



Tone Merging Patterns in Congenital Amusia in Hong Kong Cantonese

Oi-Yee Ho¹, Jing Shao¹, Jinghua Ou², Sam-Po Law³, Caicai Zhang^{1,4}

¹The Hong Kong Polytechnic University, Hong Kong

²The Chinese University of Hong Kong, Hong Kong

³The University of Hong Kong, Hong Kong

⁴Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen

joyce.oy.ho@polyu.edu.hk, jing.shao@polyu.edu.hk, jinghua.ou@cuhk.edu.hk, splaw@hku.hk,
caicai.zhang@polyu.edu.hk

Abstract

Congenital amusia is a neurodevelopmental disorder, affecting fine-grained musical pitch processing without brain injury. This disorder also affects pitch processing in speech such as lexical tone perception. On the other hand, the phenomenon of tone merging has been observed among some speakers of Hong Kong Cantonese (HKC), who exhibit confusion between certain tone pairs in perception and/or production. It has been reported that tone merging may relate to individual variation in cognitive abilities of working memory and attention. The current study is a first attempt to investigate the relationship between amusia and tone merging by examining tone merging patterns in perception of amusics in HKC and their cognitive abilities of working memory and attention as well as pitch threshold. The results revealed a different profile of amusics from that of merger groups reported in previous studies. Amusics exhibited a profound impairment in discriminating tones compared to musically intact controls, which appeared to differ from the highly selective perceptual confusion of tone pairs reported in the merger groups. Regarding cognitive measures, amusics also demonstrated broad deficits in selective attention, working memory and inhibitory control. The temporary results imply that amusia might have a limited contribution to the previously reported tone merging.

Index Terms: congenital amusia, tone merger, attention, working memory, Hong Kong Cantonese

1. Introduction

Congenital amusia (amusia hereafter), also known as tone or tune deafness, is a lifelong neurodevelopmental disorder that affects processing of musical pitch in the absence of brain injury, affecting about 1.5-4% of the population [1][2][3]. Individuals with amusia have difficulties making fine-grained pitch discrimination [4] and demonstrated impaired memory for pitch [5][6]. Evidence has shown that amusia is not a music-specific disorder, but also affects pitch processing in speech, including speech intonation and emotion prosody perception [7][8]. Amusics are also found exhibiting deficits in processing lexical tones. Shao et al. [9] found that amusic speakers of Hong Kong Cantonese (HKC) were impaired in both identification and discrimination of Cantonese tones under both clear and noisy listening conditions.

The phenomenon of confusing two or more lexical tones is called tone merging, which is a type of sound change [10-13]. Some native speakers of HKC have been reported to confuse certain tones pairs in perception and/or production of their

native tongue. HKC has a total of six contrastive tones for open syllables: 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 three pairs of tones, namely T2/T5, T3/T6 and T4/T6, are most susceptible to the phenomenon of tone merging [10-12]. Fung and Wong [11][12] found different patterns of tone merging in 120 native speaker of HKC, whose age ranged from 20 to 58. While the T2/T5 contrast was not maintained in both production and perception (i.e. [-per-pro]), the T3/T6 exhibited a contrast in perception but not in production (i.e. [+per-pro]) and the T4/T6 contrast was well-maintained in production but not in perception (i.e. [-per+pro]). Besides, another study reported that some speakers of HKC may lose the contrast for T2/T5 in production while maintaining the contrast in perception ([+per-pro]) [13]. More recently, studies have been done to investigate the relationship between tone merging and individual variation in cognitive abilities of working memory and attention, which complements the conventional sociolinguistic approach. Interestingly, it has been found that the merger groups were inferior in their ability of attention switching and working memory compared to the controls [13].

In the present study, we aim to investigate the tone merging patterns in perception in a group of amusics of HKC. We also examine their cognitive abilities in attention, working memory and pitch threshold, and by doing so to investigate to what extent the individual speakers' cognitive abilities and sensitivity to pitch change may account for their tone discrimination performance. By comparing the profile of amusics in terms of tone merging and cognitive abilities with that of the merger groups reported in previous studies [13], this study is a first attempt to explore the relationship between tone merging and amusia. Three questions will be investigated in the current study: (1) If Cantonese-speaking amusics confuse lexical tones perceptually, are their tone merging patterns similar to those reported of merger groups in previous studies, which were highly selective (T2/T5 or T4/T6, but not other tone pairs)? (2) Do Cantonese-speaking amusics perform inferiorly in cognitive tests of attention and working memory, in a way similar to the performance of merger groups reported in previous studies? (3) To what extent can the tone merging patterns of amusics be accounted for by individual variation in cognitive abilities?

2. Method

2.1. Participants

Nineteen amusics (11 F and 8 M) and 20 controls (11 F and 9 M) were recruited. All were undergraduates with HKC as their native language. The Montreal Battery of Evaluation of Amusia (MBEA; [14]) was used to measure the musical abilities of these participants. All amusics scored below 75 in the MBEA global score, while all controls scored above 85. Table 1 illustrate the performance on MBEA of the amusic and control groups. The performance of amusics on all subtests of MBEA is significantly inferior to that of controls ($p < .001$). All participants gave informed consent in compliance with experimental protocols approved by the Hong Kong Polytechnic University.

Table 1: Demographic characteristics of amusics and control.

	Amusics		Controls		<i>t</i>	<i>p</i>
	Mean	SD	Mean	SD		
Age	22.05	2.30	23.85	4.16	-1.68	.103
MBEA						
Scale	51.79	16.77	91.46	5.97	-9.74	< .001
Contour	58.66	19.68	90.30	5.42	-6.77	< .001
Interval	52.87	20.92	89.60	7.62	-7.21	< .001
Rhythm	55.37	15.78	92.73	7.49	-9.36	< .001
Metre	49.30	16.4	75.63	23.59	-4.06	< .001
Memory	61.51	24.00	97.74	3.19	-6.53	< .001
Global	54.93	16.18	89.56	5.90	-8.79	< .001

2.2. Stimuli and Procedures

2.2.1. Tone Discrimination Task

The tone discrimination task is identical to that reported in [13]. To control for any syllable effect, only one CV syllable [fʊ] was used to derive six tones. The six syllables, which are all semantically legal words in HKC, were produced and recorded by a native female Cantonese speaker. The duration of the target and comparison tone was 500 ms with inter-tone retention interval of 500ms. An AX discrimination task was adopted. Thirty-six tone pairs (6 AA pairs and 30 AB pairs) were repeated ten times, generating a total of 360 stimuli. The participants were instructed to indicate as soon as possible whether the tones presented were the same or different by clicking a mouse button. Both accuracy and reaction time (RT) were collected. All stimuli were presented binaurally through JVC HA-D610 stereo headphones at a comfortable listening level in a soundproof room.

2.2.2. Cognitive Measures of Attention and Working Memory

Following [13], the subjects' attention and working memory were tested. Six subtests of Test of Everyday Attention (TEA; [15]), the Attention Network Task (ANT; [16]), and the test of auditory attentional shifting-auditory stream segregation [17] were used to evaluate the subjects' auditory and visual

attention. The subjects' auditory and visual working memory were assessed by backward digit span and the subscales of visual processing speed, namely symbol search, coding and cancellation, in the Wechsler Adult Intelligence Scale, 4th Edition (WAIS-IV; [18]). All test instructions were given in HKC, which were initially translated from the English manuals. All tasks were administered to participants in a soundproof room.

2.2.3. Pitch Threshold Task

The stimuli were of four different types: 2 types (speech/tone) x 2 contours (discrete/gliding). The Cantonese syllable /ji/ recorded from a male Cantonese speaker was used in the speech conditions, while complex tone was used in tone conditions. The complex tones, which were generated using PRAAT, had a fundamental frequency (F_0) of 100 Hz. For discrete conditions, the stimuli consisted of a standard stimulus of 100Hz and 82 target stimuli ranging from 100.07 Hz to 178.17 Hz in steps of 0.01, 0.1 and 1 semitone. The smallest and largest pitch difference between standard and target stimuli was 0.01 and 10 semitone respectively. For gliding conditions, the stimuli were comprised of 82 rising and 82 falling glides with pitch excursion sizes between 0.01 and 10 semitones. Centred on 100 Hz, the largest rising glide started at 78.67 Hz and ended at 133.12 Hz, while the smallest rising glide started at 100.09 Hz and ended at 100.58 Hz. The discrete and glide F_0 curves were superimposed on the base syllable /ji/ in the speech conditions.

The four types of stimuli were presented in separate blocks counterbalanced in order. The task was two-alternative forced choice, where the participants were instructed to click a mouse button to indicate the pitch pattern of the stimulus pair: High-Low or Low-High for discrete pairs, and Rising-Falling or Falling-Rising for gliding pairs. Experimental trials began with a 10-semitone difference. The procedure was two-up/one-down. All stimuli were presented binaurally through JVC HA-D610 stereo headphones at a comfortable listening level in a soundproof room.

3. Results

3.1. Tone Discrimination Task

For the discrimination accuracy, a *group x tone pair* ANOVA with Greenhouse-Geisser correction for sphericity violations was conducted. The *tone pair* factor included four levels: T2/T5, T3/T6, T4/T6 and others (averaged across remaining tone pairs). There was a significant main effect of *group* ($F(1,28) = 7.855, p = .009, \eta_p^2 = .219$) and that of *tone pair* ($F(3,84) = 9.418, p < .001, \eta_p^2 = .252$), but no significant interaction between *group* and *tone pair* ($F(3,84) = 1.531, p = .212, \eta_p^2 = .052$). Post-hoc pairwise comparisons revealed significant differences between T2/T5 and others ($ps = .001$), and between T4/T6 and others ($ps = .001$). In order to investigate more closely as to in which tone pairs the amusics showed a significantly lower accuracy, a series of independent t-tests was conducted with an adjusted *p*-value of .0125 (.05/4), revealing that amusics performed worse than controls in distinguishing pairs T2/T5 ($t(37) = -3.337, p = .002$), T3/T6

($t(20.429) = -2.994, p = .007$) and T4/T6 ($t(24.469) = -3.969, p = .001$), as shown in Fig. 1. Though the amusics also exhibited lower discrimination accuracy on other pairs, the difference was not significant ($t(19.092) = -2.259, p = .036$).

As for the RT, a *group x tone pair* ANOVA demonstrated a significant main effect of *tone pair* ($F(3,84) = 13.238, p < .001, \eta_p^2 = .321$) and marginally significant main effect of *group* ($F(1,28) = 4.093, p = .053, \eta_p^2 = .128$). The interaction between *group* and *tone pair* was insignificant ($F(3,84) = .043, p = .988, \eta_p^2 = .002$). Post-hoc pairwise comparisons on *tone pair* showed significantly longer RT to T2/T5 and T4/T6 pairs compared to T3/T6 and others ($ps < .05$).

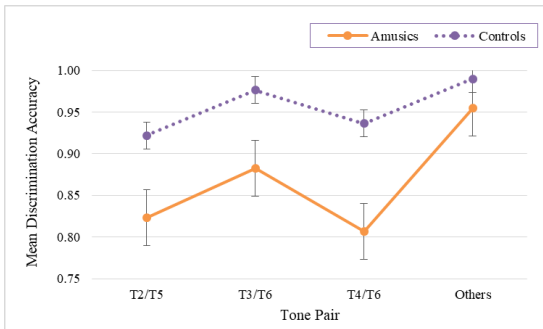


Figure 1: Discrimination accuracies of different tone pairs of amusic and control groups.

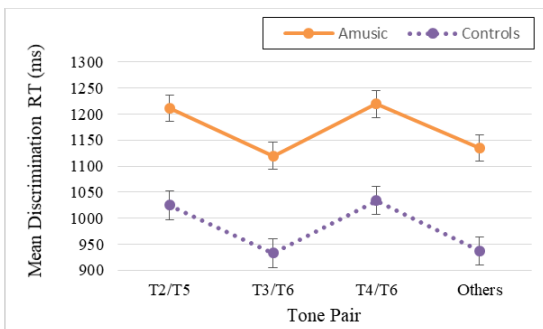


Figure 2: Mean discrimination RT (in ms) of different tone pairs of amusic and control groups.

3.2. Cognitive Measures of Attention and Working Memory

As revealed by a series of independent t-tests (significance threshold adjusted for multiple comparisons) and demonstrated in Table 2, for auditory and visual attention, amusics received lower scores on all six tests in TEA than controls, with significant group differences in tasks including Elevator Count with Distraction (ECD) ($t(37) = -3.335, p = .002$) and Elevator Count with Reversal (ECR) ($t(37) = -3.724, p = .001$). No significant group difference was found in Visual Elevator (VE), Telephone Search (TS) and Telephone Search with Counting (TSC) ($ps > .05$). There is also no significant difference between amusics and controls on ANT and Auditory Stream Segregation.

As for verbal working memory, there was no significant group difference on Backward Digit Span. In visual working memory, which was assessed by WAIS-IV, amusics scored significantly lower than controls in the Cancellation subtask ($t(37) = -5.344, p < .001$). Although amusics performed worse in Symbol Search and better in Coding than controls, these differences were not significant ($ps > .05$).

Table 2: Mean scores of all cognitive tasks of attention and working memory of amusic and control groups.

	Amusics		Controls		<i>t</i>	<i>p</i>
	Mean	SD	Mean	SD		
TEA						
MS1 ^a	9.58	2.78	11.80	2.17	-2.79	.008
MS2 ^b	6.79	2.78	9.25	2.17	-3.09	.004
ECD	8.47	3.08	11.30	2.15	-3.34	.002*
VE1 ^c	9.26	3.05	10.35	3.01	-1.12	.270
VE2 ^d	11.00	2.11	12.60	3.07	-1.89	.067
ECR	9.47	2.91	12.15	1.18	-3.72	.001*
TS	12.84	4.63	13.05	4.26	-0.15	.885
TSC	11.32	2.89	12.30	2.98	-1.05	.302
ANT ^e	82.28	37.86	71.96	25.34	1.00	.324
StrSg ^f	99.69	35.80	80.23	49.86	1.39	.172
WAIS-IV						
CA	8.74	2.45	14.00	3.57	-5.34	<.001*
SS	13.74	2.16	14.85	3.30	-1.25	.219
CD	14.42	2.89	13.95	3.07	.49	.625
BDS ^g	5.26	1.57	5.10	1.59	.32	.749

^a Correct responses in 1 min. ^b Correct response in 2 mins. ^c Number of correct responses in VE task. ^d Time-per-switch in the VE task. ^e Mean difference of RT (in ms) between incongruent and congruent tasks. ^f Attention-shifting auditory stream segregation task, SOA (in ms). ^g Backward digit span, capacity (in number). * Adjusted significance threshold equals: $p = 0.05/14 = .00357$.

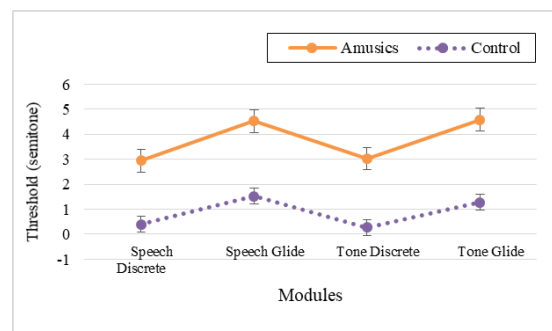


Figure 3: Pitch threshold on 4 stimulus and contour modules of amusic and control groups.

3.3. Pitch Threshold Task

Group x stimulus type x contour ANOVA conducted on pitch threshold showed a significant main effect of *group* ($F(1,37) = 15.104, p < .001, \eta_p^2 = .290$) and *contour* ($F(1,37) = 7.451, p < .001, \eta_p^2 = .167$).

.010, $\eta_p^2 = .168$). No significant main effect of *stimulus* nor any interaction between *group*, *stimulus type* and *contour* was found ($ps > .05$). The amusics demonstrated an overall larger pitch threshold than controls, meaning that amusics were less sensitive to small pitch changes. Overall, the subjects showed a larger pitch threshold in the glide condition (Rising-Falling or Falling-Rising) than in the discrete condition (High-Low or Low-High), meaning that listeners were less sensitive to small pitch changes in the glide condition.

3.4. Relationship between Tone Discrimination and Cognitive Abilities and Pitch Threshold

Bivariate correlations were first conducted between the tone discrimination performance (accuracy and RT) and cognitive measures (attention, working memory) as well as pitch threshold. Correlation analyses on accuracy revealed that the scores of Map Search (MS1: $r = .486, p = .006$; MS2: $r = .528, p = .003$), ECR ($r = .393, p = .032$), Coding (CD) ($r = .471, p = .009$) and tone discrete threshold ($r = .469, p = .009$) were significantly correlated with subjects' overall tone discrimination accuracy. Stepwise linear regression analyses were then conducted with the aforementioned five significant variables as predictors, and with their overall tone discrimination accuracy as the dependant variable to further investigate the relative contribution of these predictors to tone discrimination accuracy. Results revealed that MS2, tone discrete threshold and CD were significant predictors of the subjects' tone discrimination overall accuracy ($R^2 = .699, F(3,26) = 8.287, p < .001$), accounting for 48.9% of the variance. As for RT, bivariate correlation analyses showed that only CD was significantly correlated with the overall discrimination RT ($r = -.406, p = .026$).

4. Discussion

The overall findings revealed that amusics and controls differed not only in their musical ability and tone discrimination performance, but also in their cognitive abilities of attention and working memory. Specifically, amusics scored significantly lower in ECD, ECR and CA, showing a deficit in auditory selective attention (ECD), auditory-verbal working memory (ECD and ECR), attentional switching (ECR), inhibitory control (CA) and mental processing speed (CA). Unsurprisingly, amusics also exhibited greater (less sensitive) pitch threshold in all four types of stimuli. Importantly, the subjects' tone discrimination performance was related to visual working memory (MS and CD), speed of processing (CD), visual selective attention (MS), and discrete tone threshold.

We are aware that while a deficit in working memory was highly depicted by several cognitive tasks like ECD, ECR and CA, there is a null finding of significant inferiority of amusics in the backward digit span task, which assesses one's working memory capacity. We suggest that backward digit span, which measures one's *verbal* working memory, might be less reliable in measuring one's *pitch* memory, which might be involved in ECD and ECR as these two tasks contained different tones upon which participants had to act. In previous studies, amusics have exhibited significantly lower tone span capacity

but not digit span capacity [19], demonstrating that there is, to some degree, a distinction between verbal and pitch memory.

That the tone discrimination speed is related to the subjects' working memory and mental processing speed (CD) is in accordance with the previous finding [13]. During a speech perception task, listeners are expected to first decode auditory signal and transform the input into an accurate phonemic representation [20]. Similarly, in CD task, participants have to transform numbers into symbols according to a reference given as quickly as possible. In addition to working memory and mental processing speed, the ability of selective attention (MS) is also shown to be relevant to tone discrimination accuracy. In order to optimise speech processing, one needs to ignore irrelevant sounds so that the most important speech input or its features can be thoroughly processed. Previous studies have also shown that amusics, when presented with repeated pitch stimuli, demonstrated abnormally high activation in their middle frontal gyrus, which suggested deficits in attending to repeated pitch stimuli or encoding repeated pitch stimuli into working memory [21]. Lastly, tone discrimination accuracy is also related to discrete tone threshold. As lexical tone discrimination relies on auditory pitch processing in its early processing stages, sensitivity to discriminate small differences in auditory pitch levels (high or low) may be important for accurate tone discrimination down the processing stream.

Some differences can be observed in the profile of tone discrimination and cognitive abilities between amusic subjects examined in the current study and merger groups reported in previous studies [10-13]. While the amusics exhibited a more prevailing impairment in the discrimination of T2/T5, T3/T6 and T4/T6 in the current study, the merger groups were reported to show a highly selective merging pattern for T2/T5 or T4/T6, with the perceptual confusion of T3/T6 rarely reported. While future studies that directly compare the amusics and merger groups are required, the current results hinted at possible differences in the selectivity of tone discrimination confusion between the amusics and merger groups. In addition, while the merger groups were reported to exhibit inferior performance in only the VE task, which assesses one's attention switching [13], amusics in this study showed poor performance on several tasks, which are related to a wider array of cognitive abilities such as non-verbal working memory, selection attention and inhibitory control in both visual and auditory domains. This suggests that in addition to impairing one's tone perception, amusia, a neurodevelopmental disorder, may have a broad impact on one's cognitive abilities. Again, future studies that directly compare the amusics and merger groups are required to further examine possible differences in the cognitive profile of attention and working memory between the amusics and merger groups.

5. Acknowledgements

This work was supported by grants from the Research Grants Council of Hong Kong (ECS: 25603916), the Departmental General Research Funds, and the National Natural Science Foundation of China (NSFC: 11504400).

6. References

- [1] J. Ayotte, I. Peretz and K. Hyde, "Congenital amusia: A Group Study of Adults Afflicted with a Music-specific Disorder", *Brain*, Vol. 125, no. 2, pp. 238-251, 2002.
- [2] I. Peretz "Brain Specialization for Music", *Annals of the New York Academy of Sciences*, vol. 930, no.1, pp. 153-165, 2006.
- [3] I. Peretz, D. T. Vuvan, "Prevalence of Congenital Amusia", *Eur. J. Hum. Genet.*, 2017.
- [4] J. Foxton, "Characterization of Deficits in Pitch Perception underlying 'Tone Deafness'", *Brain*, vol. 127, no. 4, pp. 801-810, 2004.
- [5] B. Tillmann, K. Schulze and J. Foxton, "Congenital Amusia: A Short-Term Memory Deficit for Non-Verbal, but not Verbal Sounds", *Brain and Cognition*, vol. 71, no. 3, pp. 259-264, 2009.
- [6] B. Tillmann, Y. L  v  que, L. Fornoni, P. Alvo  y and A. Caclin, "Impaired Short-Term Memory for Pitch in Congenital Amusia", *Brain Research*, vol. 1640, pp. 251-263, 2016.
- [7] A. Patel, M. Wong, J. Foxton, A. Lochy and I. Peretz, "Speech Intonation Perception Deficits in Musical Tone Deafness (Congenital Amusia)", *Music Perception: An Interdisciplinary Journal*, vol. 25, no. 4, pp. 357-368, 2008.
- [8] C. Jiang, J. Hamm, V. Lim, I. Kirk and Y. Yang, "Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese", *Neuropsychologia*, vol. 48, no. 9, pp. 2630-2639, 2010.
- [9] 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. Interspeech, San Francisco, U.S.A., 8-12 September 2016.
- [10] P.P. Mok, D. Zuo and P.W. Wong, "Production and Perception of a Sound Change in Progress: Tone Merging in Hong Kong Cantonese", *Language Variation and Change*, vol. 25, no. 3, pp. 341-370, 2013.
- [11] R. Fung and C. Wong, "Mergers and near mergers in Hong Kong Cantonese tones", presented at Tone and Intonation 4, Stockholm, Sweden, 2010.
- [12] R. Fung and C. Wong, "Acoustic Analysis of the New Rising Tone in Hong Kong Cantonese", *Proceedings of the 17th International Congress of Phonetic Sciences*, pp. 716-718, 2011.
- [13] J. Ou, S.P. Law and R. Fung, "Relationship between Individual Differences in Speech Processing and Cognitive Functions", *Psychonomic Bulletin & Review*, pp. 1-8, 2015.
- [14] I. Peretz, A.S. Champod and K. Hyde, "Varieties of Musical Disorder", *Annals of the New York Academy of Sciences*, vol. 999, no. 1, pp. 58-75, 2003.
- [15] I.H. Robertson, T. Ward, V. Ridgeway and I. Nimmo-Smith, "The Test of Everyday Attention: TEA", Bury St. Edmunds, UK: Thames Valley Test Company, 1994.
- [16] J. Fan, B.D. McCandliss, T. Sommer, A. Raz and M. Posner, "Testing the Efficiency and Independence of Attention Networks", *Journal of Cognitive Neuroscience*, vol. 14, pp. 340-347, 2002.
- [17] M. Lallier, M.J. Tainturier, B. Dering, S. Donnadieu, S. Valdois and G. Thierry, "Behavioral and ERP Evidence for Amodal Sluggish Attention Shifting in Development Dyslexia", *Neuropsychologia*, vol. 48, no. 14, pp. 4125-4135, 2010.
- [18] D. Wechsler, "Wechsler Adult Intelligence Scale (4th ed.)", San Antonio, TX: Pearson/PsychoCorp, 2010.
- [19] V.J. Williamson and L. Stewart, "Memory for Pitch in Congenital Amusia: Beyond a Fine-Grained Pitch Discrimination Problem", *Memory*, vol. 18, no. 6, pp. 657-669, 2010.
- [20] A. Cutler and V. Clifton, "Comprehending Spoken Language: A Blueprint of the Listener", *The Neurocognition of Language*, New York, NY: Oxford University Press, pp. 123-166, 1999.
- [21] C. Zhang, G. Peng, J. Shao and W. S-Y Wang, "Neural bases of congenital amusia in tonal language speakers", *Neuropsychologia*, vol. 97, pp. 18-28, 2017.