THE EFFECT OF PITCH ON EMOTIONAL PROSODY PERCEPTION: EVIDENCE FROM BEHAVIORAL EXPERIMENTS ON AUTISTIC CHILDREN AND TYPICALLY DEVELOPING ADOLESCENTS

Yu Chen, Ting Wang

School of Foreign Languages, Tongji University, Shanghai 200092, China
1015112569@qq.com; 2011ting_wang@tongji.edu.cn

ABSTRACT

Acoustic cues, such as pitch, intensity and duration can be used as indicators of emotional states of speakers, among which pitch has been regarded as one of the most salient cues in reflecting different levels of emotional arousal. The current study explores the relationship between pitch and the ability of emotional prosody perception in individuals with high-functioning autism (HFA) and typically developing (TD) adolescents. An identification task was conducted using Mandarin emotional stimuli that varied in pitch level. Results showed that HFA individuals had much worse performance than TD individuals in emotional prosody identification. TD group was more sensitive to the change of pitch level than HFA group. The effect of pitch in identifying happiness is higher than that in identifying sadness.

Keywords: autism; pitch; prosody; emotional perception

1. INTRODUCTION

Emotional prosody is featured by the changes of acoustic cues, consisting of pitch, intensity, and duration. Listeners can extract emotional significance from these changeable acoustic cues to infer speakers' emotional states. This emotional prosody recognition process is believed to be mediated by a pathway that ran from the ear to several stations in the brain [1].

Specificly, pitch envelope, which can reflect different levels of emotional arousal, is among the most important parameters in differentiating basic emotions [2-4]. Justin and his colleagues [5] have found that the higher the pitch value, the higher the emotional arousal. Typically developing people have no problems in perceiving, identifying and expressing emotional states and even figure out emotional fluctuation through subtle nuances in frequency of sound. However, listeners with autism spectrum disorder (ASD) were reported to be too insensitive to decode emotional prosody. For example, some researches found that autistic people had disturbance in emotional perception, which may lead to their disability in emotional expression [6-9]. In contrast, studies on prosody production showed that individuals with ASD had increased pitch range and pitch variation in both conversation and structured communication [10-13], suggesting that they did process pitch as a salient indicator of affective differences [10].

Considering the above aspects, the current study aims to explore: 1) whether patients with high-functioning autism show impaired ability of emotional prosody perception; 2) how pitch changes affect emotional prosody perception.

2. METHOD

2.1. Participants

12 children (11 boys and 1 girl) with high-functioning autism were recruited as an experimental group (HFA group) from an autism charity organization in Shanghai. All participants were diagnosed in hospitals according to the standard clinical criteria revised by American Psychiatric Association [14]. The mean age of all participants in the HFA group was 11.83 (sd=3.460).

21 typically developing subjects (4 males and 17 females) were recruited as a control group (TD group) among postgraduate students from an university in Shanghai. The mean age of all participants in the TD group was 23(sd= 0.632).

Participants signed a consent form before the experiment and received 60 yuan as compensation for their time. They reported no problems with hearing.

2.2. Stimuli

The speech stimuli were a Mandarin sentence "我不敢相信这是真的 (I can't believe it's true)" recorded with three emotions, i.e., happiness, sadness, and neutral, by a female native speaker who had acting experience. The sentence itself was semantically neutral and suitable to convey different emotions. The sampling rate of the recording was 44.1KHz, and the sampling accuracy was 16 bits.

To rule out the possibility that other speech parameters, except for pitch, will affect the results, duration and intensity of all stimuli were normalized in accordance with that of the neutral stimulus by Praat.

Two pitch continua were synthesized using Praat on the basis of happiness and sadness templates. The pitch of each template was manipulated into 10 steps with a step size of 1 semitone, which was close to the just-noticeable difference in speech perception [15]. Table 1 and 2 explain how pitch value was adjusted. Stimulus 6 of happiness and stimulus 5 of sadness were original recordings, labeled as 0. The symbol "+" means increasing semitones and the symbol "-" means decreasing semitones. "+2" means increasing 2 semitones on the basis of the original stimulus. Similarly, "-2" means decreasing 2 semitones on the basis of the original stimulus.

Table 1: Pitch continua of happiness stimuli.

Stimuli	Semitone	Pitch(Hz)
Happy stimulus 1 (H1)	-5	245
Happy stimulus 2 (H2)	-4	260
Happy stimulus 3 (H3)	-3	274
Happy stimulus 4 (H4)	-2	291
Happy stimulus 5 (H5)	-1	305
Happy stimulus 6 (H6)	0	326
Happy stimulus 7 (H7)	+1	347
Happy stimulus 8 (H8)	+2	367
Happy stimulus 9 (H9)	+3	389
Happy stimulus10(H10)	+4	412

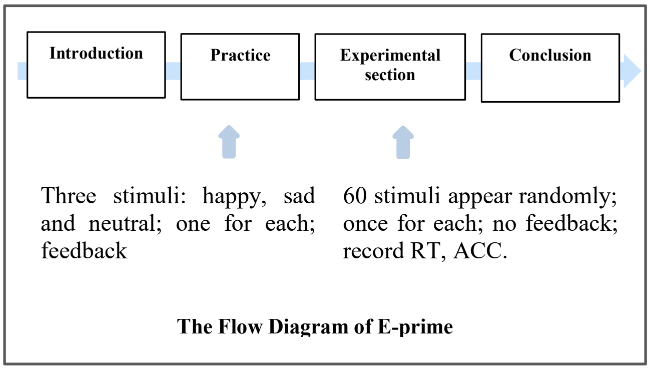
Table 2: Pitch continua of sadness stimuli.

Stimuli	Semitone	Pitch(Hz)
Sad stimulus 1 (S1)	-4	209
Sad stimulus 2 (S2)	-3	221
Sad stimulus 3 (S3)	-2	233
Sad stimulus 4 (S4)	-1	249
Sad stimulus 5 (S5)	0	263
Sad stimulus 6 (S6)	+1	277
Sad stimulus 7 (S7)	+2	295
Sad stimulus 8 (S8)	+3	309
Sad stimulus 9 (S9)	+4	330
Sad stimulus10 (S10)	+5	346

2.3. Experimental Design

An identification experiment was designed using E-prime software as shown in Figure 1. 10 happiness stimuli and 10 sadness stimuli were used in the experiment. 10 repetitions of the neutral recording were also added as fillers. All 30 stimuli were played twice randomly in one block, and 3 stimuli of each emotion were presented in a practice block with correctness feedback prior to the experimental block. A fixation using "+" was shown before each trial to draw subjects' attention. Three choices marked as happiness, neutral, and sadness in Mandarin were presented horizontally on the screen. Each subject's response and response time were recorded.

Figure 1: The experimental design using E-prime.



2.4. Procedure

Prior to testing, a brief introduction to the experiment and an explanation to the emotional prosody were made to both high-functioning autism group and typically developing group. For autistic children who faced difficulty to understand the experiment, some examples of the emotional stimuli were exhibited to them until they comprehended.

Each participant completed the experiment in a quiet room under the help with the researchers. Participants were required to sit before a laptop, listen to each stimulus carefully, and select the most appropriate emotion (i.e., happiness, neutral, and sadness) by pressing f, b, j on the keyboard respectively. The experiment lasted about 10 minutes for TD subjects and about 15 minutes for autistic subjects. Finally, 2 children with HFA failed to finish the experiment and were excluded from this study. Therefore, we collected complete results from 10 out of 12 children with high-functioning autism and from all 21 typically developing adolescents.

3. RESULTS

The identification rates of two groups for each emotion were exhibited in table 3. Generally speaking, TD group performed much better than HFA group. The identification rates of three emotions were more than 90% in TD group, whereas they were about 60% in HFA group, which was less than twice the chance level (33.3% in this study).

Table 3: Identification rates of emotions in HFA group and TD group.

	Happy	Neutral	Sad
TD	92.86%	97.62%	98.1%
HFA	64.5%	59%	64%

In order to determine whether the numerical difference between identification rates was statistically significant, we fitted a linear mixed-effect model to the identification rate using subject group (HFA and TD) and emotion (happiness, neutral, sadness) as fixed factors, and stimuli as a random factor.

The results showed a significant effect of group (X^2 = 264.5679, df = 1, p < 0.001), suggesting that the identification rate of HFA group was significantly worse than TD group. The effect of emotion was not significant (X^2 = 0.6462, df = 2, p > 0.05), nor was the interaction between group and emotion (X^2 = 2.1376, df = 2, p > 0.05), indicating that three emotions were perceived similarly across groups.

3.1. TD group

The following figures illustrate the identification rate of each pitch step for happiness and sadness continua in TD group.

As shown in Figure 2, for identifying happiness, TD group's performance got better as pitch value increased. When pitch value was up to stimulus 5 (309Hz), all TD subjects showed no trouble in identifying happiness. The results suggested that the increase of pitch may enhance TD group's perception of happiness.

In identifying sad emotion, the result did not go as our expectation, i.e., when pitch went lower, the identification of sadness should have been higher since low arousal emotion was characterized as low pitch. Listeners made most mistakes in sad stimulus 1 (209hz) and sad stimulus 2 (221hz). From sad stimulus 4 (249Hz) to sad stimulus 9 (330Hz), all participants made no mistakes.

With the data of pitch and accuracy, a simple liner regression was conducted by R. According to the fitting model, every one semitone rise in pitch value led to 3.2 percentage points significant rise in accuracy of identifying happy stimuli [F(1,8)=15.59, p=0.004, SE=0.008, t=3.948].

To sum up, typically developing adolescents showed no difficulty in identifying basic emotions. As pitch increased, TD group had a significantly better perception of happiness.

Figure 2: Identification curves of happiness in TD group.

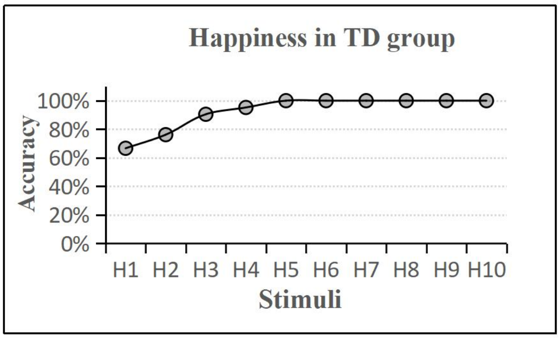
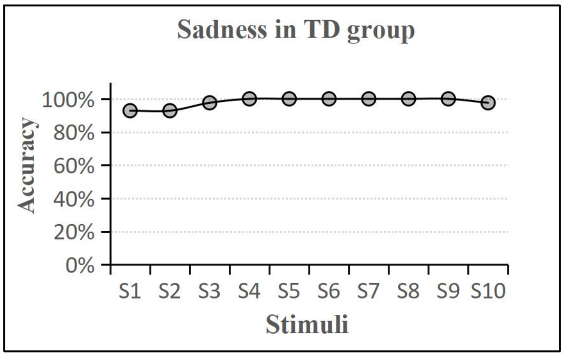


Figure 3: Identification curves of sadness in TD group.



3.2. HFA group

Figure 4 and 5 show the identification rate of each pitch step for happiness and sadness continua in HFA group. Compared with TD group, the overall identification curves were rather lower. Specifically, in figure 4, the overall trend of the curve was climbing, which was similar to that of TD group. For stimulus 1, 2, 3 with lowered pitch, the identification rates were around 40%, which was close to the chance level, indicating that HFA couldn't decode happiness successfully when pitch was less than 274 Hz in this study. We further conducted a simple liner regression by R. Results showed that every one semitone rise in pitch value lead to 4.8 percentage points significant rise in accuracy of identifying happy stimuli [F(1,8)=16.89, p=0.0034, SE=0.012, t=4.1].

In identifying sadness as shown in Figure 5, the identification curve did not show a declining trend as our expectation, which was similar to that of TD group. Except for the first two stimuli in which most participants have trouble in identifying them successfully, the rest stimuli showed a rather closely accuracy rate, fluctuating around the level of 70%.

To sum up, with the rise in pitch, participants' ability to identifying happiness will improve significantly. HFA group has better performance in identifying happiness expressed in voice than identifying sadness.

Figure 4: Identification curves of happiness in HFA group.

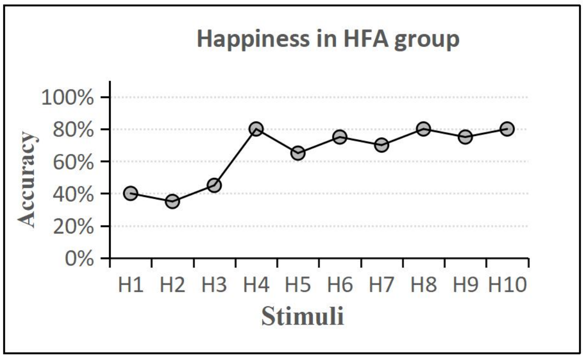
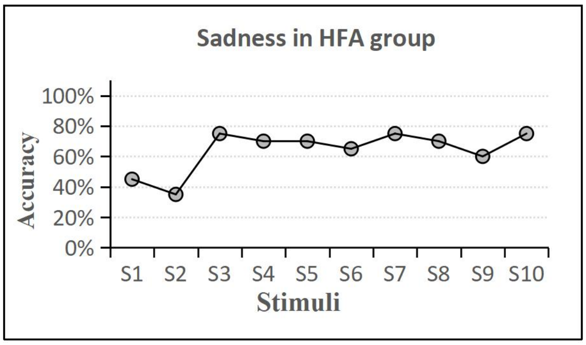


Figure 5: Identification curves of sadness in HFA group.



4. DISCUSSION

This study tried to explore the effect of pitch on the ability of emotional prosody perception between high-functioning autism group and typically developing group. Results showed that HFA children had lower identification rates than TD adolescents in decoding happiness, neutral, and sadness. TD group was more sensitive to pitch changes than HFA group.

Specifically, participants in TD group had the ability to perceive and identify emotions through pitch change. Especially in identifying happy emotion, there was a clear positive correlation between accuracy rate and pitch change. When pitch reached round 300 Hz, subjects were more likely to perceive the happy emotion.

Second, it is harder for individuals with HFA to identify sadness than identifying happiness, which may result from the fact that happiness is one of the high-arousal emotions but sadness is one of the low-arousal emotions. For children bearing emotional disorder, to perceive sadness is more difficult. Their performance in happiness identification is better, which was confirmed by a simple liner regression analysis with a significant positive correlation between identification rate and pitch value.

Third, both groups have difficulty in identifying the first two sad stimuli. Both sad emotion and neutral emotion belong to low-pitch emotions. So low accuracy rate in both groups may be ascribed to the too low pitch. When pitch value is adjusted below 233Hz, which is also neutral stimuli's pitch value, listeners are likely to be confused by the low pitch and regard it as neutral emotion.

Fourth, both group had difficulty in identifying the first two sad stimuli. Both sad emotion and neutral emotion belong to low-pitch emotions. So low accuracy rate in both group may be ascribed to the lower pitch for a female speaker. When pitch value was adjusted below 233Hz, which was close to neutral stimuli's pitch value, listeners were likely to be confused by the low pitch and regarded it as neutral emotion.

Fifth, in both groups, pitch plays an important role in identifying happiness. However, the effect of pitch seems to be reduced when identifying sadness for both groups.

As an ongoing project, the current study lies in some limitations. First, the different identification performances may be due to the aging effect of two subject groups. Further study should recruit typical developing children with matched age with the HFA children. Second, the acoustic patterns of stimuli should be better controlled so that they only vary in pitch. Therefore, future researches can use synthetic speech to generate the three emotions. Third, the pitch step could be enlarged from 1st to 5st so that the effect of pitch change in identifying emotions may be clearer especially for individuals with autism.

5. ACKNOWLEDGEMENT

This work was supported in part by the Youth Project of Humanities and Social Sciences Foundation [grant no. 18YJC740103] of the Ministry of Education in China and the Chenguang Program [grant no. 16CG21] of the Shanghai Education Development Foundation and the Shanghai Municipal Education Commission, which was awarded to the second author. Special thanks go to all the subjects for their participation.

6. REFERENCES

[1] Schirmer, A., Kotz S. A. 2006. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in cognitive sciences* 10*, 24-30.*

[2] Williams, C. E., Stevens, K. N. 1972. Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America* 52(4B), 1238-1250.

[3] Levin H., Lord W. 1975. Speech pitch frequency as an emotional state indicator. *IEEE Transactions on Systems, Man, and Cybernetics* 2, 259-273.
[4] Murray, I. R., Arnott, J. L. 1993. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America* 93(2), 1097-1108.
[5] Laukka, P., Juslin P., Bresin, R. 2006. A dimensional approach to vocal expression of emotion. *Cognition & Emotion* 19, 633-653.
[6] Hobson, R. P. 1986. The autistic child's appraisal of expressions of emotion. *Journal of Child Psychology and Psychiatry* 27(3), 321-342.
[7] Hall, G. B., Szechtman, H., Nahmias, C. 2003. Enhanced salience and emotion recognition in autism: A PET study. *American Journal of Psychiatry* 160(8), 1439-1441.
[8] Globerson E., Amir N., Kishon‐Rabin L., Golan, O. 2015. Prosody recognition in adults with high‐functioning autism spectrum disorders: From psychoacoustics to cognition[J]. *Autism Research* 8(2): 153-163.
[9] Diehl, J. J., Paul, R. 2013. Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children by children with autism spectrum disorders. *Applied Psycholinguistics* 34(1), 135-161.
[10] Hubbard, K., Trauner, D. A. 2007. Intonation and emotion in autistic spectrum disorders. *Journal of psycholinguistic research* 36(2), 159-173.
[11] Nadig, A., Shaw, H. 2012. Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners. *Journal of Autism and Developmental Disorders* 42(4), 499-511.
[12] Diehl, J. J., Watson, D., Bennetto, L., McDonough, J., Gunlogson, C. 2009. An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics* 30(3), 385-404.
[13] Bonneh, Y. S., Levanon, Y., Dean-Pardo, O., Lossos, L., Adini, Y. 2011. Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in human neuroscience* 4, 1-7.
[14] American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)* Washington, D.C: American Psychiatric Publishing.
[15] 't Hart, Johan. 1981. Differential sensitivity to pitch distance, particularly in speech. *Journal of the Acoustical Society of America* 69, 811–821.