Vocal Emotion Recognition in Mandarin-Speaking Cochlear Implanted Children

Haitao GUAN[1,2] and Wentao GU[1,2]
[1] School of Chinese Language and Literature, Nanjing Normal University, China
2 Nanjing Normal University and the George Washington University Joint Laboratory of Speech, Hearing and Rehabilitation Sciences, China

Abstract
Previous studies have shown that cochlear-implanted (CI) children have big difficulty in vocal emotion recognition in comparison to the normal-hearing peers, mainly due to the lack of efficient coding of fundamental frequency (F0) in the CI devices, but few research has been conducted on preschool CI children, and even fewer on CI children speaking Mandarin, a tone language in which F0 conveys both lexical tone and emotion related intonation. The present study compared vocal emotion recognition between two groups of 4-to-5-year-old Mandarin-speaking children, i.e., CI and normal-hearing (NH) groups. Six semantically neutral sentences were recorded by a female speaker in three basic emotions, i.e., anger, joy, and sadness, among which anger and joy showed greater mean F0, F0 variability and F0 range than sadness. The children were asked to judge the type of emotion from the recorded utterances. For all three emotions, the CI group showed significantly lower recognition accuracy and longer reaction time than the NH group. The confusion mainly occurred between anger and joy in the NH group, while all three emotions were mutually confused in the CI group, confirming that Mandarin-speaking CI children were impaired in perceiving vocal emotions due to the lack of F0 processing ability.

Keywords: vocal emotion perception, cochlear implanted children, Mandarin

1. INTRODUCTION

According to the national survey from the Ministry of Health of China in 2013, there were 27.8 million people suffering hearing loss (HL), 115,000 children under the age of 7 had severe to profound hearing loss, and 30,000 babies were born with HL each year in China (1). In recent years, we can use hearing aid (HI) or cochlear implant (CI) to help hearing loss people improve their hearing ability. CI is a surgically implanted electrical devices into cochlear that help people with severe to profound hearing loss hear and process auditory information. Many studies have documented that CI device can facilitate user’s perception and production of consonants and vowels (2), production and perception of tone (3), speech intelligibility (4) and so on. With the help of CI devices, many implanted children can acquire speech and language, and even able to attend mainstream schools (5).

However, CI device only could provide limited information about the temporal fine structure of

[1] patazhou@hotmail.com
[2] wtgu@njnu.edu.cn

speech, such as pitch and harmonics (6,7), thus CI users may have difficulty to perceive and produce important prosodic forms of speech communication such as question-statement contrasts, voice emotion (8,9).

In our daily life, emotion have an important role. The ability of expressing and understanding emotions influences social relationships. For both children and adults, problems in emotion understanding have been shown to relate with developing symptoms of psychopathology or poor social functioning (5). Due to the device limitation, CI users may have deficits on emotion perception and production and this may lead to complications in social interactions and communication, which can result in withdrawal and exclusion from social situations (10). Therefore, it is worthwhile to examine how well the CI children identify vocal emotions.

Many studies have investigated the vocal emotion recognition ability of hearing loss people who wearing hearing aid (HA) or CI. Pereira's (11) research investigated 4 emotions (angry, happy, sad, and neutral) recognition performance in CI people. The result showed mean CI accuracy was 51%, whereas mean NH accuracy was 84%. When all the sentences were amplitude normalized, NH accuracy was not affected, whereas CI accuracy was reduced to 38%, indicating that CI listeners depended strongly on intensity cues for vocal emotion recognition. Hopyan-Misakyan et al. (12) found children between 7-13 years old performed worse than NH group when identify the emotion (happy, angry, sad, fearful). Nakata et al. (13) investigated the perception and production emotion ability of Japanese CI children whose age is between 5 to 13 years old. In the perception task, the children were required to judge the sentence was spoken in a happy, sad, or angry emotion. The NH children performed better than CI children. For the CI children, happy and sad were easier to identify than angry emotion. Luo et al. (9) found CI adults performed worse than NH control group when identifying the emotion (happy, angry, sad, fearful, or neutral) of sentences. Volkova et al. (14) found that 5 to 7 years old CI children discriminated happy and sad utterances with a score above chance but less accurate than NH group. Chatterjee et al. (15) reported an experiment of vocal emotion recognition (included 5 emotions: angry, happy, neutral, sad, and scared) in NH and CI listeners, including both adult and child subjects. They found that although CI adults and CI children have similar performance, they both showed significantly poor performance than NH groups.

In short, studies based on English-speaking individuals agree that CI have deficits in vocal emotional recognition, for both child and adult CI users. But little is known about the case for CI users speaking tonal languages such as Mandarin. Mandarin is a tonal language in which F0 conveys both lexical tone and emotion related intonation. The goal of the research is to examine vocal emotion perception in child CI users and compare their performance with NH children from China. The present study was restricted to happy, sad, and angry emotions. These emotions were chosen because they meet the three emotions frequently in daily life, and it would be easy for them to understand those emotion and finish the task. We hypothesized that NH children would succeed in the task through all emotions better than CI children.

2. Methods

2.1 Participants

Children participants included 14 NH (8 boys, 6 girls), whose age varied from 4.2 to 5.11 years (mean age 5.16 years, SD = 0.55), and 14 CI (6 boys, 8girls), whose age varied from 4.2 to 5.9 years (mean age 5.05 years, SD = 0.56, mean duration of device use 2.53 years). Details of the CI participants' information are provided in Table 1.

Table 1. Background information of CI participants

Child	Age at implant (yr)	Age at test (yr)	Implant use (yr)	Hearing aid
CI-1	2.8	5.9	3.1	Y
CI-2	3	5.9	2.9	Y
CI-3	3.49	5.6	2.11	Y
CI-4	3	5.5	2.5	Y
CI-5	3	5.4	2.4	Y
CI-6	2.3	5	2.7	Y
CI-7	2.2	4.9	2.7	Y
CI-8	2.6	4.9	2.3	Y
CI-9	2.3	4.6	2.3	Y
CI-10	2.3	4.6	2.3	N
CI-11	1.6	4.5	2.9	Y
CI-12	2.1	4.4	2.3	Y
CI-13	2.09	4.2	2.11	Y
CI-14	2.5	5.3	2.8	N

2.2 Stimulus

For the emotion perception task, six semantically neutral sentences (see Table 2) were produced by the same female speaker with happy, sad, and angry in an adult-directed manner, for a total of 18 target utterances. For the selection of high quality recording material, each target sentences were produced 3 times. The speaker produced in a soundproof room and the sounds were recorded by a microphone in front of the speaker. The original audio files (44.1 kHz sampling rate, 16 bit) were edited using Adobe Audition version 1.6 software. Editing involved selecting the best of the three productions for each sentence and saving each as an individual audio file.

Table 2. Sentences for emotion perception task.

1. 今天下这么大的雪 (It’s snowing heavily today).
2. 小明把饼干吃完了 (Xiao Ming ate up all cookies).
3. 过几天就要开学了 (New term will start in a few days).
4. 明天是儿子的生日 (Tomorrow is my son’s birthday).
5. 爸爸要去北京出差 (Dad is going to Beijing on business).
6. 我们马上回家吧 (Let’s go back home right now).

2.3 Procedure

All the children were tested individually. For the CI participants, they were tested in a quiet room at the Rehabilitation Center in Nanjing. For the NH participants, they were tested in their primary school in a quiet room. We used EPRIME software on a touchscreen computer to present the audio stimuli for the task. After hearing one audio file, children would make an emotion choice for it by touching the associated picture on the touchscreen computer. Sentences and emotions were fully randomized. Prior to the formal experiment, the participants were given passive training with two sentences each for the three emotions (these sentences were not included in formal experiment) to familiarize the participants with the experiment procedure.

3. Results

3.1 Acoustic features in emotion perception task

We analyzed the mean pitch, pitch range, and pitch variability of each vocal emotion, as shown in Table 3. The mean F0, F0 range, and F0 variability were higher in anger and happy emotions comparing with sad emotion. Statistical analysis showed this is no difference($p>0.05$) between anger and happy in all parameters, while sad have significant differences ($p<0.001$) when comparing with anger or happy. Those differences in acoustic features may affect the children's performance.

Table 3. Average values of acoustic parameters of the sentence

Emotion	Mean pitch (Hz)	F0 range (Hz)	SD of F0
Anger	380	287	77
Happy	349	288	72
Sad	210	189	39

3.2 Emotion perception experiment

Figure 1 showed the mean emotion perception accuracy for NH children, as well as the CI children in three emotions. NH children significantly outperformed CI users in the perception of each motion type: happy ($p<0.044$), sad ($p<0.001$) and anger ($p<0.001$). Sad (89%) have highest accuracy in NH, followed by anger (74%) and happy (61%). The CI's accuracy only above chance level (33%), and happy have the highest accuracy (43%), followed by anger (40%) and sad (39%). In the three emotion types, sad was the easiest to identify for NH, while all emotions are hard for CI. Within group, there is no difference between the three emotion types in CI ($p>1$), while NH show significant difference between happy and sad ($p<0.001$), as well as sad and anger ($p=0.023<0.05$). We also found the CI group need more time to finish the task. The mean reaction time for CI is 3616s, and 2051s for NH group.

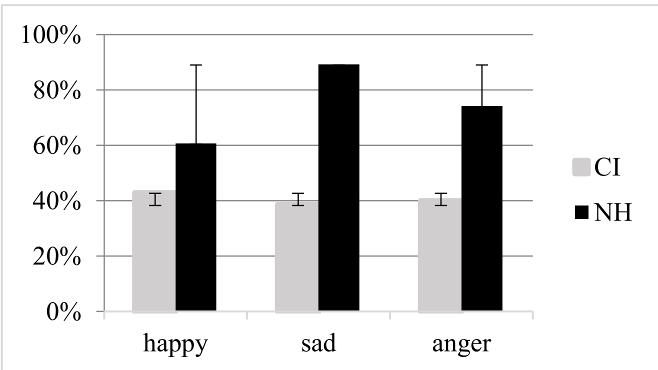

Figure 1. Accuracy of CI and NH children on emotion types. Error bars indicate standard errors.

Table 4 showed the confusion matrix of the two groups. The NH group usually confused anger with happy, while CI group confused all the three emotions mutually, among whitch they confused happy with sad most frequently.

Table 4. Confusion Matrix of the Emotion Types

Group	Emotion	Intended Emotion			
		Happy	Sad	Anger	No choice
CI	Happy	43%	**30%**	23%	4%
	Sad	27%	39%	**29%**	5%
	Anger	26%	**31%**	40%	3%
NH	Happy	61%	16%	**22%**	1%
	Sad	9%	89%	2%	0%
	Anger	**17%**	9%	74%	0%

4. DISCUSSION

The goal of the present study is to examine vocal emotion perception in Mandarin-speaking CI children and compare their performance with NH children. Result showed that CI children performed significantly worse than NH children in all emotion types, which accords with our hypothesis. The result is in accordance with previous studies for CI children speaking other languages (5,15,16).

There are some possible reasons for the differences between CI and HL children. Hearing loss have a negative impact on children language development, therefore language delay is a likely factor. Another possible reason is the duration of CI use. Children who had a longer duration of hearing deprivation or who had been using CI device for a shorter time could have lower experience and attention on emotion perception, comparing with HL children. In present research, the mean hearing age for NH is 5 years, while only 2.53 years for CI. If we control the two group's hearing age and chronological age, there could be a different result. Gray et al. (17) reported the emotions developed among deaf children: those aged 9:5–13:2 years were better at assigning emotions to story characters than younger deaf children aged 5:5–8:7 years. Future research will continue to examine how the duration of CI use affects the emotion perception performance.

In present study, NH group performed best on sad emotion perception, while CI group performed best on happy emotion perception. From acoustics analysis, we found sad emotion significantly different from happy and anger emotions on the three parameters, which may cause its highest accuracy in NH group. Most and Michaelis (16) found sadness to be the easiest to recognize by hearing impaired and normally hearing individuals in the audio-only mode. Nakata et al. (13) found Japanese children with cochlear implants succeeded in identifying happy and sad utterances which were amplitude normalized, but they failed to identify angry utterances. The different results between present study and previous ones may due to the following reason. Firstly, language effect. The tone language nature of Mandarin might be responsible for the degraded performance of the CI devices in Mandarin-speaking children than in the children speaking other languages such as English and Japanese. Secondly, participants effect. In our research the age of participants is between 4-5 years old, while in Nakata et al. (13) is 5-13 years old. In our research, all the hearing loss children are CI users, whereas the results in Most and Michaelis (16) were derived from the data of a mixture of 9 CI children and 17 children wearing HAs.

According to the confusion matrix, the NH group typically confused anger with happiness. This is probably the result of the fact that the mean pitch and pitch variability has no significant difference among those emotions. The CI group confused the three emotions mutually, though one emotion (sad) significantly differed from the other two acoustically.

CONCLUSIONS

In summary, the present study explored the emotion recognition ability of preschool Mandarin-speaking CI children. Comparing with NH children, CI children performed poorly in the task of recognizing of sad, anger and happy emotion types. It confirms that Mandarin-speaking preschool CI children were impaired in perceiving vocal emotions, which reveals their lack of ability to process complex F0 variations.

For future research, we will increase the number of participants to explore how the CI use duration affects their emotion recognition ability. Besides, the audio material will come from more speakers to test whether speaker variation affects CI children's emotion recognition.

ACKNOWLEDGEMENTS

This research was supported by the Major Program of the National Social Science Fund of China (13&ZD189) and the project for Jiangsu Higher Institutions' Excellent Innovative Team for Philosophy and Social Sciences (2017STD006). We gratefully acknowledge the cooperation of many parents and children participating in this study.

REFERENCES

1. Liang Q. Mason B. Enter the dragon – China's journey to the hearing world. Cochlear Implants International. 2013;14:26 –31.
2. Blamey PJ, Barry JG, Jacq P. Phonetic inventory development in young cochlear implant users 6 years post-operation. Journal of Speech, Language, and Hearing Research, 2001;44:73-79.
3. Peng SC, Tomblin JB, Cheung H, Lin YS, Wang LS. Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. Ear Hear. 2004;25:251–264.
4. Chin SB, Finnegan KR, Chung BA. Relationships among types of speech intelligibility in pediatric users of cochlear implants. Journal of Communication Disorders, 2001; 34:187–205.
5. Wiefferink CH, Rieffe C, Ketelaar L, De RL, Frijns JH. Emotion understanding in deaf children with a cochlear implant. Journal of Deaf Studies and Deaf Education, 2013;18:175–186.
6. Geurts L, Wouters J. Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. J Acoust Soc Am. 2001;109; 713–726.
7. Green T, Faulkner A, Rosen S. Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. J Acoust Soc Am. 2004;116: 2298–2310.
8. Peng SC, Tomblin JB, Spencer LJ, Hurtig RR. Acquisition of rising intonation in pediatric cochlear implant recipients-a longitudinal study. Int Congr Ser. 2004;1273: 336-339.
9. Luo X, Fu QJ, Galvin JJ. Vocal emotion recognition by normal-hearing listeners and cochlear implant users. Trends in Amplification. 2007;11(4): 301–15.
10. Waaramaa T, Kukonen T, Mykkänen S, Geneid A. Vocal emotion identification by children using cochlear implants, relations to voice quality and musical interests. Journal of speech, language, and hearing research. 2018;1-13.
11. Pereira C. The perception of vocal affect by cochlear implantees. In: Waltzman SB, Cohen NL, eds. Cochlear Implants. New York, NY: Thieme Medical; 2000; 343-345.
12. Hopyan-Misakyan TM, Gordon KA, Dennis M, Papsin BC. Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. Child Neuropsychology. 2009;15(2): 136–46.
13. Nakata T, Trehub SE, Kanda Y. Effect of cochlear implants on children's perception and production of speech prosody. J Acoust Soc Am. 2012;131(2):1307-1314.

14. Volkova A, Trehub SE, Schellenberg EG, Papsin BC, Gordon KA. Children with bilateral cochlear implants identify emotion in speech and music. Cochlear Implants International. 2013; 14(2): 80–91.
15. Chatterjee M, Zion DJ, Deroche ML, Burianek BA, Limb CJ, Goren AP, Kulkarni AM, Christensen JA. Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. Hearing Res. 2015; 322: 151-162.
16. Most T, Michaelis H. Auditory, visual, and auditory–visual perceptions of emotions by young children with hearing loss versus children with normal hearing. Journal of Speech, Language, and Hearing Research. 2012; 55: 1148–1162.
17. Gray C, Hosie J, Russell P, Scott C, Hunter N. Attribution of emotions to story characters by severely and profoundly deaf children. Journal of Developmental and Physical Disabilities. 2007; 19: 145–159.