The Disyllabic Tone Production and Tone Context Effect in Mandarin-speaking Children with Cochlear Implants

Jingwen Cheng¹, Yingming Gao², Yuchen Yan³, Xiaoli Feng¹, Binguai Lin³, Jinsong Zhang¹

¹School of Information Sciences, Beijing Language and Culture University, Beijing, China
²School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China
³Smart Platform Product Department, Tencent Technology Co., Ltd, Shenzhen, China

cchengjingwen98163.com, yingming.gao@bupt.edu.cn, yanyuchen1999@foxmail.com, fengxiaoli314@163.com, binghuailin@tencent.com, jinsong.zhang@blcu.edu.cn

Abstract

This study investigated the production of tones in disyllabic words by Mandarin-speaking children with cochlear implants (CIs). The speech materials consisted of 141 disyllabic and 153 monosyllabic words that were produced by 45 children with CIs. We first compared the accuracy and error patterns of disyllabic tone pairs, and then analyzed the production performance of lexical tones with regard to different contexts and positions. The results showed that: 1) The mean accuracy rate of 20 disyllabic tone pairs was 60.59%, and the tone pairs containing tone 2, tone 3, or tone 5 were produced less correctly by children with CIs. Complex error patterns were found in the present study. 2) The lexical tones had lower accuracy and different error patterns in the first syllable position of disyllabic compared with when they were in the second syllable position of disyllabic and in the monosyllable. Among the four lexical tones, tone 2 and tone 3 were more affected by the contexts and positions. This study sheds light on tone acquisition of the disyllabic words for children with CIs.

Index Terms: cochlear implant, Mandarin tones, tone context, hearing-impaired children

1. Introduction

According to the WHO’s report on deafness and hearing loss in 2021, there are 34 million children who require rehabilitation to address their ‘disabling’ hearing loss. It is well-known that hearing loss or impairment in children can delays their language and speech acquisition. Cochlear implants (CIs) are a well-known and accepted treatment method for the patients [1], which can restore and improve their perception ability to some extent. However, due to the restricted number of channels, the speech frequencies delivered by CIs are still not as effective as those by a normal ear, e.g., the fundamental frequency [2, 3]. This limitation makes it difficult for children with CIs to perceive pitch information, especially when acquiring a tonal language [4, 5, 6].

Mandarin Chinese is a tonal language, and the variations of the tone patterns of each syllable convey different lexical meanings. Compared with isolated monosyllabic words, disyllabic words account for the majority of commonly used words in Mandarin Chinese [7]. There are four lexical tones and one neutral tone in the tone system, which form four monosyllabic tones and 20 disyllabic tone pairs (4 lexical tones × 4 lexical tones, and 4 lexical tones × 1 neutral tone). Based on the pitch contours, the four lexical tones can be characterized as high level for tone 1 (T1), raising for tone 2 (T2), low dipping for tone 3 (T3) and falling for tone 4 (T4). The neutral tone is usually referred to as tone 5 (T5), and it only appears after a full lexical tone.

Previous studies on the tone production of Mandarin-speaking children with CIs have mostly focused on monosyllabic tones. The production accuracy and error patterns were examined based on the normal-hearing listener’s judgement. Peng et al. studied the production accuracy of four monosyllabic lexical tones [8]. The results showed that T1 and T4 were produced more correctly than T2 and T3. Similar results were also found in some other studies [9, 10, 11]. With regard to the error patterns, many studies found that T2, T3, and T4 were frequently mispronounced as T1 [9, 10, 11], while another study based on the subjects with a longer CI experience found that the predominant errors of four lexical tones were the confusion between T2 and T3 [12].

Tone’s realization is affected by the phonological context. For example, T3 is produced as T2 before another T3 (tone 3 sandhi rules). In disyllabic contexts, the F0 contours present more complex patterns than in isolated monosyllables. Children with CIs could face more challenges in the production of disyllabic tones than monosyllabic tones. Compared to a large number of studies on monosyllabic tone production, there are fewer studies that have investigated the disyllabic tone production in Mandarin-speaking children with CIs [13, 14, 15]. Li et al. studied the production accuracy and error patterns of 20 disyllabic tone pairs produced by 68 children with CIs [13]. The speech materials consisted of 40 disyllabic words. The results showed that the tone pairs containing T2 or T3 had lower accuracies than the tone pairs containing T1 or T4. As for the error patterns, the researchers found that many disyllabic tone pairs (e.g., T1T2, T2T2) were commonly substituted by T1T1.

More recently, Zhang et al. studied the production accuracy of four lexical tones in disyllabic and monosyllabic words [14]. The results showed that the four lexical tones had significantly lower accuracy in the disyllabic context than in the monosyllabic context. Xiao et al. studied the production of T2 in disyllabic words by 30 children with CIs. The results showed that T2 in the first syllable position of the disyllable was produced less correctly than in the second syllable position [15]. It seems that both contexts and syllable positions had an effect on the tones’ production. Tones may have a different production performance in the monosyllables and in the first syllable (S1) position and second syllable (S2) position of the disyllables. To date, no study has systematically examined the context and syllable position effects on the tone production in children with CIs.
The aim of the present study is to investigate the disyllabic tone production and the tone context and position effects in children with CIs. First, the production accuracy and error patterns of 20 disyllabic tone pairs were analyzed in detail. In contrast to the small number of materials used in the previous study [13], this study used a larger number of speech materials (141 disyllabic words). Then, the lexical tones’ performances were further examined in the monosyllable context and in the first syllable position and the second syllable position of the disyllable context.

2. Data

2.1. Subjects

This study included 45 Mandarin-speaking children (25 females and 20 males) who received cochlear implantation. Their mean chronological age at the time of recording was 6; 0 (years; months), ranging from 2; 7 to 11; 10. Their mean age at cochlear implantation was 4; 7, ranging from 1; 3 to 11; 7. And the mean duration of cochlear implants use was 1; 5, ranging from 0; 2 to 2; 11. All children were recruited from Handan, Hebei province. They had received speech and language training at professional rehabilitation centers.

2.2. Materials and data collection

The speech materials consisted of 141 disyllabic words and 153 monosyllabic words. Twenty possible tonal combinations were covered in the disyllable, and four lexical tones were covered in the monosyllable. Most of the selected words are commonly used words in Mandarin Chinese [16], which children are more familiar with. For the recording of each word, the children first heard an audio prompt that was produced by a female native Mandarin speaker, and then they were asked to repeat the word. Some words were reread or missing for some participants. Totally, there were 5643 available utterances for the disyllable and 7375 available utterances for the monosyllable.

2.3. Annotation

Ten trained native Mandarin speakers were asked to annotate the pronunciation. All the annotators reported themselves to have normal hearing. Most of the annotators are students with a major in phonetics. Tones were transcribed using numbers (‘1’, ‘2’, ‘3’, ‘4’, and ‘5’ for tone 1, tone 2, tone 3, tone 4, and tone 5, respectively). This annotation work was done with the software “Praat” [17]. In order to check transcription consistency, 698 speech records were selected to assess the agreement between annotators. The selected speech records were randomly divided into several sets. Each set was transcribed by three to four annotators. The consistency between each pair of annotators was calculated. The mean consistencies were 80.58% and 84.11% for disyllabic tone and monosyllabic tone, respectively. The low consistencies are due to the fact that the tones produced by the children with CIs are sometimes very atypical, leading to the labeling difficulty even for the trained annotators.

3. Results

3.1. The accuracy and error patterns of disyllabic tones

3.1.1. Accuracy

Figure 1 presents the accuracy rates of 20 disyllabic tone pairs produced by all the subjects. The mean accuracy rate of 20 disyllabic tone pairs was 60.59% (SD: 12.2%). The tonal combinations of T1 and T4 had the highest accuracy, including T1T4 (81.09%), T4T1 (73.3%), T4T4 (71.51%), and T1T1 (69.89%). The tone pairs containing T2 or T3 usually had lower accuracy. Except for T4T5, the other tone pairs containing the neutral tone had the lowest accuracy, which were T1T5 (51.81%), T2T5 (34.32%), and T3T5 (30.86%). It indicated that the tone pairs containing T2, T3 or T5 were relatively difficult to acquire by Mandarin-speaking children with CIs. The results are consistent with the previous study [13].

3.1.2. Error patterns

The pronunciation error patterns of 20 disyllabic tones pairs are presented in Figure 2. The confusion matrix shows the percentage of each target tone pair that was mispronounced as other tone pairs. The disyllabic tone pairs usually had one tone that was mispronounced. For the tone pairs with T1 in the second syllable, the first tone tended to be mispronounced, while the second tone was produced correctly. For example, T2T1 was often substituted by T3T1 and T4T1. Similarly, the tone pairs with T4 in the second syllable tended to substitute the first tone with T4, T3, or T1. For instance, T3T4, T2T4, and T1T4 were often substituted by T4T4. Concerning the tone pairs with T2 or T3 in the second syllable, they were showed more diverse error patterns. Both the first tone and second tone tended to be mispronounced in these tone pairs. The first tone was often substituted by T4, for example, T1T2 and T3T2 were commonly replaced by T4T2. And the second tones, T2 or T3, were frequently substituted for each other. For example, T4T3 was commonly replaced by T4T2; T2T2 was often replaced by T2T3. For the tone pairs with a neutral tone in the second syllable, the neutral tone tended to be substituted by T4 or T1. For example, T1T5 and T2T5 were commonly replaced by T1T4 and T2T4, respectively. T3T5 was often replaced by T3T1 and T4T1.

3.2. The effects of context and syllable position on the lexical tones production

In this section, we investigated the production performance of four lexical tones in different contexts and positions, including in the first syllable (S1) position of disyllable, in the second syllable (S2) position of disyllable, and in the monosyllable (MS).
context. The accuracy and error patterns were further analyzed in these three contexts. Because the neutral tone can only occur after a lexical tone, and tone sandhi occurs in the T3T3 tone pair, the disyllabic words containing neutral tone and T3T3 were excluded.

3.2.1. Accuracy

Figure 3 presents the accuracy of four lexical tones in the three different contexts and positions (MS, S1, and S2). Due to most group data violated the assumptions for parametric statistics, non-parametric statistics were performed in this analysis. Friedman’s test was conducted to examine the effects of tone context and position. Wilcoxon Signed-Ranks test was applied for the post hoc pairwise analyses.

Concerning tone 1, it had lower accuracy in the first syllable (S1) position of disyllable than in the second syllable (S2) position of disyllable and in the monosyllable (MS). Friedman’s test revealed a significant difference between these three groups ($\chi^2(2) = 10.55$, $p = 0.005$). Subsequent Wilcoxon Signed-Ranks test showed that T1 was significantly less correct in the S1 position than in the S2 position and MS context ($p = 0.00$ and $p = 0.009$, respectively). For tone 2, it was produced less accurately in the S1 position than in the S2 position and MS context. Statistical analyses revealed a significant difference between these three groups ($\chi^2(2) = 6.659$, $p = 0.036$). The pairwise comparisons indicated that T2 had a significantly lower accuracy in S1 position than in S2 and MS contexts ($p = 0.02$ and $p = 0.001$, respectively).

For tone 3, both in the S1 and S2 positions, was produced less accurately than in the MS context. Friedman’s test revealed a significant effect of tone context and position ($\chi^2(2) = 7.22$, $p = 0.027$). The Wilcoxon Signed-Ranks test showed significant differences between the S1 and MS context ($p = 0.004$) and between the S2 and MS context ($p = 0.001$). Concerning tone 4, it had lower accuracy in the S1 position than in the S2 and MS contexts. The Friedman’s test results showed a significant difference between these three groups ($\chi^2(2) = 13.223$, $p = 0.001$). The pairwise comparisons indicated that T4 had a significantly lower accuracy in S1 position than in S2 and MS contexts ($p = 0.002$ and $p = 0.001$, respectively).

The results showed that the accuracy of four lexical tones seemed affected by the contexts and positions. The four lexical tones were pronounced less correctly in the first syllable position than in the second syllable position and in the monosyllable context. Except for T3, tones in the second syllable position had approximate accuracy rates compared with when they were in the monosyllable. T3 seems to be more affected by the position effect, and it had much lower accuracy in the two positions of disyllable than in the monosyllable.

3.2.2. Error patterns

The production error patterns of four lexical tones in the monosyllable and in the first and second syllable positions of disyllable are presented in Figure 4. It showed the percentage of
each type of error produced by all the subjects. For the error patterns in the MS context, T2 and T3 were frequently replaced with each other, and T1 and T4 were often substituted for each other. In the case of tones in the S1 position, the three lexical tones T1, T2, and T3 had a large proportion of substitution by T4, and T4 was often substituted by T1. For the error patterns in the S2 position, T1 and T4 were frequently replaced with each other, T2 was commonly substituted by T1 and T3, and T3 was frequently substituted by T2.

The results showed that the error patterns of four lexical tones seemed affected by the tone contexts and positions. There were some discrepancies between the error patterns in the first syllable position of disyllable and the error patterns in the other two contexts. T2 and T3 tended to be mispronounced as T4 in the first syllable position of disyllable, while they were commonly replaced with each other in the monosyllable and in the second syllable position of disyllable.

4. Discussion

Compared to the monosyllable production in this study, the children with CIs faced more challenges in the production of disyllable tone pairs. The mean accuracy of monosyllabic tones was 82.5% (SD: 7.7%), while the mean accuracy of disyllabic tone pairs was 60.59% (SD: 12.2%). The disyllabic tone pairs containing T2, T3, and T5 were more difficult to produce correctly. The findings are consistent with the previous study [13], and similar to the production of monosyllabic tone [8, 9]. T2 and T3 were also difficult to produce in the isolated monosyllable.

The previous study found that the predominant errors of disyllabic tone pairs were the substitution of T1T1 for other tone pairs [13]. In the present study, the children with CIs produced more diverse error patterns compared with the previous research. This is probably due to the difference in the materials used or the variability of the subjects. A larger number of materials were used in the present study. Previous research found that the tone production accuracy performance correlated with the subjects’ age at implantation [8, 11, 18]. These demographic variables may also affect the production error patterns, which result in the different production performance between the subjects. In this study, we observed that some subjects who had better tone acquisition were more likely to have a confusion error between T2 and T3, while some subjects who had poor tone acquisition often substituted other tones with T1 and T4. However, the specific correlations were not examined.

In the present study, four lexical tones were pronounced less correctly in the first syllable position of disyllable than in the second syllable position of disyllable and in the monosyllable. The results are consistent with the findings of T2 in the disyllable context [15], and also similar to the findings in normal-hearing children’s research [19]. A previous study that investigated the perception character of disyllabic tones in adults with CIs found that the first syllable tone had a lower recognition score than the second syllable tone [20]. Due to the perception is harder in the first syllable, the production ability may be affected. In the previous perception study, T2 and T3 were more difficult to identify in the first syllable. This may also explain why T2 and T3 were more affected by the effects of context and position in the production performance. In this study, we also observed that the first syllable was pronounced more unclear than the second syllable sometimes. Given the importance of lexical tone in tonal languages, it is important to take phonological context into account when investigating and evaluating tone acquisition in children with CIs.

5. Conclusion

The Mandarin-speaking children with CIs had difficulties in acquiring disyllabic tones, especially for the tone pairs containing T2, T3, or T5. Diverse error patterns were found in the present study. The tone pairs usually had one tone that was misrepresented. For the tone pairs with T1 or T4 in the second syllable, the pronunciation errors were more likely to occur in the first tone, while the second tone was produced correctly. The first tone was often substituted by T4, T3, or T1. For the tone pairs with T2 or T3 in the second syllable, both the first and second tone tended to be misrepresented. The first tone was often substituted by T4, while the second tone was likely to have a confusion between T2 and T3.

The contexts and syllable positions had an effect on the lexical tone production of children with CIs. Tones in the first syllable position of disyllable were more likely to be affected. First, the accuracies of the four lexical tones were usually lower in the first syllable position than in the second syllable position and in the monosyllable. Moreover, tones in the first syllable position also had different error patterns compared with the other contexts. Among the four lexical tones, T2 and T3 were more affected by the contexts and positions. T2 and T3 tended to be misrepresented as T4 in the first syllable position of disyllable, while they were commonly replaced with each other in the monosyllable and in the second syllable position of disyllable.

6. Acknowledgements

This study was supported by Science Foundation of Beijing Language and Culture University (the Fundamental Research Funds for the Central Universities) (19PT04), Advanced Innovation Center for Language Resource and Intelligence (KYR17005)

7. References


