Congenital Amusia and Tone Merger: Perception and Production of Lexical Tones in Hong Kong Cantonese

Oi-Yee Ho[1], Jing Shao[1,4], Jinghua Ou[2], Sam-Po Law[3], Caicai Zhang[1,4]

[1]The Hong Kong Polytechnic University, Hong Kong
[2] University of Chicago, IL, United States of America
[3]The University of Hong Kong, Hong Kong
[4]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
joyce.oy.ho@polyu.edu.hk, jing.shao@polyu.edu.hk, jou@uchicago.edu, splaw@hku.hk, caicai.zhang@polyu.edu.hk

ABSTRACT

Congenital amusia is a disorder reported to affect one's pitch processing in both music and language domains, resulting in an impaired discrimination of native lexical tones. Tone merging has been observed in native speakers of Hong Kong Cantonese, where some speakers confuse certain tone pairs in perception and/or production. Existing studies have only investigated the two groups separately, leaving a gap which concerns whether amusics' profile is comparable to mergers'. The current study bridges the gap by directly comparing amusics and mergers in their ability to discriminate musical and lexical tones, plus their lexical tone production profile. Results revealed that mergers were intact in musical pitch perception and highly selective in lexical tone confusion. In contrast, amusics exhibited low sensitivity to all lexical tone pairs, and a dissociation between lexical tone perception and production. Preliminary findings suggest that congenital amusia and tone merging are inherently different.

Keywords: congenital amusia, tone merger, tone perception, tone production, Hong Kong Cantonese.

1. INTRODUCTION

Congenital amusia (hereafter amusia), a lifelong neurodevelopmental disorder, has been known to impair musical pitch processing without brain injury, and affect up to 4% of the population [2][22][23]. Apart from having difficulties making fine-grained pitch discrimination [5], individuals with amusia (hereafter amusics) also exhibited poor pitch memory [8][27][28] and inferior musical pitch production [3][10]. Previous studies have also observed a transfer to the language domain, affecting speech intonation and emotion prosody perception [11][19]. It has been found that amusics exhibited impaired identification and discrimination of lexical tones in Hong Kong Cantonese (HKC), especially for tone pairs with small acoustic differences such as T2/T5 (high-rising/low-rising) and T3/T6 (mid-level/low-level) [25], and demonstrated deficits in cognitive measures such as inhibitory control and selective attention than the control group [18].

On the other hand, a linguistic sound change called tone merging has been observed in tonal languages such as HKC. HKC has a total of six contrastive tones for open syllables, namely T1 (high-level), T2 (high-rising), T3 (mid-level), T4 (low-falling/extra-low-level), T5 (low-rising) and T6 (low-level). Previous studies have found that T2/T5, T3/T6 and T4/T6 are most likely to be confused by native speakers of HKC while other pairs were unaffected [16]. Besides, a perception/production mismatch has been found in the said three tone pairs. As shown in [6][7][16], while the T2/T5 and T3/T6 contrast were maintained in perception but not production (i.e. [+per][-pro] or 'partial merger'), the T4/T6 contrast was retained in production but not perception (i.e. [-per][+pro] or 'near merger'). Some speakers of HKC have been found to confuse T2/T5 in both perception and production (i.e. [-per][-pro] or 'full merger').

No previous studies have investigated the relationship between congenital amusia and tone merging, albeit the shared behavioural similarity in tone confusion. The present study aims to bridge the gap by directly comparing the two groups in terms of their ability to discriminate musical tones and lexical tones, as well as their production of lexical tones in HKC. Three research questions were investigated in this study: (1) Since tone merging individuals confuse lexical tones, do they show reduced ability in discriminating musical tones, compared with control participants who are non-amusic and without tone-merging? (2) Will amusics confuse lexical tones in both perception and production? If so, will the confusion patterns be highly selective, similar to those reported of tone mergers? (3) Is the discrimination ability of musical tone related to that of lexical tone?

2. METHODS

2.1. Participants

A total of seventy-two participants were recruited (Table 1), and were divided into five groups, namely

Amusics, Partial mergers, Near merger, Full mergers and Control in order to compare the difference between amusics, mergers and controls. All the participants were undergraduates in universities in Hong Kong, with HKC as their native language with no reported hearing impairment, communication disorders and history of brain injury. The Montréal Battery of Evaluation of Amusia (MBEA; see details below) [21] was used to identify amusics, who had no higher than 71% in the MBEA global accuracy. As for the recruitment of tone mergers, an AX discrimination task on lexical tones and a lexical tone production task (see details below) were conducted. 2 phoneticians with HKC as native language were asked to identify the T2, T4, T5 and T6 tones produced by the participants (κ= .893, p< .001). Participants who scored higher than 80% in perceptually discriminating T2/T5 but failed to reach 80% in producing the T2/T5 contrast distinctively were classified as 'partial mergers (PM)'; those who had an accuracy below 80% in both discrimination and production of T2/T5 were regarded as 'full mergers (FM)'; and those who scored over 80% in the production of T4/T6 but below 80% in discrimination were considered as 'near mergers (NM)'. Controls were participants that were neither amusic nor tone mergers.

Table 1: Demographic information of the five groups.

	N (M:F)	Mean age (SD)
Amusics	8:11	22.05 (2.30)
Controls	7:7	24.60 (4.10)
PM : T2/T5[+per-pro]	7:6	20.92 (2.14)
NM : T4/T6[-per+pro]	5:6	22.09 (2.55)
FM : T2/T5[-per-pro]	2:13	21.07 (1.49)

2.2. Stimuli and procedures

2.2.1. Montréal Battery of Evaluation of Amusia (MBEA)

The MBEA test consists of six subtests. Three are pitch-oriented subtests (Scale, Contour, Interval), two are rhythm-oriented (Rhythm, Metre) and one is memory related (Memory) [21]. During the first four subtests, participants were required to indicate whether the two melodies presented were the same or different, where the difference was either pitch or rhythm. For Metre, participants had to indicate whether a melody presented is a march or a waltz. Lastly, during the Memory subtest, participants were asked to answer whether they have come across the presented melody in the previous five subtests. All stimuli were presented in a soundproof room through JVC HA-D610 stereo headphones binaurally at a comfortable listening level.

2.2.2. Tone production task

The tone production task is identical to that reported in [18]. It was administered before the tone discrimination task to eliminate any priming effect.

The word stimuli used to elicit production were syllable [fu] carrying six tones. The six syllables were all meaningful words in HKC, and were embedded in different positions in two carrier phrases: [ŋɔ^{213} ji$^{21/55}$ ka:^{55} tʊk^{2} __ tsi^{22}] "I am reading the __ character", and [li^{55} kɔ^{33} tsi^{22} hɐi^{22} __] "This is the __ character". The twelve sentences (six syllables × two carriers) were repeated ten times, generating 120 trials. Participants were instructed to read the sentences out loud naturally at a normal speech rate. The speech outputs were recorded by PRAAT with a Shure SM48 microphone in a soundproof room. F0 measurements were obtained at ten time points of the vocalic portion of the target words (i.e., /u/). The F0 values were then normalised within each speaker using the log-z score method [24]. The analyses focused on the production of the two rising tones (T2 and T5), and the following acoustic parameters were calculated: *F0 offset* (F0 at the 10th point), *F0 slope* (maximal F0 – minimal F0) and *F0 height* (average F0 of ten time points), following previous studies [7][12]. The difference between T2 and T5 (T2 minus T5) in *F0 offset*, *F0 slope* and *F0 height* were obtained accordingly.

2.2.3. Tone discrimination task

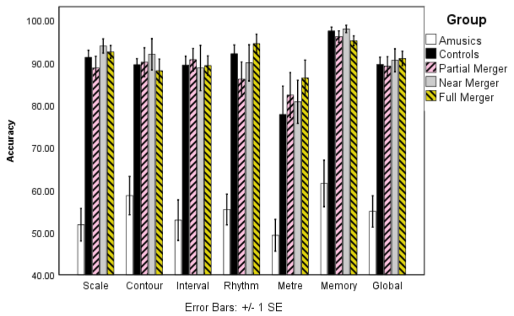
The tone discrimination task is also identical to that reported in [18]. Again, the syllable [fu] carrying six tones was used to control the syllable effect. The six words were recorded in isolation by a female native speaker of HKC. The target tones were extracted from the recordings and normalized to 500ms in duration. An AX discrimination paradigm was adopted. Two tones were paired with an inter-stimulus-interval of 500ms. A total of 360 pairs were generated by repeating 36 tone pairs (6 AA pairs and 30 AB pairs) ten times. Participants were instructed to indicate whether the tones presented were the same or different and the response and reaction times (RT) were collected. The d' score of each tone pair (z score of hit rate minus that of false alarm rate [15]) for a total of 15 different pairs was calculated to indicate the perceptual sensitivity. Only RT for correct trials was included for analysis. All stimuli were again presented in a soundproof room through JVC HA-D610 stereo headphones binaurally at a comfortable listening level.

3. RESULTS

3.1. MBEA

A *group* x *subtests* ANOVA with Greenhouse-Geisser correction for sphericity violation was conducted on the MBEA accuracies. Results showed that there were significant main effects of *group* ($F(4,67)= 37.26$, $p< .001$, $\eta_p^2= .713$) and that of *task* ($F(2.72,163.59)= 12.51$, $p< .001$, $\eta_p^2= .425$) (Fig. 1). No significant interactions were observed.

Figure 1: Mean accuracy of six MBEA subtests and Global score by 5 groups.



Post-hoc pairwise comparisons conducted to analyse the *group* effect revealed that amusics performed significantly worse than the three merger groups and controls (*ps*< .001), while merger groups performed comparably to controls. As for the *subtest* effect, post-hoc analyses revealed that participants scored higher in the Memory subtest than in the other five subtests (*ps*< .006), and were inferior in Metre than the other subtasks (*ps*.< .028) except for Interval (*p*= .236). This indicates that among the six subtests of MBEA, and the interval and metre subtests were more difficult than the others.

3.2. Tone production task

A series of one-way ANOVA with *group* as a factor were conducted on the *F0 offset difference*, *F0 slope difference* and *F0 height difference* respectively. The results showed that there was a significant group difference in *F0 offset difference* ($F(4,60)= 6.09$, $p< .001$), *F0 slope difference* ($F(4,60)= 5.41$, $p<.001$) and *F0 height difference* ($F(4,60)= 5.28$, $p= .001$) of T2/T5 (Table 2).

Post-hoc pairwise comparisons revealed that in all three parameters, the amusics' production of T2 and T5 were comparable to that of controls. For all three parameters, full mergers had significantly lower F0 differences than controls (*ps*< .002), so did partial mergers (*ps*< .011), whereas the F0 differences between the two groups were not significant.

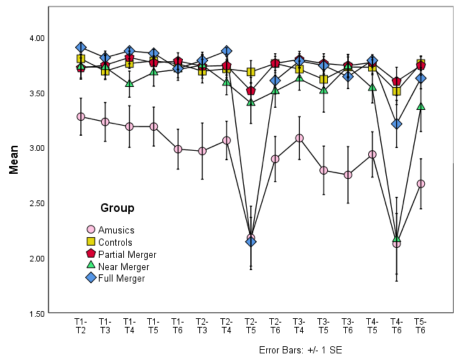
Table 2: Normalized differences in *F0 offset*, *F0 slope* and *F0 height* between T2 and T5 in the productions of five groups. [Mean(SD)]

	Offset diff	Slope diff	Height diff
Amusics	.54 (.49)	.11 (.09)	.11 (.17)
Controls	.75 (.38)	.14 (.09)	.20 (.13)
PM	.25 (.24)	.05 (.04)	.02 (.11)
NM	.47 (.34)	.09 (.06)	.14 (.11)
FM	.10 (.24)	.04 (.05)	-.01 (.11)

3.3. Tone discrimination task

A *group* x *tone pair* ANOVA on the d' scores with Greenhouse-Geisser correction demonstrated a significant main effect of *group* ($F(4,67)= 10.77$, $p< .001$, $\eta_p^2= .391$), *tone pair* ($F(5.81,389.48= 22.14$, $p<.001$, $\eta_p^2= .248$) and a significant *group:task* interaction ($F(23.25,389.48)= 4.48$, $p< .001$, $\eta_p^2= .211$) (Fig. 2).

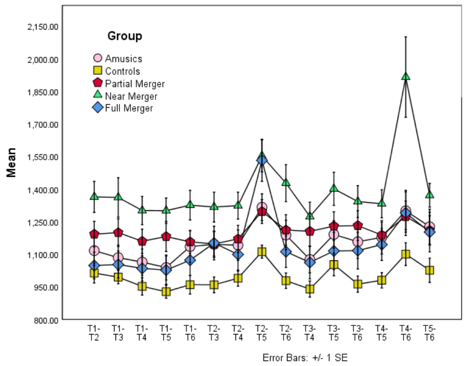
Figure 2: D' score of 15 tone pairs by 5 groups.



Post-hoc analyses with one-way ANOVA showed that amusics were less sensitive than the three merger groups and controls toward the discrimination of T1/T6, T2/T3, T2/T4, T3/T5, T3/T6, and T4/T5 (*ps*< .05), and comparable to full mergers and near mergers towards the discrimination of T2/T5 and T4/T6 respectively. Near mergers were less sensitive to T4/T6 than controls (*p*= .006) while sensitivity to other tone pairs were comparable. Similarly, full mergers were less sensitive to T2/T5 compared to other tone pairs (*ps*< .021).

As for RT of correct trials (Fig. 3), a *group* x *tone pair* ANOVA revealed a significant main effect of *group* ($F(4,67)= 4.54$, $p= .003$, $\eta_p^2= .218$) and *tone pair* ($F(5.49,356.53)= 32.99$, $p< .001$, $\eta_p^2= .337$). The interaction between *tone pair* and *group* was also significant ($F(21.95,356.53)= 3.82$, $p< .001$, $\eta_p^2= .190$). Post-hoc one-way ANOVA showed that near mergers spent a significantly longer time to discriminate two tones correctly across the board than controls (*ps*< .030). Full mergers spent more time on T2/T5 compared to on other tone pairs (*ps*< .012). No other effects were significant.

Figure 3: Reaction time for correct trials (in ms) of 15 tone pairs by 5 groups.



3.4. Relationship between musical tone perception and lexical tone perception

Using the d' score of all discrimination trials (overall d') as the dependent variable and the three pitch-related MBEA scores as predictors, results showed that 30.2% of the overall d' sensitivity can be predicted by the MBEA Contour task ($p < .001$), suggesting that the higher accuracy in the Contour task, the more sensitive one was in the lexical discrimination task. However, no significance was reached when regressions were carried out separately for amusics and merger groups.

4. DISCUSSION

The current study compared amusics and tone mergers in the perception of musical melodies and lexical tones, and the production of lexical tones, with an aim to understand the relationship between the two.

As shown by analyses of MBEA, while amusics were inferior to tone mergers and controls in their musical ability, the three tone merger groups' were comparable to controls in discriminating musical pitch in the three pitch-related subtests. This indicates that tone merging could occur in the absence of congenital amusia, suggesting a possible dissociation between tone merging and amusia.

Analyses on d' scores showed that near mergers and full mergers demonstrated a highly selective merging pattern, having only low sensitivity to T4/6 and T2/T5 respectively while leaving other tone pairs spared, which is consistent with previous studies [6][7][16]. Unlike tone mergers, amusics' inferior sensitivity was demonstrated in a wider range of tone pairs, even on those pairs with large acoustic differences such as T1/T6 and T4/T5, and they performed similarly inferiorly to the full mergers in the discrimination of T2/T5 and to the near mergers in the discrimination of T4/T6.

The correlation between musical and lexical pitch sensitivity is consistent with a number of previous studies where an influence from music to language domain was found. For instance, it was found that English musicians were better at discriminating Mandarin Chinese tones than English non-musicians [1], and that native speakers of Italian with higher melodic ability were superior in detecting tonal variation than their counterparts with lower melodic ability [4]. As to why significance was reached only in the MBEA Contour task but not the other two pitch-related subtasks (i.e. Scale, and Interval), it has been proposed that the processing of musical contour and that musical interval is governed by two distinct mechanisms [20][21]. While interval processing relies on a more local representation, contour processing is more global and 'cognitively salient' [26]. It is possible that during the lexical discrimination task, some global processing was involved in comparing the contour of stimuli (e.g., rising vs. level contour). It is also possible that since the pitch difference in the MBEA Contour subtest is larger than that in the Scale and Interval subtests, the task difficulty was moderate, hence correlations, if any, can be more easily detected. We are aware of the null finding when regression analyses were conducted separately, which suggests that the significant regression with the five groups collapsed was primarily driven by the inferior musical ability of amusics. More investigations with a larger group sample size are needed to ascertain such relationship.

It is also intriguing to observe that near mergers demonstrated the longest RT during the tone discrimination task, even for tone pairs where they did not demonstrate perceptual insensitivity. This suggests that the near merger group took a longer time to discriminate tones correctly, resulting in the comparably high sensitivity for most tone pairs to controls. In an ERP study of poor perception exhibited in the near merger group [13], the findings suggested a top-down processing in speech perception, where the listeners may rely on rich contextual information in the recognition of lexical items, and that acoustic input does not undergo complete analysis, which potentially weakens their sensitivity to speech sound distinctions, particularly of small difference such as T2/T5 and T4/T6. It is possible that near mergers may employ a longer response time to compensate for the reduced acoustic sensitivity in tone discrimination.

On the other hand, F0 analyses showed that the amusics' productions of T2 and T5 were comparable to those of controls, demonstrating that the production of lexical tones in amusics is more or less intact, in contrast to their poor discrimination of musical melodies and lexical tones. This result is consistent with previous studies, which reported intact lexical tone production in amusics [14][29]. It is also consistent with the notion of tone agnosia, namely that some amusics would exhibit exceptionally poor lexical tone perception but retain normal tone production [17] as well as higher tone contour threshold than other amusics [9].

We are aware of the gender imbalance in some groups, and that the controls were older than other groups. Given the difficulty in recruiting participants that fit our selection criteria, we had to include any available appropriate participants during the study period. Future study may investigate whether age and gender play any role in the perception and production of lexical and musical tone.

In conclusion, our findings of musical and lexical tone perception plus lexical tone production in HKC amusics, tone mergers and controls demonstrated that congenital amusia and tone merging are likely to be inherently different, and that the cause(s) of tone merging lie(s) in factors other than impaired musical pitch processing. Further investigations with a larger sample size will be worthwhile.

5. REFERENCES

[1] Alexander, J. A., Wong, P. C. M., Bradlow, A. R. 2005. Lexical tone perception in musicians and mon-musicians. *Proc. 9th European Conference on Speech Communication and Technology* Lisbon, Portugal.

[2] Ayotte, J., Peretz. I., Hyde, K. 2002. Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain* 125, 238-251.

[3] Bella, S. D., Berkowska, M., Sowinski, J. 2011. Disorders of pitch production in tone deafness. *Front Psychol*, 2, 164.

[4] Delogu, F., Lampis, G., Olivetti Belardinelli, M. 2006. Music-to-language transfer effect: may melodic ability improve learning of tonal languages by native non-tonal speakers? *Cogn Process* 7(3), 203-207.

[5] Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., Griffiths, T. D. 2004. Characterization of deficits in pitch perception underlying 'tone deafness'. *Brain* 127(4), 801-810.

[6] Fung, R., Wong, C. 2010. Mergers and near mergers in Hong Kong Cantonese tones, *Proc. Tone and Intonation in Europe* 4 Stockholm, Sweden.

[7] Fung, R., Wong, P. C. M. 2011. Acoustic analysis of the new rising tone in Hong Kong Cantonese. *Proc. 17th ICPhS* Hong Kong.

[8] Gosselin, N., Jolicœur, P., Peretz, I. 2009. Impaired memory for pitch in congenital amusia. *Annals of the New York Academy of Sciences* 1169(1), 270-272.

[9] Huang, W. T., Nan, Y., Dong, Q., & Liu, C. 2015. Just-noticeable difference of tone pitch contour change for mandarin congenital amusics. *J Acoust Soc Am* 138(1).

[10] Hutchins, S., Zarate, J. M., Zatorre, R. J., Peretz, I. 2010. An Acoustical Study of vocal pitch matching in congenital amusia. *J Acoust Soc Am* 127(1), 504-512.

[11] Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., Yang, Y. 2010. Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia* 8(9), 2630-2639.

[12] Khouw, E., Ciocca, V. 2007. Perceptual correlates of Cantonese tones. *Journal of Phonetics* 35(1), 104-117

[13] Law, S. P., Fung, R., Kung, C. 2013. An ERP study of good production vis-a-vis poor perception of tones in Cantonese: implications for top-down speech processing. *PLoS One*, 8(1).

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

[15] Macmillan, N. A., Creelman, C. D. 2005. *Detection theory: A user's guide (2nd ed.)*. New York: Psychological Press

[16] Mok, P.P., Zuo, D., Wong, P.W. 2013. Production and perception of a sound change in progress: tone merging in Hong Kong Cantonese, *Language Variation and Change* 25(3), 341-370.

[17] 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.

[18] Ou, J., Law, S.P., Fung, R. 2015. Relationship between individual differences in speech processing and cognitive functions, *Psychon Bull Rev*, 1-8.

[19] Patel, A., Wong, M., Foxton. J., Lochy, A., Peretz, I. 2008. Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception: An Interdisciplinary Journal* 25(4), 357-368.

[20] Peretz, I. 1990. Processing of local and global musical information by unilateral brain-damaged patients. *Brain* 113(4), 1185-1205.

[21] Peretz, I., Champod, A., Hyde, K. 2003. Varieties of musical disorders. *Annals of the New York Academy of Sciences* 999(1), 58-75.

[22] Peretz, I., 2006. Brain specialization for music. *Annals of the New York Academy of Sciences* 930(1), 153-165.

[23] Peretz, I., Vuvan, D. T. 2017. Prevalence of congenital amusia. *Eur J Hum Genet* 25(5), 625-630.

[24] Rose, P. 1987. Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication* 6, 343-352.

[25] Shao, J., Zhang, C., Peng, G., Yang, Y., Wang, W. S. Y. 2016. Effect of noise on lexical tone perception in cantonese-speaking amusics. *Proc. Interspeech 2016* San Francisco, USA.

[26] Stewart, L. 2011. Characterizing congenital amusia. *Q J Exp Psychol (Hove)* 64(4), 625-638.

[27] Tillmann, B., Schulze, K., Foxton, J. M. (2009). Congenital amusia: a short-term memory deficit for non-verbal, but not verbal sounds. Brain Cogn, 71(3), 259-264.

[28] Williamson, V. J., Stewart, L. 2010. Memory for pitch in congenital amusia: beyond a fine-grained pitch discrimination problem. *Memory* 18(6), 657-669.

[29] Yang, W. X., Feng, J., Huang, W. T., Zhang, C. X., Nan, Y. 2014. Perceptual pitch deficits coexist with pitch production difficulties in music but not Mandarin speech. *Front Psychol* 4, 1024.