doi.org/10.1121/10.0003330

R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Saffran, J. R. (2003). Musical learning and language development. *Annals of the New York Academy of Sciences, 999*(1), 397–401. https://doi.org/10.1196/annals.1284.050

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

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. *Journal of Speech, Language, and Hearing Research, 62*(1), 190–205. https://doi.org/10.1044/2018_JSLHR-H-17-0483

Shao, J., & Zhang, C. (2018). Context integration deficit in tone perception in Cantonese speakers with congenital amusia. *The Journal of the Acoustical Society of America, 144*(4), EL333–EL339. https://doi.org/10.1121/1.5063899

Tang, P., Yuen, I., Rattanasone, N. X., Gao, L., & Demuth, K. (2019). The acquisition of Mandarin tonal processes by children with cochlear implants. *Journal of Speech, Language, and Hearing Research, 62*(5), 1309–1325. https://doi.org/10.1044/2018_JSLHR-S-18-0304

Tang, W., Wang, X. J., Li, J. Q., Liu, C., Dong, Q., & Nan, Y. (2018). Vowel and tone recognition in quiet and in noise among Mandarin-speaking amusics. *Hearing Research, 363,* 62–69. https://doi.org/10.1016/j.heares.2018.03.004

Tillmann, B., Albouy, P., & Caclin, A. (2015). Congenital amusias. *Handbook of Clinical Neurology, 129,* 589–605. https://doi.org/10.1016/B978-0-444-62630-1.00033-0

Tillmann, B., Albouy, P., Caclin, A., & Bigand, E. (2014). Musical familiarity in congenital amusia: Evidence from a gating paradigm. *Cortex, 59,* 84–94. https://doi.org/10.1016/j.cortex.2014.07.012

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,* 120. https://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. https://doi.org/10.1016/j.brainres.2015.10.035

Vuvan, D. T., Nunes-Silva, M., & Peretz, I. (2015). Meta-analytic evidence for the non-modularity of pitch processing in congenital amusia. *Cortex, 69,* 186–200. https://doi.org/10.1016/j.cortex.2015.05.002

Wang, T., Potter, C. E., & Saffran, J. R. (2020). Plasticity in second language learning: The case of Mandarin tones. *Language Learning and Development, 16*(3), 231–243. https://doi.org/10.1080/15475441.2020.1737072

Wang, W. S. Y. (1973). The Chinese language. *Scientific American, 228*(2), 50–60. https://doi.org/10.1038/scientificamerican0273-50

Wang, X., & Peng, G. (2014). Phonological processing in mandarin speakers with congenital amusia. *The Journal of the Acoustical Society of America, 136*(6), 3360–3370. https://doi.org/10.1121/1.4900559

Wiener, S., & Ito, K. (2015). Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition & Neuroscience, 30*(9), 1048–1060. https://doi.org/10.1080/23273798.2014.946934

Wiener, S., & Ito, K. (2016). Impoverished acoustic input triggers probability-based tone processing in mono-dialectal Mandarin listeners. *Journal of Phonetics, 56,* 38–51. https://doi.org/10.1016/j.wocn.2016.02.001

Wiener, S., Ito, K., & Speer, S. R. (2018). Early L2 spoken word recognition combines input-based and knowledge-based processing. *Language and Speech, 61*(4), 632–656. https://doi.org/10.1177/0023830918761762

Wiener, S., Ito, K., & Speer, S. R. (2021). Effects of multitalker input and instructional method on the dimension-based statistical learning of syllable–tone combinations. *Studies in Second Language Acquisition, 43*(1), 155–180. https://doi.org/10.1017/S0272263120000418

Wiener, S., & Lee, C. Y. (2020). Multi-talker speech promotes greater knowledge-based spoken Mandarin word recognition in first and second language listeners. *Frontiers in Psychology, 11,* 214. https://doi.org/10.3389/fpsyg.2020.00214

Wiener, S., Lee, C. Y., & Tao, L. (2019). Statistical regularities affect the perception of second language speech: Evidence from adult classroom learners of Mandarin Chinese. *Language Learning, 69*(3), 527–558. https://doi.org/10.1111/lang.12342

Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and Speech, 59*(1), 59–82. https://doi.org/10.1177/0023830915578000

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

Wong, P. C., Ciocca, V., Chan, A. H., Ha, L. Y., Tan, L. H., & Peretz, I. (2012). Effects of culture on musical pitch perception. *PLOS ONE, 7*(4), e33424. https://doi.org/10.1371/journal.pone.0033424

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America, 120*(2), 1063–1074. https://doi.org/10.1121/1.2213572

Yang, W. X., Feng, J., Huang, W. T., Zhang, C. X., & Nan, Y. (2014). Perceptual pitch deficits coexist with pitch production difficulties in music but not Mandarin speech. *Frontiers in Psychology, 4,* 1024. https://doi.org/10.3389/fpsyg.2013.01024

You, W., Zhang, Q., & Verdonschot, R. G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PLOS ONE, 7*(10), e46595. https://doi.org/10.1371/journal.pone.0046595

Yu, K., Li, L., Chen, Y., Zhou, Y., Wang, R., Zhang, Y., & Li, P. (2019). Effects of native language experience on Mandarin lexical tone processing in proficient second language learners. *Psychophysiology, 56*(11), e13448. https://doi.org/10.1111/psyp.13448

Zhang, C., Ho, O. Y., Shao, J., Ou, J., & Law, S. P. (2021). Dissociation of tone merger and congenital amusia in Hong Kong Cantonese. *PLOS ONE, 16*(7), e0253982. https://doi.org/10.1371/journal.pone.0253982

Zhang, C., Peng, G., Shao, J., & Wang, W. S. Y. (2017). 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., & Chen, S. (2018). Impaired perceptual normalization of lexical tones in Cantonese-speaking congenital amusics. *The Journal of the Acoustical Society of America, 144*(2), 634–647. https://doi.org/10.1121/1.5049147

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. https://doi.org/10.1371/journal.pone.0183151

Zhang, H., Zhang, J., Peng, G., Ding, H., & Zhang, Y. (2020). Bimodal benefits revealed by categorical perception of lexical tones in Mandarin-speaking kindergarteners with a cochlear implant and a contralateral hearing aid. *Journal of Speech, Language, and Hearing Research, 63*(12), 4238–4251. https://doi.org/10.1044/2020_JSLHR-20-00224

Zhu, J., Chen, X., & Yang, Y. (2021). Effects of amateur musical experience on categorical perception of lexical tones by native Chinese adults: An ERP study. *Frontiers in Psychology, 12,* 611189. https://doi.org/10.3389/fpsyg.2021.611189

Appendix

Syllable–Tone Combinations

Condition	Stimulus	Condition	Stimulus
F+P+	bao2	F+P–	bao4
F+P+	da4	F+P–	da2
F+P+	gong1	F+P–	gong3
F+P+	ji1	F+P–	ji3
F+P+	qi2	F+P–	qi1
F+P+	ren2	F+P–	ren4
F+P+	shou3	F+P–	shou1
F+P+	si1	F+P–	si3
F+P+	wu2	F+P–	wu4
F+P+	xiao3	F+P–	xiao4
F+P+	yan2	F+P–	yan4
F+P+	zhu4	F+P–	zhu2
Condition	**Stimulus**	**Condition**	**Stimulus**
F–P+	bin1	F–P–	bin4
F–P+	chai1	F–P–	chai2
F–P+	feng1	F–P–	feng3
F–P+	gua4	F–P–	gua3
F–P+	hong2	F–P–	hong4
F–P+	kao3	F–P–	kao1
F–P+	leng3	F–P–	leng4
F–P+	mao4	F–P–	mao3
F–P+	niao3	F–P–	niao4
F–P+	pang2	F–P–	pang1
F–P+	tie3	F–P–	tie1
F–P+	zhou1	F–P–	zhou2