Interspeech 2022
18-22 September 2022, Incheon, Korea

A Study of Production Error Analysis for Mandarin-speaking Children with Hearing Impairment

Jingwen Cheng[1], *Yuchen Yan*[1], *Yingming Gao*[1], *Xiaoli Feng*[1], *Binghuai Lin*[2], *Jinsong Zhang*[1]

[1]School of Information Sciences, Beijing Language and Culture University, Bejing, China
[2]Smart Platform Product Department, Tencent Technology Co., Ltd, Shenzhen, China

chengjingwen19@163.com, yanyuchen1999@foxmail.com, yingming.gao@mailbox.tu-dresden.de, fengxiaoli314@163.com, binghuailin@tencent.com, jinsong.zhang@blcu.edu.cn

Abstract

Investigating the speech acquisition of hearing-impaired children attracts considerable attentions in recent years. Previous studies that investigate Mandarin-speaking children with hearing impairment mostly focus on production of some specific phonemes. Besides, the phonemes are sometimes embedded in a limited number of speech materials or uttered by only a few speakers. In this study, we analyzed the pronunciation errors of all Mandarin vowels and consonant produced by 60 pre- or post-lingually hearing-impaired children. We designed a set of speech materials that consisted of 153 monosyllable and 145 disyllable commonly used words and had a comprehensive phonetic coverage. The analysis shows that monophthongs were produced less accurately than diphthongs and triphthongs. Bilabials and nasals and plosives were relatively easier for hearing-impaired children to acquire than other consonants with respect to articulation manner and place, respectively. The Mandarin affricates had the lowest accuracy. Substitution is the most frequent error patterns for initial consonants while deletion is the common error for final consonants. The findings of this study can shed light on pronunciation teaching of hearing-impaired children. Besides, the corpus can benefit developing computer-assisted speech assessment system.

Index Terms: hearing impairment, Chinese children, Mandarin speech, pronunciation error patterns

1. Introduction

Hearing impairment is a disease that affects people's hearing function and is always accompanied by a reduced sensitivity to sounds normally perceived [1]. Children acquire language by accessing the auditory input of their native language. The language and speech development of hearing-impaired children may be delayed due to the deprivation of sufficient auditory input as perceived by healthy people [2]. Investigating the speech acquisition attracts considerable attentions in recent years.

Exploring the speech development of hearing-impaired children is usually based on the descriptive analysis of phonetic transcription. Their production data are examined in terms of pronunciation accuracy and error patterns [3, 4]. Studies that focused on the phonemic acquisition accuracy of English-speaking hearing-impaired children found that more visible consonants were likely to have a higher production accuracy. For example, bilabials and labiodentals were produced more correctly than the other consonants. With respect to articulation manner, fricatives and affricates were the least accurately produced. Vowels with a central articulation place tended to be produced more correctly than the other vowels [5, 6]. And the phonological error patterns observed in English-speaking hearing-impaired children usually include substitution, initial consonant deletion, consonant cluster reduction, final consonant deletion, diphthongization, and vowel neutralization [6, 7].

There also exist studies that investigated speech acquisition of Mandarin-speaking children with hearing impairment. For example, Peng et. al. studied the pronunciation accuracy of 21 initial consonants produced by 30 hearing-impaired children [8]. The results showed that the mean accuracy across all consonants was 57.9%. And in terms of articulation manner, plosives and nasals had significantly higher accuracies than other consonants. Yang et. al. investigated the accuracy and error patterns of four fricatives in Mandarin Chinese [9]. The researchers recruited 14 hearing-impaired children and selected 13 Mandarin words as the speech materials. The results showed that fricatives f [f] and x [ɕ] tended to have higher accuracies than s [s] and sh [ʂ]. The common error patterns include place errors and substitution of the target fricative with plosives or homorganic affricates. Similarly, Yu et. al. studied the production of six plosives produced by 20 hearing-impaired children [10]. Each child uttered 23 monosyllabic words through an imitation task. The researchers found that the plosives with a bilabial articulation place tended to have the lowest accuracy, followed by velar plosives and alveolar plosives. One more comprehensive study was conducted by Zan et. al. [11]. The researchers studied the phonological development through a phonologic speech evaluation and a picture-naming task. The results showed that the main error patterns produced by children with ages ranging from 4 to 11 years old include deletion, aspiration errors, place errors, substitution of affricates and fricatives with plosives, and diphthong or triphthong reduction. Besides that, some researchers also studied the production of Mandarin tones [12, 13].

So far, previous studies mostly focused on the Mandarin consonants or tones, whereas only a few studies focused on the vowels. Another limitation is most studies have used a few speech materials or examined speech data produced by a very small number of children. And for the limitation of methodological factors, there are some controversial conclusions [8, 10]. In this study, we first developed a speech corpus of Mandarin-speaking children with hearing impairment. And then a preliminary study was conducted to analyze the accuracy and error patterns of Mandarin consonants and vowels.

2. Corpus design

2.1. Mandarin Chinese phonology

Mandarin Chinese (Standard Chinese) is a well-known syllable-based tone language. Each syllable consists of an indispensable vowel and optional consonants that can be attached in the initial and final position of the syllable. Therefore, there are four pos-

Copyright © 2022 ISCA

10.21437/Interspeech.2022-10477

sible syllable structures: V, CV, VC, and CVC. Each syllable is also associated with one of five pitch tones, including four lexical tones and a neutral tone.

There are 22 consonants and 22 vowels, which are listed in Table 1. They are presented in Pinyin with corresponding IPA symbols in square brackets. According to articulation place, the consonants can be classified into seven categories: bilabial, labiodental, dental, alveolar, retroflex, palatal, and velar. They can also be classified into six groupings with regard to articulation manner: plosive, affricate, fricative, nasal, approximant and lateral approximant. Aspiration also serves as a distinctive feature for Mandarin consonants, forming six pairs of aspirated and unaspirated consonants. It should be noted that, except for nasal ng [ŋ], all the other consonants can occur in syllable initial position, and only the nasals n [n] and ng [ŋ] can occur in syllable-final position.

Table 1: *Mandarin consonants and vowels*

manner \ place	bilabial	labiodental	dental	alveolar	retroflex	palatal	velar
plosive	b[p]			d[t]			g[k]
	p[p^h]			t[t^h]			k[k^h]
affricative			z[ts]		zh[tʂ]	j[tɕ]	
			c[ts^h]		ch[$tʂ^h$]	q[$tɕ^h$]	
fricative		f[f]	s[s]		sh[ʂ]	x[ɕ]	h[x]
nasal	m[m]			n[n]			ng[ŋ]
approximant					r[ɻ]		
lateral				l[l]			
monophthong	a[a] i[i] u[u] o[o] e[ɤ] ü[y] er[ɚ] -i[ɿ] -i[ʅ]						
diphthong	ai[ai] ei[ei] ao[ao] ou[ou] ia[ia] ua[ua] ie[iɛ] üe[yɛ] uo[uo]						
triphthong	iao[iao] iu[iou] ui[uei] uai[uai]						

2.2. Participants

The speech samples in this corpus were collected from 60 hearing-impaired children (30 males and 30 females). Their chronological ages at the time of recording ranged from 3 to 12 years old (mean: 6.15 years). All of them had received rehabilitation at professional rehabilitation centers with the amount of rehabilitation varying. However, their demographic information was not completely collected. According to the available information, most children had a prelingual hearing loss and a few children had a postlingual hearing loss. All the children had received hearing reconstruction and worn cochlear implants or hearing aid devices.

2.3. Speech materials

The speech materials include 298 words with a length of one or two syllables. Both syllable-length words were designed to have a complete phonetic coverage, which include all the phonemes (22 consonants and 22 vowels) and four lexical tones for monosyllable and 20 tonal combinations for disyllable. Considering the difficulty of recording for children, we try to control the amount of speech material within an appropriate range and use words that children are more familiar with. The selected words are mostly common words in Chinese [14]. Finally, 153 monosyllables and 145 disyllables were selected for our recordings.

2.4. Recording

The speech recordings were recorded in quiet rooms. All the speech samples were collected by mobile phones. For each recording word, an audio prompt produced by a female native Mandarin speaker was first played to the children, then the participants were asked to repeat. Their teachers and parents were allowed to give instruction if the children failed to repeat.

2.5. Human annotation

Phonetic transcriptions were done in Pinyin, while tones were transcribed using numbers. Some diacritics were used to label sound variations, including insertion, deletion, ambiguous, and undefinable speech. Before the human annotation, all the recordings were automatically segmented into initials and finals (the Chinese traditional phonemes), using forced-alignment technique. And multi-level phonetic transcriptions were also automatically generated. Then, ten trained native Mandarin speakers were asked to check the phonetic segmentation boundaries and annotate the pronunciation. This work was done with the software "Praat" [15].

In order to check consistency in annotation, 698 speech records were selected to assess the agreement between annotators. The selected speech records were randomly divided into several sets. Each set was transcribed by three to four annotators. The consistency between each pair of annotators was calculated. The mean consistencies were 78.3%, 76.9%, and 81.8% for consonants, vowels, and tones, respectively.

3. Preliminary Analysis

3.1. Initial consonants

3.1.1. Accuracy

The mean accuracy of 21 initial consonants produced by 60 hearing-impaired children was 63.8% (SD: 15.6%). The initial consonants d, b, and m had the highest accuracy of above 80%, followed by the consonants t, p, n, h, g, j, f, k, x, q, and l with an accuracy between 62% and 74.9%. The consonants sh, zh, r, ch, s, z, and c had the lowest accuracy, which was between 28% and 57%.

With respect to articulation place, the mean score for correct consonant production was highest for bilabials (80.6%) and, in descending order, for alveolars (73.9%), velars (70.6%), palatals (66.9%), labiodentals (66.7%), retroflexes (55.2%) and dentals (34.2%). For articulation manner, nasals (77.8%) and plosives (76.5%) were produced most correctly, followed by lateral approximant (62%), fricatives (60.3%), approximant (54.9%), and affricates (51.2%).

3.1.2. Error patterns

The error patterns of initial consonants include substitution, deletion, and undefinable utterance (speech that is ambiguous or undefinable). The most common error type is substitution (mean: 26.7%), followed by deletion (mean: 6.1%) and undefinable utterance (mean: 3.4%). For all the initial consonants, substitution is the most frequent error, accounting for a much higher proportion than other error patterns. In previous studies, it has been found that hearing-impaired children show a certain regularity in substitution error. For example, they tend to use a phoneme that has a different articulation place or manner with the target consonant [6, 11]. In this section, the substitution error is analyzed in detail.

Figure 1 presents the confusion matrix of 21 initial consonants with regard to their substituted responses. In order to see the error trend clearly, the ratio on the diagonal of the matrix was set to zero. In terms of articulation place errors, dental consonants z, c, and s were predominantly substituted by

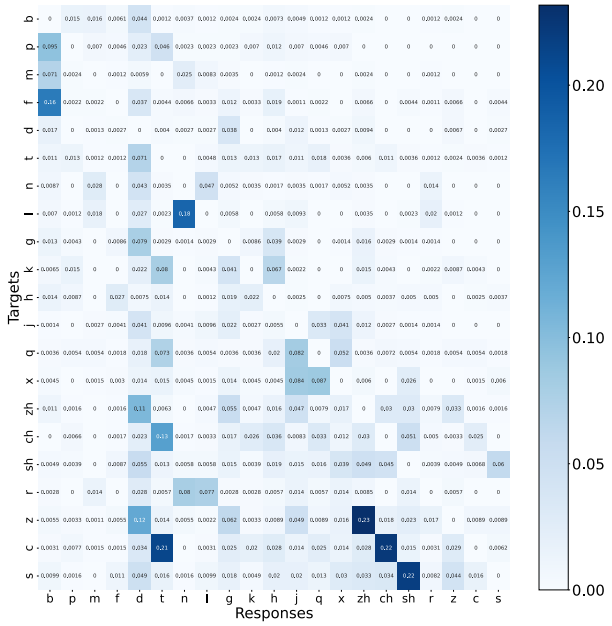


Figure 1: *The confusion matrix of initial consonants.*

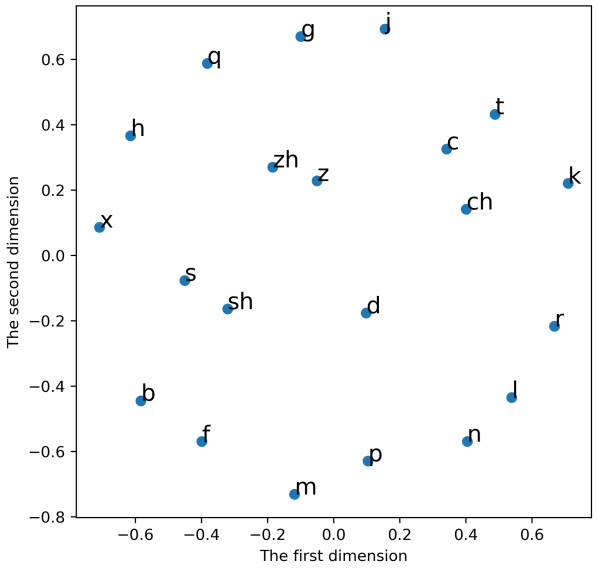


Figure 2: *The 2D distribution of initial consonants based on phonetic error*

retroflexes zh, ch, and sh, respectively. Similarly, plosives g, k, b, and d also tended to have place errors, such as velar plosives g and k were often substituted by alveolar plosives d and t. For aspiration errors, aspirated consonants p, t, and q tended to be replaced by their unaspirated counterparts b, d, and j. With respect to articulation manner errors, nasal m was often substituted by the homorganic plosive b. Lateral approximant l was often replaced by nasal n. Fricative x was commonly substituted by the homorganic affricatives j and q. Besides that, f was commonly substituted by b, and r was often substituted by n and l. Affricatives and fricatives were also frequently replaced by alveolar plosives d and t. Such as, unaspirated affricatives z, zh, and j were often substituted by d, and aspirated affricatives c, ch, and q were often substituted by t. Taken together, the most common errors produced for a particular target phoneme usually shared the same feature with the target, and this means that hearing-impaired children tend to achieve a close approximation of the target consonant.

In order to further intuitively analyze the error patterns of initial consonants, the multidimensional scaling (MDS) analysis was conducted to see the distribution feature of consonants based on the confusion matrix. It can be observed in Figure 2 that bilabial and labiodental consonants b, p, m, and f are close together in the lower left quadrant. These consonants have the same or close articulation place, and consonants f, m, and p are commonly substituted by b. In the lower right quadrant, nasal n, lateral approximant l, and approximant r are gathered together. They are all voiced consonants. In the middle of the quadrant, dentals and retroflexes are grouped into three pairs, including unaspirated affricatives z and zh, aspirated affricatives c and ch, and fricatives s and sh. Each pair has the same articulation manner but with a different articulation place, and the dentals z, c, s are often substituted by retroflexes zh, ch, sh. The plosive d has a middle position in the quadrant. This is because other consonants are very commonly substituted by it. However, for palatals j, q, x and velars g, k, h, they are not regularly gathered into groups.

Table 2: *Error patterns of final consonants*

	Deletion	**Substitution**	**Undefined**
n	20.8%(542)	5.3%(137)	3.6%(94)
ng	24.3%(579)	5.0%(119)	2.9%(69)
mean	22.6%	5.1%	3.3%

3.2. Final consonants

3.2.1. Accuracy

There are two final consonants in Mandarin Chinese, n and ng. The final consonant n was produced more accurately (70.3%) than ng (67.8%). The mean accuracy of two final consonants was 69.1%, which was higher than the initial consonants.

3.2.2. Error patterns

The error patterns of final consonants include deletion, substitution (n and ng substitute for each other), and undefinable utterance. Deletion is the most common error, followed by substitution and undefined utterance. Table 2 lists the percentage and the number of occurrences of each error pattern.

3.3. Vowels

3.3.1. Accuracy

The mean accuracy of 22 vowels produced by hearing-impaired children was 76.9%, which is higher than initial consonants and final consonants. The mean accuracies were 72.5% (SD: 11.4%), 80.6% (SD: 5.7%), and 78.5% (SD: 3.6%) for monophthongs, diphthongs, and triphthongs, respectively. The monophthongs had the lowest accuracy but the highest standard deviation compared with diphthongs and triphthongs. For some monophthongs, such as a, i, and e, their accuracy was above 80%, but for ii (refer to [ɿ]) and er, their accuracy was below 60%. It indicates that some monophthongs are still relatively difficult for hearing-impaired children.

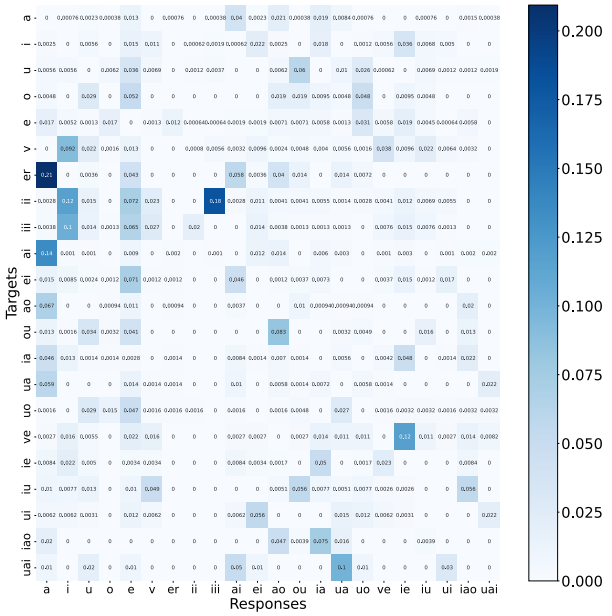


Figure 3: *The confusion matrix of vowels.*

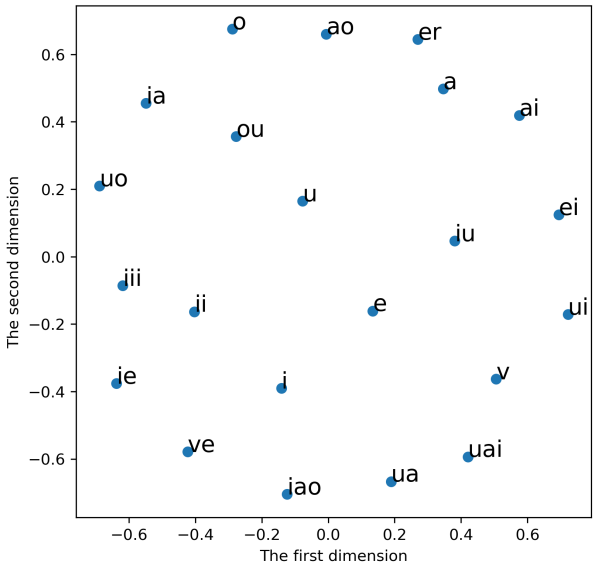


Figure 4: *The 2D distribution of vowels based on phonetic error.*

3.3.2. Error patterns

The error patterns of vowels include substitution, diphthongization, triphthongization, diphthong or triphthong reduction, and a small proportion of undefinable utterance. Except for the undefinable utterance, we analyze the other patterns in detail based on the confusion matrix.

Figure 3 presents the confusion matrix of 22 vowels. Substitution errors are common in monophthongs, diphthongs, and triphthongs. For example, monophthong er was predominantly substituted by a; monophthongs v, ii, iii (refer to [y], [ɿ], [ʅ]) were frequently substituted by i; ii was also commonly substituted by iii; diphthongs ou and ve were often substituted by ao and ie, respectively; triphthong iu was often substituted by iao.

Diphthongization and triphthongization errors are common in monophthongs and diphthongs. For instance, monophthong a was often mispronounced as the diphthongs ai, ao, ia, and ua; i was mispronounced as diphthongs ei, ie, and ia; u was mispronounced as ou and uo; and diphthongs ia and ua were mispronounced as triphthongs iao and uai, respectively. In these error patterns, there was a tendency for vowels to add another phoneme to their original element to complete diphthongization and triphthongization.

Reduction errors are common in diphthongs and triphthongs. For example, diphthongs ai, ao, ua, and ia were commonly mispronounced as the single vowel a; ei and uo were mispronounced as the single vowel e; triphthong uai was often mispronounced as diphthongs ua and ai; iao was commonly mispronounced as diphthongs ia and ao; and ui was often mispronounced as ei. There are some regularities in reduction errors, that is, compound vowels tend to maintain one vowel of their constituents and delete the others.

Similarly, multidimensional scaling analysis was also conducted on the vowels. From figure 4, we can see two group of vowels are gathered in the lower left quadrant. One group includes ii, iii, and i. The substitution error is very common in this group. Another group includes ie and ve, and ve is often substituted by ie. In the lower right quadrant, ua and uai are close together, and uai is often reduced to the diphthong ua. The triphthong ui is also reduced to the diphthong ei, and they are gathered in the middle right quadrant. For a, ai, and ao, they are gathered in the upper right quadrant. The diphthongization and diphthong reduction error are common in this group. Note that er is also very close to a, because it is most commonly substituted by a. The vowels u and ou are gathered in the upper left quadrant, and u is often replaced by the diphthong ou. And the rest of the vowels are not regularly gathered into groups.

4. Discussion

In this study, we presented a Mandarin speech corpus produced by sixty hearing-impaired children and then a preliminary analysis was conducted to study the speech production in the segmental level. The results showed that vowels were produced more correctly than final consonants and initial consonants. Monophthongs were produced less correctly than diphthongs and triphthongs, and the main error patterns of vowels include substitution, diphthongization, triphthongization, diphthong or triphthong reduction. With respect to articulation place, bilabials were produced more correctly than the others. As for articulation manner, affricates had the lowest accuracy. Substitution was the most frequent error pattern for initial consonants while deletion was the common error for final consonants. The findings of this study can benefit the pronunciation teaching of hearing-impaired children and the corpus will be useful for the development of automation technologies. In future work, we will focus on conducting acoustic analysis of production data.

5. Acknowledgements

This study was supported by the Science Foundation and Special Program for Key Basic Research fund of Beijing Language and Culture University (the Fundamental Research Funds for the Central Universities)(20YJ04002), advanced Innovation Center for Language Resource and Intelligence (KYR17005), and Wutong Innovation Platform of Beijing Language and Culture University (19PT04). Jinsong Zhang is the correspondence.

6. References

[1] S. Granberg, "Functioning and disability in adults with hearing loss: The preparatory studies in the icf core sets for hearing loss project," Ph.D. dissertation, Örebro university, 2015.

[2] J. E. Lieu, M. Kenna, S. Anne, and L. Davidson, "Hearing loss in children: A review," *Jama*, vol. 324, no. 21, pp. 2195–2205, 2020.

[3] J. Li, R. Mayr, and F. Zhao, "Speech production in Mandarin-speaking children with cochlear implants: A systematic review," *International Journal of Audiology*, pp. 1–9, 2021.

[4] M. J. Osberger and N. S. McGarr, "Speech production characteristics of the hearing impaired," *Speech and Language*, vol. 8, pp. 221–283, 1982.

[5] N. Tye-Murray, L. Spencer, and G. G. Woodworth, "Acquisition of speech by children who have prolonged cochlear implant experience," *Journal of Speech, Language, and Hearing Research*, vol. 38, no. 2, pp. 327–337, 1995.

[6] C. R. Smith, "Residual hearing and speech production in deaf children," *Journal of Speech and Hearing Research*, vol. 18, no. 4, pp. 795–811, 1975.

[7] C. V. Hudgins and F. C. Numbers, "An investigation of the intelligibility of the speech of the deaf," *Genetic Psychology Monographs*, vol. 25, pp. 289–392, 1942.

[8] S.-C. Peng, A. L. Weiss, H. Cheung, and Y.-S. Lin, "Consonant production and language skills in Mandarin-speaking children with cochlear implants," *Archives of Otolaryngology–Head & Neck Surgery*, vol. 130, no. 5, pp. 592–597, 2004.

[9] J. Yang, J. Vadlamudi, Z. Yin, C.-Y. Lee, and L. Xu, "Production of word-initial fricatives of Mandarin chinese in prelingually deafened children with cochlear implants," *International Journal of Speech-Language Pathology*, vol. 19, no. 2, pp. 153–164, 2017.

[10] J. Yu and X. Xia, "Production of Mandarin stop consonants in prelingually deaf children with cochlear implants," *Proceeding of International Congress of Phonetic Sciences*, pp. 3115–3119, 2019.

[11] F. Zan and T. Shengqin, "The study of phonetic development of hearing impaired children (in chinese)," *Chinese Journal of Special Education*, vol. 1, pp. 10–17, 1998.

[12] D. Han, N. Zhou, Y. Li, X. Chen, X. Zhao, and L. Xu, "Tone production of Mandarin chinese speaking children with cochlear implants," *International Journal of Pediatric Otorhinolaryngology*, vol. 71, no. 6, pp. 875–880, 2007.

[13] N. Zhou, J. Huang, X. Chen, and L. Xu, "Relationship between tone perception and production in prelingually-deafened children with cochlear implants," *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, vol. 34, no. 3, p. 499, 2013.

[14] L.-X. Zheng, *Three thousand common words list of Mandarin chinese*. Beijing: Mandarin Character Reform Press, 1987.

[15] P. Boersma and D. Weenik, "http://www.fon.hum.uva.nl/praat/."