Interspeech 2018
2-6 September 2018, Hyderabad

Emotional Prosody Perception in Mandarin-speaking Congenital Amusics

Yixin Zhang, Tianzhu Geng, Jinsong Zhang

Beijing Advanced Innovation Center for Language Resources, Beijing Language and Culture University, Beijing 100083, China

blcuzyx@163.com, gengtianzhu@126.com, jinsong.zhang@blcu.edu.cn

Abstract

Congenital amusia, which is a neurogenetic disorder affecting musical pitch processing, was found recently to affect not only human speech perception, but also emotional perception. Since previous studies only examined participants with non-tonal languages, they cannot easily generalize the finding to people with tonal language background, due to the fact that those people utilize pitch cues much more heavily in daily communication compared with others. To make clear the doubt, this paper investigates emotional prosody perception of Mandarin speakers with congenital amusia. We tried to recruit 19 amusics and matched control group of similar number of normal speakers, and carried out emotional perception experiments in which speech and non-speech stimuli with six kinds of emotions were used, including happy, sad, fear, angry, surprise, and neutral. Results showed that the amusics performed significantly worse than matched controls. This indicated that tone-language expertise cannot compensate for pitch deficits in amusia for emotional perception. Further analyses demonstrated that there was a positive correlation between emotion prosody performance and pitch perceptional ability. These findings further support previous hypothesis that music and language share cognitive and neural resources, and provide a new perspective on the proposition of the relation between music and language.

Index Terms: Congenital amusia, emotional prosody, perception, correlation

1. Introduction

Congenital amusia (amusia, hereafter) is a heritable neurogenic disorder characterized by a poor ability in processing fine-grained pitch in music [1], [2]. It has been suggested that music and language are intricately linked in terms of evolution and cognitive processing [3], [4], [5]. In particular, it is hypothesized that pitch processing in music and speech shares cognitive and neural mechanisms [4], [6], with increasing evidence from amusia, which is generally considered a domain-general pitch-processing in music and speech domain.

Pitch is not only an important perceptual property of music, but also encodes linguistic and emotional prosodic information [7]. For linguistic prosody, Petal et al. [8] and Liu et al. [9] examined the processing of intonation by English amusics. The results showed that amusic individuals had implications for speech intonation processing. Nguyen et al. found that French-speaking amusics performed significantly worse than control group on the discrimination of Mandarin lexical tones, indicating a transfer of musical deficits to lexical tone perception [10]. English-speaking amusics also showed impaired phonemic awareness for speech segments and lexical tones [11], [12]. For emotional prosody, Thompson et al. examined sensitivity to emotional prosody in English-speaking amusics [13], showing a worse performance than matched controls at decoding emotional prosody. Sydney et al. [14] investigated the relationship between pitch perception and emotional identification in speech of amusics from United States, found their poor pitch perception was correlated with low-pass filtered speech.

The above-mentioned studies mainly concentrated on speech processing in amusics with non-tonal language background. It is recently hypothesized that tonal-language experience might compensate for musical pitch disorder such that tonal-language speaking amusics might retain normal sensitivity to pitch changes in their native languages [15], [16]. The hypothesis was due to a fact that individuals whose native language is tonal use relatively small pitch variations to alter the meaning of words, like different tones in Mandarin, Thai, and Cantonese. They should naturally develop fine-grained pitch categories for the tones of their native languages [12]. To test this issue, a series of studies were conducted. Wong et al. found that the proportion of amusics in Cantonese was lower than that of non-tonal language background [17], providing supporting evidence for this hypothesis, whereas other studies suggested that this hypothesis was questionable. For example, Jiang et al. found that amusics with tone language background also show deficits in detecting small pitch changes in music [18]. In another research, Mandarin-speaking amusics demonstrated impaired perception of intonation and of lexical tones when pitch contrasts involved are relatively small [19], [20], [21]. Whereas investigations for emotional prosody processing of amusics with tonal-language background are still in blank. Thus, previous findings based on non-tonal language speakers cannot be easily generalized to tonal-language speakers. It will be interesting to investigate emotional prosody processing of tonal-language speaking amusics, further research is needed.

Based on this, we carried out emotional perception experiments in which speech and non-speech stimuli with six kinds of emotions were used, and explored whether there is any relation between music pitch processing and emotional speech perception. The exploration may not only provide further evidence for the hypothesis that music and language share cognitive and neural resources, but also provide a new perspective on the relationship between music and language.

2. Method

2.1. Participants

Nineteen amusics and 19 controls participated in the experiment. All subjects with Mandarin Chinese as their native language had normal hearing in both ears, right-handed, and none reported history of musical training. Participants were recruited by means of MBEA (Montreal Battery of Evaluation

 10.21437/Interspeech.2018-91

of Amusia) [22], which is widely used as a standard method to identify cases of amusia around the world. It involves six subtests that aim to assess the various components, for potential impairments in pitch perception (scale, contour, interval), temporal judgment (rhythm and meter subtests), and melodic memory (memory subtest). Those who scored 65 or under on the pitch composite score (sum of the scale, contour, and interval subtests) were classified as amusics. Totally, 246 students from Beijing Language and Culture University and Shandong University of Finance took part in the experiment and nineteen of them were diagnosed with amusics. Another nineteen non-amusics were randomly selected and assigned to the control group. Characteristics of amusics and controls are summarized in table 1:

Table 1: *Characteristics of the amusic and control groups.*

Characteristics	**Amusic mean**	**Control mean**
Age (SD)	22.6 (3.1)	23.1 (2.9)
Scale(SD)	20.3 (3.4)	28.0 (1.8)
Contour(SD)	22.2 (3.0)	27.4 (1.8)
Interval(SD)	18.6 (2.6)	27.1 (2.2)
Pitch composite (SD)	61.0 (5.1)	82.5 (3.8)

In order to examine the correlation of music pitch perception and emotional prosody perception, another twenty individuals were recruited to take part in the MBEA diagnosis and emotion identification test. They were native speaker of Mandarin, right-handed. None of them displayed any neurologic or psychiatric disorder or reported history of musical training. Their MBEA pitch composite scores ranged from 67 to 81.

2.2. Stimuli

Stimuli were chosen from corpus of CASIA—which amassed by the Institute of Automation, CAS. Each sentence was spoken with the intention to communicate each of the six emotions: happy, sad, fear, angry, surprise and neutral. Utterances were 5-6 syllables on average, and recorded by four professional speakers majoring in broadcasting. Stimuli were chosen through three steps of screening work. Firstly, three listeners who major in phonetics discriminated every sentence produced by four speakers (two males, two females) in the corpus and picked out a speaker's utterances with the best performance. The third speaker's production was checked as experimental stimuli, ultimately. Secondly, trails which contained emotional words were removed, in order to keep the materials away from the influence of word sentiment orientation. Thirdly, four listeners were invited to make an identification of all materials. If there were three or more listeners made the wrong options to the same trail, we omitted it. Finally, all stimuli were presented for both speech and non-speech conditions. In non-speech condition, the speech sounds were low-pass filtered using Praat software with the Haan band filter (0–500 Hz), in order to eliminate the intelligibility of speech but to preserve prosodic information. This resulted in 522 stimuli (29 sentences per emotion × 6 emotions × 3 repetitions) in each condition.

2.3. Procedure

Both speech and non-speech stimuli were intermixed and presented randomly in E-prime 2.0. The presentation order was counter-balanced across the participants. They were tested individually through a program in a soundproof booth and heard all stimuli at a comfortable loudness level. Before the task, a few practice trials to familiarize them with the procedures and stimuli. The subjects were tasked to identify the intended emotion from a list of the six emotion categories that were displayed on the computer screen, by pressing buttons 1-6 on a keyboard. The experiment took about an hour, precision and recall were collected.

2.4. Data analysis

In the MBEA task, pitch composite score (sum of contour, scale, and interval) was collected and represented participants' musical pitch ability. For the identification, performance of each subject was calculated with *F-score*. The formula of F-measure was shown as follows:

$$F = 2 \times \frac{P \times R}{P + R}$$

"P" refers to precision. It can be seen as a measure of exactness to emotional performance; "R" refers to recall, it is a measure of results' completeness.

Correlation analyses were conducted between each participant's MBEA melodic score and F-score of each emotion in the identification task. The *Pearson* correlation test was computed in the analyses. Moreover, we modeled the correlation by using linear regression models. These models were later verified with error mean and relative error mean.

3. Results

3.1. Emotional prosody perception

Figure1 shows the identification mean F-scores and standard error under two conditions. The results demonstrated that the F value of each emotion in amusic group was lower than that of the controls in both types of stimuli. A *paired-samples T test* was conducted to examine groups' difference, and significant differences were observed in both speech condition (*t=7.544, df=5, p<0.01*) and non-speech condition (*t=6.111, df=4, p<0.01*). The finding showed that amusic group achieved worse performance. In the non-speech condition, the recognition results of both amusic and control group decreased.

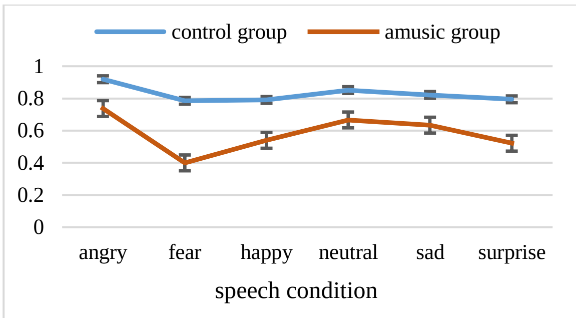

Figure 1: *F-score of emotional prosody perception*

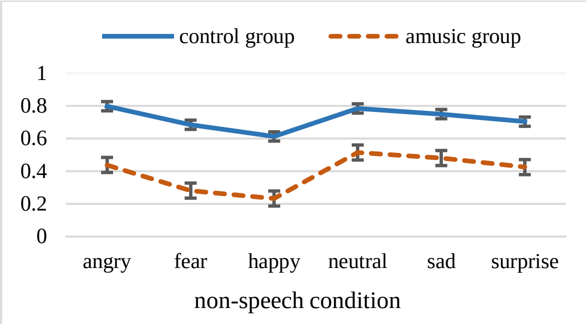

Figure 2: *F-score of emotional prosody perception (non-speech condition)*

There was a similarity between two groups' performance, namely both groups had the best recognition of "angry", and the "fear" was the worst. In non-speech condition, the error rate of "angry" and "happy" was greatly increased.

3.2. Correlation analysis between music perception and emotional prosody perception.

As can be seen from table 2, in spite of conditions, the correlation coefficient between 58 participants' (19 amusics, 19 controls and 20 middle group members) pitch composite score of MBEA and F-score of emotion task were greater than 0.6, showing a strong correlation between pitch perception and emotional prosody perception.

Table 2: *Pearson correlation coefficient analysis*

Person Correlation	Speech	Sig.	Non-speech	Sig.
Angry	.666**	.000	.788**	.000
Fear	.806**	.000	.810**	.000
Happy	.694**	.000	.824**	.000
Neutral	.631**	.000	.799**	.000
Sad	.647**	.000	.763**	.000
Surprised	.700**	.000	.680**	.000

In order to quantify the correlation, multivariate regression models were established by SPSS. In the table 3, x_1, x_2, and x_3 represent the score of scales, contour and interval, respectively. Y represents mean value of F-score in each emotion. Because of page limitation, part of the ternary regression models were shown as follows:

Table 3: *Multiple regression models (speech condition)*

Emotion types	Multiple regression model
angry	$y = 0.022x_1 + 0.003x_2 + 0.012x_3 - 0.088$
fear	$y = 0.023x_1 + 0.006x_2 + 0.022x_3 - 0.607$
happy	$y = 0.015x_1 + 0.002x_2 + 0.021x_3 - 0.250$
neutral	$y = 0.021x_1 - 0.004x_2 + 0.014x_3 - 0.001$
sad	$y = 0.015x_1 + 0.007x_2 + 0.005x_3 + 0.072$
surprise	$y = 0.016x_1 + 0.010x_2 + 0.017x_3 - 0.386$

Table 4: *Multiple regression models (non-speech condition)*

Emotion types	Multiple regression model
angry	$y = 0.020x_1 + 0.006x_2 + 0.020x_3 - 0.485$
fear	$y = 0.023x_1 + 0.006x_2 + 0.022x_3 - 0.607$
happy	$y = 0.017x_1 + 0.009x_2 + 0.020x_3 - 0.686$
neutral	$y = 0.015x_1 + 0.001x_2 + 0.010x_3 + 0.144$
sad	$y = 0.030x_1 + 0.020x_2 - 0.010x_3 - 0.327$
surprise	$y = 0.017x_1 + 0.009x_2 + 0.014x_3 - 0.424$

The fitting effect of the model was verified by means of mean and relative error mean, and the result was shown in the figure below:

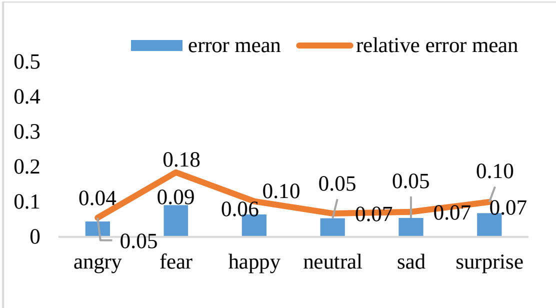

Figure 3: *error analysis of regression models (speech condition)*

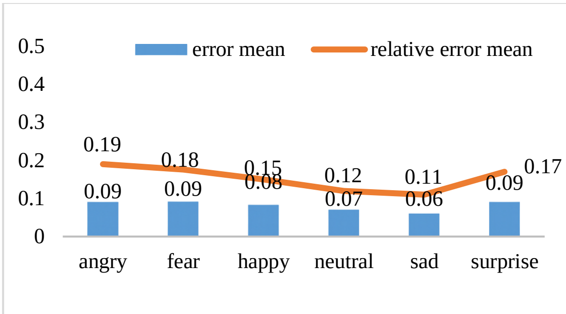

Figure 4: *error analysis of regression models (non-speech condition)*

4. Discussion

The current study assessed pitch sensitivity of 19 adults with amusia whose first language is Mandarin in the emotional identification task, and explored the correlation between musical pitch perception and emotion perception abilities in both music and speech domain. Overall, our results showed that the amusics were impaired in emotional speech perception. These findings echoes with the results found in amusics with non-tonal language background. The poor performance of Mandarin-speaking amusics confirmed that the pitch-processing deficit in amusia was not limited to music, but also transferred to language domain. It seems that tonal-language experience provides little compensation for the pitch deficits of amusics. Moreover, correlation analyses revealed that there was a positive correlation between emotional prosody capacity and pitch perception capacity, supporting the hypothesis that music and language share cognitive and neural mechanism.

In particular, there was a certain similarity of two groups in the identification task. First, both groups had the best performance of “angry”. The possible reason for the finding may be due to its pitch range, which is generally large and easy to be sensed. Another interpretation of this finding is that prosodic signals of “angry” are largely decoded using acoustic attributes other than pitch direction, like intensity and duration. So amusics were able to make the right choices with the help of other acoustic cues. Second, both groups had the worst performance of “fear”. One explanation of why “fear” got the worst identification is its non-universality in daily life, so participants may be unfamiliar with it. Another possibility may be caused by the speaker’s modest expression of this emotion.

Not surprisingly, when presented with acoustic stimuli that contained only prosodic information (i.e., low-pass filtered speech), both groups showed a significant upward trend of misjudgement rates. A reason why amusics and controls performed more poorly for these acoustically stimuli may be due to the absence of context for helping listeners to make determination. Although the two groups performed worse in non-speech condition, the group difference witnessed an expanding trend. This indicates that the amusic individuals are probably more rely on context and semantic information than controls.

It is worth noting that control group’s F-score of emotions arranged in similar orders, showing a similar perception pattern in both speech and non-speech contexts. This is in stark contrast with amusic group, which performance was less stable than the control group. It also demonstrated that amusics may rely more on semantic information than normal individuals.

We also explored the correlation between musical perception and speech prosody perception. The Pearson correlation coefficient of each emotion and music perception is greater than 0.6, suggesting a strong positive correlation between these two perception abilities in both conditions. After error analyses, we found that the error means under two conditions were within 0.2, indicating a good fitting effect of models. The fitting effect of models under the speech condition is better. The significant positive correlation between pitch perception ability and emotional prosody performance seems to suggest that emotion perceptive ability is, at least to some extent, constrained by musical perception ability. This leads to the question of a possible link between amusia and emotional prosody, and further support previous hypothesis that music and language share cognitive and neural resources. The relationship between these two abilities warrants further investigations.

5. Conclusions

In summary, this study examined musical pitch perception and speech prosody perception in music and speech in Mandarin-speaking individuals with congenital amusia. The results indicate a significant impairment in speech prosody processing for these amusics. A significant correlation was also observed between music perception and emotion prosody perception ability. These findings provide further evidence that amusia is a domain-general language-independent pitch-processing deficit, and tone language experience has no impact on pitch processing in amusia. Our results also support previous evidence that the processing of pitch within and outside language may share common mechanisms.

6. Acknowledgements

This work is supported by Advanced Innovation Center for Language Resource and Intelligence (KYR17005), Research Funds of State Language Commission (ZDI135-51), Wutong Innovation Platform of Beijing Language and Culture University (16PT05), the Fundamental Research Funds for the Central Universities（18YJ030006）The corresponding author is Jinsong Zhang.

7. References

[1] Drayna, Dennis, “Genetic correlates of musical pitch recognition in humans,” *Science* 291.5510, pp. 1969-1972, 2001.

[2] Peretz, Isabelle, Stéphanie Cummings. "The genetics of congenital amusia (tone deafness): a family-aggregation study." *The American Journal of Human Genetics* 81.3, pp. 582-588, 2007.

[3] Darwin C, “The Descent of Man, and Selection in Relation to Sex,” New York, D. Appleton and Company.

[4] Patel, Aniruddh D, “Music, language, and the brain”, Oxford university press, 2010.

[5] Arbib, Michael A, “Language, music, and the brain: a mysterious relationship”. MIT Press, 2013.

[6] Patel, Aniruddh D. "Language, music, syntax and the brain." *Nature Neuroscience* 6.7(2003):674.

[7] Jiang, Cun Mei, Y. F. Yang. “Pitch Processing of Music and Speech in Congenital Amusia.” *Advances in Psychological Science* 20.2, pp: 159-167, 2012.

[8] Patel, Aniruddh D, Wong M, Foxton J, et al. “Speech intonation perception deficits in musical tone deafness (congenital amusia).” *Music Perception: An Interdisciplinary Journal* 25.4, pp: 357-368, 2008.

[9] Liu, F, Patel A D, Fourcin A, et al. “Intonation processing in congenital amusia: discrimination, identification and imitation.” *Brain A Journal of Neurology* 133.Pt 6, pp: 1682, 2010.

[10] Nguyen, Tillmann B, Gosselin N, et al. “Tonal language processing in congenital amusia.” *Annals of the New York Academy of Sciences* 1169.1, pp: 490-493, 2009.

[11] Jones, Jennifer L, Lucker J, Zalewski C, et al. “Phonological processing in adults with deficits in musical pitch recognition.” *Journal of Communication Disorders* 42.3, pp: 226-234, 2009.

[12] Nan, Yun, Yanan Sun, and Isabelle Peretz. "Congenital amusia in speakers of a tone language: association with lexical tone agnosia." *Brain* no.133.9, pp: 2635-2642, 2010.

[13] Thompson, William F, Manuela M. Marin, et al “Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis.” *Proceedings of the National Academy of Sciences* 109.46, pp: 19027-19032, 2012.

[14] Lolli S, Lewenstein A D, Basurto J, et al. “Sound frequency affects speech emotion perception: results from congenital amusia”, *Frontiers in Psychology* pp: 1340, 2015.

[15] Stewart, Lauren. "Congenital amusia." *Current Biology* no. 16.21: pp: R904-R906, 2006

[16] Peretz, Isabelle. “Musical Disorders from Behavior to Genes.” *Current Directions in Psychological Science* no.17.5, pp: 329-333, 2008.

[17] Wong P C M, Ciocca V, Chan A H D, et al. “Effects of Culture on Musical Pitch Perception.” *Plos One* no 7.4 pp: e33424, 2012.

[18] Jiang C, Hamm J P, Lim V K, et al. “Fine-Grained Pitch Discrimination in Congenital Amusics with Mandarin Chinese.” *Music Perception an Interdisciplinary Journal,* no 28.5, pp: 519-526, 2011.

[19] Jiang C, Hamm J P, Lim V K, et al. “Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics.” *Memory & Cognition.* no: 40.7, pp: 1109-1121, 2012.

[20] Jiang C, Hamm J P, Lim V K, et al. “Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese.” *Neuropsychologia* no: 48.9 pp: 2630-2639, 2010.

[21] Liu F, Jiang C, Thompson W F, et al. "The Mechanism of Speech Processing in Congenital Amusia: Evidence from Mandarin Speakers." *Plos One*, no 7.2, pp: 1-11, 2012.

[22] Peretz I, Champod A S, Hyde K, "Varieties of musical disorders", *Annals of the New York Academy of Sciences*, no: 999(1), pp: 58-75, 2003.