Mandarin compound vowels produced by prelingually deafened children with cochlear implants

Jing Yang a, Li Xu b, *

a Communication Sciences and Disorders, University of Central Arkansas, Conway, AK 72035, USA
b Communication Sciences and Disorders, Ohio University, Athens, OH 45701, USA

ABSTRACT

Objective: Compound vowels including diphthongs and triphthongs have complex, dynamic spectral features. The production of compound vowels by children with cochlear implants (CIs) has not been studied previously. The present study examined the dynamic features of compound vowels in native Mandarin-speaking children with CIs.

Methods: Fourteen prelingually deafened children with CIs (aged 2.9 ± 8.3 years old) and 14 age-matched, normal-hearing (NH) children produced monosyllables containing six Mandarin compound vowels (i.e., /aɪ/, /aʊ/, /uo/, /i3/, /iaʊ/, /ioʊ/). The frequency values of the first two formants were measured at nine equidistant time points over the course of the vowel duration. All formant frequency values were normalized and then used to calculate vowel trajectory length and overall spectral rate of change.

Results: The results revealed that the CI children produced significantly longer durations for all six compound vowels. The CI children’s ability to produce formant movement for the compound vowels varied considerably. Some CI children produced relatively static formant trajectories for certain diphthongs, whereas others produced certain vowels with greater formant movement than did the NH children. As a group, the CI children roughly followed the NH children on the pattern of magnitude of formant movement, but they showed a slower rate of formant change than did the NH children.

Conclusions: The findings suggested that prelingually deafened children with CIs, during the early stage of speech acquisition, had not established appropriate targets and articulatory coordination for compound vowel productions. This preliminary study may shed light on rehabilitation of prelingually deafened children with CIs.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Cochlear implants (CIs) have provided unprecedented auditory sensation to patients with severe to profound hearing loss. While numerous studies have examined speech perception of the CI users, only a few studies have focused on their speech production, in particular, development of vowel and consonant inventory [1–7]. A number of acoustic studies were undertaken to examine the basic acoustic features of selected monophthongal vowels in CI children [8–15]. There was a consensus regarding the facilitating role of implantation on the acoustic development of vowel production in the hearing-impaired children. However, it remained controversial whether or not the CI children could reach the age-matched norms in the detailed acoustic-phonetic characteristics [8,14,15]. Unlike monophthongs that involve one articulatory position, diphthongs have two distinct articulatory targets connected by a marked transition within one syllable [16]. Unlike two adjacent vowels belonging to two different syllables, diphthongs involve a smooth continuous change of the position of the tongue body and a distinct rate of formant movement [17]. Therefore, production of diphthongs requires smooth vowel-to-vowel transition within one single syllable, which involves complex motor planning and coordination of articulatory gestures.

Auditory input plays an important role in the development of speech motor control [18,19]. Previous studies have shown that the absence of auditory feedback in deaf speakers causes a distortion of speech production [20,21]. Compared to deafened speakers, CI recipients regain partial auditory sensation that helps them better regulate speech motor processes. However, compared to normal-
hearing (NH) speakers, CI users still lack a full range of auditory information. Some recent studies have suggested that children with CIs who have no developmental or cognitive impairment still show delays in motor development, especially in complex motor [22] and fine motor skills [23]. The delayed motor skills in CI children may cause difficulty in compound vowel production that requires more complex planning than single vowel production. Therefore, CI children’s phonetic-acoustic features might be different from those of NH children. However, little research has been done to examine the production of compound vowels in children with CIs.

The present study aimed to expand our knowledge of speech development in CI children by comparing the acoustic features of selected Mandarin diphthongs and triphthongs in CI children with those in age-matched NH children. Mandarin differs from English on both segmental and suprasegmental levels. Regarding the vowel systems, Mandarin has a smaller inventory of monophthongal vowel phonemes than English. However, Mandarin contains a large inventory of compound vowels that includes nine diphthongs ([ai], [ei], [ia], [iu], [iou], [uai], [uai], [uei]) and four triphthongs ([iau], [iou], [uai], [uei]). So far, a number of studies have documented the vowel acquisition in normally developing Mandarin-speaking children [24–27]. However, no study has investigated the production of compound vowels in Mandarin-speaking children with CIs. Although there is a shortage of comparable studies on compound vowel production in CI children, some early research on the acoustic features of compound vowels in the deaf population were of relevance to the present study [20,28]. Monsen [20] examined the acoustic properties of the vowels /i, a, o/ and the diphthong /ai/ produced by 36 deaf adolescents and 4 NH adolescents. The authors found a noticeable reduction of the phonological space due to the lack of auditory input. In the present study, we investigated how prelingually deafened children who started obtaining relevant auditory information including the etiology of hearing loss and auditory threshold were not available. The 14 age-matched NH children were reported as having no impairments of language, speech, or hearing by their parents or teachers.

2.2. Speech materials

The speech materials included a list of nine Mandarin monosyllables (“ài, bāo, pāo, dōu, tōu, jie, qie, yāo, yǒu” in Pinyin) containing four diphthongs ([ai], [au], [uo], and [iu]) and two triphthongs ([iau] and [iou]). The target vowels either followed obstruent consonants or occurred in an environment without initial consonant. None of the recorded monosyllables had a coda. The onglides of the compound vowels covered the three corner vowels /a, i, u/ in Mandarin Chinese. Each syllable was produced once in each of the four Mandarin tones. Therefore, each participant produced $9 \times 4 = 36$ tokens.

2.3. Recording

For each participant, the recording session was conducted in a quiet room using a Sony portable DAT recorder (Model TCD-D100) connected to an ElectroVoice omnidirectional microphone (Model RE50B). The sampling rate was set at 44.1 kHz and the quantization rate was set at 16 bits. The experimenter produced each syllable in Mandarin tone 1 and required the participants to articulate the same syllable in all four tones. Each syllable in each tone was produced once and the experimenter made no corrections during the recording session.

2.4. Data analysis

All recorded tokens were transferred to a computer hard disk drive. The recorded tokens were then segmented into separate wave files with each file containing one syllable. To capture the nature of the dynamic change of vowel quality in diphthongs and triphthongs, formant frequency values were extracted at nine equidistant time locations (10–20–30–40–50–60–70–80–90% point) over the course of vowel duration using the spectrographic analysis program TF32 [31]. The landmarks of vowel onset and offset were located by hand through a visual check of the waveform accompanied with spectrographic display. In particular, vowel onset was defined at the beginning of vowel periodicity and vowel offset was defined at the end of vowel periodicity where both F1 and F2 were present.

Due to the relatively large age range of the child speakers, all measured formant frequency values were normalized prior to further analysis. Given that only selected compound vowels were investigated in the present study, a vowel-intrinsic approach of the Bark transformation was used and all formant values of each token were converted to the Bark scale following the formula of Traunmüller [32]:

$$Z_i = 26.81 \left( \frac{1 + 1960}{F_i} \right) - 0.53,$$

in which $F_i$ is the formant frequency value of a given formant $i$ and $Z_i$ is the Bark value of formant $i$. Based on the bark values at multiple
measurement points, we calculated trajectory length (TL) and spectral rate of change (TL_roc) [33]. Trajectory length defined the sum of Euclidean distance between each two consecutive time points (10–20%, 20–30%, 30–40%, etc.), which was calculated by:

\[ TL = \sum_{n=1}^{9} VSL_n. \]

The Euclidean distance of each vowel section (VSL) was calculated by:

\[ VSL_n = \sqrt{(F1n + 1 - F1n)^2 + (F2n + 1 - F2n)^2}. \]

TL_roc defined the average rate of formant change over a certain period of time. We calculated the average spectral rate of change over the tested vowel portion using the formula:

\[ TL_roc = TL/(0.8d), \]

where \( d \) is the duration of the vowel. The acoustic measures of TL and TL_roc enabled us to quantify the dynamic feature of formant trajectories in terms of the magnitude and the velocity of articulatory movement during the course of vowel production.

2.5. Statistical analysis

Preliminary analysis showed no apparent differences on formant trajectories among the four tones. In addition, for a given vowel that was produced in more than one word, the preceding consonants shared the same place of articulation. Therefore, the data were collapsed across tones and words. Subject mean values of each vowel were obtained for each acoustic measurement prior to further statistical analyses. For the comparison of formant frequency values, a three-way repeated-measures ANOVA was used on the F1 and F2 data respectively with vowel quality and measurement point as the within-subject factors and group as the between-subject factor. For the comparison of other acoustic measures (e.g., duration, TL, and roc), two-way repeated measures ANOVA tests were implemented with vowel quality as the within-subject factor and group as the between-subject factor. Because the effects of vowel quality on formant frequency values and vowel spectral change were anticipated and have been reported in previous studies, which were not of interest in the present study, only significant results related to the group effect were reported and discussed. One-way ANOVAs were conducted afterwards to compare the group difference in each vowel. The alpha level was adjusted for multiple comparisons.

3. Results

3.1. Formant trajectories in \( F1 \times F2 \) space

Fig. 1 shows the comparison of the formant trajectories between individual CI children and the group mean data of the NH children. Considerable individual variation was present in formant trajectory patterns in the CI children. Some of the CI children, such as CI8 and CI9, showed similar patterns of formant trajectories to those of the NH children. In particular, these CI children approximated the NH children in terms of the relative position of the compound vowels in the acoustic space, the direction, and the magnitude of the formant movement. However, some CI children, such as CI5 and CI7, produced all six compound vowels in a restricted, centralized region of the vowel space and demonstrated a distinctive pattern of vowel organization from that of the NH children. Many CI children differed from the NH children in one or more aspects of the formant trajectory features. In the top-left panel for the NH children, the two triphthongs /iaʊ/ and /iu/ showed curved trajectories that indicated a change of vowel quality corresponding to three vowel targets. Some CI children, such as CI7 and CI14, showed less curved formant trajectories, which suggested that these CI children did not show as much change of vowel quality as the NH children did. Some CI children, such as CI3 and CI14, showed a higher position of /iaʊ/ than /iu/ while some other CI children, such as CI6 and CI7, showed a formant trajectory moving from high back to mid-high back position for the vowel /iu/. These patterns were dramatically different from the pattern of front-to-back formant movement in the NH children, which suggested that these CI children pronounced the triphthongal vowels with different articulatory movement patterns.

In terms of the formant trajectories of diphthongs, some CI children, such as CI4, CI6, and CI12, showed very little formant movement in the acoustic space for the vowel /ai/. CI5 produced this vowel with the trajectory moving from a central position to a high front position. CI7 produced /ai/ with a “U-shape” trajectory which was located at a position similar to his /aʊ/ in the acoustic space. For the vowel /ii/, some CI children, such as CI2 and CI14, produced it with a relatively static formant trajectory which indicated that this CI child monophthongized this diphthong. In contrast, CI12 produced /ii/ with a formant trajectory moving down to a low central position which indicated that this CI child mispronounced /ii/ as /aʊ/. In addition, the formant trajectory of /ii/ produced by CI7 showed a front-to-back movement which also substantially differed from the direction of formant trajectory in the NH children. For the other two diphthongs /au/ and /uo/, some CI children, such as CI6, CI12, and CI13, produced /uo/ with the formant trajectories...
moving from a high back position to a low central position, which suggested that these CI children produced this vowel as /ua/ rather than a typical /uo/. For the vowel /aʊ/, some CI children, such as CI1, CI3, CI5, and CI6, showed little formant frequency change which indicates a static vowel quality rather than a diphthong moving from a low mid position to mid-high back position.

The three-way repeated-measures ANOVA revealed no significant main effect of group for either F1 or F2. However, a significant interaction effect of vowel by group was yielded for F1 (F(5,130) = 7.075, p < 0.001) and a significant interaction effect of measurement point by vowel by group was yielded for F2 (F(40,1040) = 4.157, p = 0.001). These significant interaction effects suggested that the two groups of children showed different formant frequency values on certain vowels at certain measurement points. In addition to the comparison of relative position defined by F1 and F2 between the NH and CI children, a measure of overall acoustic distance was calculated to show the potential positional deviation of the CI children relative to NH targets. Following the procedure described in Yang et al. [15], for each vowel, the mean F1 and F2 were calculated for each measurement point in the NH children. Then, the Euclidean distance between each participant and the NH target was calculated for each measurement point. Finally, the Euclidean distances of the nine points were summed up, which served as an index to quantify the potential positional deviation of the formant trajectories of each participant from the NH targets.

As shown in Fig. 2, the CI children produced all six compound vowels with greater acoustic distance from the NH targets. This observation was confirmed by the significant group effect on the acoustic distance (F(1, 26) = 12.158, p = 0.002). This result suggested that the CI children, as a group, deviated from the NH targets in their vowel production to a greater extent than the NH children did. Note that the CI children demonstrated a great amount of individual variability in the vowel acoustic distances. The subsequent one-way ANOVAs revealed that the CI children showed significantly greater acoustic distance for the vowels /au/ and /au/ with larger variability than the other vowels, which indicated that the NH children did not always produce these vowels in a consistent manner.

3.2. Vowel duration

Fig. 3 displays the duration of individual compound vowels. Not surprisingly, the two triphthongs showed much longer vowel durations than the diphthongs for both CI and NH children. The CI
children generally matched the NH children on the pattern of vowel duration, but they produced significantly longer vowel durations than did the NH children for all six compound vowels. The two-way repeated measures ANOVA revealed a significant main effect of group (F(1, 26) = 14.877, p = 0.001). Subsequent one-way ANOVAs confirmed our observation of longer vowel duration in the CI children than in the NH children for each compound vowel. Note that the NH children presented a bimodal distribution of duration for certain vowels. A nonparametric test was conducted to validate the ANOVA results. The Mann-Whitney U test results yielded significant longer duration in the CI children than in the NH children for all six vowels.

### 3.3. Vowel trajectory length

Fig. 4 displayed the trajectory lengths (TLs) over the 80% portion of vowel duration. Not surprisingly, the two triphthongs showed longer TLs than the four diphthongs in the NH children. In general, the CI children showed a pattern compatible with the NH children although the variation of the CI children was greater than that in the NH children for certain vowels such as /au/, /iau/, and /iou/. A two-way mixed-factor repeated measures ANOVA was implemented with vowel as the within-subject factor and group as the between-subject factor. The results revealed a significant main effect of vowel quality (F(5, 125) = 101.1, p < 0.001). No group difference or vowel by group interaction was found.

### 3.4. Vowel spectral roc

While VSLs and TLs characterize the magnitude of formant change for each vowel section and the entire tested vowel portion (80% of vowel in the present study), it fails to capture the feature of the rate of formant change over time. Fig. 5 shows overall spectral rocs (TL_rocs) over the 80% portion of vowel duration. The TL_rocs of the two triphthongs were larger than those of diphthongs in the NH children. This indicates that the articulators move faster for the triphthongs than for the diphthongs. The CI children generally followed the pattern of TL_roc over the six compound vowels in the NH children. However, they showed evidently smaller TL_rocs than did the NH children for most of the compound vowels. In addition, the CI children did not show significantly larger TL_rocs for the triphthongs, as the NH children did. The two-way mixed-factor repeated measures ANOVA revealed a significant main effect of vowel by group interaction effect (F(5,130) = 5.737, p < 0.001). Subsequent one-way ANOVAs revealed a significantly smaller TL_roc of /uo/, /iau/, and /iou/ in the CI children relative to the NH children.

### 4. Discussion

The present preliminary study examined the acoustic-phonetic
Although some CI children produced similar formant trajectories to those of the NH children, the majority of the CI group still differed from the NH children in certain aspects of acoustic features. As shown in Fig. 1, most of the CI children either undershot the formant trajectories by producing some compound vowels with much less formant movement, or overshot the formant trajectories by producing some compound vowels with greater formant movement. The relative positions of the compound vowels and the organization of the vowel space in the CI children were different from the NH targets (as shown in Figs. 1 and 2). The off-target onset and offset of the compound vowels as well as the deviation of the formant trajectories in the CI children relative to the NH peers suggested that the CI children were less likely to initiate or complete the articulatory gestures in a similar manner to the NH peers. This result indicated that the CI children experienced more difficulties in establishing precise tongue configuration for the compound vowel targets.

An early study pointed out that the hearing-impaired speakers produced vowels within a more restricted frequency range for both F1 and F2 [20]. Among the 14 CI children in the present study, CI5, CI7, and CI14 produced the compound vowels in a centralized area and presented substantial acoustic overlaps among these vowels. Notice that CI5 and CI7 had implantation after 5 years of age and had relatively short length of device use. CI14 had received implantation for less than 4 months. The highly clustered production of compound vowels in these participants suggested that they were still at the emergent stage of speech development and performed similarly to deaf children reported in Ref. [20]. In a more recent study [29], Palethorpe and colleagues reported retraction of F2 for diphthong targets in some deafened speakers. In the present study, some CI children produced certain vowels with retracted F2. However, we also observed fronted F2 in some vowels of the CI children. For example, CI2 and CI5 produced the vowel /uo/ with a fronted F2. CI5, CI6, CI7, and CI13 produced the vowel /au/ with a fronted F2. These inconsistent patterns of positional deviation in the CI children relative to the NH children revealed diverse articulatory mechanisms used by CI children.

Consistent with previous studies [15,39], the present study showed that the CI children produced all compound vowels with longer duration than the NH peers, which suggested that the CI children may need more time to form the articulatory gestures and to travel from one target to the other. Earlier research on NH speakers reported that lengthened vowel duration and greater formant frequency change corresponded to higher speech clarity [40]. The CI children might use elongated vowel durations as a compensatory strategy to increase the clarity and intelligibility of their speech. The similar overall trajectory lengths in the CI children relative to the NH children suggested that the CI children were able to move the articulators with similar magnitude of movement as the NH children did. However, the significantly lower TL_rocs in the CI children demonstrated that the CI children did not move the articulators as quickly as the NH children. Moreover, the result of significantly longer TL_rocs for both triphthongs in the CI children relative to the NH children suggested the difficulty of articulatory coordination for multi-target phonetic segments in the CI children.

The findings of the present study, although preliminary, bear important implications in clinical training and habilitation for children with CIs. Currently, diverse strategies and techniques have been proposed to improve CI user’s perceptual and recognition abilities in a variety of listening conditions. But, we still lack an effective evidence-based oral training program that improves the articulatory accuracy for pediatric CI users, especially for non-English speaking CI children. The present study demonstrated both undershooting and overshooting of the vowel targets during the process of compound vowel production in CI speakers. To
address these deficits, effort should be made to identify the specific pattern of vowel organization in individual CI children and then develop a more targeted plan to help the CI children establish appropriate vowel targets for the compound vowels. Moreover, the present study revealed that not all compound vowels showed equally poor performance in the CI children. Some CI children showed more difficulty in producing certain vowels than other vowels. Triphthongs were generally more challenging than the diphthongs for most CI children. Therefore, an overall impressionistic and acoustic evaluation of the production of individual speech sounds is needed for individual CI children.

These results are informative and might help guide rehabilita-
tion effort in prelingually deafened children with CIs. However, the results should be interpreted with caution due to the large individual variation of the CI population and small sample size of our participants. The overshooting and undershooting of formant trajectories and the slower rate of change in our CI children did not occur ubiquitously. The NH children also presented great variability in certain measures (e.g., duration). In the future research, the sample size of participants in both CI and NH groups needs to be expanded so that we can better group the children based on their developmental stages. Specifically, CI children with a wider range of age at implantation and varying lengths of device use should be included to examine how these factors affect the general shape of formant trajectories and other types of fine-grained acoustic features. In addition to recruiting NH children as the control group, we are interested in comparing CI children’s production with NH adults to better understand the deviation of CI children’s vowel produc-
tion from the adult norms. Moreover, a perception test of CI chil-
dren’s vowel production by NH adult listeners should also be carried out to examine how the acoustic deviation of CI children’s vowel production will affect the recognition of the intended vowels by the NH listeners.

5. Conclusions

In general, the findings of the present study suggested that prelingually deafened children with CIs, during the early stage of speech acquisition, had not established appropriate targets and demonstrated more difficulty of articulatory coordination for compound vowel productions.

Acknowledgements

Emily Hahn and Lauren Muscari provided editorial assistance during the preparation of the manuscript. The study was supported in part by an NIH NICHD Grant No. R15-DC014587.

References

[3] D.J. Erter, J. A. Mellon, Beginning to talk at 20 months: early vocal develop-
[5] T.A. Serry, P.J. Blamey, A 4-year investigation into phonetic inventory develop-