The Perception of Lexical Tone in Whispered Speech by Mandarin-speaking Congenital Amusics

Gaoyuan Zhang[1], Jing Shao[2,3], Lan Wang[3], Caicai Zhang[2,3]

[1]Department of Chinese Language and Literature, Peking University
[2]Department of Chinese and Bilingual Studies, the Hong Kong Polytechnic University
[3]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
gaoyuanzhang@pku.edu.cn ,jing.shao@polyu.edu.hk, lan.wang@siat.ac.cn,caicai.zhang@polyu.edu.hk

ABSTRACT

Congenital amusia is a neurodevelopment disorder of musical pitch processing, which also affects lexical tone perception in tonal languages like Mandarin Chinese. In this study we aimed to investigate how congenital amusia affects lexical tone recognition without pitch information. Nineteen Mandarin-speaking congenital amusics and 19 matched controls were tested on lexical tone identification in both phonated and whispered speech. The results revealed that the performance of congenital amusics was inferior to that of controls in lexical tone identification in both phonated and whispered speech, but the differences between the two groups were smaller in whispered speech. Moreover, the identification of Tone 3 and Tone 4 was easier than that of Tone 2 and Tone 1 in whispered tone for both groups. The results indicate that the primary disorder of amusia lies in pitch processing but the deficits of amusia also appear to extend beyond pitch processing.

Keywords: congenital amusia, lexical tone perception, pitch, whispered speech, Mandarin Chinese.

1. INTRODUCTION

Compared with phonated speech, the most obvious characteristic of whispered speech is that the dominant perceptual cue – fundamental frequency (F0, hereafter) is absent. Previous studies have showed that compared with phonated tones, listeners have difficulty in identifying lexical tone in whispered speech, but the accuracy is above chance level, meaning that lexical tones can still be recognized [1, 3, 4, 14]. For instance, Gao [1] examined tone recognition in whispered Mandarin speech (four lexical tones in Mandarin: T1-high level tone, T2-rising tone, T3-dipping tone, T4-falling tone), which found that the rank of the four tones in monosyllables in the order of increasing difficulty is T3 (52%), T4 (34%), T1 (11%) and T2 (2%). Jiao et al. [4] also showed that in Mandarin the identification rates for phonated tones were: T1 (98.9%), T2 (98.9%), T4 (94.7%) and T3 (94.2%). In contrast, the rates for whispered tones were: T3 (85%), T4 (66.9%), T2 (35.8%) and T1 (21.7%). Some researchers proposed that perceptual cues other than pitch, such as duration [1], amplitude contour [6], and formant frequency [7], could facilitate tone identification in whispered speech.

Congenital amusia (amusia, hereafter) is a life-long neurodevelopment disorder of musical pitch processing [12]. Individuals with amusia (amusics, hereafter) have difficulty acquiring basic musical skills [11]. It has been reported that amusics are not only impaired in musical pitch processing, but also in speech pitch processing such as the perception of lexical tone, intonation, and emotional prosody [2, 8, 13]. These results indicate that the pitch-processing deficit in amusia is not restricted to music, but transfers to the language domain. For example, Nan et al. [10] examined the perception of four Mandarin tones by Mandarin-speaking amusics. They found that amusics had significant deficits in the identification of Mandarin tones relative to controls.

However, amusics' inferior performance is not restricted to pitch. A few studies have reported that amusics have inferior segmental processing and frequency/spectral processing beyond pitch processing [5, 9, 15]. For instance, the results of Zhang et al. [15] showed that Cantonese-speaking amusics had impairment in the discrimination of pure tones, lexical tones and vowels, but their ability of duration processing remained intact, which indicated that the deficit of amusia might not be confined to pitch processing, but also influence frequency/spectral processing more broadly. It is likely that amusia is a syndromic disorder frequently accompanied by deficiencies of other kinds [5].

Whispered speech is an ideal case to examine whether amusics have impairment in other aspects of the linguistic domain other than pitch processing. On the one hand, amusics may have comparable or even better performance than controls in the recognition of whispered tones, for the reason that pitch cues are absent and listeners need to rely on other cues such as duration, intensity or spectral frequency information to perceive tones. The results will shed light on the potential compensation mechanism of amusics in speech comprehension. On the other hand, it is also possible that the amusics'

performance of tone recognition is worse than musically intact controls even in whispered speech. As mentioned above, amusics are impaired in formant frequency processing [15], which is believed to play some role in the perception of whispered speech [7]. If so, the amusics' perception of whispered tones may be affected. Furthermore, it has been reported that lexical tone impairment in amusics is not purely due to the low-level auditory pitch deficit, but prevails to the higher-level phonological processing, affecting the phonological representations of lexical tone [15]. According to this finding, amusics may show impairment in tone recognition despite the absence of pitch information in speech stimuli. To address these questions, this paper compared the performance of Mandarin-speaking amusics and matched controls in lexical tone recognition in whispered speech and phonated speech.

2. METHOD

2.1. Participants

Nineteen Mandarin amusics and 19 matched musically intact controls were recruited in this study. All participants are native speakers of Mandarin Chinese from northern areas, right-handed and reported no previous history of speech, hearing, neurological or psychiatric impairments. No participants had any formal musical training. All subjects were selected by the test of Montreal Battery of Evaluation of Amusia (MBEA), which consists of six subtests: scale, contour, interval, rhyme, meter and memory [11]. Those participants who scored above 85% were classified as controls, and those who scored below 70% were classified as amusics. The demographic characteristics of participants are shown in Table 1.

Table 1: Demographic characteristics of subjects.

	Amusics	Controls
Male/Female(total)	9/10(19)	9/10(19)
Mean Age(range)	24(19-30)	23.95(20-30)
MBEA		
Scale(SD)	53.33(14.24)	85.90(11.16)
Contour(SD)	58.56(15.50)	94.23(4.76)
Interval(SD)	58.57(7.85)	93.02(3.78)
Rhyme(SD)	61.13(13.75)	93.54(6.88)
Meter(SD)	50.53(10.44)	84.39(12.07)
Memory(SD)	71.06(16.12)	96.49(3.92)
Total(SD)	58.84(7.32)	91.26(4.65)

2.2. Stimuli

To assess tone identification in Mandarin Chinese, 36 words with nine base syllables (/ta/, /ti/, /tu/, /pa/, /pi/, /p^{h}u/, /a/, /i/, /u/) contrasting four lexical tones (T1: high level tone, T2: rising tone, T3: dipping tone, T4: failing tone) were selected as the stimuli. Two Mandarin speakers (1 male and 1 female) were invited to produce the isolated words both in whispered and phonated mode.

2.3. Procedure

The study included two conditions: tone identification in phonated speech and in whispered speech, both of which were implemented using E-prime 2.0. There were practices before each task to familiarize participants with the experimental procedure. The order of these two conditions was counterbalanced across the subjects as much as possible. The stimuli from the two speakers were presented in two separate blocks and the order of the two blocks was also counterbalanced across subjects.

In each block, the stimuli were repeated 3 times and presented randomly. In each trial, a fixation occurred at first for 500ms, followed by the presentation of a speech stimulus via the headphones. The subjects were asked to identify the tone of the stimulus by pressing buttons 1-4 on a computer keyboard. The experiment only proceeded to the next trial when a response was received.

2.4. Data analysis

For both identification tasks, accuracy and response time were recorded and analysed. Accuracy was the percentage of trials correctly identified for each tone per subject. Incorrect trials were disregarded in the analysis of response time. *Group* × *lexical tone* × *phonation mode* repeated measures ANOVAs were conducted on the accuracy and response time of identification tasks respectively.

3. RESULTS

Figure 1 shows the identification accuracy of lexical tone perception under phonated and whispered conditions. There were significant main effects of *phonation mode* ($F(1,74)=490.17$, $p<0.001$), *lexical tone* ($F(2.67,197.84)= 92.43$, $p<0.001$), and *group* ($F(1,74)=15.45$, $p<0.001$). There were also significant interactions between *lexical tone* and *phonation mode* ($F(2.84,209.89)=103.07$, $p<0.001$), and among *lexical tone*, *phonation mode* and *group* ($F(2.84,209.89)=3.12$, $p=0.029$). We conducted *lexical tone* × *group* analyses in each phonation

mode to analyse the three-way interaction. For the phonated mode, two-way ANOVAs showed that there were significant main effects of *lexical tone* (F(2.43,179.54)=9.81, p<0.001) and *group* (F(1,74) =16.83, p<0.001), as well as a two-way interaction between *group* and *lexical tone* (F(2.43,179.54)= 8.52, p<0.001). One-way ANOVAs with the factor of *lexical tone* within two groups revealed only a significant effect of *lexical tone* in the amusic group (F(3,148)=8.53, p<0.001), where post hoc results found that the accuracy of T2 was significantly lower than other three tones (ps<0.001).There was no significant effects in the control group. Independent samples t-tests were conducted to examine the effect of *group* within each lexical tone. Significant *group* differences were found in T2 (p<0.001) and T4 (p=0.001), while the group difference in T3 (p=0.054) showed marginal significance and no group difference was found in T1 (p=0.16). Controls performed significantly better than amusics in T2 and T4. For whispered speech, two-way ANOVAs revealed that there were significant main effects of *lexical tone* (F (2.72,201.35)=118.45, p<0.001) and *group* (F(1,74) =6.30, p=0.014). Although the two-way interaction was not significant (p=0.31), to answer the question of whether amusics performed less accurately than controls in each tone, post-hoc analyses were conducted. One-way ANOVAs were first conducted to investigate the effect of *lexical tone* in each group. For amusics, the effect of *lexical tone* was significant (F(3,148)=37.97, p<0.001). Post hoc tests showed that the accuracy of T3 was significantly higher than the other three tones (ps<0.001), and that the accuracy of T4 was also significantly higher than T2 and T1 (ps<0.001). In control group, a significant effect of *lexical tone* was observed. Post hoc results found that the accuracy of T3 was significantly higher than other three tones (ps<0.001), and that the accuracy of T4 was significantly higher than T1 and T2 (ps<0.001). Independent samples t-tests with the factor of *group* within four tones revealed significant group differences only in T4 (p<0.016), while a marginal significant effect was presented in T2 (p=0.058) and no significant effect was found in T3 (p=0.12) and T1 (p=0.64). Controls performed significantly better than amusics only in T4.

Figure 2 shows the response time (RT, hereafter) under phonated and whispered conditions. There were significant main effects of *phonation mode* (F(1,72)=151.45, p<0.001), and *lexical tone* (F(3,216)=19.59, p<0.001), as well as significant two-way interactions between *phonation mode* and *group* (F(1,72)=8,71, p=0.02), between *lexical tone* and *group* (F(3,216)=9,31, p=0.001), and between *phonation mode* and *lexical tone* (F(3,216)=1.70, p<0.001). We analysed the interaction between *phonation mode* and *group* firstly. Independent samples t-tests were conducted to analyse *phonation mode* in each group. In both groups, the accuracy of the phonated mode was significantly higher than the whispered mode (ps<0.001). The same statistic method was used to examine the effect of *group* within in each phonation mode, which suggested that there was only a significant effect of *group* in the phonated mode (p<0.001) where the RT of controls was shorter than amusics. Secondly, we analysed the *lexical tone* and *group* interaction. One-way ANOVAs were conducted to reveal the effect of *lexical tone* in two groups. A significant effect of *lexical tone* was observed in the amusic group (F(3,301)=5.89, p=0.001). Post hoc tests revealed that the RT of T4 and T3 was significantly shorter than the two other tones (ps≤0.006). A significant effect of *lexical tone* was also found in the control group (F(3,303) =4.75, p=0.003). Post hoc results suggested that the RT of T1 was significantly longer than other three tones (ps≤0.01). Independent samples t-tests were conducted to analysed the effect of *group* in each lexical tone. The group differences were significant only in T2 and T3, where controls responded faster than amusics (ps<0.025). Finally, we analysed the interaction between *phonation mode* and *lexical tone*. One-way ANOVAs were conducted to reveal the effect of *lexical tone* in each phonation mode. A significant effect of *lexical tone* was observed in the phonated mode (F(3,303)=3.56, p=0.015). Post hoc tests showed that the RT of T4 was significantly shorter than T2 and T3 (ps≤0.039). A significant effect of *lexical tone* was also found in the control group (F(3,301)=11.83, p<0.001). Post hoc results suggested that RT of T1 was significantly longer than other three tones (ps≤0.009) and the RT of T2 was significantly longer than T3 (p=0.004). Independent samples t-tests used to reveal the effect of *phonation mode* in each lexical tone. Significant effect of *phonation mode* was found in all tones (ps<0.001), where the RT in the phonation mode was faster than the whispered mode.

Figure 1: The identification accuracy.

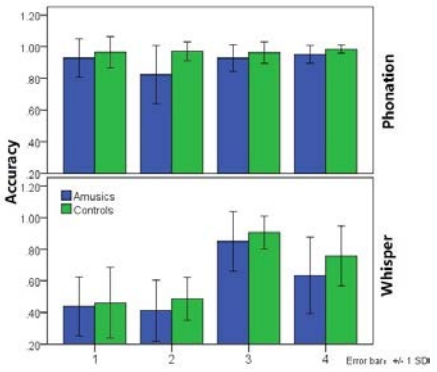
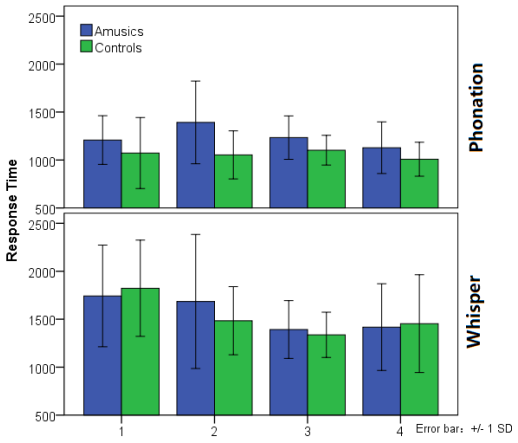


Figure 2: The identification response time.



Furthermore, the rank of the four whispered tones in the order of increasing difficulty in the control group is: T3 (90.64%), T4 (75.73%), T2 (48.64%), T1 (46%). The rates for whispered tones in the amusic group is T3 (85.09%), T4 (63.55%), T1 (43.96%), T2 (41.13%). The accuracy of T3 was the highest in both groups, followed by T4.

4. DISCUSSION

The present study examined the identification of lexical tone by Mandarin-speaking amusics and controls in phonated and whispered mode. The experimental results showed that Mandarin-speaking amusics were impaired in lexical tone identification in both phonated and whispered speech, but the group differences were smaller in whispered speech. The identification of T3 and T4 was more accurate than T2 and T1 in whispered tone for both groups.

In phonated mode, compared with the control group, the amusic group demonstrated significantly lower accuracy in the identification of T2 and T4, and overall significantly longer response time. Furthermore, T2 was the most difficult tone for amusics to identify. These results echoed with the findings in Nan et al. [10] that the pitch disorder was not confined in the music domain but also transferred to the language domain.

In whispered mode, it is worth noting that although amusics still had worse performance in accuracy, there was no significant difference in response time between controls and amusics. In the accuracy, the differences between two groups were also smaller compared to the phonated mode and post hoc analyses only revealed a significant group difference in the identification of T4. The reduced group difference in the whispered mode confirms that the primary disorder of amusia lies in pitch processing [16,17]. However, the deficits of amusia also appear to extend beyond pitch processing, as indicated by the reduced accuracy of amusics in the identification of T4 in whispered speech. There are several possible explanations for this result. First, it is possible that amusics are impaired to some extent in other perceptual cues for tone perception in whispered speech. As mentioned before, it is not yet clear as to what perceptual cues contribute to tone perception in whispered speech, but duration [14], amplitude contour [6], and formant frequency [7] are proposed to be three primary perception cues of whispered tone. It is also not very clear whether the amusics were impaired in these perceptual cues or not. To address these issues, systematic investigations are required in the future to examine how other perceptual cues such as duration, amplitude contour and formant frequency contribute to the identification of each tone and whether amusics are impaired in the perception of such cues. A second explanation for the reduced group difference is that the deficits of amusics prevail to the phonological representations of lexical tone, affecting tone recognition where pitch information is absent. However, this account cannot explain why the group difference in tone recognition was only found in T4, not in any other tones. Nonetheless, the seemingly comparable performance of amusics to the controls in the identification of T1 and T2 may be caused by a floor effect due to the greater difficulty to identify tones in whispered mode.

For the performance of both groups in whispered speech, T3 was the easiest tone to identify and T4 was the second easiest tone to identify, but the difficulties of T1 and T2 were different in two groups. The results were consistent with previous studies [1, 3, 4], which found that T3 and T4 were easier to recognize than T2 and T1. Gao [1] showed that some speakers tend to increase the falling slope for T4 in amplitude contour and to increase the falling and rising slope in amplitude contour for T3 in production. These acoustic features may aid the identifiability of T3 and T4 in whispered mode.

5. CONCLUSION

The results of this study indicated that Mandarin-speakers with amusia presented degraded performance in tone identification in both phonated and whispered modes, but the differences were smaller in whispered mode.

6. ACKONWLEDGEMENTS

This work was supported by grants from the National Natural Science Foundation of China (NSFC: 11504400), the Research Grants Council of Hong Kong (ECS: 25603916), and the PolyU Start-up Fund for New Recruits. The authors would like to thank Mr. Yulin Wen for the data collection.

7. REFERENCES

[1] Gao, M. 2002. Tones in whispered Chinese: articulatory features and perceptual cues. Master Thesis, University of Victoria, Canada.

[2] Huang, W. T., Liu, C., Dong, Q., Nan, Y. 2015. Categorical perception of lexical tones in mandarin-speaking congenital amusics. *Frontiers in psychology*, 6,829.

[3] Jensen, M.K., 1958. Recognition of Word Tones in Whispered Speech. *Word*, 14, 187-196.

[4] Jiao, L., Ma, Q., Wang, T., Xu, Y. 2015. Perceptual cues of whispered tones: Are they really special? *In Sixteenth Annual Conference of the International Speech Communication Association*.

[5] Jones, J. L., Lucker, J., Zalewski, C., Brewer, C., Drayna, D. 2009. Phonological processing in adults with deficits in musical pitch recognition. *Journal of communication disorders*, 42(3), 226-234.

[6] Li, B., Guo, Y. 2012. Mandarin tone contrast in whisper. *In Tonal Aspects of Languages-Third International Symposium*.

[7] Li, X. L., Xu, B. L. 2005. Formant comparison between whispered and voiced vowels in Mandarin. *Acta Acustica united with Acustica*, 91(6), 1079-1085.

[8] Liu, F., Patel, A. D., Fourcin, A., Stewart, L. 2010. Intonation processing in congenital amusia: Discrimination, identification and imitation, *Brain* 133, 1682–1693.

[9] Liu, F., Jiang, C., Wang, B., Xu, Y., Patel, A. D. 2015. A music perception disorder (congenital amusia) influences speech comprehension. *Neuropsychologia*, 66, 111-118.

[10] 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.

[11] Peretz, I., Champod, A. S., Hyde, K. L. 2003. Varieties of musical disorders: the montreal battery of evaluation of amusia. *Ann N Y Acad Sci*, 999, 58–75.

[12] Peretz, I., Cummings, S., Dubé, M. P. 2007. The genetics of congenital amusia (tone deafness): a family-aggregation study. *The American Journal of Human Genetics*, 81(3), 582-588.

[13] Thompson, W. F., Marin, M. M., Stewart, L. 2012. Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences*, 109(46), 19027-19032.

[14] Yang, L., Li Y.,Xu, B. 2005. The establishment of a Chinese whisper database and perceptual experiment (in Chinese). *Journal of NanJing University* (Natural Sciences), Vol, 41, No. 3.

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

[16] 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.

[17] Hyde, K. L., Peretz, I. 2004. Brains That Are Out of Tune but in Time. *Psychol Sci* 15.5:356-360.