9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

The effect of syllable variation on the perception of lexical tones in Cantonese-speaking amusics

Jing Shao[1], *Phyllis Oi Ching Tang*[1], *Caicai Zhang*[1,2]

[1]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
jingshao@polyu.edu.hk, oi-ching.tang@connect.polyu.hk, caicai.zhang@polyu.edu.hk

Abstract

Congenital amusia is a neurogenetic disorder of fine-grained pitch processing. Though there is some evidence that this disorder extends to the language domain and negatively influences lexical tone perception, its deficiency mechanism remains unclear. This study designed a series of perception tasks to probe different levels of lexical tone perception, and expected to shed light on the mechanism underlying tone perception in amusia. Sixteen Cantonese-speaking amusics and 16 matched controls were tested on the effects of syllable variations on the perception of Cantonese tones with low variations, i.e., tones were always associated with the same syllable, versus high variations, i.e., tones were always associated with different syllables. Results of the identification task showed a trend of more pronounced group differences in the low variation condition compared to the high variation condition. In the discrimination task, the group difference was larger in the low variation condition, where more acoustic constancy was provided. These findings suggested that the amusics' tone perception abilities, in terms of both domain-general pitch processing and high-level phonological processing are impaired. Furthermore, Cantonese-speaking amusics seemed to be more impaired in the low acoustic variation context, implying a possible 'anchoring deficit' in congenital amusia.

Index Terms: congenital amusia, syllable variation, lexical tone perception, Cantonese.

1. Introduction

Congenital amusia (amusia hereafter), also known as tone or tune deafness, is a lifelong disorder of musical pitch processing. It occurs without brain damage and affects about 1.5-4% of the population [1], [2]. Individuals with amusia have difficulties detecting mistuned tones and out-of-key tones as well as noticing singing that is out-of-tune [3]. The primary deficit in amusia lies in the processing of the pitch dimension [4]–[6] and impaired short-term memory for pitch [7].

Empirical evidence has revealed that the pitch deficit in amusia is domain-general. It extends to the language domain [8], and influences speech intonation processing and identification of emotion status [9]–[12], in which pitch is extensively used as a cue. Pitch is also used to systematically distinguish word meanings in tonal languages. In this regard, two studies have found that French speakers with amusia showed impairment in the perception of Mandarin and Thai tones, further suggesting that the deficit in amusia lies in the domain-general pitch processing [13], [14].

On the other hand, native tonal language speakers with amusia are also found to be impaired in lexical tone perception when the base syllable were always the same [15]–[17], further attesting the domain-general nature of amusia. For example, [15] found that Cantonese-speaking amusics were less accurate at discriminating four pairs of native Cantonese tones than musically intact controls, while their tone production ability was largely intact. [17] investigated lexical tone identification and discrimination in Cantonese-speaking amusics in quiet and noise conditions. Amusics performed less accurately in tone identification and discrimination in the clear condition. It is suggested that when the syllables are carried by the same base syllable, it did not involve much phonological manipulation and primarily tapped into the domain-general pitch processing [18].

While the aforementioned studies consistently pointed out that the impoverished lexical tone perception in tonal and non-tonal language speakers with amusia is due to the domain-general pitch deficit, some studies have suggested that high-level phonological processing of lexical tones might be impaired in native tonal language speakers with amusia. For instance, several studies provided evidence that categorical perception of native tones in tonal language speakers with amusia is impaired [19]–[21]. In contrast to the controls, Mandarin-speaking amusics showed no benefit for between-category tone discriminations, suggesting the absence or impairment of categorical perception of lexical tones, though one study found subgroup differences among Mandarin-speaking amusics [18], [19]. In another study, Zhang et al. found that Cantonese-speaking amusics exhibited less benefit in between-category discriminations than controls in speech contexts (lexical tone and vowel), suggesting reduced categorical perception; on the other hand, they performed inferiorly compared to controls regardless of between- and within-category discriminations in nonspeech contexts (pure tone), suggesting impaired auditory pitch processing [21].

Taken together, the above studies suggest that there are two main yet not conflicting findings on the deficit of amusia in speech processing. One line of research showed that the deficit of amusia primarily lies in the domain-general auditory pitch processing [13]–[15]. Another line of research confirmed that the deficit already prevails to higher-level phonological processing, affecting categorical perception of lexical tones in native tonal language speakers [19]–[21]. However, due to the scarcity of studies directly comparing lower-level pitch processing and higher-level phonological processing, the mechanism underlying the deficient tone perception performance in amusia is not yet well understood. This study attempted to probe the impairment mechanism of lexical tone

10.21437/SpeechProsody.2018-30

perception in Cantonese speakers with amusia. We designed a series of perception tasks aiming at selectively tapping into domain-general pitch processing of lexical tones and high-level phonological processing of lexical tones, and by doing so expected to shed some light on the nature of deficits in amusia.

It is well known that syllable variation influences word recognition and speech discrimination in important ways [24]–[26]. For instance, [16] examined the performance of Mandarin-speaking amusics and controls in Mandarin tone discrimination, in which half of the tone pairs were associated with the same syllables, and the other half of the tone pairs were associated with different syllables. Results revealed that amusics performed similarly to controls in discriminating Mandarin tone pairs that were associated with the same syllables, whereas their performance was impaired in the different syllable condition. These results suggested that different degrees of syllable variation tap into different levels of lexical tone processing, and Mandarin-speaking amusics appeared to be selectively impaired in the different syllable condition with a greater demand of phonological processing.

In light of the above findings, low and high syllable variation conditions offer an excellent scenario to examine the low-level *auditory pitch processing* versus higher-level *phonological processing* in tone perception in amusics. In the current study, we compared Cantonese-speaking amusics and musically intact controls on the effect of these two conditions on tone perception. We expected that the amusics' performance would be impaired in both conditions, and that the high variation condition, which is a more demanding task, would be particularly challenging for amusics. Since listeners have to extract abstract tonal representations through higher-level operations, amusics are expected to show greater impairment in the high variation condition.

2. Method

2.1. Participants

Sixteen congenital amusics and 16 musically intact controls participated in this experiment. Control participants were matched with amusic participants one by one in age, gender, and years of education. All participants were native speakers of Hong Kong Cantonese, right-handed, with no hearing impairment, and no reported history of formal musical training. Amusics and controls were identified using the Montreal Battery of Evaluation of Amusia (MBEA) [27]. The MBEA consists of six subtests: three of them are pitch-based tests (scale, contour, and interval), two of them are duration-based tests (rhythm and meter), and the last one is a memory test. All amusic participants scored below 71%, whereas all control participants scored higher than 80%. Demographic characteristics of the participants are summarized in Table 1. The experimental procedures were approved by the Human Subjects Ethics Sub-committee of The Hong Kong Polytechnic University. Informed written consent was obtained from participants in compliance with the experiment protocols.

2.2. Stimuli

The stimuli in the syllable variation condition were 24 monosyllabic words contrasting six Cantonese tones (high level tone-T1, high rising tone-T2, mid level tone-T3, extra low level/low falling tone-T4, low rising tone-T5, low level tone-T6) on syllables /ji/, /fɐn/, /fu/ and /wɐi/. One female native Cantonese speaker was recorded reading aloud these words in a carrier sentence, 呢個字係 /li55 ko33 tsi22 hɐi22/ ('This word is') for six times. For each word, one clearly produced token was selected and segmented out of the carrier sentence. All selected words were normalized in duration to 650 ms and in mean intensity to 70 dB using Praat [28].

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

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

2.3. Procedures

The same set of stimuli was presented in high and low variation conditions. The critical difference was that stimuli from multiple syllables were presented in separate blocks in the low variation condition and were intermixed in one block in the high variation condition. Each condition included an identification task and a discrimination task. The stimuli were presented using E-prime 2.0.

In the low variation condition, the stimuli of the four syllables were presented in separate blocks. In the identification task, each set of six words from one syllable was presented in a sub-block, generating four sub-blocks. Six words in a syllable set were repeated twice and presented randomly within the sub-block. Subjects were instructed to identify the tone of the word by pressing buttons 1-6 on a keyboard. In the discrimination task, six words in each syllable set were grouped into 15 different tone pairs and six same tone pairs. Each syllable set was presented in a sub-block, generating four sub-blocks in total. Within a sub-block, different tone pairs were repeated twice and same tone pairs were repeated five times, generating equal number of different and same tone pairs, which were intermixed and randomly presented. Subjects were instructed to judge whether the two words carried the same tone or different tones by pressing "left arrow" (same) and "right arrow" (different) on a keyboard within 3 seconds.

In the high variation condition, tone stimuli carried by the four syllables (/ji/, /fɐn/, /fu/ and /wɐi/) were intermixed in a block. In the identification task, the four syllable sets were repeated twice and presented randomly within a block. The procedure was the same as that described above. In the discrimination task, each tone pair (same or different) was always associated with different syllables. In order to keep the experiment duration short, the four syllables were grouped into two sets, set A including six syllable pairs (/fɐn/-/ji/, /ji/-/wɐi/, /fu/-/wɐi/, /fɐn/-/wai/, /fɐn/-/fu/ and /fu/-/ji/) and set B including the same syllable pairs in reversed order (/ji/-/fɐn, /wɐi/-/ji/, /wɐi/-/fu/, /wɐi/-/fan/, /fu/-/fɐn/ and /ji/-/fu/). Each syllable pair carried 15 different and six same tone pairs. Different tone pairs were repeated twice and same tone pairs were repeated five times, generating 60 trials for each syllable pair. Within each group, half of the participants were randomly assigned to the set A and the other half to the set B.

The presentation order of the identification and discrimination tasks was counterbalanced across the participants. In both identification and discrimination tasks, accuracy and reaction time (RT) were collected.

2.4. Data analysis

For the identification task, accuracy and RT were analyzed. Response to each trial was coded as 1 or 0 (correct or incorrect) for each participant. In order to compare the accuracy of amusics and controls in the identification task, generalized mixed-effects models were fitted on the responses to each trial (1 or 0) with *group* (amusics and controls), *variation* (low syllable variation and high syllable variation) as two fixed effects, and subjects as a random effect. When analyzing RT in the identification task, incorrect trials were disregarded. Linear mixed-effects models were fitted on the RT with *group* and *variation* as two fixed effects and *subject* as a random factor. The above two sets of analyses were performed with R (R Core Team, 2014), using the *lme4* package [30] and the *lsmeans* package [31].

For the discrimination results, the sensitivity index d' and RT were analyzed. The d' was computed as the z-score value of the hit rate ("different" responses to different tone pairs) minus that of the false alarm rate ("different" responses to same tone pairs) for each tone pair per subject [32]. Trials with null responses were disregarded. *Group* × *variation* repeated measures ANOVAs were conducted on the discrimination sensitivity d', using Statistical Package for the Social Sciences (SPSS) [33]. When analyzing RT in the discrimination task, the statistical analysis method used was same as that in the identification task.

3. Results

Figure 1 (left) illustrates the identification accuracy in each condition for the two groups in the identification task. Generalized mixed-effects models found a significant main effect of *group* ($\chi^2(1) = 7.17$, $p < 0.001$), and the effect of *variation* also approached significance ($\chi^2(1) = 2.73$, $p = 0.09$). Amusics showed lower accuracy in tone identification compared to normal controls (amusics: $M = 0.41$, $SD = 0.21$; controls: $M = 0.61$, $SD = 0.16$). Although the *group* by *variation* interaction was not significant ($\chi^2(1) = 1.335$, $p = 0.247$), the group difference in the low syllable variation condition was noticeably larger. Further planned pairwise comparisons confirmed this observation, showing that the group difference in the high variation condition (amusics: $M = 0.42$, $SD = 0.18$; controls: $M = 0.59$, $SD = 0.13$; $z = -2.43$, $p = 0.02$) was smaller than in the low syllable variation (amusics: $M = 0.41$, $SD = 0.24$; controls: $M = 0.64$, $SD = 0.18$; $z = -3.04$, $p = 0.002$). Furthermore, the amusics' performance in the low and high variation conditions was not significantly different ($z = -0.336$, $p = 0.736$), but controls obtained significantly higher identification accuracy in the low variation condition than in the high variation condition ($z = -1.99$, $p = 0.04$). These results suggest that amusics performed similarly in the high and low syllable variation conditions, whereas the controls benefited more from the low syllable variations in tone identification.

Figure 1 (right) shows the RT in the identification task in each condition for the two groups. Linear mixed-effects model found a significant main effect of *variation* ($\chi^2(1) = 40.88$, $p < 0.001$), and significant two-way interaction ($\chi^2(1) = 8.43$, $p = 0.003$). Pairwise comparisons revealed that in the control group, RT in the high variation condition was significantly longer than that in the low variation condition ($z = 6.81$, $p < 0.001$), but such effect was not significant in the amusic group. Under the high variation condition, controls exhibited significantly longer RT than amusics ($z = -2.39$, $p < 0.001$), but such difference cannot be found in the low variation condition ($z = -0.69$, $p = 0.49$). This suggests that controls were more careful in identifying the tones in the high variation condition.

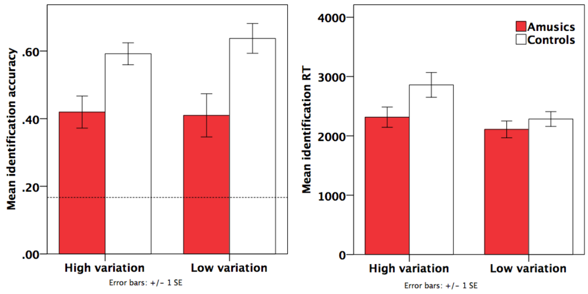

Figure 1: *Results of the identification accuracy (left) and RT (right) for the amusic and control groups in low and high variations. Dotted lines indicate the chance level accuracy (0.167).*

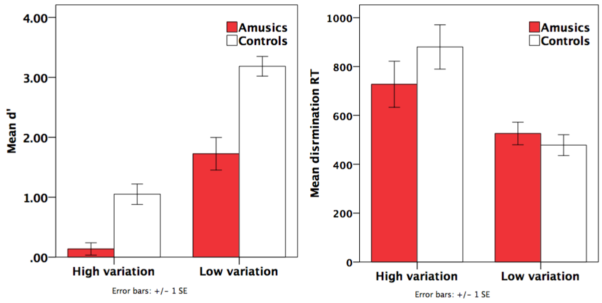

Figure 2: *Results of the discrimination sensitivity index (left) and RT (right) for amusics and controls in low and high variations.*

Figure 2 (left) displays the discrimination sensitivity d' in the two conditions for the two groups. *Group* × *condition* repeated-measures ANOVA found significant main effects of *group* ($F(1, 30) = 33.637$, $p < 0.001$, $\eta_p^2 = 0.529$) and *variation* ($F(1, 30) = 133.709$, $p < 0.001$, $\eta_p^2 = 0.817$) and significant two-way interaction ($F(1, 30) = 4.926$, $p = 0.034$, $\eta_p^2 = 0.141$). Post hoc analyses showed that although both amusic and control groups exhibited significantly higher d' scores in the low variation condition than in the high variation condition (amusics: $t(30)= -5.462$, $p < 0.001$, $d = 1.931$; controls: $t(30)= -9.017$ $p < 0.001$, $d = 3.188$), the effect was larger in the control group. Within each variation condition, the amusic group consistently demonstrated significantly poorer discrimination performance compared to controls (high variation: $t(30)= -4.585$, $p < 0.001$, $d = 1.621$; low variation: $t(30)= -4.595$ $p < 0.001$, $d = 1.624$).

Figure 2 (right) shows RT in the discrimination task. Linear mixed-effects model found a significant main effect of *variation* ($\chi^2(1) = 1385.4$, $p < 0.001$) and significant two-way interaction ($\chi^2(1) = 92.7$, $p < 0.001$). Pairwise comparisons revealed that in the high syllable variation condition, RT in the control group was longer than the amusic group, and the effect was approaching significance ($z = -1.683$, $p = 0.09$). In the low syllable variation condition, RT in the two groups was similar ($z = 0.24$, $p = 0.81$). Within each group, RT employed in the high variation condition was significantly longer than that in the low variation condition, but the effect was more evident in the control group ($z = 35.09$, $p < 0.001$) than in the amusic group ($z = 18.21$, $p < 0.001$). These RT results suggest that controls were more careful than amusics (longer RT in the high variation); and controls tended to show greater benefit from the low acoustic variation condition, where they employed significant shorter RT in the low variation condition than in the high variation condition.

To summarize, there were clear trends of larger group difference in the low variation condition in terms of the tone identification accuracy. In the discrimination task, there was a significant *group* by *variation* interaction effect on the d' score, where the group difference was more evident in the low variation condition. Moreover, the controls showed robust improvement in the low variation condition than in the high variation condition, whereas amusics did not. The discrimination RT also revealed an interaction effect, where the RT was shorter in the low variation condition, but the effect was more robust in the control group.

4. Discussion

While amusia has been consistently reported to influence tone perception negatively [15]–[17], the mechanism underlying the deficient tone perception in tonal language speakers is still not well understood. In this study, we examined this issue through a comparison of two conditions, *low* syllable variation versus *high* syllable variation condition. It is expected that tone perception in low variation context tapped into relatively low level of pitch processing, while tone perception in high variation context tapped into relatively high level of phonological processing of lexical tones and amusics would suffer more from such variations because of the larger acoustic noise present in the auditory input.

We found that amusics performed significantly worse than controls in the *low* syllable variation condition. These results are consistent with previous findings on Cantonese speakers with amusia [15], [17]. In [15], four Cantonese tone pairs were selected and presented to amusics and controls. The four tone pairs were always carried by the same syllables, which corresponded to the low syllable variation condition in the current study. The current results also echo with the finding of [17], where amusics showed deficient tone perception where the tone stimuli were carried by the same syllables.

The results also revealed that Cantonese-speaking amusics showed significantly lower d' score than controls when the lexical tones were associated with different syllables, suggesting a deficit in *high level* phonological processing of lexical tones. The tone pairs carried by different syllables are more acoustically different, increasing the difficulty of reliably extracting tone categories from acoustical signals. It thus may decrease the "quality" of extracted tone categories, compromising perception accuracy. These findings are in line with previous studies in which categorical perception of native tones is found to be impaired in tonal language speakers [18], [19], [20]. For instance, Zhang et al. (2017) [20] found that Cantonese-speaking amusics exhibited less benefit in between-category discriminations than controls in speech contexts, suggesting reduced categorical perception of tones.

It should be noted that we hypothesized that greater speaker variation increases the difficulty of extracting abstract tone categories, and thus the group difference is expected to be more pronounced in the higher variation condition, which is a phonologically more demanding task. However, we observed opposite patterns in that the group difference was generally larger in the low variation condition, especially in the discrimination task, where the syllables were consistently repeated in a trial and the acoustic variation was limited.

The finding that amusics were more impaired in the low acoustic variation condition can be explained by the 'anchoring theory' which is originally proposed to account for the phonological deficit in dyslexia [34], [35]. This hypothesis claims that the deficit of dyslexics lies in the dynamics that link perception with perceptual memory through the implicit formation of anchors. Empirical evidence showed that the normal population tunes around, or 'anchors to', incoming stimuli automatically, therefore responding more accurately when these stimuli are subsequently repeated. On the contrary, individuals with dyslexia failed to benefit from specific repetitions. It suggests that dyslexic individuals have difficulties in dynamically constructing stimulus-specific predictions, deriving from a deficient adaptation mechanism.

Our findings can be accounted for by the 'anchoring theory'. In the low variation condition, tone pairs were always associated with the same syllable. Anchoring to the same syllable in the speech stimuli possibly provided the perceptual system with better predictions, which may facilitate lexical tone perception. However, in contrast to controls, the performance of amusics improved less when the same syllable was presented across trials, especially in the discrimination task, indicating that their ability to construct and tune to an internal syllable anchor is impoverished. Therefore, amusics may have to perform more effortful processing in every trial. Without a properly functioning anchoring mechanism, the perceptual system of amusics is therefore less resilient to external noise, as in the case of dyslexics. This explains why under conditions with low acoustic variation, the perception of normal controls greatly sharpened compared with the high acoustic variation condition, whereas the perception of amusics showed less improvement.

5. Conclusion

We found that Cantonese-speaking amusics were impaired in both low-level and higher-level processes of lexical tones. Furthermore, amusics seemed to be more impaired in the low acoustic variation context, where acoustic constancy was provided to construct perceptual anchors. These findings shed some light on the nature of the deficits in congenital amusia. We propose that it reflects impaired dynamics of the stimulus anchoring mechanism, suggesting a possible 'anchoring deficit' in congenital amusia.

6. Acknowledgements

This work was supported by grants from the Research Grants Council of Hong Kong (ECS: 25603916), the National Natural Science Foundation of China (NSFC: 11504400), and the PolyU Start-up Fund for New Recruits.

7. References

[1] I. Peretz and K. L. Hyde, “What is specific to music processing? Insights from congenital amusia,” *Trends Cogn. Sci.*, vol. 7, no. 8, pp. 362–367, 2003.

[2] I. Peretz and D. T. Vuvan, “Prevalence of congenital amusia,” *Eur. J. Hum. Genet.*, 2017.

[3] J. Ayotte, I. Peretz, and K. Hyde, “Congenital amusia: A group study of adults afflicted with a music-specific disorder,” *Brain*, vol. 125, no. 2, pp. 238–251, 2002.

[4] J. M. Foxton Dean, Jennifer L., Gee, Rosemary, Peretz, Isabelle, and Griffths, Timothy D., “Characterization of deficits in pitch perception underlying ‘tone deafness,’” *Brain*, vol. 127, pp. 801–810, 2004.

[5] K. L. Hyde and I. Peretz, “Brains that are out of tune but in time,” *Psychol. Sci.*, vol. 15, no. 5, pp. 356–360, 2004.

[6] I. Peretz, J. Ayotte, R. J. Zatorre, J. Mehler, P. Ahad, V. B. Penhune, and B. Jutras, “Congenital amusia: A disorder of fine-grained pitch discrimination,” *Neuron*, vol. 33, no. 2, pp. 185–191, 2002.

[7] B. Tillmann, Y. Lévêque, L. Fornoni, P. Albouy, and A. Caclin, “Impaired short-term memory for pitch in congenital amusia,” *Brain Res.*, vol. 1640, no. Part B, pp. 251–263, 2016.

[8] A. D. Patel, J. M. Foxton, and T. D. Griffiths, “Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech,” *Brain Cogn.*, vol. 59, no. 3, pp. 310–313, Dec. 2005.

[9] C. Jiang, J. P. Hamm, V. K. Lim, I. J. Kirk, X. Chen, and Y. Yang, “Amusia results in abnormal brain activity following inappropriate intonation during speech comprehension,” *PLoS One*, vol. 7, no. 7, p. e41411, 2012.

[10] F. Liu, A. D. Patel, A. Fourcin, and L. Stewart, “Intonation processing in congenital amusia: discrimination, identification and imitation,” *Brain*, vol. 133, no. 6, pp. 1682–1693, 2010.

[11] X. Lu, H. T. Ho, F. Liu, D. Wu, and W. F. Thompson, “Intonation processing deficits of emotional words among Mandarin Chinese speakers with congenital amusia: an ERP study,” *Front. Psychol.*, vol. 6, p. 385, Apr. 2015.

[12] W. F. Thompson, M. M. Marin, and L. Stewart, “Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 46, pp. 19027–19032, 2012.

[13] S. Nguyen, B. Tillmann, N. Gosselin, and I. Peretz, “Tonal language processing in congenital amusia,” *Ann. N. Y. Acad. Sci.*, vol. 1169, no. 1, pp. 490–493, 2009.

[14] B. Tillmann, D. Burnham, S. Nguyen, N. Grimault, N. Gosselin, and I. Peretz, “Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages,” *Front. Psychol.*, vol. 2, no. JUN, 2011.

[15] F. Liu, A. H. D. Chan, V. Ciocca, C. Roquet, I. Peretz, and P. C. M. Wong, “Pitch perception and production in congenital amusia: Evidence from Cantonese speakers,” *J. Acoust. Soc. Am.*, vol. 140, no. 1, pp. 563–575, 2016.

[16] Y. Nan, Y. Sun, and I. Peretz, “Congenital amusia in speakers of a tone language: Association with lexical tone agnosia,” *Brain*, vol. 133, no. 9, pp. 2635–2642, 2010.

[17] J. Shao, C. Zhang, G. Peng, Y. Yang, and W. S.-Y. Wang, “Effect of noise on lexical tone perception in Cantonese-speaking amusics,” in *Proceedings of the Interspeech*, 2016.

[18] F. Liu, C. Jiang, W. F. Thompson, Y. Xu, Y. Yang, and L. Stewart, “The mechanism of speech processing in congenital amusia: Evidence from Mandarin speakers,” *PLoS One*, vol. 7, no. 2, p. e30374, Feb. 2012.

[19] W. T. Huang, C. Liu, Q. Dong, and Y. Nan, “Categorical perception of lexical tones in Mandarin-speaking congenital amusics,” in *Frontiers in Psychology*, 2015, vol. 6, no. 829.

[20] C. Jiang, J. P. Hamm, V. K. Lim, I. J. Kirk, and Y. Yang, “Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics.,” *Mem. Cognit.*, vol. 40, no. 7, pp. 1109–21, 2012.

[21] C. Zhang, J. Shao, and X. Huang, “Deficits of congenital amusia beyond pitch: Evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics,” *PLoS One*, vol. 12, no. 8, p. e0183151, 2017.

[22] W. T. Huang, C. Liu, Q. Dong, and Y. Nan, “Categorical perception of lexical tones in Mandarin-speaking congenital amusics,” *Front. Psychol.*, vol. 6, no. 829, 2015.

[23] C. Jiang, J. Hamm, V. Lim, I. Kirk, and Y. Yang, “Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics,” *Mem. Cognit.*, vol. 40, no. 7, pp. 1109–1121, 2012.

[24] A. R. Bradlow, L. C. Nygaard, and D. B. Pisoni, “Effects of talker, rate, and amplitude variation on recognition memory for spoken words,” *Percept. Psychophys.*, vol. 61, no. 2, pp. 206–219, 1999.

[25] D. B. Pisoni, “Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning,” *Speech Commun.*, vol. 13, no. 1–2, pp. 109–125, 1993.

[26] M. Antoniou and P. C. M. Wong, “Poor phonetic perceivers are affected by cognitive load when resolving talker variability; Poor phonetic perceivers are affected by cognitive load when resolving talker variability,” *J. Acoust. Soc. Am.*, vol. 138, no. 2, pp. 571–574, 2015.

[27] I. Peretz, A. S. Champod, and K. Hyde, “Varieties of musical disorders,” *Ann. N. Y. Acad. Sci.*, vol. 999, no. 1, pp. 58–75, 2003.

[28] P. Boersma and D. Weenink, “Praat: doing phonetics by computer,” 2010.

[29] R. C. Team, “R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.” 2014.

[30] D. Bates, M. Maechler, and B. Bolker, “lme4: linear mixed-effects models using S4 classes. R package version 0.999375-42. 2011,” 2012.

[31] R. V. Lenth, “Least-squares means: the R package lsmeans,” *J. Stat. Softw.*, vol. 69, no. 1, pp. 1–33, 2016.

[32] N. A. Macmillan and C. D. Creelman, *Detection Theory: A User's Guide*. Mahwah: Lawrence Erlbaum Associates, 2005.

[33] I. B. M. SPSS, “IBM SPSS statistics for Windows, version 20.0,” *New York IBM Corp*, 2011.

[34] M. Ahissar, “Dyslexia and the anchoring-deficit hypothesis,” *Trends Cogn. Sci.*, vol. 11, no. 11, pp. 458–465, 2007.

[35] M. Ahissar, Y. Lubin, H. Putter-Katz, and K. Banai, “Dyslexia and the failure to form a perceptual anchor,” *Nat. Neurosci.*, vol. 9, no. 12, pp. 1558–1564, 2006.