LEXICAL-TONE PRODUCTION IN PRELINGUALLY DEAF MANDARIN-SPEAKING CHILDREN WITH COCHLEAR IMPLANTS*

Jue YU Meiqi ZHANG

School of Foreign Languages, Tongji University, Shanghai, China
erinyu@126.com, maggiezhangnk@163.com

ABSTRACT

The present study examined the production performance of Mandarin lexical tones by prelingually deaf children with cochlear implants (CIs). The main purpose was to obtain a comprehensive understanding about how CI children acquired lexical tones in Mandarin Chinese. The results showed that, compared with normal-hearing (NH) children, CI children followed the universal development order to tone acquisition but displayed more articulatory errors and more diverse error patterns; successfully maintained the pitch contour of each lexical tone but at the cost of pitch height and pitch slope; strategically relied more on "creaky voice" to realize the low-dipping property of Tone 3; acoustically performed better in Tone 1 and Tone 2, yet still categorically overlapped. Significant correlation was found between Mandarin lexical-tone accuracy performance and the length of CI device use but no systematic correlation between F0 measurements and the use of CI devices.

Keywords: Cochlear implants, lexical-tone, accuracy, F0, acoustic distance

1. OBJECTIVES AND BACKGROUND

Previous studies demonstrated that cochlear implants (CIs) enabled a number of severely hearing impaired individuals to access auditory information and effectively improve speech perception as well as speech production skills to some extent (e.g. [1, 24, 29]). However, by directly converting the acoustic signals to electrical impulses, pitch information is not explicitly presented in the electrical stimulations in current cochlear implant technology [17], thus resulting in deficits in the recognition of lexical-tone, speech-intonation, melody and other pitch-related complex acoustic stimuli [9], [19], [28].

In tone languages such as Mandarin Chinese, pitch not only carries the lexical meaning of a syllable but also conveys various expressive functions of a sentence. So much information relies on the pitch variation that it becomes a particular challenge to CI users who speak tone languages. In fact, evidence has shown that their performance on tone perception and production is far from satisfactory.

Mandarin Chinese has altogether four distinct tone patterns. Each is distinguished by its distinctive pitch contour (F0 contour) as well as pitch height (F0 height). In general, Tone 1 is a level tone located in a relative high F0 region; Tone 2 is a rising tone but the onset of the rise occurs in the middle region of the F0 range and ends at a point approaching the F0 height of Tone 1; Tone 3 is a low-dipping tone, typically with a slowly falling dipping and a small rise toward the end (there are other atypical free variants: low falling and low dipping with creaky voice [4], [5]); Tone 4 is a falling tone, starting high and falling to the bottom of the range [29]. Previous studies have shown that Tone 2 and Tone 3 are more confusable than other tonal pairings to native speakers and are also the last to be acquired by Mandarin-speaking children with normal hearing [2], [12], [14].

For Mandarin-speaking children with cochlear implants (CIs), some findings are slightly controversial. For example, some previous studies demonstrated that Tone 1 and Tone 4, lack of dipping property, were mastered earlier than Tone 2 and Tone 3 (e.g. [6, 23, 25]); while some showed that Tone 4 was the most challenging to acquire, due to its shortest duration (e.g. [7]). Moreover, some studies confirmed that age at implantation and the length of CI device use were both found to be significant contributors for tone production and perception in CI children [18]; while some claimed that only age at implantation matters [13].

The present study will examine the tone production by prelingually deaf Mandarin-speaking children with CIs. Meanwhile, the effects of CI devices on the tone production will also be explored. The main purpose is to obtain a comprehensive understanding about how CI children acquire Mandarin lexical tones. Three research questions are proposed:

1. To what extent the children with CIs can accurately produce Mandarin lexical tones, in terms of accuracy and confusion pattern?
2. What are the acoustic differences in tone patterns between CI and NH children, in terms of F0 measurements?

* This paper is sponsored by the Fundamental Research Funds for the Central Universities, China (Grant No. 22120170497)

3. How are the children's post-implant tone production skills associated with the factors of CI devices?

2. METHOD

2.1. Stimuli and procedure

Each participant produced a list of 113 isolated Mandarin monosyllables that contained all the target lexical tones and were phonemically balanced. All the target words were elicited through a modified imitative task and randomised in order not to be predictable. For each target word, the participants first heard an audio prompt naturally produced in a citation form by a female native Mandarin speaker and were then asked to repeat the word once immediately after the audio prime (repetition was allowed). The reasons for using audio prompt instead of an adult clinician's utterance are first to avoid the possible exaggeration by the clinician to facilitate the child's best production [27], and second to maintain the homogeneity for each target stimuli.

2.2. Participants

Altogether, 21 prelingually deaf Mandarin-speaking children with unilateral multi-channel CIs (12 male and 9 female) were recruited and 11 NH children (4 male and 7 female) were used as a baseline. The CI participants were aged between 4.33 to 12.33 years (mean: 9.29 years) at the time of recording and their average length of CI experience was 6.37 years. All of them were non-verbal prior to implantation and were reported to have no visual, developmental or cognitive problems except for a hearing impairment. Most of them received their implants before the age of 4 years (averagely 2.92 years), which was claimed to be a crucial age for tone acquisition by previous studies [19], [31]. After the surgery, all of them received intensive speech and language training at professional rehabilitation centres in Shanghai and mainly used oral communication. Detailed demographic information for CI children is omitted due to the length limitation.

The average age of NH participants were 7.38 years (range from 3.5 to 10.58), chronologically slightly younger than the CI group, but the length of acquiring lexical tones was nearly the same as the CI group. All had been reported without language or speech impairments. All participants use spoken Mandarin at home and in their daily life.

2.3. Data analysis

Both NH and CI participants' speech samples were first transcribed by two native Mandarin speakers trained in phonetics and if there was inconsistency between the two transcribers, the first author double-checked it and made a final decision. The transcribers were instructed to code the tone production as accurate (including acceptable but distorted) or mispronounced. The mispronounced sounds were then analysed for the accuracy and confusion pattern.

After transcription, all speech samples (except those mispronounced ones), were annotated for further acoustic analysis in Praat [3]. Acoustic analysis mainly focused on the measurement of fundamental frequency (F0), as it was proved to be the primary acoustic parameter to characterize Mandarin lexical tones, though other parameters like intensity and duration also counted [11], [16], [29]. The onset and offset of vowels were manually labelled by referring to the change of F2 in the spectrogram [15]. Then, the F0 of the vowel at ten equidistant points were extracted in "ProsodyPro" [32], thus all the F0 measurements (including the minF0, maxF0 and meanF0) of vowels for all stimulus syllables were obtained. Meanwhile, all raw F0 values were converted from Hertz to log-scale 5-level values. Time-normalized F0 contours for Mandarin lexical tones were thus compared between CI and NH children.

The pitch trajectories of all the speech samples obtained were further characterized by two variables: the pitch height and the pitch slope. The acoustic distance between each tone produced by CI participants and the native norm (grand mean of a certain tone category produced by the NH participants) was also calculated based on the above two variables, according to Euclidean Distance (see details in [33]). So the discrepancy of tone production between CI and NH children can be quantified.

A one-way ANOVA was conducted to examine the effect of group on Mandarin lexical-tone production. Pearson correlation analysis was also performed, so the interrelation between the use of CI devices and Mandarin lexical-tone production can be observed.

3. RESULTS AND DISCUSSION

3.1. Accuracy metrics and confusion patterns

In order to answer Research Question 1, the overall articulatory accuracy and the confusion matrix were presented in Table 1, with the NH children as a reference. Not surprisingly, the CI group produced Mandarin tones with significantly lower accuracy than the NH group, especially in Tone 2 ($F(1, 31)=27.572$, $p<0.001$) and Tone 3 ($F(1, 31)=12.468$, $p<0.01$). But it also turned out that most CI children had established a complete tonal inventory before 4 years of hearing age and could produce most lexical tones (except T2: 70.6%) with the mastery level accuracy (above 75% accuracy) [20], [22], which is consistent with the previous studies [19].

A universal principle for the order to tone acquisition indicates that Tone 1 and Tone 4 are mastered earlier than Tone 2 and Tone 3 [23], [26]. The results revealed from the present study were consistent with this tone development order. But besides Tone 3, Tone 2 became another challenging tone for the CI children, for acoustically Tone 2 has also a slight dipping acoustical property and physiologically producing rising tones may be associated with more vocal effort [8], [10], [21].

According to the confusion matrix, the CI children made more diverse error patterns than the NH children. For example, Tone 4 was mispronounced as Tone 1, Tone 2 or Tone 3 in the CI group, while only as Tone 1 in the NH group. Tone 3 was confused with Tone 2 and Tone 4 in the CI group, while mainly confused with Tone 2 in the NH group.

Table 1: Confusion matrix for CI and NH children.

Group	Stimulus	Response (%)			
		Tone 1	Tone 2	Tone 3	Tone 4
CI	Tone 1	**95.55**	1.24	0.78	2.42
	Tone 2	12.66	**70.60**	12.88	3.86
	Tone 3	3.95	9.14	**76.30**	10.62
	Tone 4	5.00	2.38	0.24	**92.38**
	Overall accuracy	**88.10**			
NH	Tone 1	**99.29**	0.14	0.57	0.00
	Tone 2	0.00	**100.00**	0.00	0.00
	Tone 3	0.53	1.06	**97.87**	0.53
	Tone 4	0.53	0.00	0.00	**99.47**
	Overall accuracy	**99.22**			

3.2. Acoustic analysis: F0 measurements

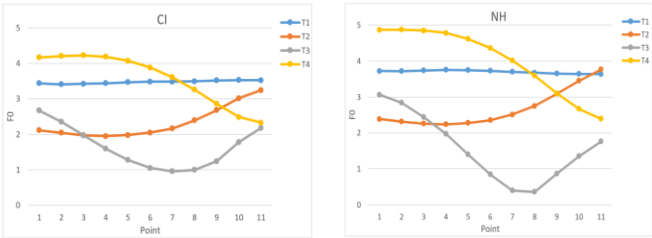
3.2.1. Pitch contour

The typical Mandarin lexical-tone contours produced by the CI children and the NH children were displayed in Figure 1. Generally speaking, the CI children successfully maintained the pitch contour of each tone, but showed great inadequacy in manipulating the pitch height and thus compressed the whole pitch range a lot, especially in the contour tones. In other words, tones with high F0 were not sufficiently high and tones with low F0 were not sufficiently low. For example, Tone 4 is supposed to be a high-falling tone, but in the CI children the onset of the falling occurred in the middle region of the F0 range; Tone 3 is supposed to be a low-dipping tone, but in the CI children, it was not low enough, thus making the dipping property not as prominent as in the NH children.

Moreover, Tone 2 and Tone 3 in the CI children occupied about the same pitch range and had a similar dipping shape, which made the two tones more confusable. So far producing the level tone pattern (Tone 1) appeared to be the easiest for the CI children.

Tone 3, as the most challenging lexical tone for both CI and NH children, was further examined for its other free variants: low falling and low dipping with creaky voice. All these variants were found in both groups, but with different incidence. The results showed that compared with the NH children, the CI children appeared to be more likely to realize the low-dipping property with creaky voice and the average incidence rate was almost 50% of their production. The low-falling pattern occurred with relatively lower incidence in both groups (CI: 19.13%; NH: 22.56%). Statistically, there was a significant difference in the incidence of creaky voice in Tone 3 between the CI group and the NH group ($F(1, 31) = 5.409$, $p < 0.05$).

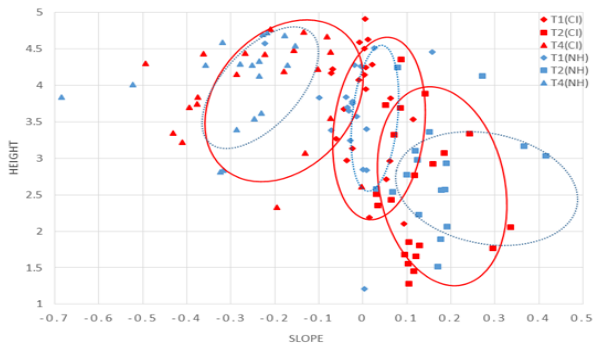
Figure 1: Time-normalized F0 contours for Mandarin lexical tones. (Left: CI group; Right: NH group)



3.2.2. Pitch slope and pitch height

The pitch trajectories of all the speech samples obtained were characterized by two variables: the pitch height and the pitch slope. An overall view of tone distribution was shown in Figure 2 (except Tone 3). Each point in the panel represented the average value of a single participant's production of one tone category. Each ellipse covered more than 90% points of the same tone category.

Figure 2: Tone charts for Mandarin lexical tones. (CI: solid line; NH: dotted line)



Firstly, the CI group showed greater dispersion for each tone category than the NH group, which indicated that individual variability was a persistent characteristic as mentioned in the previous studies. Moreover, the three lexical tones produced by the CI participants overlapped substantially (especially Tone 1 and Tone 2), while the NH participants showed clear differentiation between those tone categories. It suggested that the tone categorical boundaries were still obscure for the CI children. Studies indicated that CI children were very likely to produce most of Mandarin lexical tones as flat tone or monotone [30].

3.2.3. Acoustic distance

Acoustic distance of each lexical tone between the CI children and the NH children was calculated based on the above two variables (pitch height and pitch slope), according to Euclidean Distance (see formula (1)).

$$(1)\quad D = \sqrt{(H_{CI} - H_N)^2 + (S_{CI} \times 10 - S_N \times 10)^2}$$

H_{CI}, S_{CI}, H_N and S_N respectively stand for the pitch height and the pitch slope of the CI children and the native norm (see section 2.3.). The pitch slope, ranging between -0.5 and 0.5, was multiplied by 10 so that it was enlarged to the same range as the pitch height. Tone 3 (hereby referring to the typical low falling-rising), was handled differently from the other tones, as the turning point divided the pitch contour into two parts: the falling part and the rising part. The acoustic distance of Tone 3 was calculated by casting the two parts up with different coefficients. The coefficients were determined by the occurrence of the turning point. In the CI group, the median of the turning point occurred at the 7.5^{th} point, thus the coefficient for the falling part is about 0.68 and 0.32 for the rising part. Smaller distance means higher resemblance between the CI participant's production and the native norm.

The results showed that acoustically Tone 1 was the closest to the native norm, thus the easiest one to grasp for the CI participants; while Tone 3 exhibited the longest acoustic distance, which meant that it was the worst mastered tone category, probably due to its complicated pitch property. In addition, Tone 4 in the CI participants was mastered worse than Tone 2, which was slightly inconsistent with the findings in Section 3.1. The overall articulatory accuracy results indicated that Tone 1 and Tone 4 were articulated with higher accuracy rates than Tone 2 and Tone 3. This inconsistency may be caused by the elicitation and recording procedures. In the present study, audio prompt was used to directly elicit speech production. The participants were required to recognize the lexical tone first and then articulated it. Therefore, the highest error rate of Tone 2 may result from its perceptual difficulty rather than its acoustic property.

3.3. Interactions between CI device use and Mandarin lexical-tone production

The interrelation between tone production performance (in terms of articulatory accuracy and acoustic distance) and the use of CI devices (mainly in terms of age at implantation and length of CI device use) were examined by Pearson correlation analysis. The length of CI device use was found to be the only significant predictor for overall tone articulatory accuracy (r^2=0.514, $p<0.05$). Among all the tone categories, the accuracy performance of Tone 3 and Tone 4 was significantly correlated with the length of CI device use (R_{T3}^2=0.463, $p<0.05$; R_{T4}^2=0.492, $p<0.05$). Tone 1 (most accurately produced tone category) and Tone 2 (worst produced tone category) were neither directly influenced by the length of CI device use nor age at implantation.

As for the acoustic distance, no significant correlation was found between the use of cochlear devices and F0 measurements of Mandarin lexical-tones. The acoustic discrepancy may be relevant to other factors of CI devices, such as the intrinsic spectral resolution, which needs to be further studied.

4. CONCLUSION

The present study mainly discussed the production performance of Mandarin lexical tones in prelingually deaf children with cochlear implants (CIs). The main purpose is to reveal the underlying processes and preferred strategies in Mandarin tone production utilized by CI children.

Some interesting findings are illustrated here: 1) CI children followed the universal development order to tone acquisition and could produce most Mandarin tones with the mastery level accuracy. But compared with NH children, CI children displayed more articulatory errors and more diverse error patterns. 2) CI children were able to maintain the pitch contour of each lexical tone, but showed great inadequacy in manipulating the pitch height and thus compressed the whole pitch range a lot, especially in the contour tones. Moreover, CI children were more inclined to realize the low-dipping property of Tone 3 with creaky voice. 3) Although the acoustic performance of CI children in Tone 1 and Tone 2 was better, the categorical boundaries of the two lexical tones were still obscure. In addition, individual variability was a persistent characteristic as mentioned in most previous studies. 4) With regard to acoustic distance, Tone 1 approached the native norm, while Tone 3 was the most discrepant one for CI children. 5) Only the length of CI device use was found to be a significant contributor for accuracy performance of CI children, but no potential contributing factors were found for their acoustic performance. Further studies are targeting to take more spontaneous data to verify these results, including tonal production in continuous speech.

5. REFERENCES

[1] Blamey, P. J., Barry, J. G., Jacq, P. 2001. Phonetic inventory development in young cochlear implant users 6 years post operation. *Journal of Speech, Language, and Hearing Research*, 44, 73–79.

[2] Blicher, D. L., Diehl, R., Cohen, L. B. 1990. Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics* 18, 37-49.

[3] Boersma, P., Weenink, D. 2004. Praat (Version 4.3). Amsterdam, the Netherlands: Institute of Phonetic Sciences, University of Amsterdam.

[4] Chao, Y.R. 1956. Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. *For Roman Jakobson*, 52-59.

[5] Chao, Y.R. 1968. *A grammar of spoken Chinese*. University of California Press, Berkeley.

[6] Chen, Y., Wong, L. L. N., Chen, F., Xi, X. 2014. Tone and sentence perception in young Mandarin-speaking children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, 78, 1923–1930.

[7] Cui, L. 2011A Study on Tone Perception and Production in Mandarin-Speaking Children with cochlear implants. (Doctoral dissertation, East China Normal University)

[8] Fon, J. 1997. What are tones really like? An acoustic-based study of Taiwan Mandarin Tones. National Taiwan University: Unpublished M. A. Thesis.

[9] Han, D., Liu, B., Zhou, N., et al. 2009. Lexical tone perception with HiResolution and HiResolution 120 sound-processing strategies in pediatric Mandarin-speaking cochlear implant users. *Ear and Hearing*, 30, 169–177.

[10] Ho, A. T. 1976. The acoustic variation of Mandarin tones. *Phonetica* 33, 353-367.

[11] Howie, J. M., Howie, J. M. 1976. *Acoustical studies of Mandarin vowels and tones* (Vol. 6). Cambridge University Press.

[12] Huang, T. 2001. The interplay of perception and phonology in Tone 3 sandhi in Chinese Putonghua. *OSU forking Papers in Linguistics* 55: 23.

[13] Lee, K. Y., Van Hasselt, C. A., Tong, M. C. 2010. Age sensitivity in the acquisition of lexical tone production: Evidence from children with profound congenital hearing impairment after cochlear implantation. *Annals of Otology, Rhinology & Laryngology*, 119(4), 258-265.

[14] Li, C. N., and Thompson, S. A. 1977. The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4, 185-199.

[15] Ling, B., Liang, J. 2017. Focus encoding and prosodic structure in Shanghai Chinese. *Journal of the Acoustical Society of America,* 141(6), EL610.

[16] Moore, C.B., and Jongman, A. 1997. Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America* 102, 1864-1877.

[17] Moore, B. C. 2003. Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants. *Otology & neurotology*, 24(2), 243-254.

[18] Peng, S.C., Tomblin, J. B., Cheung, H. et al. 2004. Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. *Ear and Hearing*, 25, 251–264.

[19] Peng, S. C., Lu, H. P., Lu, N. et al. 2017. Processing of acoustic cues in lexical-tone identification by pediatric cochlear-implant recipients. *Journal of Speech, Language, and Hearing Research*, 60(5), 1223-1235.

[20] Sander, E. K. 1972. When are speech sounds learned?. *Journal of speech and hearing disorders*, 37(1), 55-63.

[21] Shih, C. 1988. Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory* 3, 83-109.

[22] Shriberg, L. D., Kwiatkowski, J. 1994. Developmental phonological disorders I: A clinical profile. *Journal of Speech, Language, and Hearing Research*, 37(5), 1100-1126.

[23] Su, A. 1985. The Acquisition of Mandarin Phonology by Taiwanese Children. Fu-Jen Catholic University: Unpublished M. A. Thesis.

[24] Svirsky, M. A., Chin, S. B. 2000. Speech production. In S. B. Waltzman & N. L. Cohen (Eds.), Cochlear Implants. New York: Thieme Medical Publishers

[25] Tao, D., Deng, R., Jiang, Y., Galvin, J. J., III, Fu, Q.-J., & Chen, B. 2015. Melodic pitch perception and lexical tone perception in Mandarin-speaking cochlear implant users. *Ear and Hearing,* 36, 102–110

[26] Tse, J. K. 1978. Tone acquisition in Cantonese: a longitudinal case study. *Journal of Child Language*, 5, 191–204.

[27] Tye-Murray, N., Kirk, K. I. 1993. Vowel and diphthong production by young users of cochlear implants and the relationship between the phonetic level evaluation and spontaneous speech. *Journal of Speech, Language, and Hearing Research*, 36(3), 488-502.

[28] [28] Wang, S., Liu, B., Dong, R., et al. 2012. Music and lexical tone perception in Chinese adult cochlear implant users. *The Laryngoscope*, 122(6), 1353-1360.

[29] Wu, Z. J. 1986. The spectrographic album of mono-syllables of standard Chinese. *Beijing, China: Chinese Academy of Social Sciences*.

[30] Xu, L., Li, Y., Hao, J., Chen, X., Xue, S. A., Han, D. 2004. Tone production in Mandarin-speaking children with cochlear implants: a preliminary study. *Acta oto-laryngologica*, 124(4), 363-367.

[31] Xu, L., Chen, X., Lu, H., Zhou, N., Wang, S., Liu, Q., et al. 2011. Tone perception and production in pediatric cochlear implants users. Acta Oto-Laryngologica, 131(4), 4.

[32] Xu, Y. 2005-2011. ProsodyPro (Version 5.7.8.1). [A Praat script for large-scale systematic analysis of continuous prosodic events].

[33] Zhang, K., Peng, G., Li, Y., Minett, J. W., Wang, W. S. 2018. The effect of speech variability on tonal language speakers' second language lexical tone learning. *Frontiers in psychology*, 9, 1982.