Context integration deficit in tone perception in Cantonese speakers with congenital amusia

Jing Shao, and Caicai Zhang

Citation: The Journal of the Acoustical Society of America **144**, EL333 (2018); doi: 10.1121/1.5063899

View online: https://doi.org/10.1121/1.5063899

View Table of Contents: https://asa.scitation.org/toc/jas/144/4

Published by the Acoustical Society of America

Context integration deficit in tone perception in Cantonese speakers with congenital amusia

Jing Shao[a)] **and Caicai Zhang**[a),b)]
Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong Special Administrative Region, China
jing.shao@polyu.edu.hk, caicai.zhang@polyu.edu.hk

Abstract: Congenital amusia is a neuro-developmental disorder of pitch processing. This study investigated how this deficit affects lexical tone perception with and without context. Twenty-three Cantonese-speaking amusics and 23 controls were tested on the identification of high-variation tone stimuli in isolation vs in a carrier sentence. The controls generally achieved a higher accuracy with context than in isolation, suggesting that speech context facilitated tone identification. In contrast, amusics generally failed to benefit from the context, despite some variation among different tones. These findings provide insights into the underlying deficits of amusia, revealing a context integration deficit of tone perception in amusia.

[RS]
Date Received: April 9, 2018 **Date Accepted:** July 19, 2018

1. Introduction

Congenital amusia (amusia hereafter) is an inborn disorder of musical pitch processing, affecting approximately 1.5%–4% of the population (Peretz and Vuvan, 2017). The primary deficit in amusia lies in the fine-grained pitch discrimination (Hyde and Peretz, 2004) and impaired short-term memory for pitch (Tillmann *et al.*, 2016).

Empirical evidence has revealed that the pitch deficit in amusia is not specific to music, but is domain general and influences speech pitch processing negatively. For instance, several studies have found that lexical tone perception is impaired in both non-tonal and tonal language speakers with amusia (Nguyen *et al.*, 2009; Liu *et al.*, 2016, 2012; Nan *et al.*, 2010; Shao *et al.*, 2016). Amusical individuals in non-tonal languages showed inferior performance on the discrimination of non-native tone pairs in Mandarin and Thai (Nguyen *et al.*, 2009). For tonal language speakers with amusia, Nan *et al.* (2010) tested a group of 22 Mandarin-speaking amusics and found that half of them showed impaired tone identification and discrimination. Liu *et al.* (2016) replicated this result in amusical speakers of another tonal language—Cantonese, showing that Cantonese-speaking amusics were less accurate at discriminating four pairs of native Cantonese tones than musically intact controls. In a recent study, Shao *et al.* (2016) found that Cantonese-speaking amusics performed less accurately than controls in tone identification and discrimination and responded more slowly. Altogether, these results suggested that the pitch-processing deficit in amusia is domain general and impoverishes lexical tone perception.

Furthermore, some studies have suggested that high-level phonological processing of lexical tones is impaired in tonal language speakers with amusia. The most substantial evidence for this claim is that categorical perception of native tones in tonal language speakers with amusia is impaired (Jiang *et al.*, 2012; Zhang *et al.*, 2017). For instance, Zhang *et al.* (2017) examined categorical perception by Cantonese-speaking amusics and found that amusics exhibited significantly less benefit in between-category discriminations of lexical tones, suggesting reduced categorical perception of native tones.

Taken together, the above studies provided corroborating evidence that lexical tone perception in tonal language speakers with amusia is impaired. However, most of the previous studies focused on tone perception when tones were presented in isolation. Little attention had been given to the context effect on tone perception in amusia. Unlike tonal languages such as Mandarin, Cantonese has a highly dense tonal system (Peng *et al.*, 2012). It has three level tones that share the same $f0$ direction but primarily differ in $f0$ level, and acoustic variations generated by different talkers or syllables

---

[a)]Also at: Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China.
[b)]Author to whom correspondence should be addressed.

could easily lead to overlap between neighboring tone categories. As a result, the accuracy of tone identification in isolation in Cantonese was reported to be low, at approximately 35% for amusics and 60% for the controls (Shao *et al.*, 2016), in stark contrast to the close-to-ceiling accuracy in the case of Mandarin tone identification (85.4% for amusics and 98.8% for controls) (Nan *et al.*, 2010). Embedding the Cantonese tones into a speech context greatly decreased the ambiguity and enhanced the identification accuracy, suggesting that the perception of Cantonese tones relies more on context than tonal languages like Mandarin (Peng *et al.*, 2012).

As a matter of fact, the mapping between the acoustic signal and perceptual representations is not always one-to-one, but complicated by acoustic variations. It has been widely reported that the perception of a target speech signal is sensitive to acoustic cues in the surrounding context (Peng *et al.*, 2012; Zhang and Chen, 2016). Along this line, it has been demonstrated that tone perception is affected by the distribution of *F*0 cues in the speech context. For instance, Zhang and Chen (2016) found that shifting the *F*0 level in meaningless and meaningful speech contexts significantly changed the perception of a target tone. The phenomenon that listeners perceive a target speech signal by integrating acoustic cues carried by the target signal itself together with those in the context is referred to as "context integration" in the current study (cf. Massaro and Cohen, 1983).

While it is widely reported that the perception of Cantonese tones relies on the speech context, it remains unclear whether amusic individuals benefit from the speech context in tone perception to a similar extent as musically intact listeners. There is still a gap in the literature on amusia to directly compare tone perception in isolation versus in a larger speech unit, i.e., in a carrier sentence. Examination of this question could lead to a better understanding of how the deficit in amusia influences speech pitch processing. To this end, we compared the performance of Cantonese-speaking amusics and matched controls in two tone identification tasks, one in isolation (isolation condition), where the perception of target tones primarily relied on the *F*0 of the target word itself, and the other task with the same set of target tones presented within a carrier sentence as context (context condition), which provided pitch references to estimate the *F*0 of the target tones. The target tones were carried by ten different base syllables to create high-variation tone stimuli. Integrating the *F*0 of the target tones with the pitch references in the context is expected to reduce the influence of acoustic variations and to increase the accuracy of tone identification. If amusics are impaired in integrating the *F*0 information of the speech context with that of the target word, we predict that amusics would show reduced or no improvement in tone perception in the context condition over the isolation condition, whereas the controls would show a significant improvement in tone perception.

2. Method

2.1 Participants

Twenty-three amusics and 23 musically intact controls matched in age, gender, and years of education participated in the experiment. All participants were native speakers of Hong Kong Cantonese studying in local universities in Hong Kong. They were all right-handed, with no hearing impairment, no reported history of neurological illness, or musical training. Amusics and controls were diagnosed using the Online Identification Test of Congenital Amusia (Peretz *et al.*, 2008), which has been used as a tool for identifying amusics in several previous studies (e.g., Zhang *et al.*, 2017). Amusics scored 73% or lower, whereas controls scored 80% or higher in the global score of this test, which is the mean accuracy of all three subtests (out-of-key, offbeat, and mistuned). All amusics scored 71% (Nan *et al.*, 2010) or lower in at least one of the two pitch-based subtests (out-of-key and mistuned) or in both. Demographic characteristics of the participants are summarized in Table 1. Informed written consent was obtained from participants in compliance with the experiment protocols approved by the Human Subjects Ethics Sub-committee of the Hong Kong Polytechnic University.

2.2 Stimuli

The stimuli were 60 monosyllabic words contrasting six Cantonese tones (high level tone (T1)-/55/, high rising tone (T2)-/25/, mid-level tone (T3)-/33/, extra low level/low falling tone (T4)-/21/, low rising tone (T5)-/23/, and low level tone (T6)-/22/) on ten base syllables (/jɐn/, /ji/, /jɐu/, /jiu/, /fɐn/, /fu/, /ŋa/, /si/, /sɛ/, and /wɐi/), all of which were meaningful words in Cantonese. Each monosyllabic target word was embedded in a carrier phrase "呢個係__ /li55 kɔ33 hɐi22__/" [This is __]. One female Cantonese speaker was asked to record the stimuli, and each target word together with its carrier phrase was repeated 6 times. To make sure the target word sounded natural in isolation and to minimize the co-articulatory effects, the speaker was instructed to pause briefly between the carrier sentence and the target word. Stimulus selection and manipulation was conducted using Praat. One clearly produced token for each target word was chosen as the stimuli by two experimenters who were native Cantonese speakers with phonetics training. The target words were normalized in duration to 631 ms, equal to the mean duration of all selected target words. The carrier phrases preceding the selected target words were also segmented out and normalized in duration to 877 ms, equal to the mean duration of all selected carrier phrases. The mean intensity of the target words and the carrier phrases was scaled to 70 dB.

2.3 Procedure

The stimuli were presented in two identification tasks. In one condition, the 60 target words were presented in isolation; in the other condition, the same set of 60 target words was presented with the corresponding carrier sentence "呢個係__ /li55 kɔ33 hɐi22__/" [This is __]. In both conditions, the 60 target words were intermixed in a block and presented randomly to the participants. The target words were repeated 3 times, generating a total of 180 trials for each isolation/context condition (30 trials per tone in each condition). Participants were instructed to identify the tone carried by the word by pressing buttons 1–6 which corresponded to the six Cantonese tones on a computer keyboard within 5 s. No feedback was given. Before the test, participants were given a few practice trials to familiarize them with the procedure. The practice trials were /fu/ carrying the six tones, which were produced by a different talker not used in the experiment. The isolation condition was presented before the context condition, for the reason that the speech context may facilitate talker adaptation and enable listeners to accurately identify the tones. This effect may transfer to a subsequently presented isolation condition, where the same set of target words was used. As a result, it is possible that listeners may also be able to correctly identify the words even without the context. To minimize the transfer of talker adaptation effects, the isolation condition was presented first.

2.4 Data analysis

The identification accuracy and reaction time (RT) were analyzed. Response to each trial was coded as 1 or 0 (correct or incorrect) for each participant. To compare the accuracy of amusics and controls in the identification task, generalized mixed-effects models were fitted on the responses to each trial (1 or 0) with *group* (amusics and controls), *condition* (isolation and context), and *tone* (T1, T2, T3, T4, T5, and T6) as three fixed effects, and subjects and items (ten base syllables) as random effects. Two- and three-way interaction terms were also included as fixed effects in the models. In order to test the significance of fixed effects, a simple model with only the intercept as a factor was first fitted, and the factors *group*, *condition*, and *tone* were added to the model consecutively. Models were compared by likelihood ratio tests, and *p*-values were obtained from those tests.

RT was measured from the offset of the target stimuli. Both the correct and incorrect trials were included in the RT analysis. Trials that exceeded 3 standard deviations (SDs) of the mean RT in each condition were disregarded (0.9% of trials in the isolation condition and 0.34% of trials in the context condition). The RTs were log-transformed. Linear mixed-effects models were fitted on the log-transformed RT with *group* (amusics and controls), *condition* (isolation and context), and *tone* (T1, T2, T3, T4, T5, and T6) as three fixed effects, and subjects and items (ten base syllables) as random factors. Two- and three-way interaction terms were also included as fixed

Table 1. Demographic characteristics of 23 amusic participants and 23 control participants. Amusic and control participants were determined according to the global score of the Online Identification Test of Congenital Amusia (Peretz *et al.*, 2008) (http://www.brams.umontreal.ca/onlinetest). M = male; F = female.

	Amusics	Controls
No. of participants	23 (11 M, 12 F)	23 (11 M, 12 F)
Age (range)	22.32 ± 1.1 years (20.2–24.6 years)	21.86 ± 1.1 years (19.8–24.5 years)
Test of Congenital Amusia		
Out-of-key (SD)	62.3 (10.6)	91.1 (8.3)
Offbeat (SD)	64.3 (14.0)	87.5 (7.1)
Mistuned (SD)	60.1 (13.6)	86.3 (10.8)
Global score (SD)	62.4 (8.8)	88.3 (4.5)

effects in the models. Models were fitted and tested using the same procedures described above. The above two sets of analyses were performed with R, using the *lme4* package and the *lsmeans* package. Only significant effects were reported.

3. Results

The identification accuracy in the two conditions for the two groups is shown in Fig. 1. The generalized mixed-effects model found significant main effects of *group* [$\chi^2(1) = 12.014$, $p < 0.001$], *condition* [$\chi^2(1) = 56.966$, $p < 0.001$], and *tone* [$\chi^2(5) = 2540.8$, $p < 0.001$]. Furthermore, there were significant two-way interactions of *group* by *condition* [$\chi^2(1) = 24.743$, $p < 0.001$], *group* by *tone* [$\chi^2(5) = 51.123$, $p < 0.001$], and *condition* by *tone* [$\chi^2(5) = 11.809$, $p = 0.037$]. The three-way interaction was also significant [$\chi^2(5) = 28.322$, $p < 0.001$].

To analyze the three-way interaction effect, further models were constructed with *group* and *condition* as the fixed effects together with the two-way interaction item for each of the six tones. In brief, significant *group* by *condition* interactions were found on T1 (high level tone), T3 (mid-level tone), T4 (extra low level/low falling tone), and T5 (low rising tone), where controls showed a significantly higher accuracy than amusics in the context condition, whereas the group difference was either reduced or not significant in the isolation condition. Furthermore, controls showed a significantly higher accuracy in the context vs isolation condition, whereas the benefit of the context condition was either reduced or absent in amusics. As for the two remaining tones—T2 (high rising tone) and T6 (low level tone)—the two-way interaction was not significant. Unsurprisingly, there was a significant main effect of *condition*, showing that the context condition elicited a higher accuracy than the isolation condition; the *group* effect was marginally significant in the case of T2 and significant in T6, with amusics obtaining a lower accuracy than controls.

For T1 (high level tone), there was a significant main effect of *group* [$\chi^2(1) = 6.89$, $p = 0.008$] and significant two-way interaction [$\chi^2(1) = 3.931 p = 0.047$]. Pairwise comparisons showed that controls showed a higher accuracy than amusics in both conditions (Context condition: $z = -2.958$, $p = 0.003$; isolation condition: $z = -2.168$, $p = 0.030$), but the group difference was more evident in the context condition. Besides, a trend of a higher accuracy in the context condition than in the isolation condition can be observed in the control group ($z = -1.779$, $p = 0.075$), but not in the amusic group ($z = 0.924$, $p = 0.355$). For T2 (high rising tone), there was a significant main effect of *condition* [$\chi^2(1) = 33.424$, $p < 0.001$], where tone identification accuracy in the context condition was significantly better than that in the isolation condition. The effect of *group* was marginally significant [$\chi^2(1) = 2.883$, $p = 0.089$], with amusics scoring lower than controls. For T3 (mid-level tone), the fixed effects of *group* [$\chi^2(1) = 3.897$, $p = 0.048$], *condition* [$\chi^2(1) = 19.446$, $p < 0.001$], and two-way interaction [$\chi^2(1) = 4.483$, $p = 0.034$] were all significant. Pairwise comparisons showed that controls performed better than amusics in the context condition ($z = -2.463$, $p = 0.013$), but not

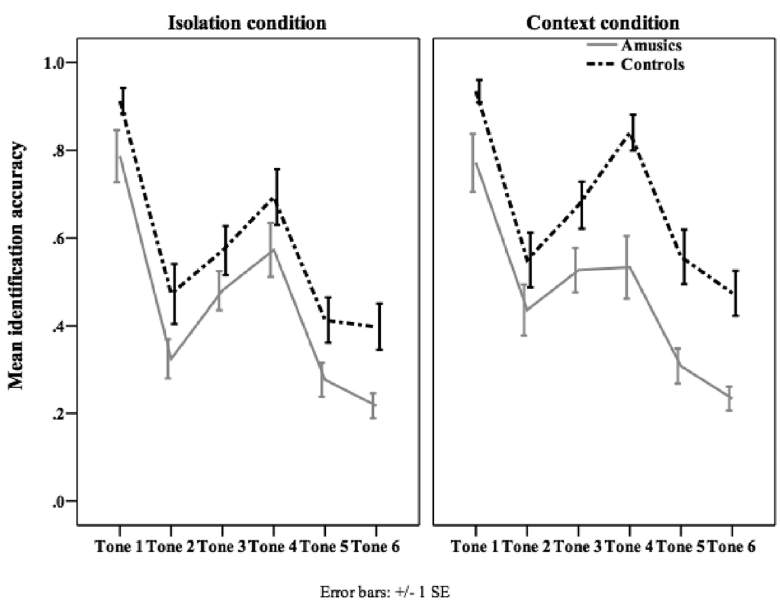

Fig. 1. Mean identification accuracy in isolation and context conditions for the amusic and control group.

in the isolation condition. For the control group, tone identification accuracy with a preceding context was significantly higher than that in the isolation condition ($z = -4.526$, $p < 0.001$), whereas for the amusic group, this effect was not significant. For T4 (extra low level/low falling tone), the fixed effects of *group* [$\chi^2(1) = 8.502$, $p = 0.003$], *condition* [$\chi^2(1) = 12.207$, $p < 0.001$] and two-way interaction [$\chi^2(1) = 45.459$, $p < 0.001$] were all significant. Pairwise comparisons showed that the group difference was significant in the context condition ($z = -4.226$, $p < 0.001$), but not in the isolation condition. Controls obtained a significantly higher accuracy in the context condition than in the isolation condition ($z = -7.214$, $p < 0.001$), whereas amusics failed to show a significant difference. For T5 (low rising tone), there were significant effects of *group* [$\chi^2(1) = 8.741$, $p = 0.003$] and *condition* [$\chi^2(1) = 21.241$, $p < 0.001$] and significant two-way interaction [$\chi^2(1) = 10.53$, $p = 0.001$]. Pairwise comparisons revealed that amusics showed a lower accuracy than controls in both the context condition ($z = -3.798$, $p < 0.001$) and isolation condition ($z = -2.004$, $p = 0.045$), but the group difference was more robust in the context condition. For controls, their tone identification accuracy in the context condition was significantly higher than that in the isolation condition ($z = -5.507$, $p < 0.001$), but amusics failed to show this effect. For T6 (low level tone), only the fixed effects of *group* [$\chi^2(1) = 12.75$, $p < 0.001$] and *condition* [$\chi^2(1) = 6.810$, $p = 0.009$] were significant, revealing that amusics performed with a lower accuracy than controls, and that the context condition elicited a higher accuracy than the isolation condition.

Figure 2 displays the RT for the two groups. Linear mixed-effects models found significant main effects of *condition* [$\chi^2(1) = 285.06142.75$, $p < 0.001$], where RT in the context condition was significantly shorter than that in the isolation condition. Moreover, there were significant main effects of *tone* [$\chi^2(5) = 3430.42970.7$, $p < 0.001$], and two-way interaction of *group* by *tone* [$\chi^2(5) = 157.93$, $p < 0.001$]. Pairwise comparisons were conducted to explore the *group* by *tone* interaction. Within the control group, the RT on T1 and T4 was significantly shorter than the other tones (ps < 0.001), while T1 showed an even shorter RT than T4 ($p < 0.001$); the RT on T5 was significantly longer than the other tones (ps < 0.01). Within the amusic group, the RT on T1 was significantly shorter than the other tones (ps < 0.001); T5 showed a significantly longer RT than the other tones (ps < 0.001). No other effects were significant. Within each tone, no significant group difference was found. This result suggests that in general for both controls and amusics, T1 (high level tone) was identified with the shortest RT, while T5 was the most difficult to identify, eliciting the longest RT.

To summarize, the accuracy results showed that except for T2 and T6, there were significant *group* by *condition* interaction effects on the identification of Cantonese tones, where controls consistently showed a significantly higher accuracy when identifying tones presented with a preceding speech context but amusics failed to demonstrate such effects. This suggests that controls benefited from the preceding context in lexical tone perception, whereas this facilitatory effect was largely absent in the performance of amusics. On the other hand, while controls achieved a higher accuracy than amusics in both isolation and context conditions, the group difference was larger in the context condition, which contained pitch references to estimate the target tones. As for the RT results, they mainly revealed that RT in the context condition was significantly shorter, implying that the identification of lexical tones was much easier with the context providing pitch references. Furthermore, T1 was identified with the shortest RT, whereas T5 was identified with the longest RT.

4. Discussion

While amusia has been consistently reported to affect lexical tone perception (Jiang *et al.*, 2012; Liu *et al.*, 2016; Nan *et al.*, 2010), it remains unclear how amusic individuals' context integration ability in tone perception is affected. In this study, we compared Cantonese-speaking amusics and controls on the identification of high-variation tone stimuli in isolation vs in a carrier sentence, with the aim of revealing more about how the deficit in amusia influences speech pitch processing.

We found that amusics showed a lower accuracy than controls in the isolation condition, which broadly corroborates previous findings that amusics are impaired in tone perception without context (Nan *et al.*, 2010; Shao *et al.*, 2016; Zhang *et al.*, 2017). In the present study, we increased the syllable variability, i.e., including ten base syllables, and thus increased the difficulty of exacting tonal categories from the speech signals. Echoing with previous findings, the current study confirmed that amusics performed worse than controls in tone identification when the tones were auditorily presented in isolation.

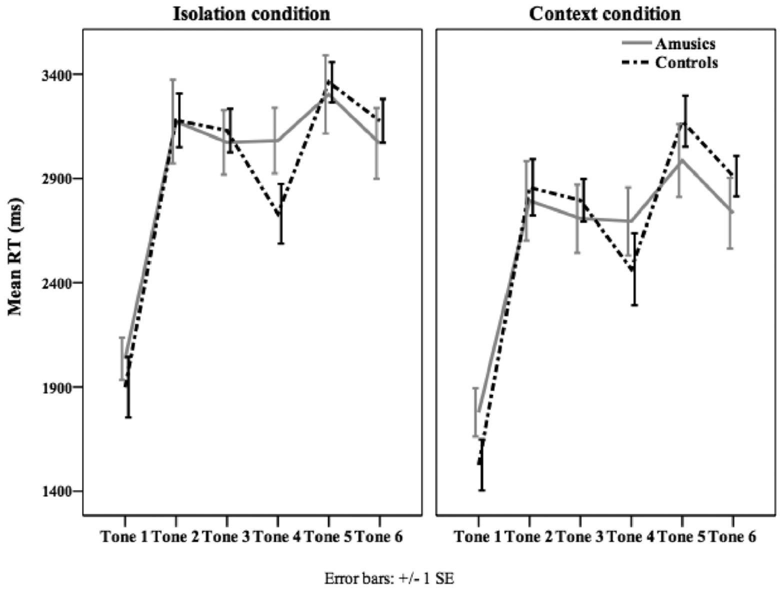

Fig. 2. Mean identification response time in isolation and context conditions for the amusic and control group.

Furthermore, controls showed a clear pattern that the identification of tones with speech context was significantly more accurate than that in isolation, implying that the carrier phrase facilitated tone perception. In contrast, the amusic group failed to show such effects, as their performance in the two conditions was not significantly different in the identification of most tones. These results suggested that amusic individuals benefited less from pitch references in speech context in tone perception. The observation that amusics are impaired in making use of the $F0$ cues in the preceding speech context can be attributed to a context integration deficit. In a previous study, it is suggested that the speech context contained phonological cues for listeners to form references of a talker's tonal space to evaluate the $F0$ of the target tone (Zhang and Chen, 2016). In the current study, it appears that the controls were able to successfully take advantage of such cues in the carrier sentence to identify the target tone. In contrast, amusics failed to make use of the phonological cues in the carrier sentence in most cases, which suggests that amusics have a context integration deficit.

Some interesting variation among the six tones can be noted in terms of the benefit of the speech context in contrast to the isolation condition. For T6, although the *group* by *condition* interaction effect was not significant [$\chi^2(1)=2.656$, $p=0.103$], a trend of larger group differences in the context condition can be observed. Pairwise comparisons with p-value corrected for multiple comparisons ($0.05/2=0.025$) confirmed this observation. The group difference was significant under both the isolation ($z=-3.120$, $p=0.001$) and context condition ($z=-4.101$, $p<0.001$), but the effect size was larger in the context condition. Furthermore, only controls showed an enhanced accuracy in the context condition over the isolation condition ($z=-3.023$, $p=0.002$), whereas amusics did not ($z=-0.549$, $p=0.583$). This result suggested that the response patterns on T6 were more or less in line with the other tones in that amusics showed no benefit from the speech context compared with controls. On the other hand, the identification of T2 showed a significant *condition* effect and a marginally significant *group* effect. The response patterns on T2 appeared to deviate from the other tones and require an explanation. One possible explanation is that T2 (high-rising tone) is acoustically similar to and often perceptually confused with T5 (low-rising tone), but T2 is usually more accurately identified than T5 (Peng *et al.*, 2012; Shao *et al.*, 2016). Therefore, the identification of T2 might benefit from the presence of speech context in the disambiguation from T5, but both groups appeared to show more or less comparable competence in the accurate identification of T2. Interestingly, while both groups showed an improvement in the identification of T2 in the context condition relative to the isolation condition, only the control group showed such an improvement in the identification of T5. It may suggest that amusics can benefit from the speech context to some extent in identifying the relatively easier tone, i.e., T2.

It is also worth noting that the *group* by *condition* interaction was only found on the identification accuracy, not on the RT data. The RT results generally revealed a longer RT in the isolation condition than the context condition and some differences among the six tones. While the lack of group differences in the RT results might hint at more or less intact processing time of amusics in context integration, this result should be interpreted with caution. For one thing, it is possible that there might be a ceiling effect on RT in the context condition, such that there is limited room for controls to demonstrate faster RT than amusics in this condition. Future studies can further explore the impairment of amusics in their ability of context integration in tone perception and processing time.

While we conclude that the impairment of Cantonese-speaking amusics in tone perception can be mainly attributed to a context integration deficit, it is possible that potential attention deficits and impairment in short-term pitch memory in some amusics may also contribute to the observed impairment to some extent. It was reported that about 40% of the amusics have an attention deficit (Jones *et al.*, 2009), which may hinder them from allocating proper attentional resources to the processing of the preceding speech context to extract useful cues for tone perception. Moreover, a number of studies have shown that amusics showed impaired short-term memory for pitch (e.g., Tillmann *et al.*, 2016). The deficient short-term pitch memory system might have increased the difficulty for amusics to maintain the pitch references contained in the proceeding speech context in memory during the process of tone identification. Future studies may further investigate the impact of attention and pitch memory deficits on tone perception in amusia.

To conclude, we found that Cantonese-speaking amusics showed degraded performance in tone identification with a preceding speech context compared to musically intact listeners. This indicates that the deficit of amusia is not confined to tone perception in isolation, but also affects tone perception in a carrier sentence. These findings can be largely accounted for by a context integration deficit in amusia. Our findings provide a fuller understanding on the nature of the deficits of amusia in speech pitch processing. Last, the isolation condition was always presented before the context condition in the current study, which might have contributed to the enhancement of performance in the context condition to some extent. While the group difference observed in the current study cannot be simply attributed to this order effect for the reason that amusics and controls received the same order of conditions, future studies may consider counterbalancing the isolation and context condition.

Acknowledgments

This work was supported by grants from the Research Grants Council of Hong Kong (ECS: Grant No. 25603916), the National Natural Science Foundation of China (NSFC: Grant No. 11504400), and the PolyU Start-up Fund for New Recruits. We thank Xiao Yao for help with data collection.

References and links

Hyde, K. L., and Peretz, I. (**2004**). "Brains that are out of tune but in lime," Psychol. Sci. **15**(5), 356–360.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., and Yang, Y. (**2012**). "Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics," Mem. Cognit. **40**(7), 1109–1121.

Jones, J. L., Zalewski, C., Brewer, C., Lucker, J., and Drayna, D. (**2009**). "Widespread auditory deficits in tune deafness," Ear Hear. **30**(1), 63–72.

Liu, F., Chan, A. H. D., Ciocca, V., Roquet, C., Peretz, I., and Wong, P. C. M. (**2016**). "Pitch perception and production in congenital amusia: Evidence from Cantonese speakers," J. Acoust. Soc. Am. **140**(1), 563–575.

Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., and Stewart, L. (**2012**). "The mechanism of speech processing in congenital amusia: Evidence from Mandarin speakers," PLoS One **7**(2), e30374.

Massaro, D. W., and Cohen, M. M. (**1983**). "Phonological context in speech perception," Percept. Psychophys. **34**(4), 338–348.

Nan, Y., Sun, Y., and Peretz, I. (**2010**). "Congenital amusia in speakers of a tone language: Association with lexical tone agnosia," Brain. **133**(9), 2635–2642.

Nguyen, S., Tillmann, B., Gosselin, N., and Peretz, I. (**2009**). "Tonal language processing in congenital amusia," Ann. N. Y. Acad. Sci. **1169**, 490–493.

Peng, G., Zhang, C., Zheng, H.-Y., Minett, J. W., and Wang, W. S.-Y. (**2012**). "The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems," J. Speech, Lang. Hear. Res. **55**(2), 579–595.

Peretz, I., Gosselin, N., Tillmann, B., Cuddy L. L., Gagnon, B., Trimmer, G. C., Paquette, S., and Bouchard, B. (**2008**). "On-line identification of congenital amusia," Music Percept. **25**(4), 331–343.

Peretz, I., and Vuvan, D. T. (**2017**). "Prevalence of congenital amusia," Eur. J. Hum. Genet. **25**, 625–630.

Shao, J., Zhang, C., Peng, G., Yang, Y., and Wang, W. S.-Y. (**2016**). "Effect of noise on lexical tone perception in Cantonese-speaking amusics," in *Proceedings of the Interspeech*, San Francisco, CA.

Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., and Caclin, A. (**2016**). "Impaired short-term memory for pitch in congenital amusia," Brain Res. **1640**(Part B), 251–263.

Zhang, C., and Chen, S. (**2016**). "Toward an integrative model of talker normalization," J. Exp. Psychol. Hum. Percept. Perform. **42**, 1252–1268.

Zhang, C., Shao, J., and 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.