DURATION AND AMPLITUDE OF TONE PRODUCTION IN CHILDREN WITH COCHLEAR IMPLANTS

YANG Jing, XU Li

Abstract The purpose of the present study was two-fold: (1) to examine whether Mandarin-speaking children with cochlear implants (CIs) show distinctive durational and amplitude features for the four lexical tones in their tone production; (2) to compare the durational and amplitude patterns of Mandarin lexical tones in monosyllables produced in citation form by children with CIs with those by age-matched normal-hearing (NH) children. The participants included 14 prelingually deafened Mandarin-speaking children with CIs and 14 NH children, all aged between 2.9 and 8.3 years old. Each participant produced five CV syllables (fa /fa/, fu /fu/, pi /pʰi/, xu /cy/, ke /kʰ e/) in four tones through a tone drill activity. The vowel duration and root-mean-squared amplitude values at nine equidistant time locations over the vowel duration were obtained. The results revealed that the children with CIs produced distinctive durational and amplitude features for the four lexical tones. Their durational pattern and amplitude contours were highly similar to the NH children on tones 1, 2 and 4 but differed from the NH children on tone 3. These findings suggested that the children with CIs may utilize the secondary features of tone duration and amplitude to realize tonal differences.

Key Words Cochlear implants, Mandarin-speaking children, Tone duration, Tone amplitude

1. INTRODUCTION

Tonal languages adopt pitch change to differentiate lexical or grammatical meanings. Mandarin is the most widely used monosyllabic tonal language. Each stressed syllable in the citation has one of the four lexical tones, namely tone 1, tone 2, tone 3, and tone 4. So far, a large number of studies have examined the acoustic correlates of Mandarin tones [1, 2, 3, 4, 5]. It has been widely acknowledged that the primary acoustic feature for tonal identity lies in the fundamental frequency (F0). The pitch patterns corresponding to tone 1, tone 2, tone 3, and tone 4 are high-level, mid-rising, low-dipping, and high-falling.

In addition to the F0 pattern, previous studies revealed that Mandarin tones also differ in their intrinsic duration and overall amplitude contour of the syllable rhymes that carry the tones [3, 6, 7, 8]. Many studies reported that in isolated Mandarin words, tone 3 has the longest duration and tone 4 has the shortest duration. Tone 1 and tone 2 just fall in between [3, 8]. Ho [3] reported that the order of tone duration in Mandarin isolated monosyllables was tone 3 > tone2 > tone1 > tone 4. Chang [9], Lee and Hung [10] confirmed this durational pattern in isolated Mandarin lexical tones. The amplitude contour varies in these four Mandarin tones. Ho [3] described the amplitude contours of Mandarin tones as level/level-falling, rising/level, double-peak, and falling/level-falling, corresponding to tones 1, 2, 3, and 4, respectively. Lin [7] divided Mandarin tone amplitude into five categories: level, higher at the onset, higher at the offset, higher in the middle, and double-peak. He summarized that tones 1, 2, and 4 normally
show an amplitude contour with a higher onset or a higher offset, while tone 3 normally presents a double-peak amplitude contour. Although previous studies might show inconsistent categorization or description of the amplitude shapes, most of them consistently demonstrated a positive correlation between the pitch contours and amplitude contours in Mandarin lexical tones of isolated monosyllables [8, 11, 12].

For normal-hearing listeners, the primary perceptual cue for tone identification lies in pitch-related properties [6, 13, 14]. Other types of co-occurring features such as duration or amplitude play a very limited role in tone perception [7, 12, 15]. In comparison to normal-hearing listeners who receive a full range of auditory information, hearing-impaired listeners with cochlear implants only obtain partial auditory information. The acoustic signals conveyed through the prosthesis are vocoder processed, in which pitch information is poorly coded. In this scenario, CI users, in general, demonstrated poorer recognition of pitch-related information (e.g., tones) than normal-hearing (NH) listeners [16, 17, 18, 19, 20]. In addition to the overall poor tone recognition performance, the cue-weighting pattern in tone recognition between the CI users and NH listeners was different [21, 22].

The deficits of pitch coding in CIs are reflected not only in speech perception, but also in speech production. Researchers found that children with CIs, in general, demonstrated monotonous pitch contours and produced all four tones with substantially lower accuracy than children with NH. Children with CIs produced less definable F0 contours and less separable F0 categories than the children with NH [23, 24, 25, 26, 27, 28]. While these studies focused on the F0 pattern of tone production, very little research has been conducted to examine the duration or amplitude of Mandarin lexical tones produced by children with CIs. Given that temporal and amplitude information is relatively well-maintained in CIs, we wonder whether children with CIs will show a NH-like pattern on the durational and amplitude features in their tone production. Deroche et al. [22] found that children with CIs exaggerated the durational contrast for tone 1 and tone 4 in comparison to the children with NH. Further, the authors reported that the children with CIs who showed greater reliance on duration in tone perception also showed poorer control in voice pitch in tone production. Extending from the Deroche et al. [22] study that only focused on tone 1 and tone 4, the present study was conducted to examine the duration and amplitude of all four lexical tones in citation-form monosyllables produced by Mandarin-speaking CI children with reference to age-matched NH children. Specifically, there are two related research questions: (1) whether the children with CIs show distinctive durational and amplitude features for the four lexical tones in their tone production; (2) whether the children with CIs show duration and amplitude patterns of lexical tones in citation-form monosyllables similar to agematched NH children.

2. METHODS

2.1 Participants

A total of 28 Mandarin-speaking children were included in the present study. All participants were from previous studies [25, 26]. Fourteen of them (8 boys and 6 girls) were prelingually deafened children with CIs, aged between 2.9 and 8.3 years old (Mean = 5.16 yrs, SD = 1.80 yrs). Fourteen age-matched children with NH (6 boys and 8 girls) (Mean = 5.15 yrs, SD = 1.77 yrs) were recruited to serve as the control group. All children were recruited in Beijing, China and raised by Mandarin-speaking parents or caregivers. The 14 children with CIs were nonverbal
prior to implantation and received unilateral cochlear implantation between 1.16 and 7.09 years of age (Mean = 3.43 yrs, SD = 1.91 yrs). The length of device use varied between 0.31 and 2.60 years (Mean = 1.73 yrs, SD = 0.76 yrs). They all received post-implantation training focusing on basic language and speech skills at a professional rehabilitation center in Beijing. The 14 children with NH reported having no speech-language or hearing impairments.

2.2 Data collection

Five CV syllables fa /fa/, fu /fu/, pi /p/i/, xu /cy/, ke /k/e/ from each participant were used for the present study. These five syllables all had voiceless obstruents at the syllable onset and contained the five basic monophthongal vowel phonemes /a, i, u, y, ye/ in Mandarin. Each participant was recorded separately in a quiet room. A tone drill activity was used to collect speech samples. The experimenter produced each syllable in tone 1 and the participants were asked to follow the experimenter and produce all four tones for the target syllable in consecutive order. Each tone of individual target syllables was produced once. All speech samples were collected through a digital recorder with a 44.1 kHz sampling rate and a 16-bit quantization rate. Due to missing tokens and peak-clipping issue, a total number of 469 tokens (223 from the children with CIs and 246 from the children with NH) were used for further analysis. Note that all measurable tokens were included, regardless of whether the token was produced correctly or not.

2.3 Acoustic measurement

Although the domain of tones remains controversial, the present study followed Howie’s [2] notion that syllable rhyme carries the tone. Because all five tested syllables were in the form of CV structure, the tone duration was defined as the duration of the vowel in each word. Specifically, the onset and offset of the vowels were determined by a visual check of waveform accompanied by a spectrogram display. The vowel onset was set at the start of vowel periodicity, and the vowel offset was set at the end of vowel periodicity with both F1 and F2 present.

The amplitude contour was extracted at nine equidistant time points (10% - 20% - 30% - 40% - 50% - 60% - 70% - 80% - 90% point) over the course of vowel duration. In particular, a root-mean-squared (rms) value was obtained at each time location. The rms value was calculated for a set of 10 ms vowel portions with the midpoints located at these nine time points, using a custom-written MATLAB program. If peak clipping occurred in certain tokens, those tokens were excluded from further analysis. Because the recording environment was not consistently or strictly controlled across all participants to better compare the amplitude contours across different speakers and maintain the relative pattern of the four tones in each speaker, the extracted rms amplitude was normalized for individual participants following the formula:

$$A'_{n[s]} = A_{n[s]} - MEAN_{n[s]}$$ (1)

Where $A'_{n[s]}$ is the normalized rms amplitude value for point n of subject S, $A_{n[s]}$ is the original rms amplitude value for the point n of subject S, and $MEAN_{[s]}$ is the grand mean of rms amplitude values for all nine points of four tones from the five tokens of subject S.

Based on the normalized amplitude, the sum of squared differences (SSD) was calculated to evaluate the similarity between the amplitude contours of each tone between the NH and CI children. This value is defined as the sum of the squares of the distance between two corresponding points in the examined groups for a given tone.

$$SSD = \sum_{n=1}^{i} (A_{xn} - A_{yn})^2$$ (2)

Where $A_{xn}$ is the rms amplitude value of
point \( n \) for a given tone in group \( x \) and \( A_{yn} \) is the rms amplitude value of point \( n \) for the same tone in group \( y \).

The F0 data from the 28 participants has been analyzed and reported in previous studies [26]. Specifically, the F0 of an articulated vowel was extracted using an autocorrelation algorithm implemented in a custom-written MATLAB program. Errors with automatic extraction were fixed with a manual check. Given that the absolute duration of the articulated vowel in each token of each participant was different, the number of F0 data points in each token varied. Therefore, the interpolation method was used to make F0 data comparable on the same normalized duration scale. F0 values selected from the same nine equidistant points over the course of vowel duration were used to plot F0 contours.

3. RESULTS

Fig. 1 shows the durations of the four tones in the NH and CI children. The children with CIs produced all four tones with significantly longer durations than the NH peers. In addition, the CI group demonstrated a different durational pattern across four lexical tones in comparison to the NH group, which was mainly caused by tone 3. While the NH children produced tone 3 and tone 4 shorter than tone 1 and tone 2, the children with CIs produced tone 4 significantly shorter than the other three tones, and tone 3 slightly longer than tone 1 and tone 2. The longest duration of tone 3, and the shortest duration of tone 4 in the children with CIs matched the durational pattern of four lexical tones in isolation reported in previous studies.

To better compare the tone duration in CI and NH groups, a repeated-measures ANOVA was conducted with the hearing group as the between-subject factor and tone type as the within-subject factor. The results revealed a significant main effect of group \( (F(1, 26) = 14.549, p = 0.001) \), tone \( (F(3, 78) = 12.054, p < 0.001) \) and a significant interaction effect of group by tone \( (F(3, 78) = 5.884, p = 0.001) \). The significant interaction effect suggested that the duration at pattern of the four tones in the children with CIs differed from that in the children with NH. To compare the duration pattern of four tones in each group, a repeated-measures ANOVA was conducted for the CI and NH groups, respectively, with the tone type as the within-subject factor. The post-hoc contrast comparison revealed that for the CI children, tone 1, tone 2, and tone 3 were all significantly longer than tone 4 (all \( p \leq 0.006 \)), but none of the tone contrast among tones 1, 2, and 3 were significantly different. For the NH children, tone 1 and tone 2 were significantly longer than tone 3 and tone 4 (both \( p \leq 0.001 \)).

![Figure 1](image-url)

**Figure 1** Box plot of tone duration for the four lexical tones in isolated monosyllables produced by the NH and CI children. Each data point represents one subject.

Fig. 2 shows the amplitude contours of the four lexical tones in individual CI and NH children and the comparison of group mean amplitude contour and F0 contour between the two groups. The F0 data was adopted from a previous study [26]. The comparison of group mean F0 data revealed that the CI children approximated the NH peers in tone 1,
but they deviated from the NH peers in the direction and/or absolute F0 height for the other three tones. Note that the NH children produced tone 3 with a falling F0 contour, not a typical low-dipping contour.

![Figure 2](image-url)

Figure 2  Top three rows: Amplitude contours of the four lexical tones in tested monosyllables by the NH and CI children and the group mean data. Each line in the top two rows represents amplitude contour data from one subject. The rms amplitude was extracted at nine equidistant time points (10% – 20% – 30% – 40% – 50% – 60% – 70% – 80% – 90%) over the vowel duration and normalized. Bottom row: Group mean F0 contour for each lexical tone. The F0 data were adopted from \[26]\.

As for the amplitude contour, it can be observed that the variability of the amplitude contours is larger for the CI than the NH groups for all tones except tone 4. In the meantime, the two groups of children demonstrated a similar overall pattern of amplitude changes for tones 1, 2, and 4 but not for tone 3. In particular, both groups of children showed a falling amplitude contour for tones 1 and 4, and a slightly rising amplitude contour for tone 2. For tone 3, the majority of NH children presented a falling contour. By contrast, some children with CIs showed a double-peak amplitude contour while the others showed a relatively flat amplitude contour. The SSD value between the CI and NH children was 36.7 for tone 1, 1.8 for tone 2, 149.9 for tone 3, and 9.1 for tone 4. The very small SSD value for tone 2 and tone 4 reflected the highly compatible and overlapped group mean amplitude contours in the NH and CI groups for these two tones. The larger SSD value in
tone 1 was caused by the consistently higher amplitude of NH group than the CI group even though the shape of the contours for this tone in these two groups was highly similar. The largest SSD value for tone 3 reflected the substantial difference in the shape of the amplitude contours between the two groups.

A repeated-measures ANOVA was conducted to compare the amplitude data of each tone between the NH and CI groups, with the nine measurement points as the within-subject factor and hearing group as the between-subject factor. The results showed that for all four tones, there was a significant main effect of the measurement point (all $p < 0.05$). As the amplitude changes as time elapses, the significant effect of the measurement point was expected and would not be discussed further. As for the group effect and group by point interaction effect, the results showed a significant group effect only for tone 1 ($F(1, 25) = 8.826, p = 0.006$), and a significant group by point interaction effect only for tone 3 ($F(8, 208) = 7.416, p = 0.001$). These results suggested that the CI and NH children showed similar magnitudes of amplitude and contour pattern for tone 2 and tone 4. However, the two groups differed in the magnitude of amplitude in tone 1 and the pattern of amplitude contour in tone 3.

![Figure 3](image.png)

**Figure 3** The histogram of correlation coefficients of F0 and amplitude for each token of tone 3 in the NH and CI children

Previous studies suggested that normal adult speakers produced positively matched F0 and amplitude for lexical tones [8, 11, 12]. In the present study, we observed that the NH children produced tones 1, 2, and 4 with similar F0 and amplitude contours as described in previous studies and highly compatible F0 and amplitude contours for tone 3. The CI group showed compatible group mean F0 and amplitude contours as the NH group for tones 1, 2, and 4 but not for tone 3. To further examine the relationship between the F0 and amplitude, the correlation coefficient between F0 contour and amplitude contour was calculated for each token of tone 3 in the CI and NH group, respectively. If the two contours move in the same direction, the coefficient value should be a positive value close to 1. Among the 70 tokens for each group, due to the missing data for F0 and/or amplitude, there were 34 tokens in the CI group and 36 tokens in the NH group with valid F0 and amplitude data. As shown in the distribution of correlation coefficients for each group (Fig. 3), the NH group showed highly matched F0 and amplitude contours, but the CI group produced a large number of tokens with unmatched F0 and amplitude contours.

### 4. DISCUSSION AND SUMMARY

The purpose of the present study was to investigate whether the children with CIs produce distinctive durational features and amplitude contours among the four lexical tones and whether they follow the NH peers on the patterns of these features in tone production. Our results showed that the CI children produced distinctive durations and amplitude contours for the four tones, which are similar to the canonical pattern reported in Mandarin-speaking adults [6, 7, 8, 9]. For tone durations, the CI children tended to produce tone 3 with the longest duration and tone 4 with the shortest duration. Regarding the amplitude contour, the CI children produced tone 1 and
tone 4 with flat or flat-falling amplitude contours, tone 2 with a rising contour, and tone 3 with a double-peak contour. Although the CI children produced tone duration and amplitude with patterns similar to adult norms reported in previous studies, they differed from the age-matched NH children in the present study on both features. At first, the CI children consistently produced all four lexical tones with significantly longer durations than the NH children. Previous literature reported that children with CIs tended to produce longer durations for speech segments than NH peers [29, 30]. Lexical tones are carried by syllable rhymes, which are mainly composed of vowels. The longer tone durations in the CI children than the NH peers provided additional evidence supporting the findings of longer speech segments in children with CIs. Additionally, the CI children demonstrated different durational and amplitude patterns across the four lexical tones in comparison to the NH peers, which was mainly caused by tone 3. The children with CIs produced tone 3 with a double-peak amplitude contour and a relatively longer duration. By contrast, the NH children produced tone 3 with a falling amplitude contour and shorter duration than tone 1 and tone 2.

It is noteworthy that the duration and amplitude patterns of the NH children in the present study did not follow the patterns of adults reported in previous studies. This may be related to the tone drill activity used in the present study in which the participants were required to produce the four tones of a given monosyllable in consecutive order. Most of the NH children did not produce a full version of tone 3 with the rising part because they did not pause enough between every two tokens and the token was not produced in citation form in a truly isolated environment. The CI children were recorded with the same procedure but they produced the duration and amplitude features with similar patterns to normally-speaking Mandarin adults. We speculate that this might be associated with the slower articulatory movements in CI children relative to NH children. Our earlier studies revealed that CI children produced vowels with longer duration than NH children [29]. As the domain of tone is over the syllable rhyme in which the nucleus vowel mainly carries the tone, the longer vowel duration corresponded to the longer tone duration in the CI children relative to the NH children. Accompanying with the slow articulatory movement, the CI children were less likely to make the transition from one token to the next token as fast as the NH children. On the other hand, although pitch information is poorly transmitted through CI devices, the temporal and amplitude information is relatively well-maintained. Therefore, the CI children were likely to perceive the differences in duration and amplitude among the four lexical tones through the language input. Previous studies reported that children with CIs could hardly produce distinctive F0 contours as children with NH did, but they relied more on the secondary cues to perceive tonal identity. Because of this, some CI children might fully produce tone 3 with a complete falling-rising amplitude contour to make distinguishable tones in their speech production.

Many previous studies revealed a positive correlation between F0 contour and amplitude contour in normally-speaking Mandarin adults [8, 11, 12]. In the present study, we observed well-matched F0 contour and amplitude contour in the NH children. However, the CI group presented unmatched F0 and amplitude contours for many tokens. These observations indicate that the amplitude contour may not necessarily be positively correlated to the F0 contour, which is consistent with the finding of Kuo et al. [31] who found that the amplitude contour of a tone and the F0 contour of the same tone were not always most highly correlated.
In general, the present study revealed that the Mandarin-speaking CI children produced distinctive durational and amplitude features for the four lexical tones. Their durational pattern and amplitude contours differed from the NH children mainly in tone 3, but they followed the patterns of these two features in normal adult speakers reported in previous studies. The difference between the CI and NH groups might be associated with the limitation of the elicitation procedure in this study. The NH children did not really produce the monosyllabic words in a true isolation environment, which might cause systematically biased tone production in the two groups. In the future, a more strictly controlled procedure should be used to collect the production data. Some earlier studies suggested that the phonetic features of Mandarin tones in spontaneous, connected speech become less distinctive [9, 32]. Correspondingly, children with CIs may show more difficulty in producing tonal differences in spontaneous speech. For future studies, we shall recruit a larger size of speakers producing isolated monosyllabic and spontaneous speech to better examine the tone production in CI users. Additionally, the acoustic analysis of tone production shall be connected with the perceptual data of the four tones or tone contrasts in children with CIs to examine whether and how different acoustic cues (i.e., F0, duration, and amplitude) are perceived in this population.

REFERENCES

rehabilitation following cochlear implantation in children. *Acta Oto-Laryngologica* 129 (2), 218–221.


**YANG Jing PhD** She is an associate professor in the Department of Communication Sciences and Disorders at the University of Wisconsin-Milwaukee. Her research focuses on the speech development of bilingual children and children with cochlear implants. E–mail: jyang888@uw.edu

**XU Li MD, PhD** He is a professor in the Division of Communication Sciences and Disorders at Ohio University. His current research focuses on speech perception in noise, tone perception and production in children with cochlear implants. E–mail: xul@ohio.edu