10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Acoustic Attributes of Mandarin Retroflex Vowel "ɚ" Produced by Prelingually Deaf Children with Cochlear Implants

Yu-xin Lin, Jue Yu, Yiyuan Liao

School of Foreign Languages, Tongji University, Shanghai, China
1930865@tongji.edu.cn, erinyu@126.com, mereditliao@tongji.educn

Abstract

The present paper examined the acoustic attributes of Mandarin retroflex vowel "ɚ" articulated by prelingually deaf children with cochlear implants (CIs). The main purpose is to obtain a comprehensive understanding about how CI children acquire this target vowel in different lexical positions. The results showed that, (1) CI children could averagely produce the retroflex vowel with the mastery level accuracy, (above 75% accuracy), but made more articulatory errors than the NH counterparts; (2) allocated different duration to the target vowel according to specific lexical position as the NH children did, but averagely tended to produce the vowel with longer duration, especially in the medial location; (3) showed the formant pattern of the target vowel with significant difference from the NH group, specifically in the F3, which determined the retroflex color; (4) only the length of CI device use was found to be a significant contributor for accuracy performance of CI children but no systematic correlation was found between the acoustic attributes and the use of cochlear implant devices.

Keywords: retroflex vowel, Mandarin, prosodic location, prelingually deaf children, children with cochlear implants

1. Introduction

Children acquire language by accessing the auditory input of their native language. For children with hearing impairment, say children with cochlear implants, the input may not be accessible, which may result in impediment to speech acquisition. For persons with profound hearing deficiency, cochlear implantation has been an approved method of treatment; the most direct benefit from cochlear implantation is improved speech perception and spoken word recognition skills [1]. The auditory advances facilitate spoken language development for deaf children. Their speech development followed a process similar to that of normal-hearing children but at a rather slower rate [2]. Distinctive features were studied previously in language production and perception shown by Mandarin-speaking children with cochlear implants, as in performance of the Mandarin vowels [3], consonants [4] [5], tone production [6], sentence production [7]. Taken together, the previous studies suggest that while an increased duration of CI use might facilitate production performance, deficits still exist in language production in CI children compared with the normal-hearing age-peers (e.g. [7]). Both non-linguistic and linguistic factors are taken into account when studying the speech performance of CI children [8]. The non-linguistic factors include age at implantation, the length of CI device use, etc. [5]. The linguistic factors are consisted of different levels of linguistic properties [8]; acoustic attributes such as duration and formant pattern [9] are examined in the present study.

Few studies have investigated the linguistic performance of the retroflex vowel "ɚ" in Mandarin articulated by children with cochlear implants. The retroflex vowel "ɚ", the target vowel in this paper, is a distinctive sound frequently pronounced in Mandarin. Its acoustic feature has been in controversy; yet it is generally accepted to be a diphthong, which is initiated by a schwa [ə], and is rapidly followed by curling up the apex of tongue towards the hard palate [10] [11]. This articulatory description has been corroborated by previous studies [12] [13].Typical production of the retroflex vowel can be observed clearly in the spectrogram where the vowel shows a significant narrowing of the F2 and F3 formant space. The F3 trajectory glides downward over the time course of the vowel; the F3 slope determines the retroflex color. At the same time, its F2 trajectory shows an upward movement. Together, the three trajectories, F1, F2 and F3, form a formant pattern characteristic of the letter "Z".

To seek a greater understanding about the prosodic location and the acoustic attributes of the Mandarin retroflex vowel "ɚ" produced by CI children, this paper aims to address the following research questions:

1. To what extend the CI children can accurately produce the Mandarin retroflex vowel "ɚ" in terms of articulatory accuracy?
2. What are the differences in vocalic duration of the Mandarin retroflex vowel in different prosodic locations between cochlear-implant and normal-hearing children?
3. What are the differences in formant structure of the Mandarin retroflex vowel in different prosodic locations between cochlear-implant and normal-hearing children?

2. Method

2.1. Data

A word list, which was composed of 10 Chinese lexical items with each containing the target syllable "儿" (/ɚ/) (whose lexical meaning is "child" or "son"), was used to be articulated twice by each subject. Of all the lexical items, seven target syllables were designed to be in the initial location and three in the medial location, aiming to examine possible different acoustic attributes assigned to the target vowel according to specific lexical position. The lexical "儿" (/ɚ/) on the wordlist all acted as content words, though it could be a suffix in certain context, which is out of the domain of this paper.

 10.21437/SpeechProsody.2020-168

Speech recording was carried out in a sound-proof room. The speech signals were recorded at a sampling frequency of 44,100 Hz.

2.2. Subjects

Altogether, 8 prelingually deaf Mandarin-speaking children with cochlear implants (4 male and 6 female) were recruited and 8 NH children were used as a baseline. All participants in this study spoke only Mandarin at home and in their daily life. Parental informed consent was obtained for all participants and the study was approved by the local research ethics committee.

The mean age of the CI subjects at the time of recording were 8.13, ranging from 6.33 to 10.16. 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. They received their implants at the age of 1.63 years averagely. Their average length of CI use was 6.5 years. After the surgery, all of them received intensive speech and language training at professional rehabilitation centers in Shanghai.

The average age of NH subjects were 8.63 years, chronologically matching the CI subjects. All had reported without language or speech impairments.

2.3. Data annotation and analysis

All the speech recordings of the subjects were annotated with the help of Praat. The segmental labelling of the target syllable "儿" (/ɚ/) was manually annotated by the first author and the annotated words were double-checked by the second author.

After phonetic annotation, acoustic analysis was conducted to all the speech samples in Praat, mainly focusing on the measurements of two parameters in the target vowel: (1) vocalic duration; (2) F1, F2 and F3 frequencies, as they are proved to be the primary acoustic parameters to characterize the Mandarin retroflex vowel /ɚ/, especially the F3 trajectory. The duration measurements were extracted from the time-stamps in the annotation files. In order to eliminate individual difference across the subjects, normalization was conducted by using z-score. The z-score measure (see Equation 1) was adopted to normalize all the duration parameters.

$$\text{Z-Score}(x) = (x - \text{mean})/SD \quad (1)$$

The measurements of each formant were extracted by a praatscript and then manually double checked. Based on the labeled onset and offset of the retroflex vowel /ɚ/ in the target syllable, each formant of the target vowel at 21equidistant points were extracted. In this way, the time-normalized formant trajectories of the three formants of each target vowel were obtained. After that, all the raw formant values were converted from Hz to Bark scale, a psycho-acoustical scale for subjective measurement of loudness, in order to normalize the possible difference in frequency range between the two genders, or caused by anatomical, physiological or sociological factors [14].

One-way ANOVA was conducted to examine group difference effect on the retroflex vowel / ɚ / production. Two-factor ANOVA was conducted when the effects of both group and prosodic position were involved, with the speaker group as the between-subject factor and the prosodic position as the within-subject factor. If there was a significant main effect of the within-subject factor, a post-hoc test with Tukey HSD was applied. Thus the group differences in acoustic attributes of Mandarin retroflex vowel /ɚ/ in different prosodic position can also be revealed.

3. Results and discussion

3.1. Accuracy metrics

To answer Research Question 1, the articulatory accuracy of each subject was first calculated. Each word of a subject was given a score by the first and third authors according to its accuracy. Zero, one and two points were given accordingly to a wrongly-pronounced word, an unclearly pronounced but understandable word, and a correctly-pronounced word. An inter-rater reliability analysis using the Kappa statistic was performed to determine inter-rater consistency. 'Almost perfect' agreement was found between the two raters (k=0.89). The final score of each word was given by averaging the scores of two raters. Then each subject's mean score of all his/her scores was calculated. The articulatory accuracy of the CI group was averagely 91.1% in the initial location of lexical words while 88.5% in the medial location. Not surprisingly, the CI group produced the retroflex vowel with lower accuracy than the NH children, who, as was expected, produced with 100% accuracy. But it also turned out that they were able to produce the target vowel with the mastery level accuracy (above 75% accuracy) [15]. Two-way ANOVA test was then conducted. Statistical analysis did reveal a significant 'group' difference effect on the articulatory accuracy of the retroflex vowel /ɚ/ ($F=11.177$, $P=0\ .002<0.05$); yet no significant prosodic 'position' difference effect was found ($F=0.112$, $p=0.740>0.05$).

Table 1: *Results of Two-way ANOVA: Tests of Between-Subjects Effects.*

Dependent Variable: accuracy

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.088[a]	2	.044	5.644	.009
Intercept	28.754	1	28.754	3702.577	.000
group	.087	1	.087	11.177	**.002**
position	.001	1	.001	.112	.740
Error	.225	29	.008		
Total	29.066	32			
Corrected Total	.313	31			

a. R Squared = .280 (Adjusted R Squared = .231)

3.2. Acoustic attributes

3.2.1 Between-group difference in vocalic duration

In this section, the vocalic duration of the target vowel will be compared between the CI and the NH children. Figure 1 gives the box plot showing vocalic duration of the target vowel in different prosodic locations between the two groups. The figure showed that the CI children allocated different duration to the target vowel according to specific lexical position as the NH children did, but averagely tended to produce the vowel with longer duration. Especially the target vowel tends to be longer in the medial location of a lexical word than in the initial one. Usually in Mandarin Chinese, when the target syllable "儿" in the medial location of the lexical word, it is prosodically unstressed; while when it occurs in the initial location, it is prosodically stressed. Therefore, the above results further indicate that CI children have some problems in manipulating the temporal differences of the target vowel according to different prosodic position, especially under the unstressed condition.

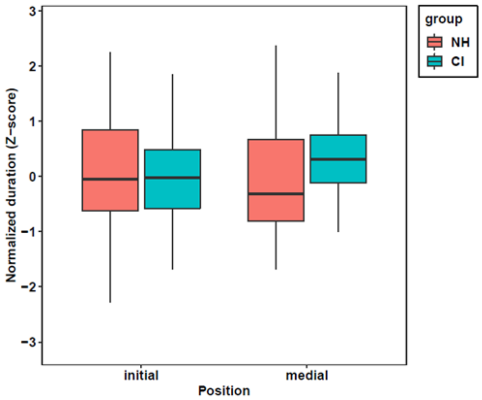

Figure 1: *Box plot showing vocalic duration of the target vowel in different prosodic location between group CI and NH.*

Table 2 reveals the results of parametric two-way ANOVA test. Statistical analysis did not reveal significant 'Group' difference effect on the vocalic duration (F=0.886, p=0.3473>0.05), nor did 'Prosodic Location' (F=0.883, p=0.3481>0.05); yet an interaction effect between Group and Location (F=4.702, P=0 .0309<0.05) can somewhat be found.

Table 2: *Results of Two-way ANOVA: Main Effect.*

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	0.76	0.757	0.886	0.3473
Location	1	0.75	0.754	0.883	0.3481
Group:Location	1	4.02	4.017	4.702	**0.0309***
Residuals	305	260.54	0.854		

3.2.2 Between-group differences in vocalic formant pattern

As is mentioned above, the first three formants of the target vowel /ɚ/ resembles the shape of letter "Z", with its F3 trajectory gliding downward and its F2 moving upward. The production of the target vowel by the CI children failed to be perceived as typically pronounced as the normal-hearing counterparts did to some extent. The perceived different is going to be corroborated by the different formant patterns of the target vowel produced by the two groups through statistical analysis.

Mean bark-values of the three formants of the target retroflex vowel /ɚ/ produced by each subject were first calculated out of his or her valid data. Thus each subject had totally 63 (21equidistant points × 3 formants) bark-values. Then the final mean bark-values of the three formants were compared between the two groups. The obtained data in the three formants were projected into a pattern of three lines in the Cartesian coordinate system with the help of Excel. The time sequence of the 21 equidistant points was on the horizontal axis, and the bark values were on the vertical axis, as shown in Figure 2. As is shown in the figure, there were great differences between the CI children and the NH children. Firstly in terms of F1 and F2, the CI group showed relatively higher values than the NH group, indicating that the CI children, when articulating the target vowel /ɚ/, tended to produce with a relatively higher and more anterior tongue position than their NH counterparts. Furthermore, the F3 trajectory displayed big difference between the two groups. The F3 trendline of the CI group started to go upward at Point 15, while the NH group slid downward from Point 1 to Point 10, and moved evenly afterwards.

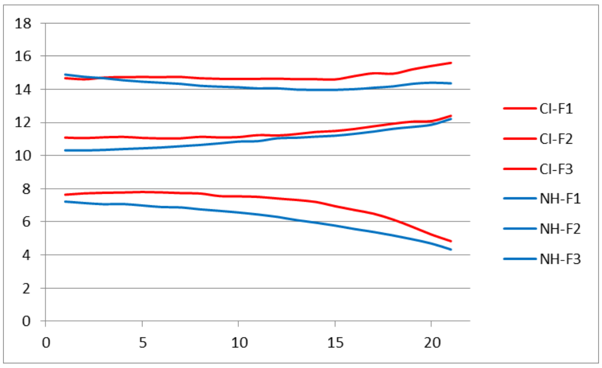

Figure 2: *Between-group formant patterns.*

Table 3 reveals the results of parametric two-way ANOVA test. Statistical analysis did not reveal significant 'Position' difference effect on the F3 slope (F=0.927, p=0.336>0.05); yet an extremely significant 'group' difference (F=54.496, P=0 .000<0.05) can be found. It suggests that significant differences exist in the formant structure, especially in the F3, of the Mandarin retroflex vowel in different prosodic locations between CI and NH children, in that CI children tend to produce the target vowel with an upward F3, which may account for their unsatisfactory performance of the retroflex vowel without typical retroflex color.

Table 3: *Results of Two-way ANOVA: Tests of Between-Subjects Effects.*

Dependent Variable: F3_slope

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.332[a]	3	.111	19.901	.000
Intercept	.010	1	.010	1.711	.192
position	.005	1	.005	.927	.336
group	.315	1	.315	56.496	**.000**
position * group	.014	1	.014	2.540	.112
Error	1.704	306	.006		
Total	2.052	310			
Corrected Total	2.037	309			

a. R Squared = .163 (Adjusted R Squared = .155)

3.3. Interactions between CI device use and Mandarin retroflex vowel production

The interrelation between the Mandarin retroflex vowel production performance (in terms of articulatory accuracy, vocalic duration and F3 frequency) and the use of CI devices (in terms of age at implantation and length of device use) were further examined by conducting the Pearson correlation analysis. The length of CI device use was found to be the only significant predicator for articulatory accuracy (P = -0.836 < -0.5). No significant difference can be found (P=-0.253> -0.5) in articulatory accuracy among different age at implantation. Vocalic duration and F3 frequency were neither influenced by the length of CI devise use nor age at implantation.

4. Conclusion

This paper mainly discusses the vocalic duration in different prosodic locations and the acoustic attributes of the Mandarin retroflex vowel "ɚ" produced by prelingually deaf Mandarin-speaking children with cochlear implants. The main purpose is to obtain a comprehensive understanding about how CI children acquire this target vowel in different lexical positions.

The results are illustrated here: (1) CI children could averagely produce the retroflex vowel with the mastery level accuracy, (above 75% accuracy), but made significantly more articulatory errors than the NH counterparts; (2) allocated different duration to the target vowel according to specific lexical position as the NH children did, but averagely tended to produce the vowel with longer duration, especially in the medial location; (3) showed the formant pattern of the target vowel with significant difference from the NH group, specifically in the F3, which determined the retroflex color; (4) only the length of CI device use was found to be a significant contributor for accuracy performance of CI children but no systematic correlation was found between the acoustic attributes and the use of cochlear implant devices.

5. Acknowledgements

The ISCA Board would like to thank the organizing committees of the past INTERSPEECH conferences for their help and for kindly providing the template files. This study was funded by the National Social Science Foundation of China, Project 15BYY183.

6. References

[1] J. K. Niparko, et al., Cochlear Implants: Principles & Practices, Philadelphia: Lippincott Williams & Wilkins, 2009.

[2] P. J. Blamey, et al. "Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing," *Journal of Speech and Hearing Research*, vol.44, pp.264–285, 2001.

[3] Y. S. Lin, 人工耳蝸植入兒童之國語輔音發音與聽辨 ("The Production and Perception of Consonants in Mandarin-speaking Children with Cochlear Implants"), Journal of Taiwan Otolaryngology Head and Neck Surgery , vol.37, no.4, pp.250-258, 2002.

[4] Yung-Song Lin and Shu-Chen Peng, "Acquisition Profiles of Syllable-initial Consonants in Mandarin-speaking Children with Cochlear Implants," Acta Oto-Laryngologica, vol.123, 1046-1053, 2003.

[5] S. Peng, et al. "Consonant production and language skills in Mandarin speaking children with cochlear implants," Archives of Otolaryngology Head & Neck Surgery, vol.130, pp.92–97, 2004.

[6] D. Han, et al., "Tone production of Mandarin Chinese speaking children with cochlear implants," International Journal of Pediatric Otorhinolaryngology, vol. 27, pp. 875-880, 2007.

[7] D. J. Ertmer and L. Goffman, "Speech Production Accuracy and Variability in Young Cochlear Implant Recipients: Comparisons With Typically Developing Age-Peers," Journal of Speech, Language and Hearing Research, vol. 54, pp. 177–189, 2011.

[8] S. C. Tseng, et al. "Acoustic characteristics of vowels and plosives/affricates of Mandarin-speaking hearing-impaired children," Clinical Linguistics & Phonetics ,

[9] D. Ling, Speech and the hearing-impaired child: Theory and practice. Washington, DC: Alexander Graham Bell Association for the Deaf and Hard of Hearing, 1976.

[10] Z. J. Wu, et al., 《实验语音学纲要》 *(Elements of Experimental Phonetics)*, Beijing: Peking University Press, 2014.

[11] T. Lin and L. J. Wang, 《语音学教程》 *(Coursebook of Phonetics)*, Beijing: Peking University Press, 2013.

[12] Eric Zee, "The phonetic value of the vowels, diphthongs and triphthongs in Beijing Mandarin," *Modern Phonetics in the New Century: Symposium of the Fifth National Academic Conference of Modern Phonetics*, Beijing: Tsinghua University Press, pp. 54-60, 2001.

[13] J. Yu, et al., "上海普通话与标准普通话卷舌元音声学特征对比研究 (A contrastive study of vowel [ɚ] between standard Putonghua and Shanghai Putonghua)," *Contemporary Linguistics*, vol.3, pp. 211-220, 2008.

[14] P. Adank, et al. "A comparison of vowel normalization procedures for language variation research," Journal of the Acoustic analysis of America, vol.116, no.5, pp.3099-3107, 2004.

[15] E. K. Sander. "When are speech sounds learned?" Journal of Speech and Hearing Disorder, no.1, pp. 55-63, 1972.