

Lower-level acoustics underlie higher-level phonological categories in lexical tone perception

Fei Chen, and Gang Peng

Citation: [The Journal of the Acoustical Society of America](#) **144**, EL158 (2018); doi: 10.1121/1.5052205

View online: <https://doi.org/10.1121/1.5052205>

View Table of Contents: <https://asa.scitation.org/toc/jas/144/3>

Published by the [Acoustical Society of America](#)

ARTICLES YOU MAY BE INTERESTED IN

[Computer-based auditory training improves second-language vowel production in spontaneous speech](#)

[The Journal of the Acoustical Society of America](#) **144**, EL165 (2018); <https://doi.org/10.1121/1.5052201>

[Differences in cue weights for speech perception are correlated for individuals within and across contrasts](#)

[The Journal of the Acoustical Society of America](#) **144**, EL172 (2018); <https://doi.org/10.1121/1.5052025>

[Modulation spectrum-constrained trajectory error training for mixture density network-based speech synthesis](#)

[The Journal of the Acoustical Society of America](#) **144**, EL151 (2018); <https://doi.org/10.1121/1.5052206>

[Lombard effect, ambient noise, and willingness to spend time and money in a restaurant](#)

[The Journal of the Acoustical Society of America](#) **144**, EL209 (2018); <https://doi.org/10.1121/1.5055018>

[Spectral change and duration as cues in Australian English listeners' front vowel categorization](#)

[The Journal of the Acoustical Society of America](#) **144**, EL215 (2018); <https://doi.org/10.1121/1.5055019>

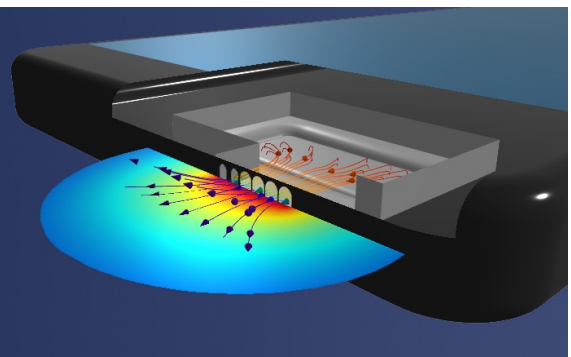
[Acoustic source characterization for a logging while drilling tool: Theoretical and experimental modeling](#)

[The Journal of the Acoustical Society of America](#) **144**, EL178 (2018); <https://doi.org/10.1121/1.5053105>

SIMULATION STORIES FOR ACOUSTICS

Acoustics engineers are using simulation for NVH testing, microphone and transducer design, and more.

[READ NOW »](#)



Lower-level acoustics underlie higher-level phonological categories in lexical tone perception

Fei Chen and Gang Peng^{a)}

*Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University,
Hong Kong Special Administrative Region
chenfeianthony@gmail.com, gpengjack@gmail.com*

Abstract: The pitch-processing deficit associated with congenital amusia has been shown to be transferable to lexical tone processing. However, it remains unclear whether the tone perception difficulties of amusics are merely due to the domain-general deficit in acoustic processing or additionally caused by impaired higher-level phonological operations. Answers to this question can shed light on the influence of lower-level acoustic processing on higher-level phonological processing. Using a modified categorical perception paradigm, the present study indicates that the acoustic processing deficit systematically extends to higher-level phonological processing. These findings suggest that lower-level acoustics underlie higher-level phonological categories in lexical tone perception.

© 2018 Acoustical Society of America

[QJF]

Date Received: May 25, 2018 Date Accepted: August 14, 2018

1. Introduction

According to neuroscience models for language processing, lexical tone perception generally involves two different neural pathways in a hierarchical manner performing lower-level acoustic processing and higher-level phonological processing, respectively (Zhang *et al.*, 2011). Thus, listeners' speech perception performance can be modulated by their experience with each type of processing (Chandrasekaran *et al.*, 2009). Categorical perception (CP) of speech sounds reflects a fundamental property of speech. The contrast between acoustic processing and phonological processing could be observed as the contrast between within-category and between-category variations by adopting a modified CP paradigm (Zhang *et al.*, 2011), which directly dissociates the acoustic and phonological processing of lexical tones for native participants. A typical CP pattern of speech sounds, which is characterized by a sharper identification boundary and a higher between-category discrimination accuracy with age, occurs only in native speakers and develops gradually due to accumulated higher-level phonological inputs (Peng *et al.*, 2010; Chen *et al.*, 2017).

Investigations into the relative influence of the acoustic and phonological processing strategies can shed light on the perceptual mechanisms involved in speech perception. Pitch perception offers a chance to examine this issue due to its important role in both the music and speech domains. In music, contour and interval are the two main types of pitch information which are used to form melodies. For lexical tone perception, differences in pitch height and direction lead to a change of lexical meanings for tone language speakers. Compared to non-musicians, individuals with extensive musical training early in life tend to be more sensitive to the pitch information of music, pure tones, and speech sounds, reflecting a domain-transfer ability of lower-level acoustic processing (Wong *et al.*, 2007; Zhao and Kuhl, 2015). It is important to explore whether and how the enhanced acoustic sensitivity facilitates higher-order phonological categories, which can help uncover the influence of the lower-level acoustic processing on the higher-level phonological processing. The CP of lexical tones in Mandarin-speaking musicians provides an optimal window for investigating this topic since both processing strategies are engaged in CP tasks. As expected, previous findings showed that within-category discrimination accuracy was enhanced in Mandarin-speaking musicians due to facilitated acoustic processing skill. However, their performance in terms of both identification steepness and between-category discrimination accuracy was similar to that of Chinese non-musicians (Wu *et al.*, 2015). These results indicate that higher-level perceptual sensitivity to the phonological contrasts of the native language is robust and less likely to be facilitated by improvement in the

^{a)} Author to whom correspondence should be addressed. Also at: CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Shenzhen 518055, China.

lower-level acoustic processing, a view that is also supported by physiological findings in a neuroimaging study (Elmer *et al.*, 2012).

Exploring the influence of lower-level acoustic processing on higher-level phonological processing can also be turned in the other direction: whether and how reduced acoustic sensitivity could negatively influence phonological processing in native speakers. Some individuals with inborn musical deficits suffer from lifelong problems in perceiving and producing music in the absence of hearing loss or brain injury. This developmental disorder is termed congenital amusia (amusia hereinafter). Accumulating evidence has suggested that the perceptual deficit in musical melody in tone-language speakers with amusia transfers to pitch perception in speech, such as aberrant lexical tone perception either at a group level (Jiang *et al.*, 2012; Zhang *et al.*, 2017) or in a subgroup of amusics (Huang *et al.*, 2015a; Huang *et al.*, 2015b; Nan *et al.*, 2010). However, it is still unclear whether the negative domain-transfer effects of musical processing deficit on native lexical tone perception can be attributed to amusics' lower sensitivity to acoustic pitch information and/or reduced internal processing of phonological categories.

In this study, a modified CP paradigm was employed to compare the behavioral performance of Mandarin-speaking amusics and non-amusics in terms of their fine-grained perception of a continuum ranging from Mandarin Tone 1 (high-level tone) to Tone 2 (mid-rising tone). Moreover, since lexical tones are defined primarily in terms of pitch contour, it is natural to ask whether native Mandarin amusics respond to the pitch contours of nonspeech sounds in the same way as controls. Both natural speech and nonlinguistic analogues were used as the perceptual materials. We hypothesized that amusics would be less accurate in discriminating the within-category pairs due to their inferior skills in acoustic processing. Moreover, if higher-level phonological perception—developed early in native Mandarin-speaking amusics—is robust enough to resist the influence of impaired acoustic sensitivity, the sharpness of the identification curve and the between-category accuracy in the amusic group would be similar to that in the control group. Alternatively, if lower-level acoustics lay the foundation of the higher-level phonological categories in lexical tone processing, the amusic participants would perform systematically worse in all the CP measurements.

2. Method

2.1 Participants

Thirty amusic participants and 30 musically-intact control participants matched in terms of chronological age, gender, and level of education were recruited for the experiment. Although no power analysis was performed for calculation of sample size, the sample size of the amusic and control participants in the current study was more than twice that of subjects in previous studies on lexical tone perception in amusics (cf. Huang *et al.*, 2015a; Jiang *et al.*, 2012; Zhang *et al.*, 2017). To identify the presence or absence of amusia, the online Montreal Battery of Evaluation of Amusia (<http://www.brams.org/amusia-public/?lang=en>) was used in the screening stage, with the cut-off average score set to 71% (Peretz *et al.*, 2008). The demographic characteristics of the participants are presented in Table 1. All the participants were native Mandarin speakers with normal hearing, and none of them had received formal musical training. Approval of the study was granted by the Behavioral Research Ethics Committee of the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences.

2.2 Materials and procedure

The Mandarin monosyllable /i/ with the high-level tone (Tone 1, around 290 Hz) was recorded by a native female speaker (22 050 Hz sampling rate, 16-bit resolution). On

Table 1. Participants' characteristics and scores of the Online Identification Test of Congenital Amusia. The results (*p*-value) of independent-samples *t*-tests are also reported.

	Controls	Amusics	<i>p</i> -value
No. of participants	30 (16 M, 14 F)	30 (15 M, 15 F)	N/A
Age (SD)	23.10 years (4.04)	22.13 years (2.37)	<i>p</i> = 0.26
Online Identification Test of Congenital Amusia			
Out-of-key (SD)	87.67% (8.66)	65.80% (11.18)	<i>p</i> < 0.001
Offbeat (SD)	83.13% (8.31)	65.63% (10.69)	<i>p</i> < 0.001
Mistuned (SD)	78.47% (12.70)	54.77% (9.21)	<i>p</i> < 0.001
Mean score (SD)	83.36% (6.44)	62.20% (5.28)	<i>p</i> < 0.001

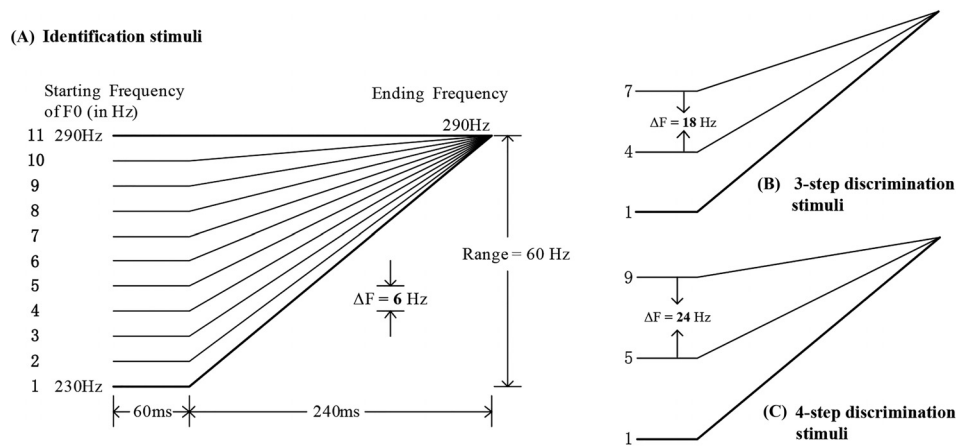


Fig. 1. (A) The 11 identification stimuli along the tone continuum ranging from Tone 2 to Tone 1, with a step size of 6 Hz. (B) The stimuli in the 3-step discrimination task. (C) The stimuli in the 4-step discrimination task.

the basis of the natural speech template with Tone 1, the 11 300-ms tone stimuli were resynthesized by Praat, and the procedures for synthesizing these stimuli followed those described in Peng *et al.* (2010). Figure 1 shows the schematic diagram of the pitch contours of the tone continuum from typical Tone 2 (stimulus #1, meaning “aunt”) to Tone 1 (stimulus #11, meaning “clothes”), with an equal step size of 6 Hz. Moreover, to investigate the influence of acoustic pitch differences on discrimination performance, two sets of discrimination stimuli were chosen, each with an 18-Hz onset difference (3-step discrimination) and a 24-Hz difference between the pair (4-step discrimination), respectively. The rationale for adopting both 18- and 24-Hz step sizes was based on just-noticeable differences (JNDs) in tone contour changes in controls (Liu, 2013) and two amusia subgroups: pure amusia and tone agnosia (Huang *et al.*, 2015a; Huang *et al.*, 2015b). To rule out a possible confounding effect of the step size, the larger acoustic difference in the present discrimination task was specifically manipulated to be 24 Hz, higher than the JNDs of all the participants. Using triangle waves (cf. Chen and Peng, 2016), an equal number of nonspeech analogues were constructed with exactly the same F_0 contour as that in the speech materials. To further match the loudness level, the intensity level of the nonspeech stimuli was set to 80 dB, 15 dB higher than that of the speech counterparts. Three native listeners rated that the nonspeech stimuli [80 dB sound pressure level (SPL)] sounded similar in loudness level to the speech stimuli (65 dB SPL).

All 60 amusic and control participants were first asked to conduct an identification task. Each participant gave their responses by clicking the corresponding button labeled “Tone 1” or “Tone 2.” The 11 stimuli of each speech or nonspeech continuum were repeated nine times, making a total of 99 identification trials in one block. Listeners completed two blocks of tone identification, speech and nonspeech, with the order of the two blocks randomized. Afterwards, half of the amusic and control participants (15 subjects in each group) performed the 3-step discrimination task and the other half performed the 4-step discrimination task. These two discrimination tasks shared the same experimental procedure, except for the different stimulus pairs used. For example, in the 3-step discrimination task, seven testing pairs were utilized: four pairs (different pairs) consisting of two different stimuli in either forward (1–4, 4–7) or reverse order (4–1, 7–4), and three pairs (same pairs) each paired with itself (1–1, 4–4, 7–7). These seven pairs were repeated randomly nine times, making a total of 63 pairs in each block, with a 500 ms inter-stimulus interval. The two blocks of tone discrimination (speech and nonspeech) were also presented to each participant separately and randomly. After each pair was presented, the participants needed to judge whether the two stimuli were the same or different and to respond quickly by pressing the corresponding key. The differences between the current study and the previous CP studies in Mandarin-speaking amusics (Huang *et al.*, 2015a; Jiang *et al.*, 2012) include sample size, step size, and the number of comparison units. Also, different from Huang *et al.* (2015a), the amusics in this study were not divided into two subgroups of pure amusia and tone agnosia.

2.3 Data analyses

The identification score was calculated as the percentage of responses with which participants identified a particular sound as being either Tone 1 or Tone 2. Boundary

position and boundary width were obtained by using Probit analyses (Finney, 1971). The boundary position, defined as the corresponding 50% crossover point in a continuum, and the boundary width, defined as the linear distance in the stimulus step between the 25th and 75th percentiles, was analyzed. The boundary position indicates the perceptual boundary straddling the two categories, and the boundary width is an index of the sharpness of the response shift around the categorical boundary.

To calculate the obtained discrimination scores, we divided the seven discrimination pairs into two comparison units (units 1–4 and 4–7, or units 1–5 and 5–9), each consisting of pairs of four types (AB, BA, AA, and BB). Then, the discrimination accuracy was analyzed using the sensitivity index d' (Macmillan and Creelman, 2008), which was computed as the z -score of the hit rate (“different” responses to different pairs: AB and BA) minus that of the false alarm rate (different responses to identical pairs: AA and BB) for each comparison unit. Moreover, in regard to the boundary position, we further subdivided the two comparison units into a between-category comparison unit and a within-category unit for each subject.

3. Results

3.1 Identification results

Figure 2 shows the identification curves for the amusics and the controls. All the data on position and width values in speech and nonspeech continua satisfied the homogeneity of variance and normality assumptions ($p > 0.05$). The mean boundary positions [standard deviations (SDs)] were 5.24 (0.83) and 5.26 (0.65) in speech and 5.47 (0.84) and 5.23 (1.01) in nonspeech for the controls and the amusics, respectively. The boundary position was analyzed by a two-way analysis of variance (ANOVA), with *group* (control vs amusic) as a between-subject factor and *stimulus type* (speech vs nonspeech) as a within-subject factor. Greenhouse–Geisser corrections were conducted when appropriate. Neither main effects nor the interaction effect were significant ($p > 0.05$), indicating that the boundary position straddling the two categories remained constant between the two subject groups and between different stimulus types.

To compare the higher-level phonological performance of the amusic and control groups in the identification task, the index of boundary width—reflecting the steepness of the category boundary—was calculated. The mean boundary widths (SDs) were 1.44 (0.61) and 1.86 (0.75) in speech and 1.43 (0.53) and 2.17 (1.28) in nonspeech for the controls and the amusics, respectively. A two-way ANOVA revealed that neither the main effect of *stimulus type* ($F(1, 58) = 1.56, p = 0.22, \eta_p^2 = 0.02$) nor the interaction between *group* and *stimulus type* ($F(1, 58) = 1.77, p = 0.19, \eta_p^2 = 0.03$) reached significance. Nevertheless, the ANOVA yielded a significant main effect for *group* on boundary width ($F(1, 58) = 10.39, p < 0.01, \eta_p^2 = 0.15$). This finding indicated that the control participants exhibited a much narrower width (i.e., a sharper slope) compared to the amusics pooled across speech and nonspeech continua (see Fig. 2). The *post hoc* power to detect the effect of *group* was 0.89 with an alpha of 0.05.

3.2 Discrimination results

The d' values of the between-category and within-category comparison units under different conditions are displayed in Fig. 3. The d' values satisfied the homogeneity of variance and normality assumptions ($p > 0.05$). A four-way repeated-measure ANOVA was conducted on the sensitivity index d' , with *stimulus type* (speech vs nonspeech) and *category type* (within-category vs between-category) as two within-subject factors and *group* (control vs amusic) and *step size* (3-step vs 4-step) as two between-subject

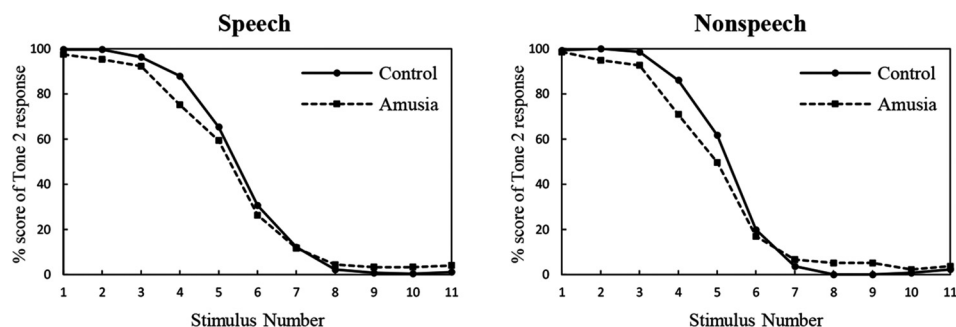


Fig. 2. The identification curves in two groups, where the solid lines with round dots represent Tone 2 responses in controls and the dashed lines with square dots represent Tone 2 responses in amusics.

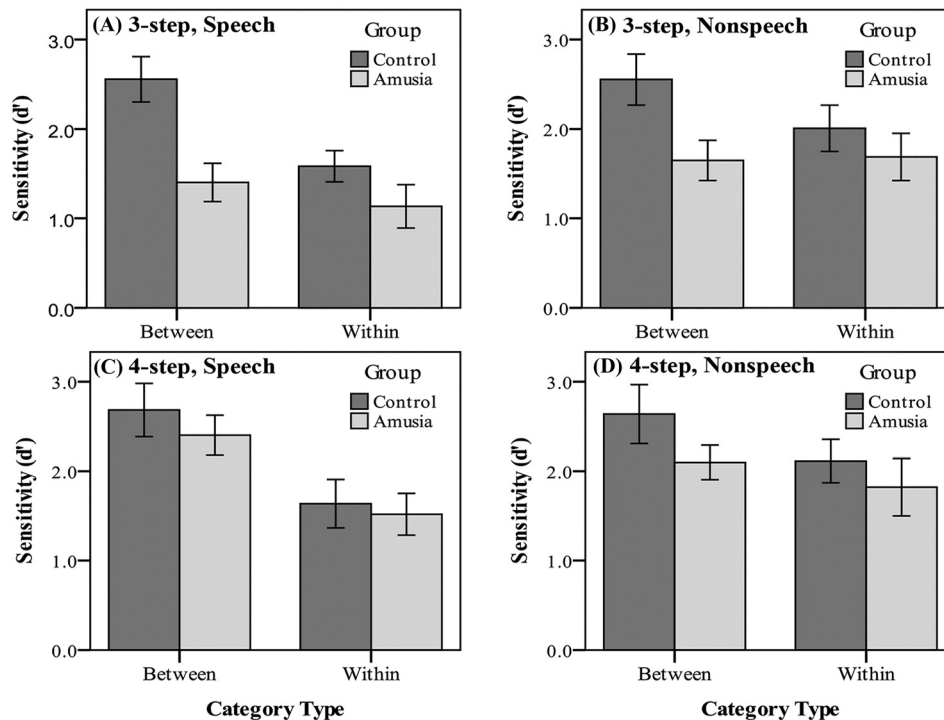


Fig. 3. The d' of the between- and within-category units for the control and amusic participants under different conditions: (A) 3-step, Speech, (B) 3-step, Nonspeech, (C) 4-step, Speech, (D) 4-step, Nonspeech. Error bars: ± 1 standard error.

factors. There were significant main effects of *group*, $F(1, 56) = 17.06$, $p < 0.001$, $\eta_p^2 = 0.23$, and *step size*, $F(1, 56) = 5.66$, $p < 0.05$, $\eta_p^2 = 0.09$, on d' values. The *post hoc* power to detect the effect of *group* and *step size* with an alpha of 0.05 was 0.98 and 0.65, respectively. The presence of the significant main effect of *group* on d' values and the absence of its interaction with the other three factors ($p > 0.05$) indicated that amusics were systematically inferior to the controls in discriminating both within-category and between-category pitch differences embedded in both speech and nonspeech.

In addition, a significant main effect of *category type*, $F(1, 56) = 12.37$, $p < 0.01$, $\eta_p^2 = 0.18$, and *category type* \times *stimulus type* interaction, $F(1, 56) = 4.46$, $p < 0.05$, $\eta_p^2 = 0.07$, were observed. Next, a simple main effect analysis of the *category type* \times *stimulus type* interaction was conducted with Bonferroni adjustment. These results indicated that for both groups, the d' value of the between-category unit tended to be higher than that of the within-category unit in the speech stimuli ($F(1, 56) = 19.46$, $p < 0.001$, $\eta_p^2 = 0.26$, *post hoc* power = 0.99), while the d' values of the two category types were not different from each other in the nonspeech stimuli ($F(1, 56) = 2.50$, $p = 0.12$, $\eta_p^2 = 0.04$). Furthermore, for the between-category unit, the d' value in speech (mean = 2.26) was similar to that in nonspeech (mean = 2.23) for both groups, $F(1, 56) = 0.03$, $p = 0.87$, $\eta_p^2 = 0.001$. However, nonspeech (mean = 1.91) yielded a much higher d' value of within-category unit than speech (mean = 1.47) for both groups, $F(1, 56) = 8.32$, $p < 0.01$, $\eta_p^2 = 0.13$, *post hoc* power = 0.81.

4. Discussion

It has been well established that pitch-processing deficits in music do stretch into a broader domain of influence, such as an inferior performance in the perception of pitch-based lexical tone (e.g., Jiang *et al.*, 2012; Zhang *et al.*, 2017) in tone-language speakers with amusia. The current study links and extends previous results by directly investigating whether and how the acoustic processing deficit in amusia extends to the higher-level phonological perception of native tone categories. The sharpness of the identification boundary reflects the listeners' native language experiences and is influenced by their long-term phonological memory (Peng *et al.*, 2010; Xu *et al.*, 2006). The present identification results show that the amusic group exhibited a much shallower identification boundary compared with the musically intact controls. Moreover, the participants with amusia performed systematically worse across the board, regardless of the between- and within-category discrimination sensitivity, in both the 3- and 4-step discrimination tasks. Since the acoustic distance in the 24-Hz discrimination

task was even higher than the JNDs of tone agnostics (above 20 Hz) (Huang *et al.*, 2015a; Huang *et al.*, 2015b), it is more likely that the inferior discrimination performance of amusics might happen at a whole-group level. On the basis of the above findings, the domain-transferred acoustic processing deficit in amusics (reflected by an inferior within-category discrimination sensitivity) does negatively influence the identification and discrimination of phonological contrasts in both speech and nonspeech stimuli (reflected by a broader boundary width and a reduced between-category sensitivity).

By using computational strategies to uncover the statistical patterns in social language input, it has been proved that the developmental trajectory of native phonological categories starts early, from around 6 months in infants (Kuhl, 2004). Two studies (Wu *et al.*, 2015; Zhao and Kuhl, 2015) have shown that the facilitatory effect of acoustic processing in Mandarin-speaking musicians does not extend to the higher-order phonological operation. Thus, it seems that the internal processing of phonological categories is so robust that it is less likely to be affected by the lower-level processing. However, the results of the present study offer a different and supplementary perspective regarding the relative influence of the two strategies. As both the control and amusic groups consisted of healthy Mandarin adults with comparable chronological ages, it was reasonable to attribute their different CP performances to the different levels of acoustic processing capacity *per se*. The current results imply that the lower-level acoustics lay the foundation of higher-level phonological processing in lexical tone perception since the reduced acoustic processing skill severely drags down the phonological processing skill in native listeners. The discrepancy between the null effect of enhanced acoustic processing induced by music training (Wu *et al.*, 2015) and the significant detrimental effect of inferior acoustic processing accompanied with amusia on phonological processing may be partially explained by the following reasoning: The onset of musical instruction always occurs much later than that of language-specific phonological perception (as early as 6 months). As congenital amusia is regarded as being inborn (Peretz, 2008), it is possible that the reduced lower-level acoustic processing in this cohort interferes with the formation of native phonological categories from the very beginning.

Furthermore, although a wider boundary width and an inferior between-category sensitivity were observed for the amusic group in this study, the cross-boundary benefit was preserved in this group, as indicated by the higher sensitivity to the between-category speech unit. This perceptual pattern in amusics is essentially different from that found in non-tone language speaking individuals who fail to show a cross-boundary benefit in their discrimination of Mandarin lexical tones (Peng *et al.*, 2010; Xu *et al.*, 2006). Consequently, long-term exposure to a native tone language environment guarantees the CP pattern of lexical tones, although reduced, even in amusics who show a phonological processing deficit. The preserved CP pattern of higher sensitivity to between-category pairs in Mandarin-speaking amusics found in this study is not consistent with the findings of a previous study (Jiang *et al.*, 2012), which indicated a similar sensitivity to between- and within-category discrimination pairs in the Mandarin-speaking amusic group. It is important to note that the step size of the discrimination pairs used in Jiang *et al.* (2012) was only 6 Hz, which falls in the range of JNDs for F_0 discrimination (4–8 Hz) among normal Mandarin-speaking individuals (Liu, 2013). Given that the amusics showed a fine-grained pitch processing deficit, the 6-Hz acoustic difference might be too small to reveal the categorical nature in this group, with both the between- and within-category accuracies being close to the chance level (Jiang *et al.*, 2012). The essential characteristic of the CP pattern re-emerged in the amusic group when the step size was greatly enlarged. Finally, in the current experiment design, both the control and amusic groups showed no cross-boundary benefit in the nonspeech context; this was caused by the much higher within-category discrimination accuracy in the nonspeech material (also seen in Wu *et al.*, 2015). The pattern of nonspeech yielding higher within-category accuracy than speech is generalizable to all native tone-language speakers, irrespective of whether they are controls (Xu *et al.*, 2006), musicians (Wu *et al.*, 2015), or amusics (Zhang *et al.*, 2017).

From the perspective of Fujisaki and Kawashima's (1971) model, long-term phonetic memory is employed to conduct identification and between-category discrimination tasks, while discriminating the acoustic difference of within-category discrimination pairs requires the short-term auditory memory code. Due to the absence of memory capacity testing in the current study, we cannot conclude whether the processing deficits of acoustic information as observed in this study are mainly caused by a compromised retrieval of auditory code in short-term memory of amusics. The underlying memory mechanisms during lexical tone perception in tone language speakers with amusia require further investigations.

To summarize, the current findings deepen our understanding of the two levels of processing by indicating that impaired lower-level acoustic processing in amusia extends to the higher-level phonological processing deficit in lexical tones. Thus, internal processing of native phonological categories may be negatively impacted by impaired acoustic processing. The initial bottom-up acoustic analysis of speech sounds lays the foundation of the top-down phonological processing of linguistic elements. We conclude that lower-level acoustics underlie higher-level phonological categories in lexical tone perception.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No. 11474300 and by the General Research Fund of the Research Grants Council of Hong Kong under Grant No. 14408914.

References and links

- Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (2009). "Relative influence of musical and linguistic experience on early cortical processing of pitch contours," *Brain Lang.* **108**(1), 1–9.
- Chen, F., and Peng, G. (2016). "Context effect in the categorical perception of Mandarin tones," *J. Signal Process. Syst.* **82**(2), 253–261.
- Chen, F., Peng, G., Yan, N., and Wang, L. (2017). "The development of categorical perception of Mandarin tones in four-to seven-year-old children," *J. Child Lang.* **44**(6), 1413–1434.
- Elmer, S., Meyer, M., and Jäncke, L. (2012). "Neuro functional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects," *Cereb. Cortex* **22**, 650–658.
- Finney D. J. (1971). *Probit Analysis*, 3rd ed. (Cambridge University Press, Cambridge).
- Fujisaki, H., and Kawashima, T. (1971). "A model of the mechanisms for speech perception-quantitative analysis of categorical effects in discrimination," Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo, Vol. 30, pp. 59–68.
- Huang, W. T., Liu, C., Dong, Q., and Nan, Y. (2015a). "Categorical perception of lexical tones in Mandarin-speaking congenital amusics," *Front. Psychol.* **6**, 829.
- Huang, W. T., Nan, Y., Dong, Q., and Liu, C. (2015b). "Just-noticeable difference of tone pitch contour change for Mandarin congenital amusics," *J. Acoust. Soc. Am.* **138**(1), EL99–EL104.
- 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**, 1109–1121.
- Kuhl, P. K. (2004). "Early language acquisition: Cracking the speech code," *Nat. Rev. Neurosci.* **5**(11), 831–843.
- Liu, C. (2013). "Just noticeable difference of tone pitch contour change for English- and Chinese-native listeners," *J. Acoust. Soc. Am.* **134**(4), 3011–3020.
- Macmillan, N. A., and Creelman, C. D. (2008). *Detection Theory: A User's Guide* (Psychology Press, New York).
- Nan, Y., Sun, Y., and Peretz, I. (2010). "Congenital amusia in speakers of a tone language: Association with lexical tone agnosia," *Brain* **133**, 2635–2642.
- Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., and Wang, W. S. Y. (2010). "The influence of language experience on categorical perception of pitch contours," *J. Phon.* **38**, 616–624.
- Peretz, I. (2008). "Musical disorders: From behavior to genes," *Curr. Dir. Psychol.* **17**, 329–333.
- Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., Paquette, S., and Bouchard, B. (2008). "On-line identification of congenital amusia," *Music Percept.* **25**(4), 331–343.
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). "Musical experience shapes human brain stem encoding of linguistic pitch patterns," *Nat. Neurosci.* **10**, 420–422.
- Wu, H., Ma, X., Zhang, L., Liu, Y., Zhang, Y., and Shu, H. (2015). "Musical experience modulates categorical perception of lexical tones by native Chinese speakers," *Front. Psychol.* **6**, 436.
- Xu, Y. S., Gandour, J. T., and Francis, A. L. (2006). "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," *J. Acoust. Soc. Am.* **120**(2), 1063–1074.
- Zhang, C., Shao, J., and Huang, X. N. (2017). "Deficits of congenital amusia beyond pitch: Evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics," *PLoS One* **12**(8), e0183151.
- Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., and Li, P. (2011). "Cortical dynamics of acoustic and phonological processing in speech perception," *PLoS One* **6**(6), e20963.
- Zhao, T. C., and Kuhl, P. K. (2015). "Higher-level linguistic categories dominate lower-level acoustics in lexical tone processing," *J. Acoust. Soc. Am.* **138**(2), EL133–EL137.