Tone Deafness in Music Does Not Preclude Distributional Learning of Non-Native Tonal Languages in Individuals With Congenital Amusia

Jiaqiang Zhu [a], Xiaoxiang Chen [b*], Fei Chen [b], Caicai Zhang [a*], Jing Shao [c], Seth Wiener [d]

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

[b] School of Foreign Languages, Hunan University, Changsha, China

[c] Department of English Language and Literature, Hong Kong Baptist University, Hong Kong SAR, China

[d] Department of Modern Languages, Carnegie Mellon University, Pittsburgh, United States

*** Correspondence:**

Xiaoxiang Chen: xiaoxiangchensophy@hotmail.com

Caicai Zhang: caicai.zhang@polyu.edu.hk

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

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

Ethics Approval Statement: This study involving human participants was reviewed and approved by the ethics review board at the School of Foreign Languages of Hunan University. Written informed consent was signed by each participant.

Funding Statement: This study was supported by grants from the Humanities and Social Science Project of Ministry of Education of China (22YJC740008; 22YJC740093), and the Fundamental Research Funds for the Central Universities of China (531118010660).

Tone Deafness in Music Does Not Preclude Distributional Learning of Non-Native Tonal Languages in Individuals With Congenital Amusia

Abstract

Purpose: Previous studies have shown that individuals with congenital amusia exhibit deficient pitch processing across music and language domains. This study investigated whether adult Chinese-speaking listeners with amusia were still able to learn Thai lexical tones based on stimulus frequency of statistical distribution via distributional learning, despite their degraded lexical tone perception.

Method: Following a pretest-training-posttest design, twenty-one amusics and 23 typical, musically intact listeners were assigned into bimodal and unimodal distribution conditions. Listeners were asked to discriminate the minimal pair of Thai mid-level tone and falling tone superimposed on variable base syllables and uttered by different speakers. The perceptual accuracy for each test session and improvement from pretest to posttest were collected and analyzed between the two groups using generalized mixed-effects models.

Results: When discriminating Thai lexical tones, amusics were less accurate than typical listeners. Nonetheless, similarly to control listeners, perceptual gains from pretest to posttest were observed in bimodally rather than unimodally trained amusics, as evidenced by both trained and non-trained test words.

Conclusions: Amusics are able to learn lexical tones in a second or foreign context of speech. This extends previous research by showing that amusics' distributional learning of linguistic pitch remains largely preserved in spite of their degraded pitch processing. It is thus likely that manifestations of amusia in speech could not result from their abnormal statistical learning mechanism. This study meanwhile provides a heuristic approach for future studies to apply this paradigm into amusics' treatment to mitigate their pitch-processing disorder.

Keywords: congenital amusia; distributional learning; Thai lexical tones; perceptual training; speech perception

1 Introduction

Distributional learning, a statistical learning mechanism, refers to the learning of speech sounds through relative frequency of exposure to different tokens in one's surroundings (Escudero & Williams, 2014). Distributional learning has been proposed to explain the perceptual attunement through which phonetic categories are acquired among both infant and adult learners (Maye et al., 2002, 2008; Werker et al., 2012). Whereas previous distributional learning studies have primarily examined typical, healthy infants and adults (e.g., Escudero et al., 2011; Wanrooij et al., 2013), here we examine atypical adult listeners with congenital amusia, i.e., "tone deafness" (Peretz & Vuvan, 2017). We build on prior phonological tone perception research, to investigate whether individuals with tone deafness still show evidence of distributional learning of phonological tone.

1.1 Degraded Processing of Musical and Linguistic Pitch in Congenital Amusics

Congenital amusia is an umbrella term for musical disabilities which cannot be attributed to brain damage, hearing loss, lack of music exposure, or cognitive disabilities after birth (Peretz, 2001). The atypical population afflicted with amusia ("amusics" thereafter) is roughly 1.5%–4% of the general population (Peretz & Vuvan, 2017). Research on amusics' processing of music has shown that amusics struggle to identify whether presented musical melodies contain a note that was modified (Ayotte et al., 2002), a pitch that was changed (Hyde & Peretz, 2003, 2004), fine-grained pitch differences among notes and melodies, as well as complex pitch patterns (Foxton et al., 2004).

Amusics also struggle to identify pitch when it is used in speech communication, such as rising and falling intonational pitch contours. Patel et al. (2005) tested amusics' and typical listeners' processing of intonation contours within the original sentences or in an analog form (i.e., gliding-pitch analogs devoid of any linguistic context). The authors found that typical listeners discriminated the original sentences and analogs equally well, whereas amusics showed poorer discrimination of the analogs (see also Patel et al., 2008; Hutchins et al., 2010).

Amusics additionally struggle to identify pitch when it is used as a phonological cue as in languages with lexical tones. Changes in fundamental frequency (F0, the physical correlate of pitch) alter the lexical semantics of tone languages like Mandarin, Cantonese, and Thai—the present study's target language. Tillmann et al. (2011) tested French-speaking amusics and controls in their discrimination of Mandarin and Thai lexical tone pairs. Results showed that amusics' performance was inferior to controls in their discrimination of all tone pairs (see also Jiang et al., 2012; Zhang et al., 2018). Thus, prior research has established that congenital amusia is a domain-general pitch-processing deficit that affects musical, sentence-level, and word-level pitch processing (Vuvan et al., 2015).

1.2 Statistical Learning of Pitch in Congenital Amusics

To date, only a handful of studies have explored amusics' statistical learning of speech. Both Omigie and Stewart (2011) and Peretz et al. (2012) reported that amusics were capable of extracting three-syllable nonce words from a continuous speech stream in which there were no acoustic or prosodic cues presented to mark word boundaries. That is, amusics were able to use transitional probabilities (see Saffran et al., 1996) to learn patterns from continuous speech. Nevertheless, because Omigie and Stewart (2011) removed all pitch information from their stimuli and Peretz et al. (2012) did not cue F0 to alter lexical semantics, neither study fully evaluated amusics' statistical learning of speech in which pitch plays a *phonological* role. More recently, Zhu et al. (2022) made use of the gating paradigm to present Mandarin-speaking control listeners and amusics increasingly longer fragments of syllable-tone words. The authors found that amusics' identification of both segments and tones was degraded relative to control listeners (e.g., Shao et al., 2019; Zhang et al., 2017a). More importantly, the authors found that in early gates containing truncated speech (e.g., consonant onset and up to 120 ms of the vowel) amusics—like control listeners—identified frequent syllables more accurately than less frequent syllables and more probable tones more accurately than less probable tones (given the syllable). In other words, Zhu et al. (2022) demonstrated that amusics tracked statistical regularities of segments and tones embedded in their native speech in a manner similar to control listeners (Wiener & Ito, 2016; Wiener & Lee, 2020). What remains unclear, and is the focus of the present study, is whether amusics can learn statistical regularities involving phonological tone in a *non-native* language within a brief training period involving distributional learning.

1.3 Distributional Learning of Non-Native Speech

In distributional learning paradigms, participants are presented with auditory tokens from a particular acoustic-phonetic continuum. The equidistant stimuli along this continuum are played with frequencies that either constitute a bimodal (tokens near the endpoints occur most frequently) or a unimodal (tokens in the middle of the continuum occur most frequently) distribution in training (see Figure 1). After the passive exposure to the tokens differing in statistical distribution, participants are then tested on their discrimination of stimuli from the continuum. Maye et al. (2002) generated an eight-step continuum on the dimension of voice onset time. Infant learners exposed to the bimodal distribution reliably discriminated stimuli from the endpoints of the continuum in post-training test better than those exposed to the unimodal distribution. Escudero et al. (2011) recruited Spanish speakers to discriminate Dutch vowels, and found that bimodally trained learners improved their performance from pretest to posttest. Escudero and Williams (2014) further showed that distributional learning has long-lasting effects; learners' accuracy rates in the immediate posttest and posttests after six and twelve months, respectively, were all higher than pretest accuracy rates.

<Insert Figure 1 around here>

Ong and colleagues have conducted a series of experiments of distributional learning of Thai lexical tones over recent years (Ong et al., 2015a, 2015b, 2016, 2017a). Different from Mandarin with four phonological tones including one static (high-level) and three dynamic (mid-rising, low-dipping, and high-falling) tones, Thai has five phonological tones consisting of three static (high-level, mid-level, and low-level) and two dynamic (falling and rising) tones (Abramson, 1978; Gandour, 1983). For example, the syllable /na:/ can mean "rice field" (mid-level), "custard apple" (low-level), "face" (falling), "aunt" (high-level), and "thick" (rising) in Thai (Zsiga & Nitisaroj, 2007). Ong et al. tested Australian English-speaking participants (2015a, 2017a) and Mandarin-speaking participants (2015b, 2016, 2017a) on a Thai tone minimal pair discrimination task using the mid-level and falling tones (which were found in previous research to be the most difficult for non-native listeners to discriminate given their similar starting pitch contours; Burnham et al., 2015). Ong et al. repeatedly observed greater learning in participants trained with bimodal distributions relative to participants trained with unimodal distributions. Ong et al.'s results highlight that listeners unfamiliar with lexical tone

(Australian English speakers) and listeners familiar with lexical tone, albeit a different tonal system (Mandarin speakers), can both learn non-native lexical tones through distributional learning.

1.4 The Present Study

In the present study, we make use of Thai distributional learning stimuli from Ong and colleagues (Ong et al., 2015a, 2015b, 2016, 2017a). Our research goal is mainly to examine whether individuals with amusia who are familiar with lexical tone in their native language are able to learn non-native tones through distributional learning. Previous work related to amusics' perception of lexical tones consistently revealed amusics' degraded performance as compared with musically intact listeners (Jiang et al., 2012; Tillmann et al., 2011; Zhang et al., 2018), it was hence hypothesized that amusics' perceptual accuracy of Thai lexical tones would be generally lower than typical individuals. Pertaining to the learning outcome, amusics could show no improvement after training in either distribution condition given that they were afflicted with the pitch-processing disorder of amusia (Vuvan et al., 2015). Meanwhile, as mentioned earlier, because several studies did uncover that amusics were able to learn speech sounds in a statistical manner (Omigie & Stewart, 2011; Peretz et al., 2012; Zhu et al. 2022), it was also expected that for non-native lexical tones, the perceptual gains would be observed in bimodally rather than unimodally trained amusics from pretest to posttest, which resembles the performance profile in typical listeners (Erickson & Thiessen, 2015; Ong et al., 2015a, 2016, 2017a; Thiessen, 2017).

2 Method

2.1 Participants

Twenty-one congenital amusics and twenty-three musically intact listeners participated in this study. All participants were native speakers of Mandarin Chinese living in China. Initially, each group consisted of 24 participants, but four of them (three amusics and one typical listeners) dropped out from the experiment due to their unavailability.

Amusics and control listeners were identified following the guidelines set by Peretz and colleagues (Peretz & Vuvan, 2017; Peretz et al., 2003, 2008). College students who spoke Mandarin Chinese as their first language, did not have any musical training, and reported difficulties in experiencing music, such as the inability to freely sing a song, were invited to participate in the study. This yielded around 400 participants who were tested via the online Montreal Battery of Evaluation of Amusia (Peretz et al., 2008). The experimenter monitored the whole testing and listeners did not take the battery repeatedly after they had finished all the three subtests (Peretz & Vuvan, 2017), encompassing rhythm-focused program of offbeat subtest and pitch-focused program of out-of-key and mistuned subtests. The cutoff mean score of 71% was set as the standard for amusics' identification in line with previous studies with tone language speakers (e.g., Chen & Peng, 2018, 2020; Huang et al., 2016; Shao & Zhang, 2018; Wang & Peng, 2014; Wong et al., 2012; Zhang et al., 2017a, 2017b). Table 1 reports subtest scores as well as overall scores. Independent-samples *t*-tests revealed that amusics were significantly lower than the matched controls in global mean accuracy and scores for the three subtests (*p*s < .001).

<Insert Table 1 around here>

Working memory was also measured. All participants performed the digit span test extracted from the Wechsler Adult Intelligence Scale-Revised by China (Gong, 1992), which contained forward digit span and backward digit span tasks. This scale has been used in previous studies of congenital amusia (e.g., Nan et al., 2016; Tang et al., 2018; Yang et al., 2014). As shown in Table 1, amusics and musically intact listeners performed similarly (p = .62), which supports the claim that amusics have similar general memory capacity to typical listeners (e.g., Williamson & Stewart, 2010; Yang et al., 2014; Tillmann et al., 2016). Using a modified Chinese version of the Edinburgh Handedness Inventory (Oldfield, 1971), each participant was confirmed to be right-handed. None of the participants reported having hearing loss, brain injuries, or neurological disabilities. Importantly, all participants had no prior experience learning Thai. All listeners were paid for their participation. The experiment was approved by the ethics review board at the School of Foreign Languages of Hunan University, with each participant signing the written informed consent.

2.2 Materials

All auditory materials were adapted from Ong and colleagues' studies (Ong et al., 2015a, 2015b, 2016, 2017a). The Thai syllables /na:/ and /k^{h}a:/ with mid-level and falling tones were recorded by two native Thai speakers (male, female). As noted earlier, this Thai tone contrast is the most difficult one for Mandarin-speaking listeners to discriminate (Burnham et al., 2015). The two Thai speakers were both young adults with minimal experience of Chinese, and they had no known speech, language, or hearing disorders. They were instructed to read aloud the target words embedded within a Thai sentence "ผมอยากบอกว่า ____" or "ฉันอยากบอกว่า ____" (with appropriate personal pronoun of gender, "ผม" for male and "ฉัน" for female), which meant, "I want to read ____" (Wu et al., 2014). The recordings were sampled at 44.1 kHz and digitized at 16 bits using Praat (Boersma & Weenink, 2018). These materials were recorded ten times repeatedly by each Thai speaker, with the most fluent and natural tokens selected. The words with monosyllables /na:/ and /k^ha:/ were then excised from the sentences, with three additional native Thai speakers confirming the naturalness of these exemplars. From these recordings, four minimal tone pairs were developed, i.e., each syllable with each speaker producing both tones for a total of eight utterances (female /na:/ and /k^ha:/ in mid-level and falling tones, and male /na:/ and /k^ha:/ in mid-level and falling tones).

2.2.1 Training Stimuli

Of these minimal tone pairs, the contrast of male /na:/ in mid-level and falling tones was used for training since this tone pair was more difficult than other pairs in tone discrimination in line with Ong et al. (2015a, 2017a). The tonal continuum based on this tone pair generated eight stimuli with the equidistant pitch difference, with Stimulus 1 representing the mid-level tone and Stimulus 8 representing the falling tone. These stimuli in the continuum were synthesized using the pitch synchronous overlap and add algorithm (Moulines & Laroche, 1995) implemented in Praat, following Peng et al. (2010) and Shen and Froud (2016, 2019). The values of F0 for the eight tokens were interpolated alone, while other acoustic cues were kept identical including duration and intensity. The training stimuli were confirmed by three native Thai speakers who exhibited a decline in the percentage of mid-level tone responses from the first to the last steps along the continuum, suggesting that the perception of the eight tokens by native Thai speakers changed from the mid-level tone category to the falling tone category, as intended (i.e., Ong et al., 2015a; also see similar methods in Xi et al., 2010, Yu et al., 2014). Table 2 shows a summary of pitch frequency information of the eight training stimuli extracted from the continuum.

<Insert Table 2 around here>

2.2.2 Test Stimuli

The test words encompassed not only the tokens appearing in training but also the non-trained ones, in order to explore the potential generalization effects of statistical learning (Maye et al., 2008). In other words, the introduction of non-trained stimuli into the test was beneficial to examine whether participants, including both amusics and typical listeners, were capable of extending what they had acquired to novel items that were primarily characterized by the similar cue or dimension, and how far they could generalize the learned categories at an abstract level across variable conditions in speech perception (Erickson & Thiessen, 2015).

<Insert Figure 2 around here>

To ensure that only F0 was different among the test stimuli, the naturalistically produced lexical tones were extracted and superimposed on a common base sound for each minimal pair following Ong and colleagues (Ong et al., 2015a, 2015b, 2016, 2017a). For each minimal pair, a base waveform of the monosyllable was firstly chosen; next, the pitch contour was extracted from each class of that minimal pair and further superimposed on this chosen base waveform using Praat. Two more exemplars were created with other naturalistically produced pitch contours superimposed on the same base waveform for each minimal pair, resulting in three exemplars with the sole difference in F0 for pairs of male /na:/ mid-level to falling, male /k^{h}a/

mid-level to falling, female /na:/ mid-level to falling, and female /k^{h}a/ mid-level to falling, individually. These exemplars would be used in the test. Figures 2 and 3 plot the pitch contours of Thai lexical tones produced by male and female speakers.

<Insert Figure 3 around here>

All stimuli were normalized at 70 dB in intensity, with duration exactly the same within each minimal pair based on the shared base waveform of the monosyllable (i.e., 567 ms for male /na:/ mid-level to falling, 508 ms for male /k^{h}a/ mid-level to falling, 658 ms for female /na:/ mid-level to falling, and 770 ms for female /k^{h}a/ mid-level to falling), which were all similar to Ong et al. (2017a). Table 3 displays the detailed information of all test words. Once again, three additional native Thai speakers verified the naturalness of these auditory tokens and correctly identified them as either mid-level or falling out of the choices of the five Thai lexical tones.

<Insert Table 3 around here>

An additional 32 440-Hz pure tones were synthesized via Praat and used as beeps during the training. Each beep tone was 800 ms in duration and 70 dB in intensity. These beeps were randomly interspersed among the training tokens, serving as an extra vigilance task for each participant (Ong et al., 2015a).

2.3 Procedures

This study followed the pretest-training-posttest design as broadly used in the distributional learning literature (Escudero & Williams, 2014; Maye et al., 2002, 2008; Wanrooij et al., 2013). The assignment of participants to distribution conditions was achieved in a pseudo-random manner by keeping the bimodal group comparable to the unimodal counterpart (e.g., age, gender, working memory, and pitch-related composite test scores in the screening battery), aiming to maximally exclude any potential effects other than stimulus statistical distribution on learning. Correspondingly, the four sub-groups were developed, including amusics' bimodal (*n* = 11) and unimodal (*n* = 10) groups and controls' bimodal (*n* = 12) and unimodal (*n* = 11) groups. The posttest was identical to the pretest.

First, the training phase involved mere exposure of the eight tokens generated from the male /na:/ mid-level to falling lexical tone continuum. The critical manipulation pertinent to the two types of statistical distribution was the number of times that each token would be heard. In bimodal condition, Stimuli 2 and 7 were played most frequently; however, in unimodal condition, Stimuli 4 and 5 occurred more often than other sounds (see Figure 1). Importantly, participants listened to the two endpoints of Stimuli 1 and 8 (representing prototypical mid-level and falling tones, respectively) an equal number of times in both conditions. Figure 4 exhibits stimulus pitch contours and occurrences in training. There were 256 training tokens in both distribution conditions. In brief, the occurrence frequencies of Stimuli 1 to 8 in bimodal distribution are 16, 64, 32, 16, 16, 32, 64, and 16; the occurrence frequencies of Stimuli 1 to 8 in unimodal distribution are 16, 16, 32, 64, 64, 32, 16, and 16 (Maye et al., 2002; Ong et al., 2017a, 2017b). Besides, an additional 32 beeps were randomly interspersed among these training items (Ong et al., 2017a). Participants in both conditions hence needed to listen to 288 sounds in all; moreover, each of them was guided to mark on a paper handout whenever the beep was heard. Their performance in the auditory vigilance task was later checked to ensure that they were indeed paying attention in the training phase.

<Insert Figure 4 around here>

The immediate posttest began after the training session (Escudero & Williams, 2014; Ong et al., 2017a). Both the pretest and posttest were ABX discrimination tasks which required participants to judge whether the third sound, X, was similar to the first sound, A, or the second one, B. Participants gave their answers by pressing the corresponding keys on the computer keyboard. Given amusics' degraded lexical tone perception (Chen & Peng, 2018; Nguyen et al., 2009; Zhang et al., 2017b), there was no time limit set for each trial but all participants were encouraged to respond as quickly as possible. The inter-stimulus-interval was 1,000 ms in line with previous studies (Ong et al., 2015a, 2016, 2017a). Recall that there were three exemplars generated for each minimal tone pair. The two endpoints of Stimuli 1 and 8 in the continuum based on an exemplar were always used as the first and second sounds in the ABX format, while the third sound was one of the two other exemplars generated within each minimal pair. This ensures that the present experiment remains comparable with other distributional learning studies (e.g., Escudero et al., 2011; Escudero & Williams, 2014; Ong et al., 2015a, 2017a). Each minimal tone pair was presented eight times in a random order, leading to a total of 32 trials for the pretest and posttest.

All stimuli across test sessions were presented via E-Prime 2.0 (Schneider et al., 2002). Prior to the pretest, participants needed to familiarize themselves with the experimental protocols. The practice trials did not contain the stimuli used in the formal test. Participants started the main experiment after they were familiar with the procedures. It took approximately 30 minutes, including the five-minute training, for participants to complete the entire experiment.

2.4 Data analysis

Accuracy rates were analyzed in the pretest and posttest. For each listener, the response to each trial was coded as 0 or 1 (incorrect or correct). The statistical analysis was conducted using generalized mixed-effects models with the lme4 package (Bates et al., 2015) in R (R Core Team, 2014). The models with group (amusics and musically intact listeners), test session (pretest and posttest), modal distribution (bimodal and unimodal), base syllable (trained and untrained), and speaker gender (trained and untrained) revealed an interaction among these five variables, [χ^2 (1) = 5.55, $p < .05$]. Given that there were two parameters manipulated in stimulus characteristics, responses to base syllable and speaker gender were further separately analyzed in order to estimate whether perceptual gains from pretest to posttest could emerge on these two test dimensions, following previous studies (Maye et al., 2002, 2008; Ong et al., 2015a, 2015b, 2016, 2017a).

To be specific, the models were established with group (amusics and musically intact listeners), test session (pretest and posttest), modal distribution (bimodal and unimodal), and base syllable (trained and untrained) acting as fixed factors with the dependent variable of accuracy rates; meanwhile, similar models were built with group (amusics and musically intact listeners), test session (pretest and posttest), modal distribution (bimodal and unimodal), and speaker gender (trained and untrained) acting as fixed factors considering participants' perceptual accuracy. Two-way, three-way, and four-way interaction terms were also included as fixed effects in the models. Working memory was the controlled covariate when fitting the models. For all possible fixed factors, by-subject and by-item random slopes and intercepts were included in the initial model (Barr et al., 2013). The initial model was compared to a simplified model which excluded a specific fixed factor using the anova function in lmerTest package (Kuznetsova et al., 2017). The *p* values for the main and interaction effects were obtained by applying likelihood ratio tests via the afex function in R (Winter, 2019). Pairwise comparisons were implemented with Tukey adjustment using the esmeans package (Lenth & Lenth, 2018).

3 Results

All participants succeeded in identifying the 32 beeps with 100% accuracy in the concurrent auditory vigilance task, hence none of them were excluded from data analysis. Table 4 documents the accuracy rates obtained by amusics and musically intact listeners in pretest and posttest. Amusics were, on average, outperformed by typical individuals as their scores in pretest and posttest were lower than the matched controls.

<Insert Table 4 around here>

3.1 Perceptual Accuracy on Test Dimension of Base Syllable

Figure 5 plots accuracy in pretest and posttest by amusics and musically intact individuals as faceted by distribution condition and test dimension of base syllable, with error bars representing one standard error (SE). This figure shows that amusics were scored lower in general than typical listeners; besides, in contrast to unimodal condition, both groups demonstrated an increase in accuracy in posttest than pretest in bimodal condition.

<Insert Figure 5 around here>

The mixed-effects models firstly revealed a significant main effect of test session [χ^2 (1) = 12.04, $p < .001$] and an interaction between test session and modal distribution [χ^2 (1) = 10.93,

$p < .001$]. Further analysis of this interaction showed that in pretest, the bimodal group did not differ from the unimodal group formed by amusic or control listeners ($\beta = -0.32$, SE = 0.20, t = -1.57, $p = .12$), suggesting that the bimodal group had similar performance as the unimodal counterpart before training. In posttest, the bimodal group achieved higher scores than the unimodal counterpart ($\beta = 0.59$, SE = 0.23, t = 2.52, $p < .05$). Crucially, regardless of typical or amusic participants, the scores obtained by bimodally trained listeners in pretest were significantly lower than in posttest ($\beta = -0.93$, SE = 0.19, t = -4.84, $p < .001$), but the pretest scores did not differ from the posttest scores among unimodally trained listeners ($\beta = -0.02$, SE = 0.19, t = -0.10, $p = .92$). This implied perceptual gains for amusics and controls who were exposed to bimodal relative to unimodal distributions. Besides, the interaction between test session and group was insignificant [χ^2 (1) = 1.08, $p = .30$], which suggested parallel improvement between amusics and controls.

There were also a main effect of group [χ^2 (1) = 31.19, $p < .001$] and a three-way interaction between group, modal distribution, and base syllable [χ^2 (1) = 4.66, $p < .05$]. This three-way interaction was further analyzed by keeping the fixed factors of group and modal distribution with the dependent variable being respective accuracies on trained or untrained syllable. The models revealed a significant main effect of group [χ^2 (1) = 26.09, $p < .001$] for the trained syllable, indicating that amusics ($M = 0.83$, $SD = 0.38$) were outperformed by controls ($M = 0.96$, $SD = 0.20$) either in bimodal or unimodal condition. For the untrained syllable, the main effect of group [χ^2 (1) = 10.09, $p < .01$] again suggested that typical listeners ($M = 0.87$, $SD = 0.34$) obtained higher scores than control listeners ($M = 0.80$, $SD = 0.40$) regardless of modal distribution.

In summary, on the test dimension of base syllable, amusics were graded lower than the matched controls, implicating their poorer lexical tone perception than typical listeners which was in line with preceding studies on amusics' perception of tone (e.g., Jiang et al., 2012; Tillmann et al., 2011; Zhang et al., 2018). Nonetheless, similar to typical listeners, amusics' accuracy rates in lexical tone perception for both trained and untrained syllables improved from pretest to posttest as being bimodally trained, indicative of their possibly spared mechanism of distributional learning for linguistic pitch information even though it was from the second language with a novel tone inventory (Escudero & Williams, 2014). Next, accuracy rates on the test dimension of speaker gender were analyzed.

3.2 Perceptual Accuracy on Test Dimension of Speaker Gender

Figure 6 depicts accuracy in pretest and posttest by amusics and typical listeners as faceted by distribution condition and test dimension of speaker gender, with error bars representing one SE. This figure exhibits amusics' degraded speech processing as their accuracy generally remained lower than controls; meanwhile, in bimodal condition, accuracies obtained by both groups were higher in the posttest than in the pretest, yet this was not obvious in unimodal condition.

<Insert Figure 6 around here>

Similar to the test dimension of base syllable, the mixed-effects models uncovered a significant main effect of test session [χ^2 (1) = 13.50, $p < .001$] and an interaction between test session and modal distribution [χ^2 (1) = 12.34, $p < .001$]. Further analysis of this interaction showed that in pretest, listeners in bimodal condition did not differ from those in unimodal condition regardless of being amusics or typical individuals (β = -0.30, SE = 0.20, t = -1.49, *p*

= .14), while in posttest, listeners in bimodal condition outperformed those in unimodal condition ($\beta = 0.70$, SE = 0.25, t = 2.79, $p < .01$). Noteworthy was that pretest scores were significantly lower than posttest scores for bimodally trained amusic or control listeners ($\beta = -1.02$, SE = 0.21, t = -4.79, $p < .001$), yet there was no difference between pretest and posttest scores for unimodally trained amusic or control listeners ($\beta = -0.02$, SE = 0.19, t = -0.1, $p = .92$).

The models also revealed the significant main effects of group [χ^2 (1) = 33.01, $p < .001$] and speaker gender [χ^2 (1) = 7.06, $p < .01$], and their two-way interaction [χ^2 (1) = 17.46, $p < .001$]. Further analysis of this interaction showed that control listeners obtained lower scores for the tokens uttered by the speaker with the novel gender than the trained gender ($\beta = 0.70$, SE = 0.29, t = 2.39, $p < .05$), suggesting typical listeners were more familiar with Thai tones recorded by the speaker with the trained gender. Notably, amusics ($M = 0.83$, $SD = 0.37$) were graded lower than controls ($M = 0.96$, $SD = 0.19$) for the tokens uttered by the speaker with the trained gender ($\beta = -1.74$, SE = 0.26, t = 6.62, $p < .001$); moreover, amusics ($M = 0.79$, $SD = 0.40$) obtained lower scores than controls ($M = 0.86$, $SD = 0.34$) for the tokens recorded by the speaker with the novel gender ($\beta = -0.58$, SE = 0.19, t = -3.09, $p < .001$). In addition, the interaction between group and test session was insignificant [χ^2 (1) = 1.62, $p = .20$], which indicated the similar magnitude of improvement between amusics and controls. Other effects were non-significant (ps > .05).

In summary, on the test dimension of speaker gender, amusics performed less accurately than their matched counterparts across distribution conditions. This suggested that amusics' lexical tone perception in speech was reduced than typical listeners in accordance with the well-documented finding that amusia negatively affects speech perception (Liu et al., 2021).

However, as test session effects were disclosed in bimodal condition in all of the aforementioned analyses, amusics' distributional learning involving linguistic pitch information was likely to be preserved.

4 Discussion

The current study investigated whether bimodally trained individuals with amusia, a musical disorder affecting pitch processing across domains of music and language, would be able to learn phonological pitch information stemming from a novel lexical tone inventory. By presenting Chinese-speaking amusics with real Thai monosyllabic words (Ong et al., 2017a), it was found that stimulus frequency of statistical distribution influenced listeners' learning outcome; that is, when amusics were exposed to the bimodal condition with two sounds near the endpoints occurring most frequently, scores in the posttest were higher than the pretest. This result was not likely due to practice or repetition effects, given that similar to the trained tokens, the scores in non-trained novel stimuli (different from the trained tokens in test syllable or test speaker gender) were also improved from pretest to posttest. This was in accordance with Maye et al. (2008), suggesting that listeners had acquired the target dimension via distributional learning, which referred to lexical pitch herein. Listeners in the unimodal condition did not increase their scores after being trained, presumably due to a single category they built in the training phase (Maye et al., 2002, 2008; Ong et al., 2017a, 2017b; Thiessen & Erickson, 2013).

4.1 Reduced Processing of Thai Lexical Tones in Amusics

To begin with, we corroborated the well-established finding that amusics' perception of lexical tone was degraded as compared with musically intact listeners (Chen & Peng, 2018, 2020; Nguyen et al., 2009; Tillmann et al., 2011), even though amusics were trained with Thai stimuli in a distributional learning experiment.

The current study followed the pretest-training-posttest design, with all procedures being identical between test sessions including exactly the same test manner and test items (e.g., discrimination in ABX format and two test dimensions of the stimuli). Results of pretest showed that amusics obtained lower accuracy than musically intact listeners. Analogous to pretest, results of posttest showed that amusics performed worse than controls. This study hence replicated and reinforced the prior finding that amusics' lexical tone perception was degraded as compared with typical listeners (e.g., Shao et al., 2019; Tillmann et al., 2011; Zhang et al., 2017b). Moreover, this study firstly verified that amusics' lexical tone perception could not reach a level similar to typical listeners, even though they have finished perceptual training in distributional learning. This could result from two factors including the deficit of amusia and the training process. Firstly, with regard to the deficit of congenital amusia, prior studies have validated that amusia is an inborn pitch-processing disorder (Vuvan et al., 2015). It has been proven that lexical tone consists of two levels of information, namely, low-level acoustic cues and high-level tonal categories (Xi et al., 2010; Yu et al., 2017, 2019), and perception of lexical tone depends on listeners' encoding of these two types of information (Chen & Peng, 2018; Peng et al., 2010; Zhao & Kuhl, 2015). Amusics' degraded low-level pitch processing is assumed to extend to high-level phonological representations, which as a consequence become weakened or less stable than normal individuals (Chen & Peng, 2018, 2020). In the process of distributional learning of novel linguistic items of Thai tones, because of impairments in acoustic processing of pitch (Jiang et al., 2012; Vuvan et al., 2015), amusics may develop less robust phonological representations of tonal categories in comparison to musically intact listeners (Jiang et al., 2012; Chen & Peng, 2018, 2020). Presumably, deficiencies in both acoustic processing and phonological representations of lexical tones integrally led to amusics' poorer performance than controls despite distributional learning.

The second factor concerning amusics' reduced accuracy than typical listeners in posttest possibly stems from the absence of feedback in the training process. Peretz (2016) pointed out that any training programs to improve amusia would be more positive when amusics were provided with external feedback. Although the current study was not aimed at amusics' intervention but focused on the question of whether amusics' distributional learning of pitch could be spared, future studies related to amusics' learning/treatment could attempt to provide feedback in the training process in the hope of appraising whether amusia as a special developmental disorder can be mitigated. Meanwhile, one study by Escudero and Williams (2014) found that learning effects elicited by a modified distributional learning paradigm were far-reaching and apparent even after six and 12 months. In their training session, those listening to exaggerated vowels, i.e., in infant-directed speech style, reached a perceptual plateau earlier than those listening to normal vowels with original acoustic parameters, indicative of more rapid or efficient distributional learning effects. Previous literature shows that amusics' insensitivity to pitch information is tightly connected with abnormally elevated pitch thresholds (Foxton et al., 2004; Huang et al., 2015a, 2015b; Vuvan et al., 2015). In light of the above discussion, future studies could adopt infant- or foreigner-directed speech as the training tokens, in an effort to further investigate whether the modified paradigm, i.e., the combination of the exaggerated speech and distributional learning, would contribute to amusics' greater perceptual gains approaching control listeners in distributional learning.

4.2 Preserved Distributional Learning of Thai Tones in Amusics

One more intriguing finding in this study involved the preserved distributional learning of Thai in amusics. Given that all participants were native Mandarin Chinese speakers who had no prior knowledge of Thai, this finding extended the existing literature by showing that amusics were able to learn a second language with novel tone inventory through distributional learning despite amusia.

Firstly, in line with the studies by Ong and colleagues (Ong et al., 2015a, 2015b, 2016, 2017a), bimodally trained listeners in the two groups improved their scores in contrast to their unimodally trained counterparts. This was thought to result from the tracking of stimulus frequency of statistical distribution (Thiessen, 2017). In the training phase, the distribution of auditory items reflected the intended categories of the continuum, thereby bimodal training led listeners to establish two distinct clusters of tokens along the continuum, yet only one cluster was presented through unimodal training (Hayes-Harb, 2007). With this regard, a bimodal distribution elicited listeners' perception of the test stimuli as exemplars from two disparate sound categories in speech, whereas a unimodal distribution induced listeners' perception of the test stimuli as exemplars from a single speech sound category (Erickson & Thiessen, 2015; Wanrooij et al., 2013). Similar to typical listeners, amusics who received bimodal rather than unimodal training increased their scores from pretest to posttest. This improvement was found not only for the trained tokens (male /na:/ tone pair), but also for the non-trained novel ones whose base syllable was shifted (i.e., the test dimension of base syllable: /na:/ and /k^ha:/) or which were recorded by the speaker with the different gender (i.e., the test dimension of speaker gender: male and female). This offered persuasive evidence that instead of practice or repetition effects from training, amusics were capable of extending what they had learned in "old" tokens (lexical tone categories) to those "new" ones (with similar lexical pitch register and contour), which shared the common critical feature in perception (Maye et al., 2008).

The generalization effects found from the current study were not unexpected. In a similar vein, prior studies of distributional learning have widely reported the generalization effects in auditory perception, for example, from trained to novel Hindi consonants with variable voice onset times (Maye et al., 2008), from synthetic to naturally produced Dutch vowels (Escudero et al., 2011), and from trained to novel Thai lexical tones characterized by different F0 akin to the current study (Ong et al., 2015a, 2015b, 2016, 2017a).

A line of novel evidence for amusics' statistical learning was therefore provided by the current study. Several previous studies claimed that amusics' statistical learning of language was preserved since amusics could track regularities of transitional probabilities that distinguished within-word and between-word syllables similarly to control listeners; nevertheless, the training and test stimuli they exploited involved words whose pitch information was removed (Omigie & Stewart, 2011) or did not play a phonological role (Peretz et al., 2012). Besides, although Zhu et al. (2022) revealed amusics' successful computation of speech when pitch was involved, their tokens that amusic participants heard were however from their native language. As a result, it was unclear whether amusics were truly able to compute the statistics embedded in the linguistic input with novel tone categories in the process of learning. Using real Thai monosyllabic words as new materials to participants who had no pre-existing knowledge of Thai, the current study revealed that amusic and control listeners showed a consistent pattern in performance profiles in that bimodally trained listeners obtained improvement while the unimodally trained counterparts did not (Maye et al., 2002; Escudero et al., 2011; Wanrooij et al., 2013). Moreover, as mentioned in the section of Method, the test stimuli for each minimal pair were solely discrepant in F0 because a common base waveform was chosen to bear the contrastive lexical tones, following Ong and colleagues (Ong et al., 2015a, 2017a). For that reason, amusics' statistical learning of pitch which plays a *phonological* role in a second language was considered preserved despite the pitch-processing disorder of amusia.

One potential explanation for amusics' consistent performance as typical listeners in distributional learning of speech might be the hierarchy of impairments of amusia. Namely, as a developmental disorder of pitch processing (Vuvan et al., 2015), amusia leads to severe impairments in music processing but only mild impairments in speech processing (Ayotte et al., 2002; Hutchins et al., 2010; Liu et al., 2021). According to previous questionnaire-based data, congenital amusics are always struggling to experience music such as their inability to recognize familiar music without the help of lyrics (Tillmann et al., 2014), which was also reported by the current amusic sample. However, amusics seldom report having difficulties in speech (e.g., in Mandarin, Jiang et al., 2010; in English, Liu et al., 2010). This was possibly because music and speech are different in certain aspects. For instance, although both music and speech involve pitch for communicative purposes, the use of pitch in music can express composition (e.g., the key forming a piece) and affect (e.g., "sadness" conveyed by a minor chord), yet pitch shifts in speech depict pragmatic (e.g., intonation) and semantic (e.g., lexical tone) meanings (Alexander et al., 2008). The execution of pitch patterns mainly needs to satisfy contrastive adequacy for representing functional contrasts in speech (Xu, 2005), which are however required to be perfected in music for musical performance (Patel, 2008; Liu et al., 2013). Given music and speech as form-driven and function-driven respectively (Patel, 2008), amusics may have less difficulty in processing linguistic pitch variations and can be aided by cues other than linguistic pitch in context (Nan et al., 2010). In this regard, Chinese-speaking amusics in the present study were capable of learning a second language with novel tone inventory via distributional learning, even if they were afflicted with the pitch disorder of amusia.

An alternative account regarding amusics' preserved distributional learning concerns their spared abilities of categorization across cognitive domains (Hutchins et al., 2010; Jiang et al., 2012; Pralus et al., 2019). Categorization is central for successful distributional learning since category membership developed in training impacts on listeners' sound classification in posttest after exposure (Erickson & Thiessen, 2015). In spite of impaired pitch processing, amusics can establish mental linguistic or musical lexicon relating to pitch. For example, in the domain of language, Chen and Peng (2018) found that amusics were outperformed by typical listeners in identifying and discriminating lexical tones; nevertheless, similar to typical individuals, amusics obtained higher accuracy in perceiving between-category lexical tone comparisons than within-category lexical tone counterparts, which serves as a milestone of categorical perception in line with Xu et al.'s (2006) model of lexical tone categorization. With the aid of the technique of event-related potentials, Chen and Peng (2020) further revealed amusics' degraded lexical tone perception but comparable perceptual patterns to categorize lexical tones as controls. Likewise, being an important pitch-related unit in speech, intonation was perceived with lower accuracy in amusics than controls; however, despite amusics' insensitivity to minor pitch variations of intonations, they could readily identify whether the pitch contour belonged to the type of statement or question (e.g., Hutchins et al., 2010; Patel et al., 2005, 2008). This is similar to amusics' categorization of musical materials. Tillmann et al. (2014) played instrumental musical excerpts to amusics and typical listeners and asked them to provide relevant familiarity judgement; notably, these musical pieces had been segmented into different lengths, hence listeners needed to rate the familiarity of the stimuli with variable acoustic information which could be sufficient or insufficient for them to access the mental musical lexicon. Results showed that amusics were able to judge the stimulus familiarity as controls even though minimal acoustic information was heard, suggesting amusics' formation of the mental musical lexicon despite amusia profoundly interfering with their musical pitch processing in everyday life. In brief, the aforementioned studies point to an agreement that amusics are able to form high-level mental representations about pitch across cognitive domains despite their abnormal pitch processing (Chen & Peng, 2018, 2020; Cheung et al., 2021; Hutchins et al., 2010; Jiang et al., 2012; Pralus et al., 2019), which is necessary for them to establish sound categories and complete the classification task of distributional learning (Erickson & Thiessen, 2015; Thiessen, 2017).

We also notice a potential account regarding the native language's influence on the non-native language's learning (e.g., Best, 1995; So & Best, 2014). The test lexical tones were mid-level and falling tones in Thai, which can be transcribed as Thai Tone 33 and Tone 241, individually (Ong et al., 2017a), with the numerals representing the relative pitch level within a talker's frequency range (1 as the lowest and 5 as the highest, Chao, 1930). Our amusics spoke

Mandarin Chinese which contains four lexical tones as high-level (Tone 55), mid-rising (Tone 35), low-dipping (Tone 214), and high-falling (Tone 51) tones (Burnham et al., 2015). A recent study by Best and colleagues has found that for Chinese-speaking listeners, Thai Tone 33 was categorized as Chinese Tone 55, since both of them belong to the level tones; however, Thai Tone 241 remained uncategorized, because this tone is phonetically rising-falling and Chinese does not have such a phonological category (Chen et al., 2020). We additionally analyzed accuracies of respective mid-level and falling tones and found that controls were more sensitive to pitch information in these two tones than amusics, but both groups perceived the mid-level tone more accurately than the falling tone. It was likely that the language experience of Chinese in our listeners affected how they recognized Thai tones; namely, they might perceive Thai Tone 33 as Chinese Tone 55, but Thai Tone 241 as a new tone or a tone candidate between Chinese Tone 55 and Tone 51 (Chen et al., 2020). This could lead to the higher accuracy for the mid-level tone than the falling tone. Based on the current study as a preliminary endeavor, future studies may want to shed light on the question of how similarity to native tonal categories affects the perception and learning of L2 tonal categories. Related to this point, it may also be interesting to compare amusics with tonal and non-tonal language backgrounds in distributional learning of non-native lexical tones.

In addition, we note that for bimodally trained listeners, amusics showed parallel improvement as typical, musically intact individuals. A caveat is that this finding does not eliminate possibly different results in future studies, especially those with a more complex design than the current experiment. For example, in the amusia literature, previous studies have identified that amusics were poorer than controls in lexical tone perception (Chen & Peng, 2018;

Jiang et al., 2012; Tillmann et al., 2011). However, most of these studies examined amusics with lexical tones in a low-variability context (a small number of types of lexical tones, base syllables, speakers, etc.). More recently, by using high-variability tokens, amusics are further revealed with a possible “anchoring” deficit (Liu et al., 2021; Shao & Zhang, 2018; Shao et al., 2019), a deficit hypothesis originally proposed to explain dyslexia (Ahissar, 2007; Ahissar et al., 2006). Shao et al. (2019) designed low-variation condition of syllable, high-variation condition of syllable, low-variation condition of talker, and high-variation condition of talker to examine amusics’ lexical tone perception. Their results showed that as compared with amusics, typical listeners displayed a greater magnitude of increase in discrimination sensitivity from high-variation to low-variation conditions. Notably, Shao et al. (2019) employed 24 words carrying all the six Cantonese tones. These stimuli created a high-variability context that possibly triggers group differences in the magnitude of increase between amusics and controls, which was not always observed in previous studies whose stimuli were relatively simple. With this respect, although our study did not find the magnitude differences between groups, future experiments with a different and more sophisticated method could likely trigger any possible magnitude differences in terms of distributional learning between amusic and typical individuals.

To summarize, amusics’ improvement of accuracy rates from pretest to posttest manifested their preserved distributional learning mechanism. Notably, this development was observed in both trained and non-trained novel Thai lexical tones carried by different base syllable and different speaker gender. This was in line with the transfer effects among featurally analogous speech contrasts as reported in prior studies (Maye et al., 2008; Ong et al., 2015a, 2015b). The observed performance was largely due to the built categories in training, which allowed amusics to classify different sound identities in posttest.

4.3 Limitations and Future Directions

We conclude by noting the limitations of the study. First of all, it remains less clear as to whether our findings would hold to different sub-types of amusics. The amusia literature documents two sub-types of amusics in terms of their lexical tone processing abilities, known as pure amusics and tone agnosics (Nan et al., 2010). Tone agnosics consistently showed their insensitivity in lexical tone perception than pure amusics and typical, musically intact individuals (Huang et al., 2015a, 2015b; Nan et al., 2016; Tang et al., 2018). Because the current study did not separate amusics as pure amusics and tone agnosics, future experiments with this special group (i.e., tone agnosics diagnosed following Nan et al., 2010) would further contribute to the understanding of distributional learning as a function of amusia. As a young and active research area, this topic of processing statistical information and acoustic cues of speech from the unique window of amusia is likely to prompt the development of research on statistical learning (Saffran, 2003; Erickson & Thiessen, 2015; Thiessen, 2017).

Second, although the current results were principally attributed to distributional learning in line with previous studies (e.g., Escudero & Williams, 2014; Maye et al., 2002, 2008; Ong et al., 2017a), a very recent exploratory study has identified that listeners' prior language experience might affect their performance in the post-training test (Chládková & Šimáčková, 2021). With this respect, as discussed earlier, it is yet unclear about whether findings from amusics as native and skillful lexical tone users could be extended to those amusics who speak non-tonal languages. Future studies may recruit English-speaking amusics, for example, to investigate second language learners' distributional learning influenced by amusia. Third, a more ecological environment is preferable than a laboratory setting. Future experiments could adopt multi-talker tokens or tones with various base syllables to develop the high-variability context among amusic and control listeners (e.g., Shao et al., 2019). Any group differences in the magnitude of improvement of distributional learning are likely to be triggered by this challenging auditory environment. Fourth, expect for the short-term effects of distributional learning found in this study, future studies may include the delayed posttest(s) to further evaluate whether long-term effects of distributional learning reported in typical listeners (e.g., after six months, Escudero & Williams, 2014) would also be present in amusics. Meanwhile, future studies may also translate the current results to real-life learning situations in which more dynamic factors are at interplay. This would not only advance the understanding of amusia but also pave the way for researchers, pathologists, and educators to support amusics in acquiring a second tone language, earning a meaningful and happy life at their will.

5 References

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech, 21*(4), 319–325. https://doi.org/10.1111/cogs.13198

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

Alexander, J. A., Bradlow, A. R., Ashley, R., & Wong, P. C. M. (2008). Music melody perception in tone-language and nontone-language speakers. Poster presented at the 156th Meeting of Acoustical Society of America, Miami, FL.

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/awf028

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 using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Best, C. T. (1995). *A direct realist view of cross-language speech perception* (W. Strange, Ed.). York Press.

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer. Version 6.0.37. Available at: http://www.praat.org.

Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., Rattanasone, N. A., Schwarz, I. C., Webster, D. (2015). Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics, 36*(6), 1459–1491. https://doi.org/10.1017/S0142716414000496

Chao, Y. R. (1930). A system of tone-letters. *Le Maître Phonétique, 45*, 24–27.

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, J., Best, C. T., & Antoniou, M. (2020). Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai lexical tones by Mandarin and Vietnamese listeners. *Journal of Phonetics, 83*, 101013. https://doi.org/10.1016/j.wocn.2020.101013

Cheung, Y. L., Zhang, C., & Zhang, Y. (2021). Emotion processing in congenital amusia: The deficits do not generalize to written emotion words. *Clinical Linguistics & Phonetics, 35*(2),101–116. https://doi.org/10.1080/02699206.2020.1719209

Chládková, K., & Šimáčková, Š. (2021). Distributional learning of speech sounds: An exploratory study into the effects of prior language experience. *Language Learning, 71*(1), 131–161. https://doi.org/10.1111/lang.12432

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review, 37*, 66–108. https://doi.org/10.1016/j.dr.2015.05.002

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America*, *130*(4). EL206–EL212.

Escudero, P., & Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition, 133*(2), 408–413. https://doi.org/10.1121/1.3629144

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

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*. Changsha: Hunan Medical College.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research, 23*(1), 65–94. https://doi.org/10.1177/0267658307071601

Huang, W. T., Liu, C., Dong, Q., & Nan, Y. (2015a). 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. (2015b). 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

Huang, X., Zhang, C., Shi, F., Yan, N., & Wang, L. (2016). Impaired vowel discrimination in Mandarin-speaking congenital amusics. *Proceedings of the 5th International Symposium on Tonal Aspects of Languages*. 138–141.

Hutchins, S., Gosselin, N., & Peretz, I. (2010). Identification of changes along a continuum of speech intonation is impaired in congenital amusia. *Frontiers in Psychology, 1*, 236. https://doi.org/10.3389/fpsyg.2010.00236

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

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

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

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

Lenth, R., & Lenth, M. R. (2018). Package 'lsmeans.' *The American Statistician, 34*(4), 216–221.

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

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

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*, 104908. https://doi.org/10.1016/j.bandl.2021.104908

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*(1), 122–134. https://doi.org/10.1111/j.1467-7687.2007.00653.x

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101–B111. https://doi.org/10.1016/S0010-0277(01)00157-3

Moulines, E., & Laroche, J. (1995). Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Communication, 16*(2), 175–205. https://doi.org/10.1016/0167-6393(94)00054-E

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

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

Ong, J. H., Burnham, D., & Escudero, P. (2015a). Distributional learning of lexical tones: A comparison of attended vs. unattended listening. *PLoS One, 10*(7), e0133446. https://doi.org/10.1371/journal.pone.0133446

Ong, J. H., Burnham, D., & Escudero, P. (2015b). Mandarin listeners can learn non-native lexical tones through distributional learning. *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, Scotland: International Phonetic Association.

Ong, J. H., Burnham, D., Escudero, P., & Stevens, C. J. (2017a). Effect of linguistic and musical experience on distributional learning of nonnative lexical tones. *Journal of Speech, Language, and Hearing Research, 60*(10), 2769–2780. https://doi.org/10.1044/2016_JSLHR-S-16-0080

Ong, J. H., Burnham, D., & Stevens, C. J. (2017b). Learning novel musical pitch via distributional learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(1), 150–157. https://doi.org/10.1037/xlm0000286

Ong, J. H., Burnham, D., Stevens, C. J., & Escudero, P. (2016). Naïve learners show cross-domain transfer after distributional learning: The case of lexical and musical pitch. *Frontiers in Psychology, 7*, 1189. https://doi.org/10.3389/fpsyg.2016.01189

Patel, A. D. (2008). *Music, Language, and the Brain*. New York, NY: Oxford University Press.

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

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

Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., & Wang, W. S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics, 38*(4), 616–624. https://doi.org/10.1016/j.wocn.2010.09.003

Peretz I. (2001). Brain specialization for music: New evidence from congenital amusia. *Annals of the New York Academy of Sciences, 930*, 153–165. https://doi.org/10.1093/acprof:oso/9780198525202.003.0013

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., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences, 999*(1), 58–75. https://doi.org/10.1196/annals.1284.006

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., 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), 361–367. 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

Pralus, A., Fornoni, L., Bouet, R., Gomot, M., Bhatara, A., Tillmann, B., & Caclin, A. (2019). Emotional prosody in congenital amusia: Impaired and spared processes. *Neuropsychologia, 134*, 107234. https://doi.org/10.1016/j.neuropsychologia.2019.107234

R Core Team. (2014). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: 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

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's Guide*. Pittsburgh, PA: Psychlogy Software Incorporated.

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

Shen, G., & Froud, K. (2016). Categorical perception of lexical tones by English learners of Mandarin Chinese. *The Journal of the Acoustical Society of America, 140*(6), 4396–4403. https://doi.org/10.1121/1.4971765

Shen, G., & Froud, K. (2019). Electrophysiological correlates of categorical perception of lexical tones by English learners of Mandarin Chinese: An ERP study. *Bilingualism: Language and Cognition, 22*(2), 253–265. https://doi.org/10.1017/S136672891800038X

So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition, 36*(2), 195–221. https://doi.org/10.1017/S0272263114000047

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

Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1711), 20160056. https://doi.org/10.1098/rstb.2016.0056

Thiessen, E. D., & Erickson, L. C. (2013). Beyond word segmentation: A two-process account of statistical learning. *Current Directions in Psychological Science, 22*(3), 239–243. https://doi.org/10.1177/0963721413476035

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, 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

Wanrooij, K., Escudero, P., & Raijmakers, M. E. J. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics, 41*(5), 307–319. https://doi.org/10.1016/j.wocn.2013.03.005

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science, 21*(4), 221–226. https://doi.org/10.1177/0963721412449459

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., & 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

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

Winter, B. (2019). *Statistics for linguistics: An introduction using R.* Routledge. https://doi.org/10.4324/9781315165547

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

Wu, X., Munro, M. J., & Wang, Y. (2014). Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *Journal of Phonetics, 46*, 86–100. https://doi.org/10.1016/j.wocn.2014.06.005

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience, 170*(1), 223–231. https://doi.org/10.1016/j.neuroscience.2010.06.077

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication, 46*, 220–251. https://doi.org/10.1016/j.specom.2005.02.014

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

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

Yu, K., Zhou, Y., Li, L., Su, J., Wang, R., & Li, P. (2017). The interaction between phonological information and pitch type at pre-attentive stage: An ERP study of lexical tones. *Language, Cognition and Neuroscience, 32*(9), 1164–1175. https://doi.org/10.1080/23273798.2017.1310909

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

Zhang, C., Shao, J., & Huang, X. (2017b). Deficits of congenital amusia beyond pitch: Evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics. *PLoS One, 12*(8), e0183151. https://doi.org/10.1371/journal.pone.0183151

Zhang, G., Shao, J., Huang, X., Wang, L., & Zhang, C. (2018). Unequal impairment of native and non-native tone perception in Cantonese speakers with congenital amusia. *Proceedings of the 9th International Conference on Speech Prosody*, Poznań, Poland. 562–566. https://doi.org/10.21437/SpeechProsody.2018-114

Zhao, T. C., & Kuhl, P. K. (2015). Higher-level linguistic categories dominate lower-level acoustics in lexical tone processing. *The Journal of the Acoustical Society of America, 138*(2), EL133–EL137. https://doi.org/10.1121/1.4927632

Zhu, J., Chen, X., Chen. F., & Wiener, S. (2022). Individuals with congenital amusia show impaired speech perception but preserved statistical learning for tone languages. *Journal of Speech, Language, and Hearing Research, 65*(1), 53–69. https://doi.org/10.1044/2021_JSLHR-21-00383

Zsiga, E., & Nitisaroj, R. (2007). Tone features, tone perception, and peak alignment in Thai. *Language and Speech, 50*(3), 343–383. https://doi.org/10.1177/00238309070500030301

Tables

Table 1 Demographic characteristics of amusics and controls.

Subject information	**Amusics**	**Controls**	***p* values**
No. of participants (Female)	21 (12)	23 (12)	N/A
Age in years (SD)	19.38 (0.86)	19.39 (0.78)	*p*s > .05
Working memory (SD)	14.24 (1.09)	14.39 (0.94)	
Online identification test of congenital amusia (SD)			
Out-of-key	66.24 (8.61)	88.17 (6.91)	*p*s < .001
Offbeat	64.33 (14.10)	80.09 (8.91)	
Mistuned	52.95 (7.14)	85.74 (8.61)	
Global score	60.86 (6.44)	84.78 (5.42)	

Note. The *p* values indicate results from independent-samples *t*-tests between the two groups.

Table 2 Summary of eight training stimuli from the continuum of mid-level tone and falling tone of Thai. Steps 1 and 8 denote mid-level and falling tones, respectively. Duration (in ms) and fundamental frequency (F0, in Hz) at three normalized time points as well as the overall average of the tone space are tabulated.

Training Stimulus		Tone Space F0 (Hz)			
Step	**Duration (ms)**	**10%**	**50%**	**100%**	**Mean**
1	567	95	94	91	95
2	567	97	99	90	97
3	567	99	104	88	100
4	567	101	109	87	102
5	567	102	113	86	105
6	567	104	118	84	107
7	567	105	123	83	110
8	567	107	128	81	112

Table 3 Summary of all experimental items of Thai monosyllabic words used in pretest and posttest. Duration (in ms) and fundamental frequency (F0, in Hz) at three normalized time points as well as the overall average of the tone space are tabulated. The duration remains constant in each minimal pair though it differs across minimal pairs.

Speaker/Syllable	Test Item			Tone Space F0 (Hz)			
	Tone	Exemplar	Duration (ms)	10%	50%	100%	Mean
Male /na:/	Mid-level tone	1	567	95	94	91	94
	Mid-level tone	2	567	91	94	84	91
	Mid-level tone	3	567	93	95	93	94
	Falling tone	1	567	105	128	82	112
	Falling tone	2	567	102	120	92	114
	Falling tone	3	567	102	124	88	112
Male /k^ha:/	Mid-level tone	1	508	100	98	98	98
	Mid-level tone	2	508	101	97	96	97
	Mid-level tone	3	508	96	95	94	95
	Falling tone	1	508	123	126	85	115
	Falling tone	2	508	121	125	100	120
	Falling tone	3	508	121	126	78	110
Female /na:/	Mid-level tone	1	658	212	210	176	204
	Mid-level tone	2	658	215	204	180	202
	Mid-level tone	3	658	207	203	181	200
	Falling tone	1	658	239	251	182	232
	Falling tone	2	658	244	262	177	239
	Falling tone	3	658	252	278	169	245
Female /k^ha:/	Mid-level tone	1	770	231	216	216	218
	Mid-level tone	2	770	243	225	220	227
	Mid-level tone	3	770	239	217	211	220
	Falling tone	1	770	271	280	179	242
	Falling tone	2	770	270	264	182	235
	Falling tone	3	770	271	256	185	233

Table 4 Mean accuracy rates (SD) of Thai lexical tones in pretest and posttest by bimodally and unimodally trained amusics and musically intact listeners.

Test	Amusics		Controls	
	Bimodal	**Unimodal**	**Bimodal**	**Unimodal**
Pretest	0.77 (0.42)	0.82 (0.39)	0.89 (0.32)	0.90 (0.30)
Posttest	0.86 (0.35)	0.81 (0.39)	0.96 (0.19)	0.90 (0.30)

Figures

Figure 1 The schematic illustration of bimodal and unimodal distributions used in distributional learning studies. In bimodal distribution, Token 2 and Token 7 are heard more often, yet in unimodal distribution, Token 4 and Token 5 occur most frequently. Two endpoints have the same occurrence frequency across bimodal and unimodal distributions.

Figure 2 F0 curves of Thai mid-level tone and falling tone uttered by the male speaker.

Figure 3 F0 curves of Thai mid-level tone and falling tone uttered by the female speaker.

Figure 4 The (a) bimodal distribution and (b) unimodal distribution of Thai lexical tones settled in distributional learning. The eight training stimuli, generated from the pitch continuum, were reflected by different colors, with line types representing different frequencies of statistical distribution. The solid lines (on the left: Token 2 and Token 7; on the right: Token 4 and Token 5) denote the stimuli that were heard more often than other tokens.

Figure 5 Mean correct responses as measured on the test dimension of base syllable in different distribution conditions and test sessions by amusics and controls, with error bars showing one standard error of the mean.

Figure 6 Mean correct responses as measured on the test dimension of speaker gender in different distribution conditions and test sessions by amusics and controls, with error bars showing one standard error of the mean.

Figure 1

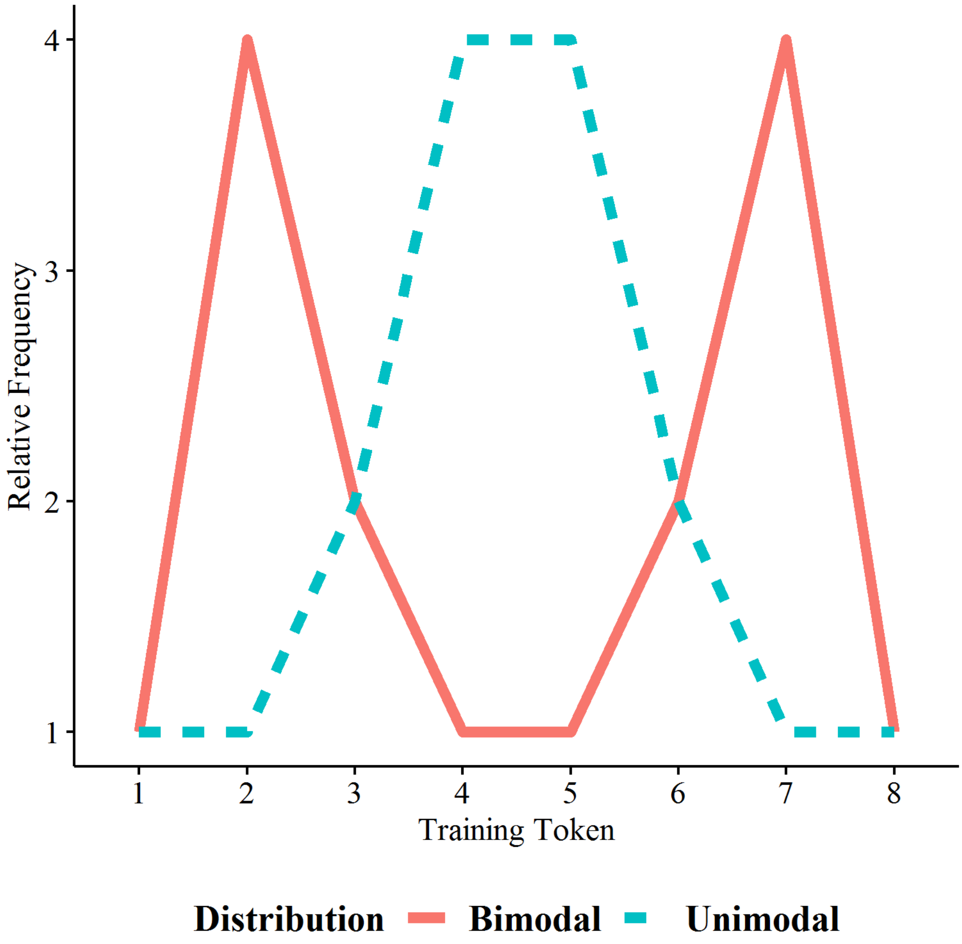

Relative Frequency
4
3
2
1
1
2
3
4
5
6
7
8
Training Token
Distribution
Bimodal
Unimodal

Figure 2

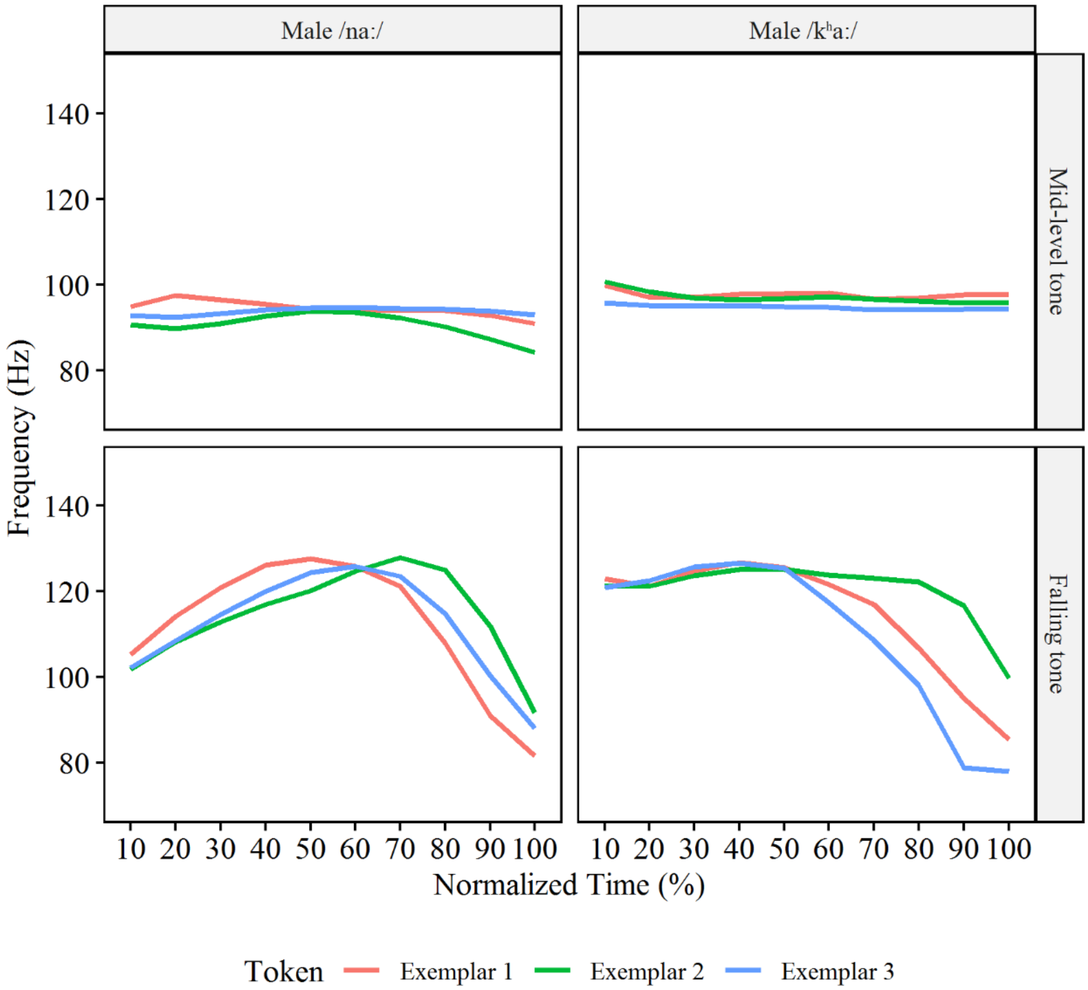

Male /na:/
Male /kʰa:/
Mid-level tone
Falling tone
Frequency (Hz)
140
120
100
80
10 20 30 40 50 60 70 80 90 100
Normalized Time (%)
Token
Exemplar 1
Exemplar 2
Exemplar 3

Figure 3

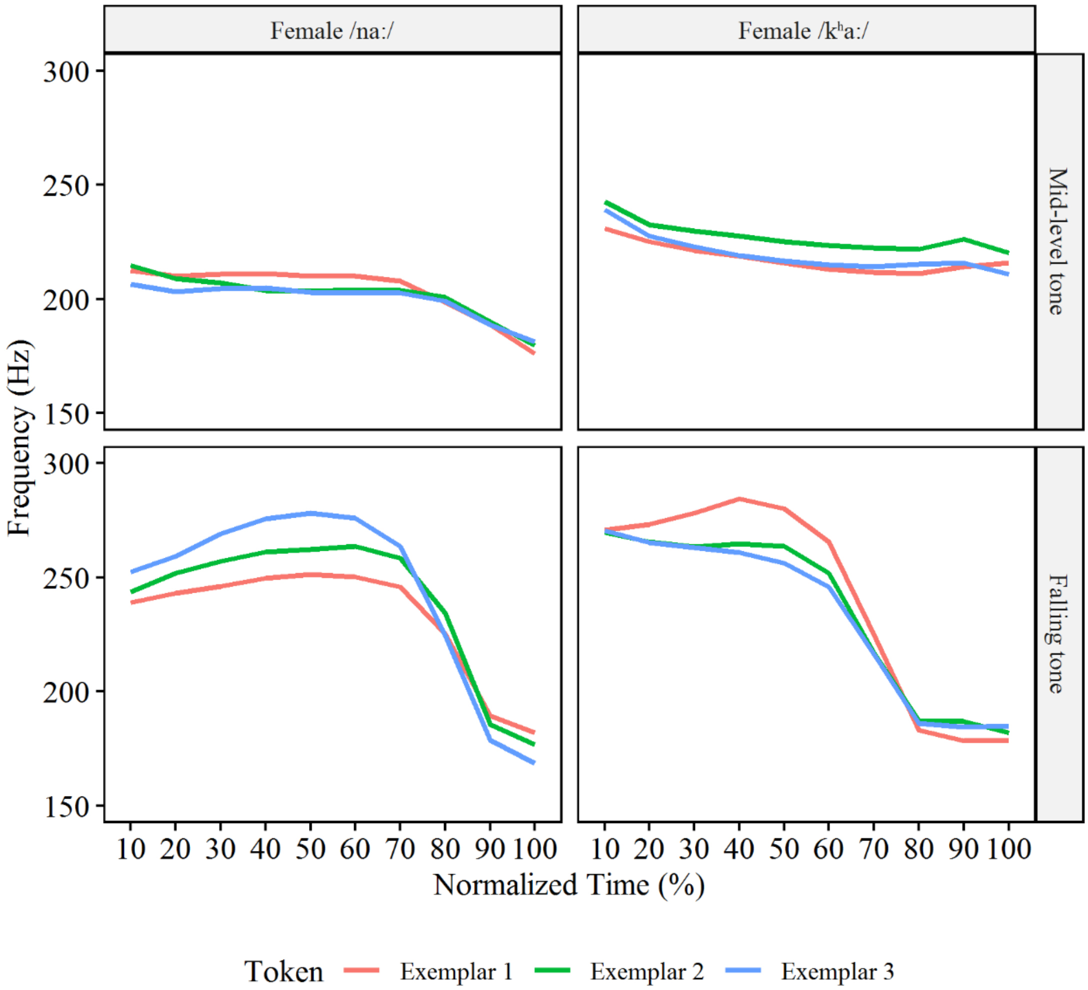
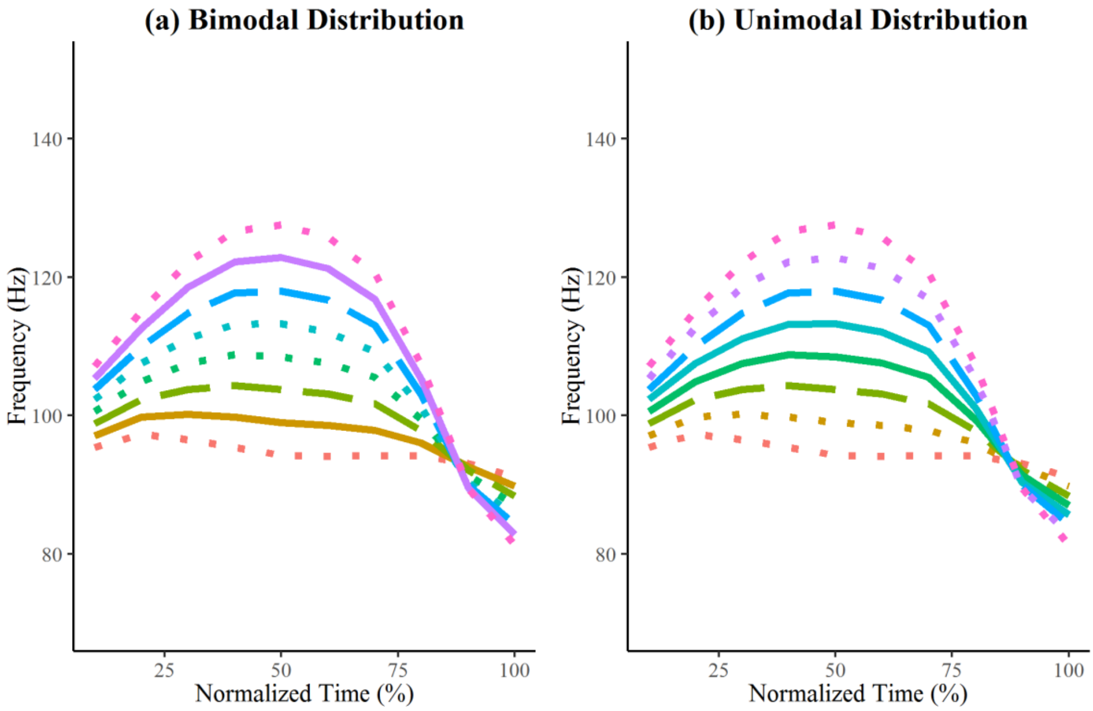
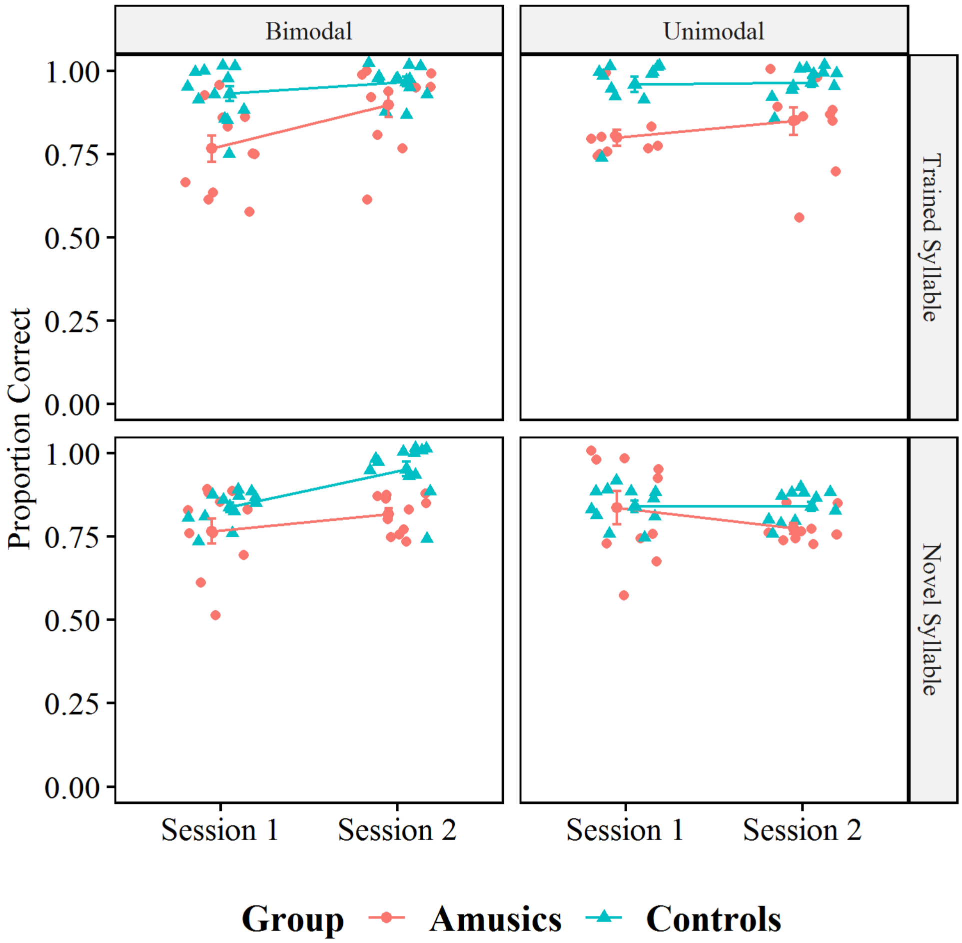

Female /na:/
Female /kʰa:/
Mid-level tone
Falling tone
Frequency (Hz)
300
250
200
150
10 20 30 40 50 60 70 80 90 100
Normalized Time (%)
Token
Exemplar 1
Exemplar 2
Exemplar 3

Figure 4


(a) Bimodal Distribution
(b) Unimodal Distribution
Frequency (Hz)
140
120
100
80
25
50
75
100
Normalized Time (%)

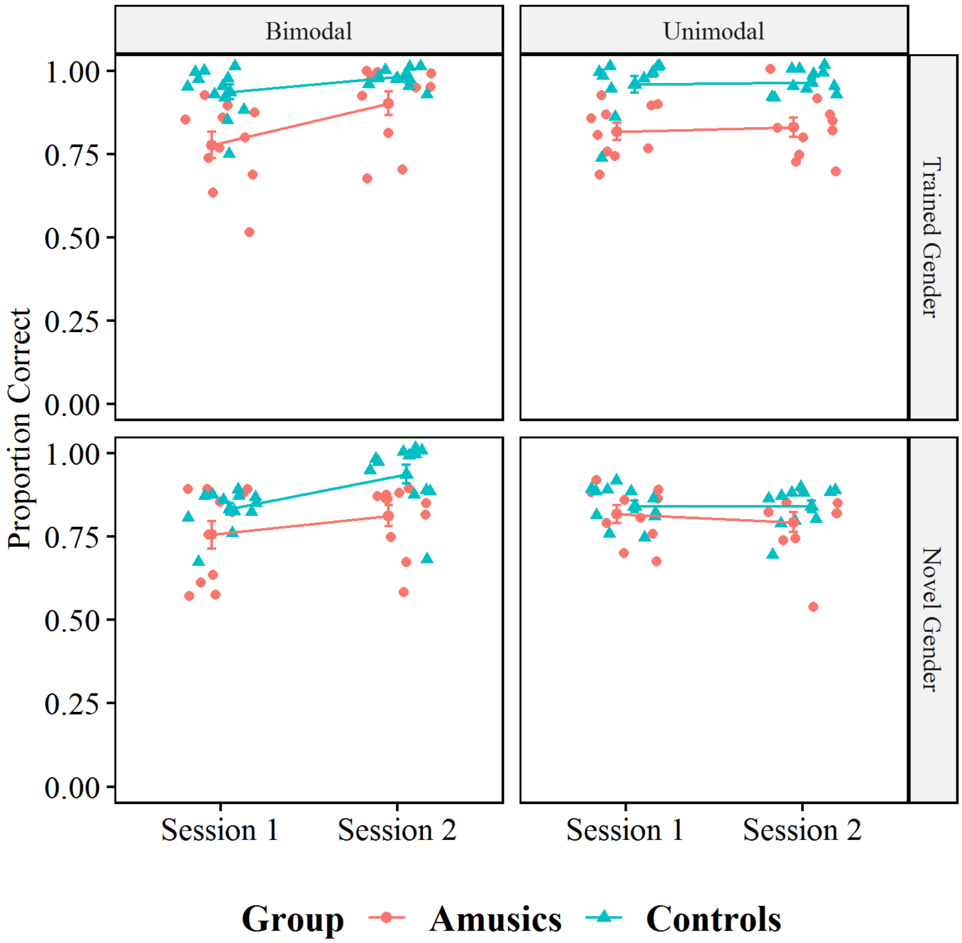
960 **Figure 5**



961

962

963 **Figure 6**



964