Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Production and perception of emotional intonation among preschool children with cochlear implants

Fang Zhang[1], *Hongtao Li*[1], *Ao Chen*[2]

[1]China Rehabilitation Research Center for Hearing and Speech Impairment
[2] School of Psychology, Beijing Language and Culture University
229695707@qq.com, irischen71@hotmail.com

Abstract

Children with cochlear implants (CI) often have difficulties processing low frequency due to insufficient resolution of frequency filters of the CI, leading to inaccurate pitch perception. The current study investigated intonation perception and production of Mandarin Chinese preschool children (3-5-year-olds) with CI. In Experiment 1, 55 children with CI and 44 children with normal hearing (NH) were auditorily presented with six pairs of semantically neutral sentences, with each pair including a happy and a sad intonation. The children were asked to identify the emotion of each sentence by selecting cards with corresponding facial expressions. Neither the NH children nor the children with CI were able to identify the emotions accurately. In Experiment 2, using a sentence repetition paradigm, another 51 children with CI and 32 NH children produced the six pairs of sentences used Experiment 1. Compared to the happy sentences, both groups showed a lower mean f0 for the sad sentences. Nevertheless, only the NH children made use of f0 range and intensity cues when differentiating the two emotions. These results indicate late development of perceptual emotional intonation representation, and the CI children made use of less cues compared to NH children when marking sentential emotion.

Index Terms: emotion, intonation, cochlear implants, children

1. Introduction

Cochlear implants (CI) make use of safe electric stimulation to provide or restore functional hearing among people with severe hearing impairment. With CI, a large amount of pediatric users are able to develop near normal oral language. However, pitch perception is challenging for CI users because CI is not yet able to provide high spectral resolution especially in the low frequency range, and the users often have difficulties extracting temporal pitch from temporal modulations [1][2][3][4].

Pitch variation abounds in human speech, with the acoustic correlates of pitch being fundamental frequency (f0). For example, intonation is the use of pitch over the domain of an utterance, and it provides crucial paralinguistic cues in communication in that it can convey intention of speakers. Across many different languages, questions tend to end with a pitch rise while statements often have a fall at the boundary of the utterance [5]. Although to what extent different emotions show qualitatively different pitch contours remains debated [6], based on analysis of acted database, sentence level f0 measures such as mean f0, minimum f0, f0 lower and upper quartile are informative cues for differentiating different emotions (e.g., joy, anger, fear, disgust, etc). Compared to neutral intonation, happy intonation tends to show a global rise in f0 as well as extended accent while sad intonation tends to show a flat f0 contour across different syllables in the sentence [7][8]. Analyses of actor portrayed emotional expressions show that mean f0 and range of f0 at sentence level vary strongly as a function of the degree of activation of the portrayed emotions.

Seeing the important role of pitch in marking different intonations, and since CI degrades spectral resolution in low frequency range, it can be hypothesized that children with CI may have difficulties acquiring intonation. Indeed, previous studies have found that even adult CI users were less accurate when identifying interrogative versus declarative intonation, and large individual variation was observed [9]. Other studies show that school-aged children and teenagers with CI have difficulties perceiving, producing, as well imitating these two intonations as well [10][11][12][13]. Identifying emotional content of natural utterances has also been found to be difficult for CI users [14].

The current study tested identification as well as production of emotional (happy and sad) intonation of Mandarin Chinese preschool children with CI. Proper use of vocal emotion is fundamental for emotion regulation since it plays a crucial role for children's social-psychological development. Mandarin Chinese is a lexical tone language in which pitch variation is used to contrast meaning lexically. Therefore, compared to non-tone languages, Mandarin Chinese speech prosody is more complex in that sentential level intonation interacts with word level lexical tones. Phonetically, previous studies have shown that in Mandarin Chinese, the lexical tones produced in in happy sentences show more pronounced pitch contours (i.e., larger pitch range) compared to those in sad sentences [15]. Research with Chinese CI users shows that Chinese preschool and school aged children are less capable of identifying lexical tones as well as emotional intonation [16][17]. By including both identification and sentence repetition experiments, we aim to examine whether the perception and production of intonation developed in parallel among children with CI, and whether and how children with CI differed from normal hearing (NH) children. Seeing the complex speech prosody of Mandarin Chinese, we predict that children with CI will perform less well than NH children in both identification and sentence repetition tasks.

2. Emotional intonation identification experiment

The testing materials and procedures were approved by Ethics Committee in China Rehabilitation Research Center for

10.21437/SpeechProsody.2024-15

Hearing and Speech Impairment. Consent was obtained from the parents of all the participants.

2.1. Participants

Fifty-five children (20 three-year-olds, 21 four-year-olds, and 14 five-year-olds) with CI implants participated in the experiment. The control group were 44 normal hearing children (10 three-year-olds, 12 four-year-olds, and 22 five-year-olds). All children spoke Mandarin natively, and were attending kindergarten at the time of the experiment. The mean hearing threshold of the children with CI was 33 dB HL (range 18-43 dB HL) after implantation. All the children with CI had CI implanted unilaterally, and received speech and language therapy for at least one year by the time of the experiment, and were able to communicate verbally. No parent reported other developmental or neurological disorders for any of the participants.

2.2. Materials

Six semantically neutral sentences were constructed, each comprising six syllables. The sentences are listed in Table 1. A female native Mandarin speaker, who was a trained stage performer, produced the six sentences with both a happy and a sad voice, thus twelve sentences in total. The mean duration of the happy and sad sentences was 1.21 (SD = 0.12) and 1.55 (SD = 0.14) seconds respectively. The happy sentences had a mean (SD) minimum pitch of 163 (23) Hz and a mean (SD) maximum pitch of 287 (27) Hz , while the sad sentences had a mean (SD) minimum pitch of 120 (53) Hz and a mean (SD) maximum pitch of 263 (17) Hz. The mean (SD) pitch range of the sad and happy intonations was 125 (28) Hz and 143 (49) Hz. The mean intensity of the happy and sad intonations were 71 (1.26) dB and 67 (1.72) dB. The pitch contours of the happy and sad intonations of one sentence are illustrated in Figure 1. As can be seen in Figure 1, sad intonation tended to have a flat pitch contour while the happy intonation always had a rise at the boundary.

Table 1: *Sentences used in the identification experiment*

No.	sentence	translation
1	小猫咪喝牛奶	Little cat drinks milk
2	我和小海挖土	I dig with *Xiaohai*
3	妈妈买豆腐脑	Mom buys soft tofu
4	那边有一朵花	There is a flower over there
5	那有很多动物	There are many animals
6	我和弟弟游泳	I swim with my little brother

Fifteen adult native Mandarin listeners listened to the sentences, and were able to recognize all the syllables as well as the emotions of all the sentences correctly. These twelve sentences were used as stimuli in the current experiment (two for practice trials and ten for the experiment, see below).

Two cards were created, one with a smiling face and the other with a frowning face, representing the happy and sad emotions respectively.

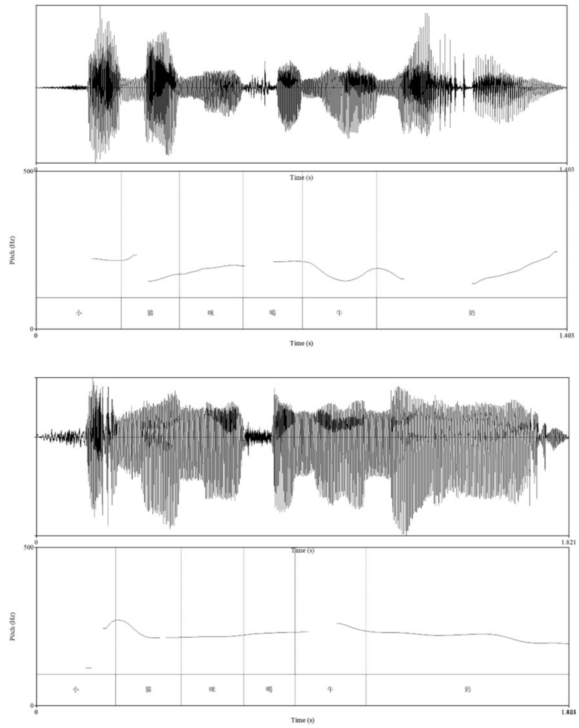

Figure 1: *Example of happy (upper panel) and sad (lower panel) intonations.*

2.3. Methods

The experiment was conducted in a quiet room, and the background noise was below 45 dBA. The stimuli were presented by a CD player in front of the children. The distance between the participants and the CD player was 1 m. The sound level of the stimuli was between 65 and 70 dB SPL.

The experiment started with a practice trial. For each child, the experimenter played one happy and one sad sentence, and asked him/her to identify the emotion of the sentence by selecting the card with the corresponding facial expression. If the children identified the emotion correctly, the experiment proceeded. Otherwise the experimenter would show the child the correct card. All the children had to finish both of the practice trials. In the experiment, the remaining ten sentences were played randomly to the children, and one point was awarded for every correct answer. No feedback was given during the experiment. For each emotion condition, the recognition score was calculated as number of correctly identified sentence divided by the total number (10) of sentences.

A mixed linear ANOVA was applied to analyze group (CI or NH), age (3-, 4-, 5-year-olds), and emotion condition (happy or sad) effects.

2.4. Results and discussion

The recognition scores of the children with CI and NH children were listed in Table 2, and the results of the ANOVA were listed in Table 2.

Table 2: *Mean (standard deviation, SD)Recognition score of the children with CI and NH children.*

Age	Group	N	Happy sentence	Sad sentence

3	CI	20	0.36(0.26)	0.33(0.30)
	NH	10	0.42(0.30)	0.52(0.25)
4	CI	21	0.34(0.24)	0.36(0.27)
	NH	12	0.52(0.34)	0.55(0.27)
5	CI	14	0.46(0.36)	0.51(0.24)
	NH	22	0.55(0.28)	0.59(0.27)

Although the NH children tended to show higher accuracy in both conditions, the ANOVA analysis failed to find any significant main effect or interaction. In other words, no significant difference was observed between the NH and CI children, nor across the difference age groups or across the happy/sad conditions. Although the adult judges were able to identify the emotion of the stimuli successfully, as can be seen from Table 2, preschool children had difficulties doing so. Even the oldest NH children were just slightly better than chance when identifying the emotions, and the differences in identification accuracy across different ages were minimal. Thus, it seems that preschool children have not fully acquired the function of intonation in terms of emotion expression. Seeing that the children failed to recognize the intonations successfully, we conducted a second experiment testing the children's production of the happy and sad intonations.

3. Emotional intonation production experiment

3.1. Participants

Fifty-one children with CI (19 three-year-olds, 17 four-year-olds, 15 five-year-olds) and 17 NH children (10 three-year-olds, 9 four-year-olds, 13 five-year-olds) participated in this experiment. Among these children, 37 also participated in the previous identification experiment. The participants were included in the experiment using the same criteria as described in the identification experiment. The mean hearing threshold of the children with CI was 32 dB HL (range 18-43 dB HL) after implantation.

3.2. Materials

The same sentences listed in Experiment 1 were used.

3.3. Methods

A sentence repetition paradigm was used. The same sentences as described in Experiment 1 were played to the participants via speakers. The participants were asked to repeat the sentence he or she had heard, and the children were instructed to pay attention to the emotion of the sentences, and to use the same emotion when repeating. Two practice trials, one with happy intonation and the other with sad intonation, preceded the experiment for the children to understand the procedure. In the experiment, the ten sentences were presented to the children randomly, with the limitation that no more than two sentences with the same emotion were presented in a row. If a child made mistakes in repetition, he or she was allowed to repeat again.

The productions of the children were recorded and analyzed with Computerized Speech Lab of KayPentax, which included a stand-alone microphone and an analyzing computer. The microphone was placed about 30 cm from the children on a table. For the analysis, the individual sentences produced by each child were saved separately, and then further analysis of f0 as well as intensity features was conducted. For F0 measurements, mean (SD) f0 and f0 range were calculated. For intensity measurements, mean (SD) intensity and intensity range in dB were calculated. We further measured the mean duration of the productions.

For each measurement, an ANOVA was conducted with intonation condition (happy or sad) as within-subject variable, and hearing status (CI or NH) and age (3-, 4-, 5-year-olds) as between-subject variable.

3.4. Results

The descriptive analyses of the children's production of each intonation were presented in Table 3.

Table 3: *Mean (standard deviation, SD) f0, f0 range, intensity, intensity range and duration of the productions of children with CI and NH children.*

	Age	Group	Happy sentence	Sad sentence
f0 (Hz)	3	CI	294.61 (12.91)	244.65 (11.32)
		NH	296.78 (12.76)	251.09 (25.40)
	4	CI	300.48 (13.42)	249.11 (16.55)
		NH	297.66 (11.45)	252.98 (17.19)
	5	CI	299.81 (12.06)	248.37 (13.13)
		NH	296.08 (7.87)	250.10 (14.82)
f0 range (Hz)	3	CI	198.20 (20.17)	205.54 (10.81)
		NH	213.75 (24.24)	211.19 (15.02)
	4	CI	185.15 (29.41)	180.83 (50.44)
		NH	215.79 (17.84)	209.20 (13.52)
	5	CI	212.15 (8.39)	208.33 (21.24)
		NH	214.73 (12.98)	208.56 (8.49)
intensity (dB)	3	CI	53.75 (3.60)	59.28 (5.40)
		NH	60.53 (1.87)	52.62 (2.33)
	4	CI	54.40 (2.25)	59.40 (3.80)
		NH	60.13 (1.51)	54.98 (3.01)
	5	CI	55.00 (3.11)	59.40 (3.90)
		NH	60.65 (3.42)	53.24 (2.51)
intensity range(dB)	3	CI	41.98 (6.97)	39.75 (6.64)
		NH	35.47 (4.94)	34.67 (6.82)
	4	CI	39.87 (6.87)	37.37 (8.25)
		NH	38.21 (8.10)	36.37 (7.78)
	5	CI	41.09	37.80

			(6.39)	(9.61)
		NH	36.83	36.31
			(3.07)	(3.51)
duration (s)	3	CI	1.75	1.96
			(0.47)	(0.45)
		NH	1.88	1.96
			(0.51)	(0.58)
	4	CI	2.03	2.17
			(0.68)	(0.66)
		NH	1.69	1.84
			(0.49)	(0.33)
	5	CI	2.27	2.37
			(1.80)	(1.70)
		NH	1.96	2.07
			(0.93)	(0.86)

For mean f0, the ANOVA showed a significant main effect of intonation type, $F(1, 62) = 510.87$, $p < .001$, where the happy intonations had a higher mean f0 than the sad intonations. No other main effect or interaction turned out to be significant. For f0 range, no main effect or interaction turned out to be significant. For mean intensity, intonation showed a significant main effect, $F(1, 62) = 70.84$, $p < .001$, no other main effect or interaction turned out to be significant. As can be seen in Table 3, the mean intensity was lower for the sad than for the happy intonation. For mean duration, no significant main effect or interaction was found for any of the variables.

These results indicate that, overall, regardless of hearing status and age, the children were able to make distinctions between happy and sad intonation by increasing the mean f0 and intensity of the former.

4. Discussion

The current study tested 3-, 4-, and 5-year-old children with CI and NH children on their identification as well as production of happy and sad intonations. The results showed that both groups of children, regardless of age group, had increased f0 when producing happy compared to sad intonations. Also, regardless of hearing status, the children produced the happy sentences with higher intensity. Interestingly, neither NH children nor children with CI were successful at identifying emotional intonation.

Previous research has shown that the production of functional emotions is largely established by five years [18]. Five-year-olds were able to produce affective prosody (liking versus reservation) with a mean accuracy of 71.4% and to identify these with a mean accuracy of 85.8%. Our results are in a way consistent with these findings in that even the youngest group were able make use of mean f0 when repeating happy versus sad sentences. Increase f0 is also consistent with previously reported acoustic features for happy intonation[19]. Therefore, children use global pitch at the sentence level efficiently from early on, and CI does not seem to hinder children from doing this.

Interestingly and surprisingly, although the children made distinction between happy and sad intonation in the sentence repetition task, they failed to identify the emotional content of the same sentences in the perception task. This holds true for both the NH children and the children with CI. Seeing that the children were able to manipulate the pitch cues in production, it is unlikely that they failed to perceive the pitch difference between the happy and sad intonations. It is more likely that they were not yet able to match pitch patterns to a particular emotional function. In the production experiment, the intonations were elicited by asking the children to repeat the sentences modeled by an adult, thus the children did not produce happy or sad sentences on their own. It is likely that they were able to perceive and produce the pitch pattern of the sentences, yet they were unaware of the meaning of these pitch patterns in terms of emotion. In other words, the children were not able to assign emotional meaning to sentential prosodic patterns. Furthermore, the model sentences were produced by an adult speaker, seeing the difference voice pitch range between adult and children, it is hard to ascertain how well the children were able to perceive and produce a target intonation. It will be informative if future studies can manipulate f0 features (e.g., f0 range, f0 maximum, f0 minimum, etc) of different emotional intonations and compare how NH children and children with CI make use of these different cues when discriminating intonations. In addition, older children should be tested to ascertain at what age children acquire the emotional meaning of different intonations. Whether children with CI go through the same developmental trajectory as NH children in learning emotional intonation needs further investigation. It should be acknowledged that different children participated in the identification and production experiment, and for direct comparison, it would be worthwhile to test same children on these two tasks.

5. Conclusions

Preschool children with CI produced happy intonation by increasing the pitch level and intensity in a similar way to their NH peers. Yet neither NH children nor children with CI were able to match emotional function to sentential prosody at preschool ages.

6. Acknowledgements

We thank the children and the parents who participated in this study. The study was supported by Beijing Social Science Fund, Grant/Award, Number: 18YYC013.

7. References

[1] B. C. J. Moore, “Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants, ” *Otology & Neurotology*, vol. 24, pp. 243–54, 2003.

[2] C. M. McKay, H. J. McDermott and G. M. Clark. “Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees,” *Journal of the Acoustical Society of America*, vol. 96, pp. 2664–73, 1994.

[3] F. G. Zeng. “Temporal pitch in electric hearing,” *Hearing Research*, vol. 174, pp. 101–106, 2004.

[4] K. Gfeller, C. Turner, J. Oleson, X. Zhang, B. Gantz, R. Froman and R. C. Olszewski. “Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise,” *Ear and Hearing,* vol. 28, pp. 412-423, 2007.

[5] Nolan, F. *Intonation*. In B. Aarts, A. McMahon, & L. Hinrichs (Eds.), *The Handbook of English Linguistics* (pp. 385–405). Wiley, 2020.

[6] Bänziger and K. R. Scherer. “The role of intonation in emotional expressions,” *Speech Communication*, vol. 46, pp. 252–267, 2005.

[7] C. Busso, S. Lee and S. Narayanan, “Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection,” *IEEE transactions on audio, speech, and language processing*, vol. 17, 2009.

[8] C. Pereira, and Watson, C. “Some acoustic characteristics of emotions”, *Proceedings of International Conference on Spoken Language Processing*, 1998.

[9] X. Luo, Q. J. Fu, C. G. Wei and L. Cao, "Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users,", *Ear and Hearing*, vol.29, pp. 957–970, 2008.

[10] S. C. Peng, J. B. Tomblin, L. J. and Spencer, R. R. Hurtig. "Imitative production of rising speech intonation in pediatric cochlear implant recipients," *Journal of Speech Lang and Hearing Research*, vol 50, pp. 1210-27, 2007.

[11] S. C. Peng, B. J. Tomblin and C. W. Turner, "production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing," *Ear and Hearing,* vol.29: pp. 336-351, 2008.

[12] D. J. Wang, S., Trehub, A. Volkova, P. van Lieshout, "child implant users' imitation of happy- and sad-sounding speech," Frontiers in Psychology, vol.4, 2013.

[13] M. Chatterjee, A. M. Kulkarni, R. M. Siddiqui, J. A. Christensen, M. Hozan, J. L. Sis, S. A. Dam. "Acoustics of emotional prosody produced by prelingually deaf children with cochlear implants," *Frontiers in Psychology*, vol. 10, 2019.

[14] S. C. Peng, M. Chatterjee and N. Lu, "Acoustic cue integration in speech intonation recognition with cochlear implants," *Trends in Amplification*, vol. 16, pp. 67-82, 2012.

[15] A. Li, F. Qiang, and J. Dang . "Emotional intonation in a tone language: experimental evidence from Chinese", *Proceedings of International Congress on Phonetic Science XVII*, 2011, pp 1198-1201.

[16] H. Wang, Y. Wang and Y. Hu, "Emotional Understanding in Children with A Cochlear Implant", *The Journal of Deaf Studies and Deaf Education*, vol. 24, pp. 65–73, 2019.

[17] V. Ciocca, A. L. Francis, R. Aisha and L. Wong. "The perception of Cantonese lexical tones by early-deafened cochlear implantees," *the Journal of the Acoustical Society of America*, vol. 111, pp. 2250–2256, 2002.

[18] B. Wells, S. Peppé, and N. Goulandris, "Intonation development from five to thirteen," *Journal of Child Language*, vol. 34, pp. 749-778, 2004.

[19] C. Pereira, and Watson, C. "Some acoustic characteristics of emotions", *Proceedings of International Conference on Spoken Language Processing*, 1998