Acoustic properties of vowel production in prelingually deafened Mandarin-speaking children with cochlear implants

Jing Yang,[1] Emily Brown,[2] Robert A. Fox,[3] and Li Xu[2,a)]
[1]*Department of Communication Sciences and Disorders, University of Central Arkansas, Conway, Arkansas 72035, USA*
[2]*School of Rehabilitation and Communication Sciences, Ohio University, Athens, Ohio 45701, USA*
[3]*Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43210, USA*

(Received 2 February 2014; revised 27 August 2015; accepted 18 September 2015; published online 5 November 2015)

The present study examined the acoustic features of vowel production in Mandarin-speaking children with cochlear implants (CIs). The subjects included 14 native Mandarin-speaking, prelingually deafened children with CIs (2.9–8.3 yr old) and 60 age-matched, normal-hearing (NH) children (3.1–9.0 years old). Each subject produced a list of monosyllables containing seven Mandarin vowels: [i, a, u, y, ɤ, ʅ, ɿ]. Midpoint F1 and F2 of each vowel token were extracted and normalized to eliminate the effects of different vocal tract sizes. Results showed that the CI children produced significantly longer vowels and less compact vowel categories than the NH children did. The CI children's acoustic vowel space was reduced due to a retracted production of the vowel [i]. The vowel space area showed a strong negative correlation with age at implantation ($r = -0.80$). The analysis of acoustic distance showed that the CI children produced corner vowels [a, u] similarly to the NH children, but other vowels (e.g., [ʅ, ɿ]) differently from the NH children, which suggests that CI children generally follow a similar developmental path of vowel acquisition as NH children. These findings highlight the importance of early implantation and have implications in clinical aural habilitation in young children with CIs. © *2015 Acoustical Society of America.*
[http://dx.doi.org/10.1121/1.4932165]

[LK] Pages: 2791–2799

I. INTRODUCTION

A great deal of previous research has found that in addition to the severe deficiency of auditory sensation, severe-to-profound hearing-impaired children demonstrate greatly compromised oral speech and language skills. Specifically, researchers have shown that speech of severe to profound hearing-impaired children is characterized by a higher fundamental frequency (F0), greater F0 and amplitude range, largely overlapped acoustic vowel categories, reduced phonological space, weak breath control, and errors in the suprasegmental level of speech (Angelocci *et al.*, 1964; Monsen, 1976).

Contemporary multichannel cochlear implants (CIs) allow prelingually deaf listeners to gain hearing sensation and substantially change how they produce speech sounds (Manning *et al.*, 1992; Uchanski and Geers, 2003; Löfqvist *et al.*, 2010; Neumeyer *et al.*, 2010). Manning *et al.* (1992) performed a longitudinal study to examine the frequency change of vocalic formants pre- and post-implantation at 6, 12, 18, 24, and 36 months in a prelingually deafened English-speaking child implanted at five years of age. The results demonstrated a gradual approximation to age-matched norms during the three-year period. In a more recent study that recruited a larger number of CI participants, Uchanski and Geers (2003) compared acoustic measurements of English-speaking CI children's production of selected consonants and vowels to those of age-matched normal-hearing (NH) children. The average length of implant use of the CI children was 5.5 years. The authors measured voice onset time (VOT) of [t] and [d], F2 values for the vowels [i] and [ɔ], spectral moments of [s] and [ʃ], nasal manner metrics for [m] and [n], and duration at the segmental level and word or sentence level. The results showed that the majority of these measures of the CI children reached or approximated the values of the NH children. In particular, more than 70% of the CI children produced the VOT of [t] and [d] and more than 85% of the CI children produced the F2 of [i] and [ɔ] within the normal range. Many CI children (62%, 71%, and 82%) also showed fricative spectral moments (kurtosis, spectral mean, and skewness) within the normal range. In contrast, many CI children failed to reach the normal range on the nasal manner metric, sentence duration, and vowel duration.

In CI stimulations, a certain amount of temporal and spectral information is conveyed through the electrical pulses. With such auditory feedback, prelingually deafened children can better form articulatory gestures, produce speech events, and monitor their own speech (Tye-Murray, 1992). However, due to the limited spectral resolution of CI devices (as a consequence of a restricted number of channels used to deliver a wide range of speech frequencies), the reduced neural survival of the auditory system, and the impaired storage of working memory, the perceptual and cognitive abilities of CI children might still differ from those of NH listeners (e.g., Connor *et al.*, 2006; Geers *et al.*, 2008;

[a)] Electronic mail: xul@ohio.edu

Nittrouer *et al.*, 2013; Ruffin *et al.*, 2013). As a consequence, the speech production of CI children may also differ from that of NH children. For example, Horga and Liker (2006) examined a set of acoustic measures associated with voice and speech production in native Croatian children with CIs and hearing aids with reference to NH controls. These measures included the size of vowel space defined by F1 and F2 of [i], [a], and [u], the voiced versus voiceless contrast perceived by adult listeners, closure duration and VOT, word accent, sentence stress, voice quality [using the GRBAS (Grade, Roughness, Breathiness, Asthenia, Strain) scale], and pronunciation quality (using a 1–7 scale to indicate good or poor pronunciation). While the children with CIs performed in a manner more similar to the NH children than the children with hearing aids, the CI children still showed significant differences from the NH children in almost every aspect of voice and speech production. In certain measures, such as VOT and closure duration, the CI children showed poorer voiced-voiceless distinctions than the children with hearing aids.

While there are a number of factors affecting the extent to which CI children's speech production approximates that of NH children (Tobey *et al.*, 2003), the general developmental path of children's sound system should also be considered. According to previous studies (Hare, 1983; Stoel-Gammon and Herrington, 1990) with normally developing English-speaking children, vowel acquisition generally follows the pattern that corner vowels are acquired earlier than non-corner vowels and tense vowels are acquired earlier than lax vowels. If children with CIs follow this pattern of vowel development, they may show better mastery of early acquired vowels and show more difficulty in later acquired vowels. Acoustically, the early acquired vowels will show greater approximation to NH norms while the later acquired vowels will show less approximation to the NH norms. Ertmer (2001) reported the vowel development in a prelingually deafened child over 12 months after implantation. During the first four months post-implantation, this child mainly produced the corner vowels [a, u] and the central vowel [ə]. In the fifth month, this child started to produce more types of vowels such as [o, ϵ, ɪ]. This child's vowel development was generally in accordance with the path of normally developing children. Although a longitudinal, single-case study can document the developmental path of vowel acquisition for an individual child, a group study of the acoustic-phonetic features of the entire vowel system from CI children will enable us to gain "snapshot" knowledge of the acoustic development of the CI population.

To date, there has been an increasing body of research on the speech and language development in prelingually deafened English-speaking children who have received CIs (e.g., Blamey *et al.*, 2001; Tobey *et al.*, 2003; Geers *et al.*, 2008; Ruffin *et al.*, 2013); however, relatively few studies have addressed the speech production in CI users from other language backgrounds such as Croatian (Horga and Liker, 2006), Swedish (Löfqvist *et al.*, 2010), German (Neumeyer *et al.*, 2010), and Spanish (Perkell *et al.*, 2001). Very little research has been done on Mandarin. To address this gap, the present study aims to examine the acoustic properties of vowel production in prelingually deafened, Mandarin-speaking children with CIs. Of particular interest is the extent to which the fine phonetic features of the CI children's vowel production are similar to or different from those of the NH peers. In this study, a comprehensive acoustic profile (i.e., duration, formant frequency values, vowel space area, and measures of acoustic similarity to the age-matched NH children) of the entire Mandarin vowel system will be provided.

Mandarin differs from English not only in that Mandarin is a tonal language, but also in that Mandarin has a distinct set of phonetic segments compared to English. With regard to the vowel system, Mandarin has a relatively small vowel inventory with five basic monophthongal vowel phonemes (i.e., /a, i, u, ɤ, y/) (Duanmu, 2007). The vowel phoneme /i/ has three allophonic variants [i], [ɿ] and [ʅ]. The high front apical vowel [ɿ] only occurs after [ts, tsʰ, s] while the high back apical vowel [ʅ] that is produced with the tongue tip raised to the post-alveolar region (described as a retroflexed vowel in traditional Chinese linguistics) only occurs after retroflexed consonants [tʂ, tʂʰ, ʂ]. Note that the vowels [ɿ] and [ʅ] only occur in restricted phonetic contexts, but the consonants they follow ([ts, tsʰ, s, tʂ, tʂʰ, ʂ]) can combine with other vowels. The vowel [i], a high front laminal vowel, occurs in all other consonant environments. Unlike the quadrilateral vowel space in English, the Mandarin vowel space is shaped like a triangle with [a, i, u] occupying the corner positions. Among the seven vowels [i, ɿ, ʅ, a, u, y, ɤ], a majority of them are distributed at relatively high positions in the vowel space while the vowel [a] alone is located at a low position. In addition, the Mandarin vowel system lacks a tense-lax distinction. With regard to the vowel development in Mandarin-speaking children (Zhu and Dodd, 2000; Shi and Wen, 2007), it has been reported that normally developing children acquire [a, i, u] at a relatively early age, around 2 years old, but the rounded high front vowel [y] and the two allophonic vowels [ɿ] and [ʅ] at a relatively late age, around 5 years old. In a recent report on speech characteristics of Mandarin-speaking CI children, Chuang and colleagues measured the absolute differences of formant frequencies between each two of the Mandarin corner vowels ([a], [i], and [u]) in 24 children with CIs (Chuang *et al.*, 2012). They found a significantly reduced vowel space area in the CI group as compared to the age-matched NH group in sustained vowel phonation but not in sentence-reading production. The present study will extend beyond these three corner vowels by investigating all seven Mandarin monophthongal vowels so as to provide a more comprehensive profile of the acoustic characteristics of the Mandarin vowel system in CI children.

The prelingually deafened children in the present study were nonverbal before surgery. They started to receive language input and acquire the speech skills only after the device was implanted. Taking into consideration the relatively short duration of implant use of these CI children in the present study, we tentatively predict that the CI children will show vowel features close to those of age-matched NH children for early acquired Mandarin vowels. We expect that the CI children will show less approximation to the NH children for late acquired vowels. On the other hand, considering great individual differences in the CI population, we cannot rule out two other possibilities: (1) these CI children may reach normal values for all seven Mandarin vowels, or (2) they may show substantial differences from the NH children for all seven Mandarin vowels.

II. METHODS

A. Participants

The data were obtained from the participants in a previous study on tone production in Mandarin-speaking children (Zhou and Xu, 2008). There were 14 prelingually deafened, Mandarin-speaking children (2.9–8.3 years old) who had received cochlear implantation and 60 Mandarin-speaking NH children of similar ages (3.1–9.0 years old). All participants resided and were recruited in Beijing, China. Both parents of these children spoke Mandarin and interacted with the children in Mandarin at home. All CI children were profoundly deaf and nonverbal prior to implantation, and they all received rehabilitation and basic speech and language service after the surgery at a professional rehabilitation center in Beijing. The age at implantation for the CI children ranged from 1.16 to 7.09 years old [mean ± standard deviation (SD): 3.43 ± 1.91 yr] and the duration of implant use varied from 0.31 to 2.60 yr (mean ± SD: 1.73 ± 0.76 yr). The CI devices included three Clarion CII (Valencia, CA) implants, 4 Nucleus 24M (Melbourne, Australia) implants, and 7 Nucleus 24R implants. Detailed individual demographic information is available in Zhou and Xu (2008). However, detailed information related to the etiology of the hearing loss and the pre- and post-implantation thresholds for the CI children is not available. A common practice in CI mapping is that the post-implantation thresholds for the CI children were set around 30 dB hearing level (HL), which represented a mild hearing loss in the users. All the NH participants had a pure-tone average threshold at 500, 1000, and 2000 Hz of $\leq$20 dB HL and were reported as having no speech-language impairments and/or otological disorders by their parents or teachers.

B. Speech materials and data collection

Each participant produced a list of 23 Mandarin monosyllables that contained seven Mandarin vowels [i, ɿ, ʅ, a, u, y, ɤ] (see the Appendix for details). The vowel /o/ is phonetically represented as a diphthong (Zee, 2001) and thus was not included in the present study. The word list was originally designed to elicit speech samples for tone production of Mandarin-speaking children with CIs (Zhou and Xu, 2008). In the present study, to ensure a sufficient number of tokens for each vowel, the monosyllabic words containing target vowels in tone 1 and tone 2 were selected for further analysis ("si" and "ci" were available in tone 1 only). Because the effect of tone on vowel production of CI children is not of particular concern in the present study, not all four Mandarin tones for each word were used. The consonants preceding the target vowels were not controlled. Since the CI group only had 14 participants and the two syllables ("si" and "ci") containing the vowel [ɿ] were available in tone 1 only, in order to ensure sufficient number of tokens for acoustic analysis in the CI group, three disyllabic words ("fangzi," "zongzi," and "shizi") that also contain the vowel [ɿ] were selected for the CI children. Therefore, there were 47 tokens in the word list for the CI children and 44 tokens in the word list for the NH children (see the Appendix). During the recording session, each participant was seated in a quiet room and was instructed to produce the target words in different tones as syllable-and-tone drills that children learned in kindergartens or rehabilitation centers. The experimenter articulated these words in tone 1 and each child was required to produce different tones of the same consonant-vowel structure. The experimenter did not make corrections to the children during recording. Self-corrections were allowed, but rarely occurred during the recording sessions. All speech samples were recorded through an ElectroVoice (Grasbrunn, Germany) omnidirectional microphone (Model RE50B) to a Sony (Tokyo, Japan) portable DAT recorder (Model TCD-D100) with a 44.1 kHz sampling rate. Ultimately, by excluding the missing data and the tokens with poor sound quality, a total number of 622 out of 658 tokens from the CI children and 2566 out of 2640 tokens from the NH children were used for the acoustic analyses.

C. Acoustic analysis

The recorded speech samples were transferred to a computer hard disk with the same sampling rate at 44.1 kHz and a 16-bit quantization rate. These recorded materials were then segmented into separate syllables and saved as individual wave files using CoolEdit 2000 (Syntrillium Software, Scottsdale, AZ).

The speech analysis program TF32 (Milenkovic, 2003) was used to determine the frequencies of the first two formants, F1 and F2, of each monosyllable. The formant extraction was based on Linear predictive coding analysis and the extracted formant tracks were displayed on the spectrogram. Due to the high F0 of children's speech, a large analysis bandwidth (450 Hz) was used for the spectrographical analysis. The midpoint formant frequency values were extracted on the basis of temporal locations of each vowel's onset and offset. The landmark locations of vowel onset and offset were determined primarily on the basis of the waveform, accompanied with a visual check of the spectrogram and auditory check of the vowel quality. Vowel onset following oral stops, fricatives, and affricates was defined as the zero crossing point of the first period of the voicing following the stop burst or cessation of frication. Vowel onset following the nasals [m, n] was determined at the point of significant increase in amplitude of the waveform with visual check of the end of the nasal murmur in the spectrogram. Vowel onset following the lateral [l] was set at the point of significant increase in amplitude of the waveform with visual check of the end of weak energy in the spectrogram due to the lateral zero. Vowel onset in the syllables "wu" and "yi" starting with semivowels was set at the point of attainment of the relatively steady portion on the spectrogram. Given that no consonant appeared in the coda position, the vowel offset was determined at the zero crossing point of the last glottal pulse that was visually checked with energy extended through both F1 and F2.

In order to eliminate the effects of different vocal tract sizes as a function of chronological age of the participants, all formant frequency values were normalized using the Lobanov (1971) method (recommended as one of the most effective normalization approaches by Adank *et al.*, 2004). The Lobanov normalization process converted all formant values to *z*-scores for each subject. To facilitate further interpretation, all normalized *z*-scores were then rescaled to Hz-like values using the following formulas as proposed by Thomas and Kendall (2007):

$$F'_1 = 250 + 500(F^N{}_1 - F^N{}_{1\mathrm{MIN}})/(F^N{}_{1\mathrm{MAX}} - F^N{}_{1\mathrm{MIN}}),$$

$$F'_2 = 850 + 1400(F^N{}_2 - F^N{}_{2\mathrm{MIN}})/(F^N{}_{2\mathrm{MAX}} - F^N{}_{2\mathrm{MIN}}),$$

where F'_i is a rescaled normalized formant; $F^N{}_i$ is a Lobanov normalized formant value for an individual speaker; and $F^N{}_{i\mathrm{MIN}}$ and $F^N{}_{i\mathrm{MAX}}$ are the minimum and maximum values, respectively, of Lobanov normalized $F^N{}_i$ across the entire data-set. All of the following acoustic measures and analyses were based on the rescaled normalized formant frequency values.

To represent the distribution of each vowel for a subject group, the rescaled normalized F1 (ordinate) of a particular vowel of all subjects in the group were plotted against the rescaled normalized F2 (abscissa) of the same vowel of all subjects in the group. A vowel ellipse was then plotted. First, the rescaled normalized F1 × F2 scatter plot was fitted linearly and the positive angle of the linear fit was taken as the direction of the semimajor axis of the ellipse. The ellipse center was determined by the mean of the rescaled normalized F2 values along the fitting line. A perpendicular line to the fitting line defined the direction of the semiminor axis. The lengths of the semimajor and the semiminor axes were determined by two standard deviations of all data points away from the center along the respective lines. Thus, each vowel ellipse encompassed ∼95% of the data points in the rescaled normalized F1 × F2 scatter plot.

In order to quantitatively describe the extent to which the CI children's vowel acoustic features were similar to or different from those of the NH children, a measure of acoustic distance (AD) was derived on the basis of rescaled normalized formant values. First, for each vowel, the mean F1 and F2 were calculated across all the NH children. These values served as the NH target formant frequency values. Next, the Euclidean distance between each CI and NH child's vowel production and the corresponding NH target was calculated using the formula

$$\mathrm{AD}_{jk} = \sqrt{(\mathrm{F1}_{jk} - \mathrm{F1}'{}_j)^2 + (\mathrm{F2}_{jk} - \mathrm{F2}'{}_j)^2},$$

where $\mathrm{F1}'_j$ is the group mean of F1 for all NH children for vowel *j* and $\mathrm{F1}_{jk}$ is the mean F1 of vowel *j* for the *k*th subject. $\mathrm{F2}'_j$ is the group mean of F2 for all NH children for vowel *j* and $\mathrm{F2}_{jk}$ is the mean F2 of vowel *j* for the *k*th subject.

III. RESULTS

Figure 1 shows vowel durations of the CI and NH children for each of the seven Mandarin vowels. In general, the CI children produced all seven vowels with longer durations than did the NH children. This finding is consistent with previous findings of segmental lengthening in speech of hearing-impaired speakers (Pratt and Tye-Murray, 1997) and children with CIs (Uchanski and Geers, 2003). In our data, both the NH children and the CI children produced the vowel [y] with the longest duration and the vowel [ɿ] with the shortest duration. A two-way mixed-design analysis of variance (ANOVA) was conducted to examine the effects of group (the between-subject effect) and vowel (the within-subject effect) on the vowel duration. The results revealed significant main effects of group [$F(1,72) = 34.97$, $p < 0.001$] and vowel [$F(6,432) = 5.22$, $p = 0.001$]. The effect of vowel quality on duration was reported in previous literature (e.g., Jacewicz *et al.*, 2007) and therefore was not specifically addressed here. The results revealed no significant interaction effect, which indicates that the CI children exhibited a pattern of vowel duration similar to the NH children.

Figure 2 shows the scatter plots of rescaled normalized F1 by F2 at the vowel midpoint with each ellipse encircling ∼95% of the tokens of each vowel category. As shown in the left panel for the NH children, among these seven Mandarin vowels, [i, a, u, y, ɤ] were relatively well separated in the F1 × F2 acoustic vowel space. However, the other two vowels [ɿ, ʅ] showed some degree of acoustic overlap. Note that this overlap is also present in Mandarin-speaking adults (see Wu, 1986). Thus, the similar positional distribution of the vowels [ɿ, ʅ] in the NH children may reflect the standard form of acoustical features in these two vowels. The CI children acoustically separated the three peripheral vowels [i, a, u] very well (Fig. 2, right panel). However, the other non-peripheral vowels [y, ɤ, ɿ, ʅ] displayed substantial overlaps in the acoustic vowel space. In particular, the vowels [y] and [ʅ] were almost completely overlapped.

The sizes of vowel ellipses of the CI children were substantially larger than those of the NH children. Except for the vowel [a], the ellipse areas of the CI children were 2–6

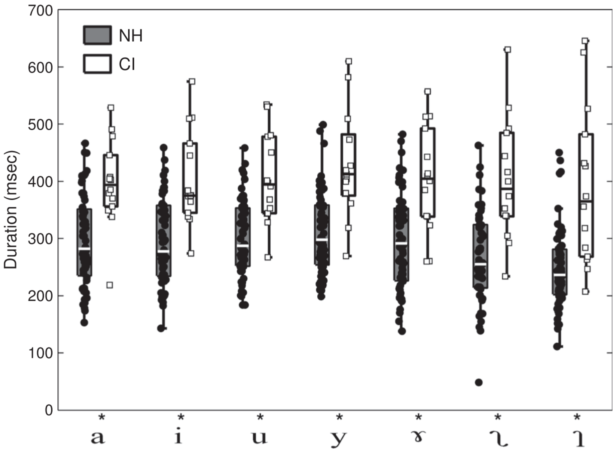

FIG. 1. Box plot showing vowel duration produced by NH and CI children for each of the seven Mandarin vowels. Each data point represents one subject. Some jitters along the abscissa were applied to the data for better visual representation. Each box shows horizontal lines at the lower quartile, median, and upper quartile values. The whiskers show 99.3% of the data and the data points out of the whiskers were outliers. The asterisks indicate statistically significant differences at the level of $p < 0.05$.

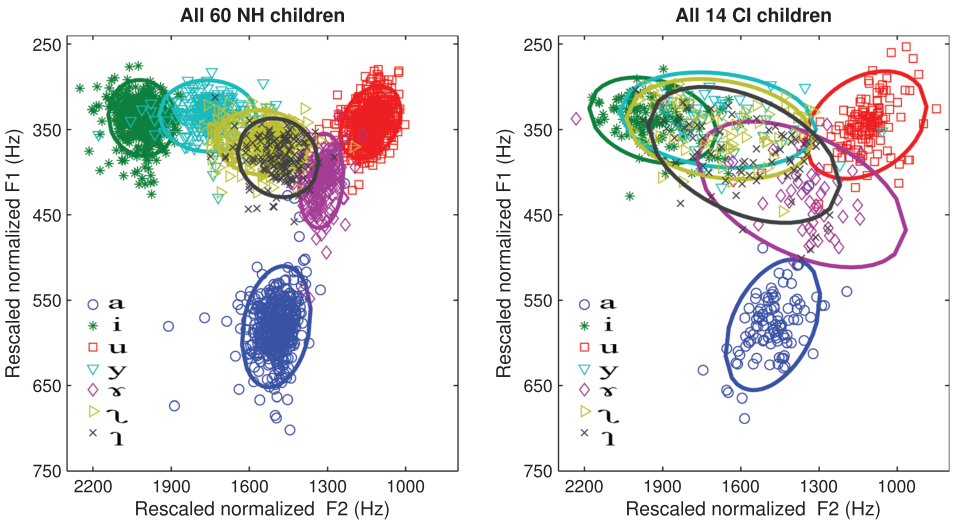

FIG. 2. Scatter plots based on rescaled normalized midpoint F1 and F2 values of seven Mandarin vowels. Data from the NH and CI children are shown on the left and right panel, respectively. Different symbols and colors represent different vowel categories and each data point represents one production. Each ellipse encompasses ∼95% of data points for each vowel category.

times larger than those of the NH children (Fig. 3). These results demonstrated that the CI children produced Mandarin vowels less consistently than did the NH children. When we further examined the spread pattern of each vowel category in the CI children, we found that compared to the NH children, the variation of the CI children's acoustic vowel categories was mainly reflected along the F2 axis (Fig. 2). Note that F2 is associated with the articulatory feature of tongue advancement. Therefore, the CI children's variation in vowel production resulted primarily from their inconsistency in the tongue position along the front-back dimension.

The overall size of the vowel space has been used to index speech intelligibility (Bradlow *et al.*, 1996), as well as developmental change as a function of vocal tract lengthening in children (Vorperian and Kent, 2007). In addition, it serves as an important acoustic measure in assessing the speech competency of speakers with various speech-language disorders because the boundary vowels used to define the acoustic vowel space represent the maximum articulatory positions that are reached by a speaker (Higgins and Hodge, 2001; Liu *et al.*, 2005). Following the common approach of defining vowel space as the area surrounded by boundary vowels (e.g., [i, æ, ɑ, u] in English), we selected the three Mandarin corner vowels [i, a, u] to calculate the Mandarin vowel space area (Yang, 2014) for both the CI and the NH children. Figure 4 (left panel) displays the mean formant values of each corner vowel for each participant in the CI and NH groups. A single triangle was plotted to connect the group mean formant values for each of the two groups of participants. It is evident from the position of the two triangles that the CI children, as a group, produced [i] at a more retracted position than did the NH children, although certain CI children demonstrated mean formant values within the normal range of NH children. One-way ANOVAs were conducted separately for each of the two formants to test whether the CI children differed from the NH children for these three corner vowels. The results confirmed that the CI children produced a significantly lower F2 compared to the NH children for the vowel [i] [$F(1,72) = 30.6$, $p < 0.001$]. Figure 4 (middle panel) shows the boxplots of the vowel space areas for the CI and the NH children. Because the assumption of homogeneity of variance was not satisfied, a two-sample Kolmogorov-Smirnov test was used to examine if the two groups differed in the vowel space area. The result showed that the CI group had a significantly smaller vowel space area than that of the NH group ($p = 0.013$), consistent with previous studies on the size of the vowel space in people with hearing loss (Monsen, 1976). However, when we observe each participant's data (Fig. 4, right panel), there was substantial individual variability in the vowel space areas among the CI children with areas differing by a factor of 2. We also noticed that 10 out of 14 CI children's vowel space areas were within the minimum-to-maximum range of values in the NH children. A strong negative correlation was found between the vowel space area and age at implantation ($r = -0.80$; z test, $p < 0.001$). However, no significant correlation was found between the length of device use and the vowel space area.

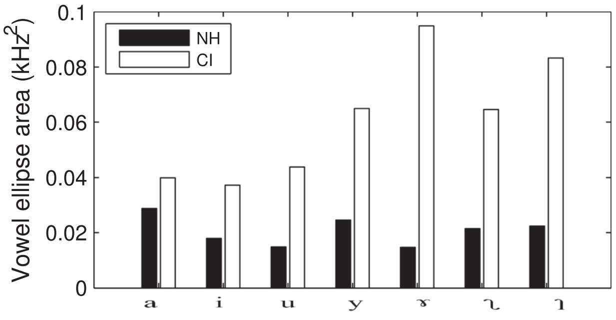

FIG. 3. Vowel ellipse area for individual Mandarin vowel categories in NH and CI groups.

Using the group mean formant values of the NH children as the formant target for each vowel, the AD between each CI or NH subject and the target is shown in Fig. 5. Smaller ADs represent greater positional proximity between the subject's vowel and the target, indicating greater similarity of the child's phonetic features relative to the target. Our results showed that as a group, the CI children produced larger ADs with greater variation for the vowels [i, u, y, ɤ, ɿ, ʅ] than did the NH children. In particular, for the NH children, the production of [i, a, u, ɤ] yielded a mean

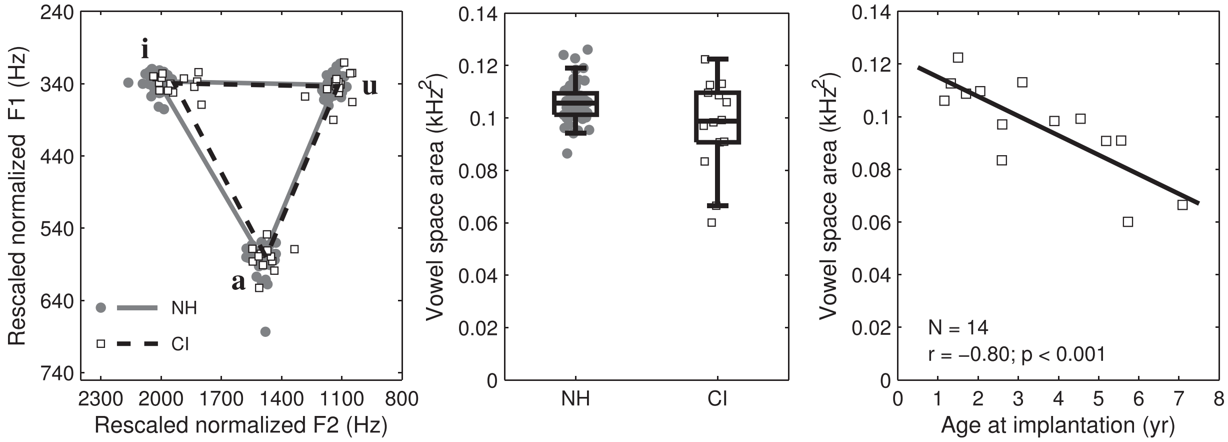

FIG. 4. Mandarin vowel space area of the NH and CI children. The left panel shows the mean F1 and F2 (rescaled normalized values) of the three corner vowels [a, i, u]. The filled circles and open squares represent data from the NH and CI children, respectively. Each data point (circle or square) represents the mean F1 and F2 for one subject. Solid and dashed lines represent the average vowel space formed by the mean of rescaled normalized F1 and F2 of the three corner vowels for the NH and CI groups, respectively. The middle panel shows the vowel space area of each subject in the NH and CI groups. Each data point represents one subject. Each box shows horizontal lines at the lower quartile, median, and upper quartile values. The whiskers show 99.3% of the data and the data points out of the whiskers are outliers. Some jitters along the abscissa were applied to the data for better visual representation. The right panel shows the correlation between vowel space area of the CI subjects and their age at implantation. Each data point represents one subject. The solid line is the least-squared fitting line of the data.

AD < 40 Hz and the production of [y, ɿ, ʅ] yielded a mean AD < 60 Hz. The shorter ADs for the vowels [i, a, u, ɤ] relative to those of the vowels [y, ɿ, ʅ] indicated that the NH children produced the vowels [i, a, u, ɤ] closer to the target. This result matches the acquisition order of vowels in normally developing Mandarin-speaking children. For the CI children, their production of [a, u] yielded a mean AD < 50 Hz. The production of the vowels [i, ɤ] yielded a mean AD between 50 and 100 Hz, whereas that of the vowels [y, ɿ, ʅ] yielded a large mean AD > 100 Hz (note that the mean AD of [y] was pulled up by one outlier). These observations were consistent with what has been demonstrated in Figs. 2 and 4. A two-way mixed-design ANOVA on these values yielded a significant difference between groups [F(1,72) = 46.42, $p < 0.001$] and among vowel categories [F(6,432) = 27.91, $p < 0.001$] as well as a significant interaction effect [F(6,432) = 11.45, $p < 0.001$]. *Post hoc* pair-wise comparisons of the vowel categories demonstrated that other than the vowel contrasts of [a]-[u], [i]-[ɤ], [y]-[ɤ], [y]-[ʅ], and [ʅ]-[ɿ], all other vowel contrasts showed significant differences in AD ($p < 0.001$). The significant interaction effect indicated that the CI children demonstrated a different pattern of AD across the vowel categories from the NH children. To further compare the difference between the CI and the NH children on each vowel, Student's *t*-tests were conducted on the AD for group differences of individual vowels. The results revealed a significant difference between the CI and NH children for the vowels [i] ($t = -2.64$, $p = 0.02$), [ɤ] ($t = -2.66$, $p = 0.019$), [ʅ] ($t = -3.36$, $p = 0.005$), and [ɿ] ($t = -3.72$, $p = 0.002$). While we observed a significant difference between the CI and NH groups, we also noticed that for certain vowels, such as [u] and [y], around 70% of data points in the CI group were located within the range of values of the NH group. For the vowel [a], all data points except for the one outlier in the CI group were located within the minimum-to-maximum range of values of the NH group. These observations revealed that there were a number of CI children who approximated the target value of NH children for certain vowels.

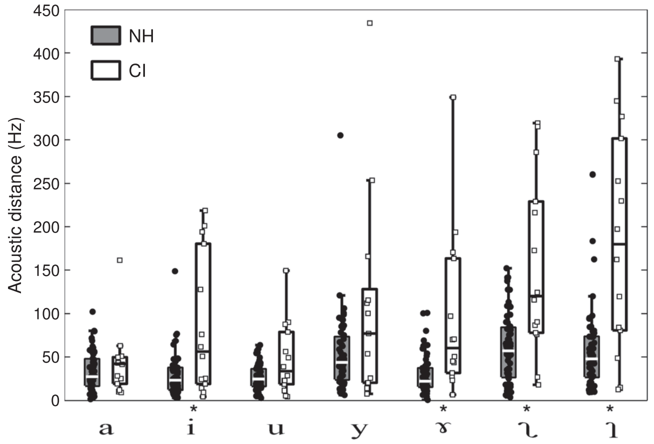

FIG. 5. Box plot showing AD of the CI and NH children for individual Mandarin vowel categories. The AD refers to the Euclidean distance between the coordinates of a CI or NH child's vowel production and the group mean coordinates of the corresponding vowel in the NH children in the F1 × F2 space. Each data point represents one subject. Some jitters along the abscissa were applied to the data for better visual representation. Each box shows horizontal lines at the lower quartile, median, and upper quartile values. The whiskers show 99.3% of the data and the data points out of the whiskers are outliers. The asterisks indicate statistically significant differences at the level of $p < 0.05$.

IV. DISCUSSION

The present study compared temporal and spectral measures of vowel production in Mandarin-speaking, prelingually deafened children with CIs and the corresponding acoustic measures in similarly aged NH children. As a group, the CI children showed a significantly different vowel acoustic profile from the age-matched NH children. First, the CI children produced a significantly longer duration for all seven Mandarin vowels than the NH children (Fig. 1). Second, the CI children's acoustic vowel categories were less defined and the vowel productions in each category except for the vowel [a] were highly scattered compared to those produced by the NH children (Figs. 2 and 3). Third, the CI children showed significantly reduced vowel spaces compared to the NH children (Fig. 4). Fourth, based on the measurement of AD, the CI children produced the vowels [a] and [u] closest to those of the NH children and produced the vowels [ʅ] and [ɿ] most differently from those of the NH children (Fig. 5).

In contrast to the findings of Uchanski and Geers (2003) that English-speaking CI children's vowel production reached or closely approximated the NH norms, our data showed differences of vowel acoustic features between the Mandarin-speaking CI children and the NH peers. Uchanski and Geers measured the F2 value of two selected vowels [i] and [ɔ], as well as different vowels' duration at various sentence positions. Our study, instead, included more acoustic measures for a larger number of vowel sounds. We did find some CI children within the range of the NH children on certain measures (F1, F2, vowel space area, or AD) for certain vowels ([a] and [u]). However, we also found clear group differences between the CI and NH children on the acoustic measures for the majority of the seven vowels. As shown in Fig. 5, the significant interaction effect between the group and vowel categories on the measure of AD indicated that the CI children demonstrated a different pattern of AD across the vowel categories from the NH children. The discrepancy between our results and the findings in Uchanski and Geers (2003) may be partly attributed to the length of CI use and rehabilitation. The participants in Uchanski and Geers (2003) had at least 3 yr length of device use and some children had 7 yr of CI experience. In the present study, our participants had an average of 1.7 yr of CI experience. Therefore, the present study provides evidence of early effects of CI on speech production in prelingually deafened children. That is, while a short period of CI experience can sufficiently change the deaf children from nonverbal to verbal and modify the acoustic characteristics of their speech production, sustained implant use experience combined with auditory-oral rehabilitation plays an important role in refining their production.

Compared to NH children, CI children receive the auditory experience provided by cochlear implants at a relatively later developmental period and thus have considerably fewer opportunities to practice speech production. In addition, CI children do not gain access to the full range of auditory cues available to NH children. Given the important role of auditory feedback in speech motor programming (Perkell *et al.*, 2000; Goffman *et al.*, 2002) and the relatively short experiences with CIs for the participants in the present study, the development of speech motor control in the CI children was most likely to be delayed relative to the NH children. In this case, the CI children may take a longer time to arrive at the target gesture of the intended vowel even though they may show a similar distributional pattern of overall vowel duration as NH children.

While the CI children produced Mandarin vowels differently from the NH children, they were not consistently different from the NH children in each of the Mandarin vowels. Indeed, they produced some vowels such as [a], [i], and [u] more similarly to the NH norms than other vowels although the vowel [i] showed relatively large individual differences in certain CI children who produced this vowel with a more scattered distribution. As shown in Figs. 2 and 4, the three peripheral vowels ([i, a, u]) produced by the CI children were clearly separated and organized in a triangular shape in the acoustic vowel space, which contributed to a roughly comparable shape and position of the vowel frame to that of the NH children. Moreover, as shown in Fig. 5 for the AD, the CI children's productions of vowels [a] and [u] were very similar to those of the NH children. These results suggested that the CI children approximated the NH norms and had mastered the Mandarin corner vowels relatively well.

In contrast, the other four non-corner vowels showed a substantial deviation from the NH norms. As shown in Fig. 2, the introduction of non-corner vowels into the acoustic space caused substantial overlaps among vowels. These overlaps, to a large extent, were associated with variation in the F2 dimension. The unstable F2 in the CI children reflected their inconsistency in the tongue position along the front-back dimension. This finding may reflect the availability of fewer visible cues associated with F2 variation as opposed to F1 variation (Neumeyer *et al.*, 2010). F1 variation involves the change of tongue movement along the high-low dimension that is associated with the degree of mouth openness and mandible movement. These features are relatively easy to infer visually. In contrast, F2 variation that is substantially associated with the change of tongue movement along the front-back dimension is less visible. Unlike NH children who can rely on audible cues to compensate for the less visible cues for F2 variation, CI children are less likely to accurately perceive F2 changes. However, for the vowel [u] that is characterized by the feature of lip-rounding, which also affects F2, the CI children showed more consistent production along F2 dimension than the non-corner vowels because they can access the visible cue of lip-rounding. On the other hand, we also observed that a large number of vowel tokens (except for the vowels [a] and [u]) produced by the CI children were located in a high and front-to-central region (Fig. 2) as opposed to relatively well separated vowel categories in the NH children. This result suggested that unlike the NH children who used well differentiated tongue shapes to produce these vowels, the CI children used similar tongue shapes, in particular, a relatively centralized tongue position when producing these vowels. It is worth noting that previous research has found that hearing-impaired children tend to have a more centralized and less differentiated tongue position in the production of English vowels as compared to their NH peers (Dagenais and Critz-Crosby, 1992). Therefore, given the duration of CI use (on average 1.7 yr in the present study), the prelingually deafened children had not been able to acquire correct tongue positions for many of those vowels. Since the acoustic development of vowel production in normally developing children is a long-term process, which does not reach the adult form even at 6 years of age (cf. Yang, 2014, for the acoustic development of Mandarin-speaking children; see also To *et al.*, 2013 for

Cantonese vowel development), it would be interesting to examine a group of CI children's vowel development in a longitudinal study to document the developmental trajectory of acoustic features in the CI population.

Other than the large variation along the F2 axis, the substantially greater vowel ellipse areas (Figs. 2 and 3) and the significantly greater ADs (Fig. 5) for the vowels [y, ɤ, ʅ, ɿ] in the CI children also demonstrated greater deviation in the production of non-corner vowels relative to the NH children. The more scattered vowel dispersion in the acoustic space suggests less defined vowel categories. This finding indicates that the CI children, as a group, were inconsistent in their production of these non-corner vowels. Correspondingly, the ADs of these vowels from the NH norms were larger and showed greater variability compared to the ADs of the corner vowels [a] and [u]. These findings support our prediction that CI children would also demonstrate a certain order of vowel development even though this order may not completely match with that of normally developing children.

The later development of non-corner vowels can be partly explained by the less salient perceptual cues in the non-corner vowels relative to those of corner vowels. The corner vowels are relatively well separated in the vowel space and are characterized by distinctive formant patterns in terms of F1 and F2 values. Perceptually, the corner vowels are easier to differentiate and perceive than the non-corner vowels (Schwartz *et al.*, 2005). In addition, the frequency of occurrence of each vowel also varies in Mandarin. In general, the non-corner Mandarin vowels have relatively low frequency of occurrence compared to the corner vowels (Suen, 1982; Thomas, 2005). For instance, the two allophones of the vowel /i/ (i.e., [ʅ] and [ɿ]) in Mandarin Chinese occur in a restricted consonant environment. In contrast, the corner vowels [a, i, u] not only occur independently in the form of monophthongs in Mandarin syllables, they are also widely used to form diphthongs and to combine with nasal codas in Mandarin syllables. Therefore, CI children may have less opportunity to gain language input containing the less frequent non-corner vowels relative to the corner vowels. In addition, from the signal processing perspective, CI stimulations provide coarse spectral contrasts (e.g., Xu and Pfingst, 2008). Such coarse spectral cues appear to be sufficient to provide perception of corner vowels. Non-peripheral vowels with less differentiated formants have less salient perceptual contrasts under CI stimulations.

Overall, the present study provides data to supplement previous studies and expands our understanding of the vowel development of prelingually deafened children with CIs in a Mandarin-speaking environment. In addition, it has implications in guiding clinical aural habilitation in young children with CIs. That is, aural habilitation should focus more on those vowels that show greater dissimilarity from the NH norms. It is noteworthy that we found a strong negative correlation between the vowel space area and age at implantation. Given the association between the size of the vowel space area and speech intelligibility (Bradlow *et al.*, 1996), this finding highlights the role of early intervention with CIs in improving the speech intelligibility in prelingually deaf children. Meanwhile, we also realize the limitations involved with the small sample size and the design of the word list. The present study did not control the consonant environment. However, a recent study (Warner-Czyz *et al.*, 2010) found that children (both NH and CI) produced CV syllables with articulatory compatibility (compatible place of articulation between the consonant and the vowel) with higher accuracy than those with articulatory diversity. This finding suggests that CI children's vowel production may also be affected by the adjacent consonant due to coarticulation effects. Clearly, future work is needed to recruit a larger number of CI participants and to develop a better controlled word list. The coarticulation between consonants and vowels is worth further exploration because it involves the coordination of articulators that reflects a speaker's motor planning and timing control skills. In addition, the vowel dynamic spectral change is of interest because it may show distinctive formant movement patterns over the course of vowel duration (Fox and Jacewicz, 2009).

ACKNOWLEDGMENTS

Xiuwu Chen and Xiaoyang Zhao provided assistance in data collection in Beijing, China. Alex Fickey, Robert Garrett, Yitao Mao, Tiffany McDonald, Alexa Patton, and Brianna Smith provided technical support in data analysis and in preparation of the manuscript. The study was supported in part by an NIH NIDCD Grant No. R15-DC014587.

APPENDIX

The word list used for Mandarin vowel production is shown. Each word was produced in tone 1 and tone 2 except for "si" and "ci" (only available in tone 1). An additional three disyllabic words "fangzi," "shizi," and "zongzi" containing the vowel /ɿ/ in "zi" were used for the CI children.

Vowel	Word (pinyin)	Word (International Phonetic Alphabet)
/i/	bi	/pi/
	pi	/p^{h}i/
	ji	/ɕi/
	qi	/tɕ^{h}i/
	yi	/ji/
/ɿ/	si	/sɿ/
	ci	/ts^{h}ɿ/
/ʅ/	zhi	/tʂʅ/
	chi	/tʂ^{h}ʅ/
	shi	/ʂʅ/
/a/	fa	/fa/
	la	/la/
	ma	/ma/
	na	/na/
/u/	du	/tu/
	tu	/t^{h}u/
	wu	/wu/
	fu	/fu/
	hu	/hu/
/y/	xu	/ɕy/
	yu	/jy/
/ɤ/	ge	/kɤ/
	ke	/k^{h}ɤ/

Adank, P., Smits, R., and van Hout, R. (**2004**). "A comparison of vowel normalization procedures for language variation research," J. Acoust. Soc. Am. **116**, 3099–3107.

Angelocci, A. A., Kopp, G. A., and Holbrook, A. (**1964**). "The vowel formants of deaf and normal-hearing eleven- to-fourteen-year-old boys," J. Speech Hear. Disord. **29**, 156–160.

Blamey, P. J., Barry, J. G., and Jacq, P. (**2001**). "Phonetic inventory development in young cochlear implant users 6 years postoperation," J. Speech Lang. Hear. Res. **44**, 73–79.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech. I. Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Chuang, H. F., Yang, C. C., Chi, L. Y., Weismer, G., and Wang, Y. T. (**2012**). "Speech intelligibility, speaking rate, and vowel formant characteristics in Mandarin-speaking children with cochlear implant," Int. J. Speech Lang. Pathol. **14**, 119–129.

Connor, C. M., Craig, H. K., Raudenbush, S. W., Heavner, K., and Zwolan, T. A. (**2006**). "The age at which young deaf children receive cochlear implants and their vocabulary and speech-production growth: Is there an added value for early implantation?," Ear Hear. **27**, 628–644.

Dagenais, P. A., and Critz-Crosby, P. (**1992**). "Comparing tongue positioning by normal-hearing and hearing-impaired children during vowel production," J. Speech Hear. Res. **35**, 35–44.

Duanmu, S. (**2007**). *The Phonology of Standard Chinese* (Oxford University Press, New York), pp. 35–39.

Ertmer, D. J. (**2001**). "Emergence of a vowel system in a young cochlear implant recipient," J. Speech Lang. Hear. Res. **44**, 803–813.

Fox, R. A., and Jacewicz, E. (**2009**). "Cross-dialectal variation in formant dynamics of American English vowels," J. Acoust. Soc. Am. **126**, 2603–2618.

Geers, A., Tobey, E., Moog, J., and Brenner, C. (**2008**). "Long-term outcomes of cochlear implantation in the preschool years: From elementary grades to high school," Int. J. Audiol. **47**, S21–S30.

Goffman, L., Ertmer, D. J., and Erdle, C. (**2002**). "Changes in speech production in a child with a cochlear implant: Acoustic and kinematic evidence," J. Speech Lang. Hear. Res. **45**, 891–901.

Hare, G. (**1983**). "Development at 2 years," in *Phonological Development in Children: 18–72 months*, edited by J. V. Irwin and S. P. Wong (Southern Illinois University Press, Carbondale, IL), pp. 55–88.

Higgins, C. M., and Hodge, M. M. (**2001**). "F2/F1 vowel quadrilateral area in young children with and without dysarthria," Can. Acoust. **29**, 66–68.

Horga, D., and Liker, M. (**2006**). "Voice and pronunciation of cochlear implant speakers," Clin. Linguist. Phonet. **20**, 211–217.

Jacewicz, E., Fox, R. A., and Salmons, J. (**2007**). "Vowel duration in three American English dialects," Am. Speech. **82**, 367–385.

Liu, H.-M., Tsao, F-M., and Kuhl, P. K. (**2005**). "The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy," J. Acoust. Soc. Am. **117**, 3879–3889.

Lobanov, B. M. (**1971**). "Classification of Russian vowels spoken by different listeners," J. Acoust. Soc. Am. **49**, 606–608.

Löfqvist, A., Sahlen, B., and Ibertsson, T. (**2010**). "Vowel space in Swedish adolescents with cochlear implants," J. Acoust. Soc. Am. **128**, 3064–3069.

Manning, W. H., Moore, J. N., Dunham, M. J., Lu, F. L., and Domico, E. (**1992**). "Vowel production in prelinguistic child following cochlear implantation," J. Am. Acad. Audiol. **3**, 16–21.

Milenkovic, P. (**2003**). "TF32 software program" (University of Wisconsin, Madison, WI), http://userpages.chorus.net/cspeech/ (Last viewed 30 September 2015).

Monsen, R. (**1976**). "Normal and reduced phonological space: The production of English vowels by deaf adolescents," J. Phonet. **4**, 189–198.

Neumeyer, V., Harrington, J., and Draxler, C. (**2010**). "An acoustic analysis of the vowel space in young and old cochlear-implant speakers," Clin. Linguist. Phonet. **24**, 734–741.

Nittrouer, S., Caldwell-Tarr, A., and Lowenstein, J. H. (**2013**). "Working memory in children with cochlear implants: Problems are in storage, not processing," Int. J. Pediatr. Otorhinolaryngol. **77**, 1886–1898.

Perkell, J. S., Guenther, F. H., Lan, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., and Zandipour, M. (**2000**). "A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss," J. Phonet. **28**, 233–272.

Perkell, J., Numa, W., Vick, J., Lane, H., Balkany, T., and Gould, J. (**2001**). "Language specific, hearing-related changes in vowel spaces: A preliminary study of English- and Spanish-speaking cochlear implant users," Ear Hear. **22**, 461–469.

Pratt, S., and Tye-Murray, N. (**1997**). "Speech impairment secondary to hearing loss," in *Clinical Management of Sensorimotor Speech Disorders*, edited by M. McNeil (Thieme Medical Publishers, Inc., New York), pp. 345–387.

Ruffin, C. V., Kronenberger, W. G., Colson, B. G., Henning, S. C., and Pisoni, D. B. (**2013**). "Long-term speech and language outcomes in prelingually deaf children, adolescents and young adults who received cochlear implants in childhood," Audiol. Neurotol. **18**, 289–296.

Schwartz, J. L., Abry, C., Boë, L. J., Menard, L., and Vallee, N. (**2005**). "Asymmetries in vowel perception, in the context of the Dispersion–Focalisation Theory," Speech Commun. **45**(4), 425–434.

Shi, F., and Wen, B. (**2007**). "Vowel development of Mandarin-speaking children," Zhongguo Yuwen **5**, 444–454.

Stoel-Gammon, C., and Herrington, P. B. (**1990**). "Vowel systems of normally developing and phonologically disordered children," Clin. Ling. Phonet. **4**, 145–160.

Suen, C.-Y. (**1982**). "Computational analysis of Mandarin sounds with reference to the English language," in *Proceedings of the 9th Conference on Computational Linguistics*, Vol. 1, pp. 371–376.

Thomas, E. R., and Kendall, T. (**2007**). "NORM: The vowel normalization and plotting suite," http://lvc.uoregon.edu/norm/about_norm1.php#scaling (Last viewed 27 August 2015).

Thomas, T. W. C. (**2005**). "The effects of occurrence frequency of phonemes on second language acquisition: A quantitative comparison of Cantonese, Mandarin, Italian, German and American English," Chinese University of Hong Kong. Available at http://www.thomastsoi.com/wp-content/downloads/The%20Effects%20of%20Occurrence%20Frequency%20of%20Phonemes%20on%20SLA.pdf (Last viewed 30 September 2015).

To, C. K., Cheung, P. S., and McLeod, S. (**2013**). "A population study of children's acquisition of Hong Kong Cantonese consonants, vowels, and tones," J. Speech Lang. Hear. Res. **56**, 103–122.

Tobey, E. A., Geers, A. E., Brenner, C. B., Altuna, D., and Gabbert, G. (**2003**). "Factors associated with development of speech production skills in children implanted by age five," Ear Hear. **24**, 36S–45S.

Tye-Murray, N. (**1992**). "Articulatory organizational strategies and the roles of audition," Volta Rev. **94**, 243–260.

Uchanski, R. M., and Geers, A. E. (**2003**). "Acoustic characteristics of the speech of young cochlear implant users: A comparison with normal-hearing age-mates," Ear Hear. **24**, 90S–105S.

Vorperian, H. K., and Kent, R. D. (**2007**). "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data," J. Speech Hear. Res. **50**(6), 1510–1545.

Warner-Czyz, A. D., Davis, B., and MacNeilage, P. F. (**2010**). "Accuracy of consonant-vowel syllables in young cochlear implant recipients and hearing children in the single-word period," J. Speech Lang. Hear. Res. **53**, 2–17.

Wu, Z. J. (**1986**). *The Spectrographic Album of Monosyllables of Standard Chinese* (Chinese Academy of Social Sciences, Beijing), pp. 6–33.

Xu, L., and Pfingst, B. E. (**2008**). "Spectral and temporal cues for speech recognition: Implications for auditory prostheses," Hear. Res. **242**, 132–140.

Yang, J. (**2014**). "Acoustic properties of vowel production in Mandarin-English bilingual and corresponding monolingual children," Doctorate dissertation, The Ohio State University.

Zee, E. (**2001**). "The phonetic value of the vowels, diphthongs and triphthongs in Beijing Mandarin," in *Proceeding of 5th National Conference on Modern Phonetics*, Beijing, pp. 54–60.

Zhou, N., and Xu, L. (**2008**). "Development and evaluation of methods for assessing tone production skills in Mandarin-speaking children with cochlear implants," J. Acoust. Soc. Am. **123**, 1653–1664.

Zhu, H., and Dodd, B. (**2000**). "The phonological acquisition of Putonghua (Modern Standard Chinese)," J. Child Lang. **27**, 3–42.