Discrimination and identification of lexical tones and consonants in Mandarin-speaking children using cochlear implants

Laurianne Cabrera, Huei-Mei Liu, Lionel Granjon, Chieh Kao, Feng-Ming Tsao

Discrimination and identification of lexical tones and consonants in Mandarin-speaking children using cochlear implants

Laurianne Cabrera,[1,a)] Huei-Mei Liu,[2,b)] Lionel Granjon,[1,a)] Chieh Kao,[3,c)] and Feng-Ming Tsao[3,d)]

[1]Integrative Neuroscience and Cognition Center, Université Paris Descartes, 45 rue des saints-pères, 75006, Paris, France

[2]Department of Special Education, National Taiwan Normal University, 162, Section 1, Heping E. Road, Taipei City 106, Taiwan

[3]Department of Psychology, National Taiwan University, Number 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan

[a)]Also at: Integrative Neuroscience and Cognition Center, CNRS, 45 rue des saints pères, 75006, Paris, France.

[b)]Also at: Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, 162, Section 1, Heping E. Road, Taipei City 106, Taiwan.

[c)]Also at: Department of Speech-Language-Hearing Sciences, University of Minnesota Twin Cities, 115 Shevlin Hall, 164 Pillsbury Drive, Southeast Minneapolis, MN 55455, USA.

[d)]Also at: Imagining Center for Integrated Body, Mind and Culture Research, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan. Electronic mail: tsaosph@mail2000.com.tw

Received 8 January 2019; revised 27 August 2019; accepted 3 September 2019; published online 10 October 2019

The Journal of the Acoustical Society of America

https://doi.org/10.1121/1.5126941

Abstract

Mandarin-speaking adults using cochlear implants (CI) experience more difficulties in perceiving lexical tones than consonants. This problem may result from the fact that CIs provide relatively sufficient temporal envelope information for consonant perception in quiet environments, but do not convey the fine spectro-temporal information considered to be necessary for accurate pitch perception. Another possibility is that Mandarin speakers with post-lingual hearing loss have developed language-specific use of these acoustic cues, impeding lexical tone processing under CI conditions. To investigate this latter hypothesis, syllable discrimination and word identification abilities for Mandarin consonants (place and manner) and lexical-tone contrasts (tones 1 vs. 3 and 1 vs. 2) were measured in 15 Mandarin-speaking children using CIs and age-matched children with normal hearing (NH).

In the discrimination task, only children using CIs exhibited significantly lower scores for consonant place contrasts compared to other contrasts, including lexical tones. In the word identification task, children using CIs showed lower performance for all contrasts compared to children with NH, but they both showed specific difficulties with tone 1 vs. 2 contrasts. This study suggests that Mandarin-speaking children using CIs are able to discriminate and identify lexical tones and, perhaps more surprisingly, have more difficulties when discriminating consonants.

Keywords: cochlear implants, lexical tones, consonants, discrimination, word identification

I. INTRODUCTION

A cochlear implant (CI) is an electronic device implanted in the peripheral auditory system of people with severe-to-profound sensorineural hearing loss. Currently, CIs are fitted to children with congenital hearing loss before one year of age (Lammers *et al*., 2015; Miyamoto *et al*., 2008; Tomblin *et al*., 2005). CIs undoubtedly benefit users as speech perception improves on average after implantation (Blamey *et al*., 2001; Sarant *et al*., 2001; Svirsky *et al*., 2000; Uhler *et al*., 2011). However, CI processors do not convey well the fine spectro-temporal information of speech sounds (Shannon, 2012). CI users experience great difficulties in perceiving some phonetic contrasts as well as the pitch of speech (Zeng *et al*., 2005). How do children using CIs learn the fine phonetic details of their native language in such degraded conditions? The present study focused on the perception of consonants and lexical tones by Mandarin-speaking children using CIs aged between 4 and 7 years.

A. Lexical-tone and consonant perception for adult listeners using CIs

Current CI processors convey only the slow amplitude modulations (AM or temporal envelope) of the original external signal via a limited number of electrodes corresponding to relatively broad frequency bands (Shannon, 2012). The temporal fine structure (TFS) of the original signal, which corresponds to the fastest temporal fluctuations and are essential to convey pitch information (Moore, 2008; Rosen, 1992; Smith *et al*., 2002; Xu and Pfingst, 2003), is not transmitted by CI processors. Instead, it is replaced by a fixed train of pulses in which amplitude is modulated by the original temporal envelope. Thus, fine pitch information is not well transmitted by CIs. Although pitch variations are exhibited in all languages, as they can convey syntactic units (*e.g.*, upward inflection for questions), attract attention, or express emotional information (Collier, 1975; Nooteboom, 1997), they are particularly crucial for manifesting lexical meanings of syllables in tone languages such as Mandarin Chinese. One of the current challenges for CI technology is to better encode this information in order to improve speech perception for listeners using tone languages.

The two primary acoustic cues differentiating lexical tones in those languages are variations in the fundamental frequency level (F0; high, middle, low), and F0 contour variations (*e.g.*, steady, rising, and falling; Gandour, 1981; Gandour and Harshman, 1978; Khouw and Ciocca, 2007; Vance, 1976). Other acoustic cues such as duration or voice quality may play a secondary role in lexical tone perception (Kuo *et al*., 2008; Whalen and Xu, 1992; Xu *et al*., 2002). Temporal envelope has also been found to be highly correlated with F0 contours (Fu and Zeng, 2000; Whalen and Xu, 1992). Despite their lack of fine spectro-temporal cues, adults using CIs and speaking a tone language have been shown to perform relatively successfully in lexical-tone recognition, probably because they are able to rely on these secondary cues and on linguistic knowledge (Fu and Zeng, 2000; Kuo *et al*., 2008; Whalen and Xu, 1992; Xu and Zhou, 2012). The lexical-tone identification of CI users speaking Mandarin is on average above the chance level, but some Mandarin-speaking adults using CIs exhibit poor identification scores compared to listeners with normal hearing (NH) (Zeng *et al*., 2005). CI users show better performance for consonant contrasts, suggesting that CIs may convey sufficient acoustic information for consonant perception in quiet environments (*i.e.*, slow envelope cues) and perhaps less so for tone contrasts (Rosen, 1992; Zeng *et al*., 2005).

However, it is important to note that the adult participants of those studies had developed language before their hearing loss. These adult participants with post-lingual hearing loss had already mastered their native language. Studying children who have pre-lingual hearing loss and thus do not have any prior knowledge about their linguistic system will help us to understand how listeners rely on an impoverished speech signal in order to learn a language.

B. Lexical-tone and consonant perception for child listeners using CIs

Previous studies exploring speech perception in children with pre-lingual hearing loss and using CIs showed rather good performance despite large variability among them. Children with hearing loss speaking a non-tone language such as French or English and using CIs show poorer discrimination and identification of consonants than children with NH, but on average perform above chance (Bouton *et al.*, 2012; Havy *et al.*, 2013; Medina and Serniclaes, 2009). Children with hearing loss using CIs and speaking a tone language such as Cantonese or Mandarin Chinese show language-specific difficulties in lexical-tone production and more perceptual confusions for some lexical-tone contrasts compared to their peers with NH.

Studies on Cantonese have indicated that lexical-tone identification was difficult, with an average of 50% to 61% correct, for 17 children aged between 4 and 9 years and using their CIs for 11 to 41 months (Ciocca *et al.*, 2002). In addition to acoustical parameters (*e.g.*, F0 height and contour distance between tone pairs), extended duration of CI use and earlier ages of implantation were found to positively predict lexical-tone identification in children (Lee *et al.*, 2002). In another study, 17 children aged 4–6 years who had been using their CIs for 1 to 3 years were asked to identify a target word /ji/ produced with the six Cantonese lexical tones in a word–picture identification task, and to discriminate the syllable /wai/ produced with either the same tone or with two contrasting tones (Wong and Wong, 2004). Performance in the tone-discrimination task did not predict performance in the word-identification task. Discrimination and identification of lexical tones seem to involve different skills. More precisely, discrimination of syllables may not involve lexical processing of word-identification tasks.

Altogether, these studies revealed relatively poor scores, just above chance level, in Cantonese children using CIs.

Children's tone identification difficulties may be related to the small sample size or to the tasks and stimuli used (for instance, only one syllable was used in each task by Wong & Wong, (2004)). For Mandarin Chinese, Peng *et al*. (2004) showed that listeners (N = 41) aged between 6 and 12.6 years demonstrated relatively better identification scores (>73%). A large-scale study (N = 107) also showed that the identification scores for lexical tones in Mandarin ranged from chance level to perfect scores, with an average of 67% correct (Zhou *et al*., 2013). This study did not find any difference between the four Mandarin lexical tones. Other studies observed differences in identification performance according to lexical contrasts. The contrasts between the Mandarin tones 1 (high level) vs. 2 (rising), sharing similar pitch *height*, and tone 2 (rising) vs. 3 (dipping), sharing similar pitch *contour* after the mid-point of the tone, result in the lowest identification scores in children using CIs (Han *et al*., 2009; Peng *et al*., 2004). The most salient contrast is tone 1 vs. tone 3, which differ in both pitch height and contour, and it generally produces higher scores than other contrasts. Thus, lexical-tone perception seems to be difficult for Mandarin-speaking children using CIs and is dependent on the acoustic characteristics of the lexical tones. Regarding the perception of other speech contrasts like consonants by Mandarin-speaking children using CIs, to our knowledge only one study (N = 41) has been published to date and it showed that identification scores varied from chance level to almost perfect scores with an average of 86% (Liu *et al*., 2013). When comparing those independent studies, it may appear that consonants are somewhat easier than lexical tones when identifying words for children using CIs. However, no study has directly compared the perception of lexical tones and consonants in children with pre-lingual hearing loss and using CIs by using the same design.

C. CI-simulation studies and lexical tone perception

The perception of lexical tones and consonants via CIs is thought to be predicted by vocoder-simulation studies with listeners with NH using speech analysis and synthesis systems called "vocoders". These systems simulate CI processors by manipulating the modulation components of the speech signal in a given number of frequency bands (Friesen *et al*., 2001; Shannon *et al*., 1995; Zeng *et al*., 2005). In those simulation conditions, the perception of lexical tones in quiet is more adversely affected by the reduction of the TFS (that is not conveyed by current CIs), than by a reduction of the temporal envelope (Xu and Pfingst, 2003; Zeng *et al*., 2005). For consonants, the opposite pattern has been observed; listeners with NH are less affected by the reduction of the TFS than by the reduction of the temporal envelope (Shannon *et al*., 1995; Smith *et al*., 2002; Xu *et al*., 2005). Nevertheless, in such degraded conditions, the perception of some consonant contrasts, such as place of articulation, is relatively more difficult than the perception of voicing contrasts. From those stimulation studies, it was hypothesized that lexical tones are more difficult to perceive than consonants for CI users in quiet conditions.

Recent simulation studies have also shown that reliance on the modulation components of speech is affected by language exposure/experience. Mandarin-speaking adults with NH rely more on the voice-pitch information conveyed by the fine spectro-temporal cues, as they are more affected by its degradation when discriminating lexical tones compared to French-speaking adults with NH (Cabrera *et al*., 2014). This cross-linguistic difference is also observed at 10 months of age; however, it is not observed at 6 months. This observation suggests that the duration of exposure to the native language does influence the use of the acoustic information of the speech signal (Cabrera *et al*., 2015). Interestingly, in those studies, French listeners (infants at 10 months and adults) were shown to be better able to use the remaining temporal envelope to discriminate non-native lexical tones compared to Mandarin listeners. Therefore, the discrimination of lexical tones in such CI-simulation conditions is influenced by language experience. Together, these simulation studies entail two hypotheses regarding the perception of consonants vs. lexical tones by Mandarin-speaking children using CIs. First, like adults with post-lingual hearing loss using CIs, these children may exhibit poorer perception of lexical tones compared to consonants because CI processors convey only the temporal envelope of speech. Second, children with pre-lingual hearing loss may have developed more efficient perceptual strategies (*e.g.*, relying on temporal envelope information) as a result of being exposed to degraded acoustic information while acquiring their native language. If this is the case, we might expect them to perceive lexical tones as well as consonants.

D. Comparison of children's abilities to discriminate and identify words with lexical tone and consonant contrasts

The aim of this study was to compare how well children with little access to pitch information with CIs, who were exposed to a tone language, such as Mandarin, processed lexical tones and consonants when discriminating syllables and identifying words. This investigation will help to determine how children using CIs learn their native language, and thus, how to better plan future auditory rehabilitation programs for this pediatric population. Two speech perception tasks—phonetic discrimination and word identification—were designed to assess perceptual processing abilities of consonants (change in place of articulation or in manner) and lexical tones (change in tones 1 (high level) vs. 3 (dipping) and tones 1 vs. 2 (rising)). It was expected that place-of-articulation contrasts, mainly conveyed by the fine spectro-temporal information of the speech signal (Rosen, 1992), may be more difficult to perceive than manner by children using CIs

(Bouton *et al*., 2012; Shannon *et al*., 1995). Based on previous studies with children using CIs, it was predicted that the contrast of tone 1 vs. tone 3 would be easier to perceive because of its larger acoustic distance relative to the tone 1 vs. tone 2 contrast, which shows similar pitch height (Han *et al*., 2009; Peng *et al*., 2004). We also expected that perceptual differences between consonants and lexical tones would be task-dependent. A phonetic discrimination task relies more on an acoustic/phonetic level of processing while a word identification task requires access to lexical representations (Wong & Wong, 2004).

II. METHOD

A. Participants

Fifteen children aged 4–7 years (mean = 5.9 years, sd = 1.0, range = 4.1 to 7.0 years) and learning Mandarin (boys N = 8) with severe-to-profound sensorineural hearing loss and who used CIs (duration of CI use: mean = 3.5 years, sd = 1.1, range = 1.1 to 4.9 years) participated in this study. Children with CIs were recruited via auditory rehabilitation centers across Taiwan and were tested in a sound-treated room at these centers. An additional four children were excluded from the final data analyses because they were not able to perform the tasks (did not pass the criteria of training sessions, see below, N =3) or failed a nonverbal ability screening test (N = 1, see below). Participant characteristics are presented in **Table 1**. Note that the etiology was not known for all children except one who had Waardenburg syndrome. All children used oral/aural communication modes only, *i.e.*, they were not using sign language.

Another group of 37 children with NH (mean = 5.3 years, sd = 0.6, range= 4.1 to 6.7 years; boys N = 20) were tested in quiet rooms at kindergartens in Taiwan. Three additional children were tested in this group but not included because of inability to perform the task (N = 2) or mild hearing loss (N = 1). This group of children had received regular health checks since infancy and exhibited typical developmental history.

For all the children, except one child using CIs excluded because of time constraints, nonverbal abilities were screened using the Test of Nonverbal Intelligence, Fourth Edition, TONI-4 (Brown *et al*., 2010). All children included in the following analyses had scores above 70 points (*i.e.*, the inclusion criterion for children with typical non-verbal intelligence; the range of scores was 74–122 (mean = 98.6, sd = 14.1) for CI users and 81–142 (mean = 105, sd = 14) for children with NH).

Ethical approval for this study was obtained from the National Taiwan University Research Ethics Committee, and informed written consent was obtained from the parent/guardian of each child. Questionnaires regarding familiarity of the words used in the identification task were also obtained from the parents.

B. Stimuli

Words and pseudo-words were recorded by two native-Mandarin female speakers. Native Mandarin speakers rated the speech intelligibility of several occurrences of each word and pseudo-word and then the best three tokens of each stimulus were selected from each speaker. For both words and pseudo-words, two consonant features and two-tone contrasts were used.

The pseudo-words were Consonant-Vowel-Consonant-Vowel (CVCV) and were contrasted only on the second syllable. The aim of using bisyllabic pseudo-words was to avoid difficulties hearing the onset of the target syllable. The first CV always included the unaspirated labial stop consonant /p/ and the vowel /a/ produced with tone 4 (falling). For the consonant contrasts, the syllable /pa4/ was followed by a syllable produced with the vowel /a/ and tone 1 (high) and the consonant was varied between /p/, /t/, /ʦ/, /s/,

/ʂ/, /tʂ/, /x/, or /k/. As shown in **Table 2**, three minimal pairs of consonants differing on manner of articulation and three pairs differing on place of articulation were selected. For the tone contrasts, the first syllable /pa4/ was followed by the consonant /p/ and the vowel was varied between /u/, /ao/, or /i/ produced with either tone 1 (high), tone 2 (rising), or tone 3 (dipping). As shown in **Table 2**, these three vowels were paired to differ on tones 1 vs. 2 or on tones 1 vs. 3. Here, different vocalic contexts were used in an attempt to maximize generalizability of the results and tone 4 (falling) was not selected to contrast with others because of its shorter duration in natural occurrences (Whalen and Xu, 1992; Xu *et al.*, 2002; Yang, 1989). Pitch contours could be observed in **Figure 1** representing the spectrograms of one CVCV produced with tone 1, 2, or 3 on the last CV by one speaker. Finally, Mandarin word pairs differing on the same consonant and lexical-tone contrasts were selected (see **Table 2**). These words correspond to Mandarin words usually known by the majority of 4-year-old Mandarin-speaking children.

C. Procedure

Children with NH completed all tasks in approximately 40 min and children with hearing loss were tested within one session of approximately 2 hours including breaks. For both groups, sounds were presented at a level of around 70 dB SPL via loudspeakers, located on each side of the computer monitor in front of the child at approximately 30–40 cm. CI processors were adjusted to each participant's daily use setting. The experimenter sat next to the child. The phonetic discrimination task was conducted on a laptop computer using E-Prime. In this task, each trial was composed of four spoken pseudo-words while two pictures were displayed on the screen. Children were asked to point at a picture depicting four cows when the sounds were the same and at a picture depicting two cows and two frogs when the sounds were different. Half of the trials were 'same' trials (*i.e.*, AA–AA) and the other half were 'change' trials (*i.e.*, AA–BB). Two short familiarization blocks were designed to familiarize the children with the task and the setup. The first familiarization block included up to four trials consisting of either only cow sounds or both cow and frog sounds instead of speech sounds. This was done to make sure that children understood the concepts of 'same' and 'different' (Holt and Lalonde, 2012). The second familiarization block included up to eight trials consisting of CVCV pseudowords and same or change trials varying on non-minimal phonetic pairs (consonants and vowels not used in the testing blocks). Feedback, in the form of a positive or negative emoticon, was displayed on the screen after each trial during familiarization. If children could not perform above chance in this second familiarization block within a maximum of three attempts, they were judged to be unable to complete the tasks (two children with CIs were excluded following this criterion). Following familiarization, all children performed a total of four testing blocks, without feedback, comprising 24 trials (12 change and 12 no-change) for each speech contrast (manner, place, tones 1 vs. 2, and tones 1 vs.3) for a total of 96 trials.

For each CVCV pair (*i.e.*, pa4xa1 pa4ka1), two change trials (pa4xa1 pa4xa1 pa4ka1 pa4ka1) and two no-change trials (pa4xa1 pa4xa1 pa4xa1 pa4xa1) were presented randomly within one block. In each block, six CVCV pairs were presented, selected pseudo-randomly from the set of consonant and lexical-tone pairs. Half of the trials used tokens produced by one speaker and the other half tokens by the second speaker, and the order of presentation was randomized within blocks. Increasing the acoustic variability by using two different speakers and their different tokens also aimed to maximize generalizability.

Children were encouraged to point at the 'different' or 'same' picture on the screen to indicate their response. The experimenter recorded the child's response using the computer mouse. In addition, a puzzle piece appeared on the screen after children completed each trial in an effort to maintain interest and let participants track their progress in the task. The correct responses on change and no-change trials were recorded and averaged for each of the four speech contrasts (manner, place, tones 1 vs. 2, and tones 1 vs. 3). Participants' accuracy was estimated by a d' score where $d' = Z[p(\text{hit}) - Z[p(\text{false alarms})]$ (Macmillan and Creelman, 1991). If the hit rate was 1 or the false alarm rate was 0, then d' was calculated after adjusting the hit rate or false alarm rate by the reciprocal of the number of trials. The maximum d' score in this task was 2.77.

Children were also asked to perform a word identification task. For each trial, two pictures of objects were presented on the screen and children were asked in Mandarin to "show the –target object". The experimenter recorded the child's response using the computer mouse. Three training trials, with feedback, were used to familiarize the children with the procedure by presenting words differing in non-minimal phonetic contrasts (that were not used later in the task). This training was repeated a second time if the child was not able to get two correct answers out of three (none of the children were excluded based on this criterion). Once the training was completed, the pictures of two objects whose names differed in minimal pairs from **Table 2** were presented. During the test, no feedback was provided. A total of 12 word pairs was presented four times within two blocks with a break in between. For each word pair (A–B), two trials presented the words produced by speaker 1, and the other two presented the words produced by speaker 2. The presentation side of the target picture (left or right) was counterbalanced for each word pair. Consonant and lexical-tone trials were presented randomly within each block. Correct identification scores were computed for each pair. Half of the children completed the discrimination task first, and the other half completed the identification task first. The order of the blocks was also counterbalanced between children. Finally, a standardized assessment was used to evaluate the receptive vocabulary skills of all children with NH and 11 children using CIs, *i.e.*, the Mandarin version of the Peabody Picture Vocabulary Test (Lu and Liu, 1988).

III. RESULTS

A. Phonetic discrimination task

1. Performance comparison between children with NH and children using CIs

The performance of the 15 CI users was compared to a sub-group of 15 children with NH who were matched in chronological age [mean age for children with NH = 5.5 years, sd = 0.7; mean age for children using CIs = 5.9 years, sd = 1.0; $t(28) = 1.18$, $p = .25$] and non-verbal IQ [mean non-verbal IQ for children with NH = 99.0, sd = 11.2; mean nonverbal IQ for children using CIs = 98.6, sd = 14.1; $t(27) = -0.158$, $p = .88$; all η^2 reported here are partial]. **Figure 2** shows the *d*' scores of these two groups for each of the speech contrasts tested. A mixed analysis of variance (ANOVA) with Hearing Status (NH vs. CI) as the between-subject factor and Speech contrast (4) as the within-subject factor was performed on the *d*' scores. This analysis revealed no main effect of Speech Contrast [$F(3,84) = 0.67$, $p = .53$, $\eta^2 = .023$] but a main effect of Hearing Status [$F(1,28) = 7.09$, $p = .013$, $\eta^2 = .20$]. CI users showed overall lower *d*' scores than controls with NH. This analysis also revealed a significant interaction between these two factors [$F(3,84) = 5.89$, $p = .003$, $\eta^2 = .17$]. Pairwise comparisons using Bonferroni corrections indicated that children with NH do not show any difference in their *d*' scores across the speech contrasts (*p*s > .45), but that CI users had poorer scores for the place of articulation contrast compared to tones 1 vs. 2 ($p = .001$). Moreover, the children with NH showed significantly higher scores than CI users for both manner and place of articulation ($p = .008$ and $p < .001$, respectively), but equivalent scores for tones 1 vs. 2 and 1 vs. 3 ($p = .77$, $p = .24$, respectively). One-sample *t*-tests revealed that the averaged *d*' scores in each condition were above chance level (= 0) for the group of CI users (p = .001, .001, < .001, and .001, respectively for manner, place, tones 1 vs. 2, and tones 1 vs. 3 contrasts) and for the group of children with NH (ps < .001).

Thus, our findings suggest that CI users may have more difficulties than children with NH in distinguishing both manner and place contrasts for Mandarin consonants. Our findings also suggest that, unlike children with NH, children with CIs may find changes in place-of-articulation to be more difficult to detect than other contrasts.

2. Exploration of potential factors influencing performance of children using CIs

A second set of analyses was performed for CI users only to assess the relationship between duration of CI use, age of implantation, pre-implant unaided residual hearing (pure-tone thresholds averaged for the left and right ears), and age-normed vocabulary with children's discrimination performance. In this group of 15 children, performance was normally distributed in each speech contrast condition (p >.05 for Shapiro–Wilk tests). Pearson correlation coefficients are shown in **Table 3** and they do not reveal any significant relationship between the discrimination scores and those four variables after correction for multiple comparisons (α = .0025). The correlation between the duration of CI use and d' scores for place of articulation and tone 1 vs. 2 contrasts approached significance (uncorrected p values = .024 and .046, respectively). Note that we could not assess the effect of having a contra-lateral hearing aid because only one child did not use an additional hearing aid. Moreover, only one child used bilateral implants. The scores of these two children were similar to the group average scores in all conditions.

3. Generalization of the results for children with NH

Finally, a supplementary analysis was performed for the whole group of children with NH speaking Mandarin. This analysis aimed to assess whether similar results for children with NH would be observed from a larger group of 37 children aged 4–7 years. Their mean scores for manner, place, tones 1 vs. 2, and tones 1 vs. 3 were 1.79 (sd = 0.94), 1.53 (sd = 0.82), 1.61 (sd = 0.94), and 1.77 (sd = 0.77), respectively. As the previous analysis with the sub-group of 15 children, a repeated-measures ANOVA performed on the *d*' scores with Speech contrast as the within-subject factor did not reveal any difference between the four contrasts [$F(3,108) = 2.16$, $p = .097$, $\eta^2 = .057$].

B. Word identification task

1. Performance comparison between children with NH and those using CIs

The identification scores of the CI users were also compared with those of the same sub-group of children with NH matched in chronological age and non-verbal IQ. **Figure 3** shows the correct identification scores for these two groups as a function of Speech contrast. A mixed ANOVA assessed the effect of Speech contrast (4) as the within-subject factor and Hearing status (2) as the between-subject factor. There were significant effects for both Speech contrast [$F(3,84) = 9.08$, $p < .001$, $\eta^2 = .25$] and Hearing status [$F(1,28) = 7.41$, $p = .011$, $\eta^2 = .21$], but no Speech contrast vs. Hearing status interaction [$F(3, 84) = 0.75$, $p = .52$, $\eta^2 = .026$]. Pairwise comparisons with Bonferroni corrections indicated that, for both the CI users and children with NH, the identification scores for tones 1 vs. 2 contrasts were significantly lower (poorer) than those of manner ($p < .001$), place of articulation ($p = .029$) and tones 1 vs. 3 ($p < .001$). In addition, children with CIs exhibited lower (poorer) word-identification scores than the matched NH group. This is consistent with the fact that CI users showed significantly lower vocabulary scores than controls with NH [raw score range was 16–90 for CI users (mean = 50.5, sd= 33.1) and 29–99 for children with NH (mean = 59.7, sd = 17.9); $t(28) = 3.44$, $p = .002$, $\eta^2 = .30$. One-sample t-tests revealed that, even though some children using CIs showed scores below or chance (0.50), the averaged identification score for the group of CI users was above chance for manner, place, and tones 1 vs. 3, but not for tones 1 vs. 2 contrasts (ps $< .001$ and $= .13$) and were above chance for each contrast for the group of children with NH (ps $< .001$). These analyses thus suggested that, even though CI users had overall poorer word-identification scores than controls with NH, these two groups exhibited the same pattern of response among speech contrasts. Both CI users and children with NH had more difficulties in identifying words differing on tones 1 vs. 2 than on consonants.

2. Exploration of potential factors influencing performance of children using CIs

For CI users, no significant correlation, once corrected for multiple comparisons, was observed between identification scores and age of implantation, duration of CI use, pre-implant unaided residual hearing, or vocabulary level (see **Table 3**). Pre-implant unaided residual hearing in the left ear was slightly correlated with identification scores in the tones 1 vs. 3 contrast (uncorrected $p = .010$ and $.024$, respectively).

3. Generalization of the results for children with NH

Finally, the identification scores for the larger cohort of 37 children with NH were also compared between speech contrast. For manner, place, tones 1 vs. 2, and tones 1 vs. 3, the mean identification scores were 86 (sd = 10), 79 (sd = 14), 70 (sd = 14), and 81 (sd = 14), respectively. A repeated-measures ANOVA was performed on the percentage of correct identification scores with Speech contrast as the within-subject factor and showed the same main effect of Speech contrast as observed with the sub-group of 15 children with NH [$F(3,108) = 10.63$, $p < .001$; $\eta^2 = 22.8$]. *Post hoc* tests with Bonferroni corrections revealed that the scores for the contrast of tones 1 vs. 2 were significantly *lower* than for manner, place, and tones 1 vs. 3 ($p < .001$, $p = .04$, and $p = .002$, respectively).

4. Identification scores and parents' questionnaires

As expected, the contrast with similar pitch heights (*i.e.*, high-level tone 1 and high-rising tone 2) was more difficult than the contrast with distinct pitch heights (*i.e.*, high-level tone 1 and low-dipping tone 3). To further explore the reasons for these low scores with the tones 1 vs. 2 contrast, the parents' responses to the questionnaire regarding whether the words used in the identification task were known by their children were compared to the children's responses for each word pair. According to the parents' questionnaire, the words used in the test of tones 1 vs. 2 were not well-known by both children using CIs and those with NH. On average, parents reported that 72% of the words in the test of tones 1 vs. 2 were not known by the children using CIs (compared to 26% for manner, 38% for place, and 48% for tones 1 vs. 3). For the children with NH, 64% of words were not known in the test of tones 1 vs. 2 (compared to 20% for manner, 33% for place, and 32% for tones 1 vs. 3). The lowest identification scores observed in the tones 1 vs. 2 test might be the results of both perceptual difficulties with this tonal contrast and the word pairs selected for this tone test probably being less common to children aged 4–7 years. However, it is important to note that the overall proportion of words known per pair as indicated by the parental questionnaire and the actual proportion of children's correct responses did not correlate with each other [$r = -0.056$, $p = .17$]. For this reason, we preferred not to exclude the trials with words supposed to be not well-known by the children.

C. Relationship between phonetic discrimination and word identification tasks

Correlation analyses were also used to assess the relationship between the speech perception scores in the phonetic discrimination task and in the word identification task for the group of 37 children with NH and 15 children using CIs. A significant correlation between discrimination and identification scores was observed in the NH group for the tone 1 vs. 2 contrast. Results for the CI users indicated that word identification scores for the tone 1 vs. 3 word pairs correlated with discrimination scores for the same lexical tone contrast. Moreover, identification scores for the tone 1 vs. 2 contrast correlated with discrimination scores for manner contrasts (see **Table 4**).

IV. GENERAL DISCUSSION

This study assessed the effect of cochlear implantation on the perception of lexical tones and consonants for Mandarin-speaking children. The results of these experiments revealed that, overall, the group of children aged 4 to 7 years who were learning Mandarin and using CIs was able to perceive fine differences in lexical tones and consonants in syllables at above chance. Surprisingly, children were better at detecting changes in lexical tones than changes in some consonants.

A. Perception of lexical tones and consonants in children with NH

For the group of children with NH, no differences were observed in the phonetic discrimination task between consonant or lexical-tone contrasts, suggesting that Mandarin-speaking children with NH seemed to have developed a phonological system for both consonants and lexical tones by 4 years of age. However, some perceptual difficulties were observed in the word identification task when words shared similar pitch height but not contour, *i.e.*, tone 1 vs. tone 2.

For consonant perception, the Mandarin manner contrast, retroflex fricative–affricate, is less accurately discriminated even for adults (Tsao *et al*., 2006). However, in the present tasks, children aged 4–7 years with NH exhibited relatively accurate discrimination and identification scores with no difference between the Mandarin consonant contrasts of place and manner, even for the word pairs with retroflex fricative and affricate consonants.

For lexical-tone perception, a previous study found that tones 1 vs. 3 is the easiest contrast to discriminate for 12-month-old Mandarin-learning infants with NH (Tsao, 2008), possibly because of the larger pitch-height differences for this contrast. Nonetheless, in the present study, older children did not differ in their discrimination scores for the two lexical-tone contrasts (tone 1 vs. 2 and tone 1 vs. 3). This finding is consistent with data obtained from a large group of Mandarin-speaking children with NH aged 3–10 years (Zhou *et al*., 2013). When identifying lexical tones, tone 3 was the most difficult for 3-year-old Mandarin-speaking children with NH when segmental cues were reduced by low-pass filtering (Wong *et al*., 2005). In the present word-identification task, tones 1 vs. 2 contrasts led to the lowest scores for children with NH. The similar pitch-height between tones 1 and 2 could make this contrast more difficult to process. It is also possible that our choice of stimuli may have accounted for the lower score in this condition, as parent reports suggested that children may have been less familiar with these words.

One of the strengths of this study was designing a phonetic discrimination task and a word identification task using the same speech contrasts. Nevertheless, finding contrasting words on these specific speech contrasts was challenging and some of the words were less likely to be known by young Mandarin-speaking children. The unknown words could have been removed individually from the final data analysis. This, however, would have made it necessary to show the pictures to the children either before the task started, and thus potentially influence their performance in the following identification task, or after the task finished, and thus potentially influence the children's judgment of the picture–word pairs by being exposed to word stimuli in the identification task. Other types of word identification tasks may be developed in future studies to better evaluate children's prior knowledge of the words tested.

B. Perception of lexical tones and consonants using CIs

The goal of this study was to assess the phonetic discrimination and word identification abilities of Mandarin-speaking children using CIs in order to explore perceptual difficulties under CI conditions. The use of three pairs of CV syllables per phonetic feature as well as two different speakers allowed us to explore thoroughly the perception of consonants and lexical tones in the same children. Moreover, the use of CVCV in the discrimination task and a carrier sentence in the identification task could enhance speech intelligibility for CI users who may experience difficulties hearing speech onsets (Koning and Wouters, 2016; Stilp *et al*., 2013). Using these designs, the speech perception scores for the group of children with CIs for phonetic discrimination and word identification were above chance for all the speech features, except for the identification of tones 1 vs. 2.

Nevertheless, the results of the phonetic discrimination task revealed some differences between children with NH and children using CIs. These two groups of children did not perform similarly in detecting changes in consonants and lexical tones. Although children using CIs were able to detect changes in consonants and lexical tones above chance, their discrimination scores for consonants, for both manner and place changes, were significantly lower than their peers with NH. In particular, the CI users showed specific difficulties with the place-of-articulation contrasts. However, they were able to discriminate lexical tones as well as their peers with NH. Thus, this study supports the hypothesis that children using CIs may perform less accurately in discriminating consonant contrasts such as place of articulation, which are mainly conveyed by fine spectro-temporal information, compared to manner, which relies more on envelope information (Rosen, 1992; Bouton *et al*., 2012). These results also suggest that the CI processors used by participants in this study were sufficient to convey gross pitch information and secondary cues such as duration and amplitude for lexical-tone discrimination (Kuo *et al*., 2008; Xu *et al*., 2002). Previous acoustic analysis of the naturally produced Mandarin syllables showed that duration is similar among tones 1, 2, and 3 (Liu, Tsao, & Kuhl, 2007); thus, the secondary duration cue was not considered to be effective for discriminating the present contrasts. The duration of lexical tones might not account for the differences between the NH and CI groups in the present lexical-tone discrimination task. Amplitude cues, however, are likely to play a greater role for CI users, who have better access to temporal cues than to spectral cues (Fu & Zeng, 2000). In the present discrimination task, children using CIs might have benefited more from this secondary amplitude cue for the lexical-tone contrasts than for the consonant contrast of place of articulation, for which amplitude plays a small role (Rosen, 1992).

The current difference observed between consonants and lexical tones for children using CIs does not appear to be consistent with perceptual data obtained from Mandarin-speaking adults using CIs (Zeng *et al*., 2005) or from simulation studies using vocoders, which have shown that lexical tones were more difficult to distinguish than consonants when using CIs. Adults in those studies had developed proficient skills in Mandarin before the onset of hearing loss. It is possible that adults may have learned to rely on specific acoustic information such as the fine spectro-temporal cues conveying fine voice-pitch variations in order to perceive lexical tones. Although Mandarin-speaking adults with NH experience difficulties in perceiving lexical tones when the fine spectro-temporal cues are degraded, in non-tone languages, such as French, listeners who have learned to rely on the envelope cues to detect tone differences do not have the same difficulties (Cabrera *et al*., 2014, 2015). The children in the present study had congenital hearing loss and therefore had been listening to speech only through CI processors after implantation. In these specific listening conditions, children may have learned to rely on the remaining acoustic information (*i.e.*, envelope cues) that conveys some pitch information (Cabrera *et al*., 2015; Wong *et al*., 2005). Children with CIs exposed to Mandarin may thus have developed efficient perceptual strategies to process lexical tones. Future studies should directly compare perceptual difficulties of children with pre-lingual hearing loss and adults with post-lingual hearing loss using CIs to test this hypothesis.

Finally, the results from our identification task also provide more information about the use of consonants and lexical tones in children's speech perception. Although the CI users in this study showed poorer identification scores than the children with NH overall, both groups exhibited the same pattern of responses. Both children with NH and children using CIs had significantly lower scores for the lexical-tone contrast of tones 1 vs. 2. Moreover, when comparing the results of the two perceptual tasks, phonetic discrimination abilities did not always associate with word identification performance (the only significant correlation was for tones 1 vs. 3 contrasts for children using CIs). The specific difficulty observed with place of articulation contrasts for CI users in the phonetic discrimination task was not reflected in the word identification task. This result may demonstrate that the reliance on acoustic/phonetic cues is task-dependent, as suggested by previous studies (Wong and Wong, 2004). In the word identification task, corresponding pictures of word stimuli were shown to the children to depict their perceptual judgments. This visual help, however, might have primed children not only to access the mental lexicon of the target words, but also to activate the phonological representations associated with the mental lexicon. In other words, CI users in the word-identification task could have employed top-down levels of lexical processing when perceiving speech sounds. They may have used this information to compensate for acoustic/phonetic perceptual difficulties observed in the discrimination task (where no visual cues for speech stimuli were displayed). Using other tasks, previous studies have shown that children using CIs have poorer phonetic discrimination abilities than their peers with NH, but they demonstrate the same lexicality effect. That is, both children with

NH and those with CIs discriminate phonetic contrasts better when they are within words than when they are in pseudo-words (Bouton *et al*., 2012; Kirk *et al*., 1995). These studies suggest that children using CIs have a limited access to the acoustic cues of the speech signal, but they compensate with a greater use of lexical information to assist phonetic feature identification.

C. Relationship between hearing experience, verbal skills, and speech perception for children using CIs

Previous studies have shown that better cognitive skills (*i.e.*, attention) may predict better performance in speech tasks (Geers *et al*., 2003; Geers, 2003). Moreover, age of implantation and duration of CI use have been found to be good predictors of tone perception in Mandarin-speaking children using CIs (Zhou & Xu, 2013), but only for discrimination of manner consonant contrast (Lee *et al*., 2002). In the present study, no significant relationship was observed between age of implantation, duration of CI use, pre-implant residual hearing, or vocabulary scores with the ability of children using CIs to discriminate phonetic contrasts or identify words in quiet environments. There were trends suggesting that duration of CI use and residual hearing could be somewhat related to discrimination and identification scores of consonants and lexical tones. The sample size of the CI group was statistically small and thus the limited variations of age of implantation and duration of CI use would not be optimal to assess whether earlier age of implantation and longer duration of CI use would benefit speech perception development. Although the age of implantation was not found to correlate with any of the perceptual measures, children in this study were implanted with CIs between 1.4 and 3.3 years of age, and previous studies suggested that implantation before 3 years may favor later speech perception abilities in children with hearing loss (Sharma *et al*., 2002).

Additionally, consistent with many previous studies, we found that children using CIs exhibited lower vocabulary scores than their peers with NH, suggesting that auditory deprivation in early ages had impeded language acquisition (Lund, 2015; Välimaa *et al*., 2018). Moreover, there is a general trend that the performance variability of the CI group for discrimination and identification of speech sounds is greater (as shown in Figures 1 and 2) than that of the NH group, showing that Mandarin-speaking children with similar severity of hearing loss before CI implantation do not perceive consonants and lexical tones in similar ways after CI implantation. Longitudinal and large-scale studies would help to evaluate further the role of hearing experience with CIs on speech and language development.

V. CONCLUSIONS

The present study used both consonant feature contrasts and lexical-tone contrasts to assess speech perception in Mandarin-speaking children with hearing loss who used cochlear implants (CI). Results showed that these children were able to distinguish consonant and lexical-tone changes. However, for these children, some consonant contrasts, such as place of articulation, were more difficult to distinguish than lexical tones. Our results also suggested that children using CIs correctly identified words that contained speech contrasts that were nonetheless difficult to discriminate. In the present identification task, children using CIs were probably adopting strategies for lexical access (*e.g.*, the help of visual context) rather than exclusively depending on phonetic/acoustic processing. This study therefore contributes to existing findings that children using CIs are able to use the acoustic cues provided by their devices in order to both discriminate and identify words based on tone contrasts above chance in quiet conditions. This highlights the benefits of cochlear implantation in children with severe-to-profound hearing loss, even for those learning a tone language.

Acknowledgments

F.M. Tsao's work was supported by the Ministry of Science and Technology of Taiwan under grant no. 102-2923-H-002-001-MY3. L. Cabrera was supported by the Agence Nationale de la Recherche of France, under grant no. ANR-12-ISH2-0001-01. We wish to thank Dr Josiane Bertoncini, who initiated this research program and provided helpful comments on the manuscript. We also thank Dr Aurore Gautreau, who helped to design the experiment, and Dr Lorna Halliday for helpful comments on the manuscript. Finally, we give warm thanks to all the participants and schools and the NWL Foundation for the Hearing Impaired in Taiwan who took part in this research study.

Declaration of interests

The authors report no conflicts of interest.

REFERENCES

Blamey, P. J., Sarant, J. Z., Paatsch, L. E., Barry, J. G., Bow, C. P., Wales, R. J., Wright, M., et al. (**2001**). “Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing,” J. Speech Lang. Hear. Res., **44**, 264–285.

Bouton, S., Serniclaes, W., Bertoncini, J., and Colé, P. (**2012**). “Perception of speech features by French-speaking children with cochlear implants,” J. Speech Lang. Hear. Res. JSLHR, **55**, 139–153.

Brown, L., Sherbenou, R. J., and Johnsen, S. K. (**2010**). *Test of Nonverbal Intelligence: TONI-4*, Pro-ed.

Cabrera, L., Tsao, F.-M., Gnansia, D., Bertoncini, J., and Lorenzi, C. (**2014**). “The role of spectro-temporal fine structure cues in lexical-tone discrimination for French and Mandarin listeners,” J. Acoust. Soc. Am., **136**, 877–882.

Cabrera, L., Tsao, F.-M., Liu, H.-M., Li, L.-Y., Hu, Y.-H., Lorenzi, C., and Bertoncini, J. (**2015**). “The perception of speech modulation cues in lexical tones is guided by early language-specific experience,” *Frontiers in psychology*, **6**, 1290.

Ciocca, V., Francis, A. L., Aisha, R., and Wong, L. (**2002**). “The perception of Cantonese lexical tones by early-deafened cochlear implantees,” J. Acoust. Soc. Am., **111**, 2250-2256.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). “Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants,” J. Acoust. Soc. Am., **110**, 1150–1163.

Fu, Q.-J., and Zeng, F.-G. (**2000**). “Identification of temporal envelope cues in Chinese tone recognition,” Asia Pac. J. Speech Lang. Hear., **5**, 45–58.

Gandour, J. (**1981**). “Perceptual dimensions of tone: Evidence from Cantonese,” J. Chin. Linguist., 20-36.

Gandour, J. T., and Harshman, R. A. (**1978**). “Crosslanguage differences in tone perception: A multidimensional scaling investigation,” Lang. Speech, **21**, 1–33.

Geers, A., Brenner, C., and Davidson, L. (**2003**). “Factors associated with development of speech perception skills in children implanted by age five,” Ear Hear., **24**, 24S–35S.

Geers, A. E. (**2003**). “Predictors of reading skill development in children with early cochlear implantation,” Ear Hear., **24**, 59S–68S.

Han, Demin, Bo Liu, Ning Zhou, Xueqing Chen, Ying Kong, Haihong Liu, Yan Zheng, and Li Xu. "Lexical tone perception with HiResolution and HiResolution 120 sound-processing strategies in pediatric Mandarin-speaking cochlear implant users." Ear and hearing 30, no. 2 (2009): 169.

Havy, M., Nazzi, T., and Bertoncini, J. (**2013**). “Phonetic processing during the acquisition of new words in 3-to-6-year-old French-speaking deaf children with cochlear implants,” J. Commun. Disord., **46**, 181–192.

Holt, R. F., and Lalonde, K. (**2012**). “Assessing toddlers’ speech-sound discrimination,” Int. J. Pediatr. Otorhinolaryngol., **76**, 680–692.

Khouw, E., and Ciocca, V. (**2007**). “Perceptual correlates of Cantonese tones,” J. Phon., **35**, 104–117.

Kirk, K. I., Pisoni, D. B., and Osberger, M. J. (**1995**). “Lexical effects on spoken word recognition by pediatric cochlear implant users,” Ear Hear., **16**, 470.

Koning, R., and Wouters, J. (**2016**). “Speech onset enhancement improves intelligibility in adverse listening conditions for cochlear implant users,” Hear. Res., **342**, 13–22.

Kuo, Y.-C., Rosen, S., and Faulkner, A. (**2008**). “Acoustic cues to tonal contrasts in Mandarin: Implications for cochlear implants,” J. Acoust. Soc. Am., **123**, 2815–2864.

Lammers, Marc JW, Thijs TG Jansen, Wilko Grolman, Thomas Lenarz, Huib Versnel, Gijsbert A. Van Zanten, Vedat Topsakal, and Anke Lesinski- Schiedat. (**2015**) "The influence of newborn hearing screening on the age at cochlear implantation in children." The Laryngoscope **125**, no. 4: 985-990.

Lee, K. Y., Van Hasselt, C. A., Chiu, S. N., and Cheung, D. M. (**2002**). “Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children,” Int. J. Pediatr. Otorhinolaryngol., **63**, 137–147.

Liu, Q., Zhou, N., Berger, R., Huang, D., and Xu, L. (**2013**) "Mandarin consonant contrast recognition among children with cochlear implants or hearing aids and normal-hearing children." *Otology & Neurotology,* **34**, 471-476.

Lu, L., and Liu, H. (**1988**). “Peabody Picture Vocabulary Test—Mandarin Version, Revised Form A,” Taipei Taiwan Psychol. Publ.

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. Developmental Psychology, 43(4), 912.

Lund, E. (**2015**). “Vocabulary knowledge of children with cochlear implants: A meta-analysis,” J. Deaf Stud. Deaf Educ., **21**, 107–121.

Macmillan, N. A., and Creelman, C. D. (**1991**). “Detection theory: A user’s guideCambridge University Press,” N. Y.

Medina, Victoria, and Willy Serniclaes. (**2009**) "Development of voicing categorization in deaf children with cochlear implant." In Tenth Annual Conference of the International Speech Communication Association.

Miyamoto, R. T., Hay-McCutcheon, M. J., Kirk, K. I., Houston, D. M., and Bergeson-Dana, T. (**2008**). “Language skills of profoundly deaf children who received cochlear implants under 12 months of age: a preliminary study,” Acta Otolaryngol. (Stockh.), **128**, 373–377.

Moore, B. C. (**2008**). “The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people,” J. Assoc. Res. Otolaryngol., 9, 399–406.

Nooteboom, S. (**1997**). The prosody of speech: melody and rhythm. The handbook of phonetic sciences, 5, 640-673.

Peng, S.-C., Tomblin, J. B., Cheung, H., Lin, Y.-S., and Wang, L.-S. (**2004**). “Perception and production of Mandarin tones in prelingually deaf children with cochlear implants,” Ear Hear., **25**, 251–264.

Rosen, S. (**1992**). “Temporal information in speech: acoustic, auditory and linguistic aspects,” Philos. Trans. R. Soc. Lond. B. Biol. Sci., **336**, 367–373.

Sarant, J. Z., Blamey, P. J., Dowell, R. C., Clark, G. M., and Gibson, W. P. R. (**2001**). “Variation in speech perception scores among children with cochlear implants,” Ear Hear., **22**, 18–28.

Shannon, R. V. (**2012**). “Advances in auditory prostheses,” Curr. Opin. Neurol., **25**, 61–66.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). “Speech recognition with primarily temporal cues,” Science, **270**, 303–304.

Sharma, A., Dorman, M. F., and Spahr, A. J. (**2002**). "A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation." Ear Hear., **23**, 532-539.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). “Chimaeric sounds reveal dichotomies in auditory perception,” Nature, **416**, 87–90.

Stilp, C. E., Goupell, M. J., and Kluender, K. R. (**2013**). “Speech perception in simulated electric hearing exploits information-bearing acoustic change,” J. Acoust. Soc. Am., **133**, EL136–EL141.

Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B., and Miyamoto, R. T. (**2000**). "Language development in profoundly deaf children with cochlear implants," Psychol. Sci., **11**, 153–158.

Tomblin, J. B., Barker, B. A., Spencer, L. J., Zhang, X., and Gantz, B. J. (**2005**). "The effect of age at cochlear implant initial stimulation on expressive language growth in infants and toddlers," J. Speech Lang. Hear. Res., **48**, 853–867.

Tsao, F.-M. (**2008**). "The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants," Chin. J. Psychol, **50**, 111–124.

Tsao, F.-M., Liu, H.-M., and Kuhl, P. K. (**2006**). "Perception of native and non-native affricate-fricative contrasts: cross-language tests on adults and infants," J. Acoust. Soc. Am., **120**, 2285–2294.

Uhler, K., Yoshinaga-Itano, C., Gabbard, S. A., Rothpletz, A. M., and Jenkins, H. (**2011**). "Longitudinal infant speech perception in young cochlear implant users," J. Am. Acad. Audiol., **22**, 129–142.

Välimaa, T., Kunnari, S., Laukkanen-Nevala, P., and Lonka, E. (**2018**). "Early vocabulary development in children with bilateral cochlear implants," Int. J. Lang. Commun. Disord., **53**, 3–15.

Vance, T. J. (**1976**). "An experimental investigation of tone and intonation in Cantonese," Phonetica, **33**, 368–392.

Whalen, D. H., and Xu, Y. (**1992**). "Information for Mandarin tones in the amplitude contour and in brief segments," Phonetica, **49**, 25–47.

Wong, A. O., and Wong, L. L. (**2004**). "Tone perception of Cantonese-speaking prelingually hearing-impaired children with cochlear implants," Otolaryngol. Neck Surg., **130**, 751–758.

Wong, P., Schwartz, R. G., and Jenkins, J. J. (**2005**). “Perception and production of lexical tones by 3-year-old, Mandarin-speaking children,” J. Speech Lang. Hear. Res., **48**, 1065–1079.

Xu, L., and Pfingst, B. E. (**2003**). “Relative importance of temporal envelope and fine structure in lexical-tone perception,” J. Acoust. Soc. Am., **114**, 3024–3027.

Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). “Relative contributions of spectral and temporal cues for phoneme recognition,” J. Acoust. Soc. Am., **117**, 3255–3267.

Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). “Features of stimulation affecting tonal-speech perception: implications for cochlear prostheses,” J. Acoust. Soc. Am., **112**, 247–258.

Xu, L., and Zhou, N. (**2012**). “Tonal languages and cochlear implants,” Audit. Prostheses, Springer, pp. 341–364.

Yang, Y. F. (**1989**). “The vowels and the perception of Chinese tones,” Acta Psychol. Sin., **34**, 29–34.

Zeng, Fan-Gang, Kaibao Nie, Ginger S. Stickney, Ying-Yee Kong, Michael Vongphoe, Ashish Bhargave, Chaogang Wei, and Keli Cao. (**2005**) "Speech recognition with amplitude and frequency modulations." Proceedings of the National Academy of Sciences **102**(7), 2293-2298.

Zhou, N., Huang, J., Chen, X., and Xu, L. (**2013**). “Relationship Between Tone Perception and Production in Prelingually Deafened Children With Cochlear Implants,” Otol. Neurotol., **34**, 499–506.

Zhou, N., & Xu, L. (**2008**). Development and evaluation of methods for assessing tone production skills in Mandarin-speaking children with cochlear implants. The Journal of the Acoustical Society of America, **123**(3), 1653-1664.

Participant	Chronological age (years)	Age of implantation (years)	Hearing age (years)	Gender	CI position	Contralateral hearing aid	Hearing Threshold Left ear	Hearing Threshold Right ear
1	4.5	1.4	3.0	Female	Left	Yes	100	100
2	6.5	2.4	4.2	Male	Left	Yes	106	107
3	5.0	2.3	2.7	Male	Right	Yes	100	100
4	4.1	1.6	2.5	Male	Left	Yes	100	100
5	6.5	2.0	4.5	Male	Bilateral	No	90	110
6	6.8	1.9	4.9	Female	Left	Yes	100	100
7	5.3	2.6	2.7	Female	Right	Yes	103	108
8	4.5	3.3	1.1	Female	Left	Yes	100	95
9	7.0	2.0	4.9	Male	Left	Yes	105	105
10	6.1	1.7	4.4	Female	Right	Yes	120	110
11	6.9	3.2	3.7	Female	Right	Yes	95	115
12	6.5	2.4	4.1	Female	Right	No	110	110
13	5.2	2.8	2.4	Male	Right	Yes	95	112
14	6.4	2.7	3.8	Male	Left	Yes	105	110
15	6.9	3.3	3.6	Male	Left	Yes	90	90

Note: The hearing thresholds (PTA3, pure-tone average for three frequencies) are before implantation. All children except one were using CI device from Cochlear Company, one from Advanced Bionics

Table 1. Description of the group of children using CIs.

CONSONANTS Manner contrasts	Discrimination	Word Identification	TONES Tones 1 vs. 3	Discrimination	Word Identification
Unaspirated, Alveolar Affricate vs. Fricative [ʦ] vs. [s]	pa4ʦa1–pa4sa1	走 zou3 (to walk) 叟 sou3 (elder people)	[u]	pa4pu1–pa4pu3	菇 gu1 (mushroom) 鼓 gu3 (drum)
Unaspirated retroflex Fricative vs. affricate [ʂ] vs. [tʂ]	pa4ʂa1–pa4tʂa1	書 shu1 (book) 豬 zhu1 (pig)	[ao]	pa4pao1–pa4pao3	刀 dao1 (knife) 島 dao3 (island)
Unaspirated Velar Fricative vs. Stop [x] vs. [k]	pa4xa1–pa4ka1	火 huo3 (flame) 果 guo3 (fruit)	[i]	pa4pi1–pa4pi3	溪 xi1 (small river) 屣 xi3 (shoes)
Place contrasts			**Tones 1 vs. 2**		
Unaspirated Stop Labial vs. Alveolar [p] vs. [t]	pa4pa1–pa4ta1	綳 beng1 (bandage) 燈 deng1 (lamp)	[u]	pa4pu1–pa4pu2	驹 jyu1 (foal) 桔 jyu2 (mandarin)
Unaspirated Affricate Alveolar vs. Retroflex [ʦ] vs. [tʂ]	pa4ʦa1–pa4tʂa1	支 zhi1 (branch) 髭 zi1 (moustache)	[ao]	pa4pao1–pa4pao2	筲 shao1 (basket) 勺 shao2 (spoon)
Unaspirated Fricative Retroflex vs. Velar [ʂ] vs. [x]	pa4ʂa1–pa4xa1	虎 hu3 (tiger) 鼠 shu3 (mouse)	[i]	pa4pi1–pa4pi2	滴 di1 (drop) 笛 di2 (flute)

Table 2. In the discrimination task, the first syllable of the CVCV pseudo-words was always /pa/ produced with tone 4. For the consonant condition, the consonant of the second syllable was varied, according to a change in manner or place of articulation and tone 1 was used. For the tone condition, the consonant of the second syllable was always /p/ produced with three different vowels and the lexical tone was varied. In the word identification task, the same speech contrasts were used with Mandarin words usually known by 4-year-olds.

Speech contrast	Discrimination scores: Vocabulary level	Age at implantation	Duration of CI use	Hearing Threshold Left ear	Hearing Threshold Right ear	Identification scores: Vocabulary level	Age at implantation	Duration of CI use	Hearing Threshold Left ear	Hearing Threshold Right ear
Manner	.406 *.216*	-.102 *.717*	.346 *.207*	-.326 *.235*	-.100 *.724*	-.053 *.878*	-.078 *.782*	.257 *.355*	-.241 *.387*	-.051 *.858*
Place	-.169 *.620*	-.265 *.340*	.577 *.024**	.191 *.495*	-.056 *.844*	.464 *.150*	-.084 *.767*	.375 *.168*	-.369 *.176*	.158 *.575*
Tones 1 vs. 2	-.219 *.517*	.300 *.278*	.521 *.046**	-.022 *.938*	-.080 *.778*	.435 *.181*	-.036 *.899*	.258 *.352*	-.444 *.098*	-.270 *.331*
Tones 1 vs. 3	.046 *.893*	.317 *.250*	.435 *.105*	-.373 *.171*	-.181 *.518*	.285 *.396*	.279 *.313*	.199 *.477*	-.643 *.010**	-.036 *.898*

Note: n = 15 except for correlations with vocabulary level (n = 11). *. Correlation is significant at the .05 level (two-tailed).

Table 3. Pearson correlation coefficients (first row) and *p* values (second row) between discrimination scores and identification scores for each speech contrast and child's characteristics.

	CI group, N =15				NH group, N = 37			
Correlations between discrimination and identification scores	**Manner**	**Place**	**Tones 1 vs. 2**	**Tones 1 vs. 3**	**Manner**	**Place**	**Tones 1 vs. 2**	**Tones 1 vs. 3**
Manner	.462	.584*	.504	.507	.166	.310	.174	.117
	.083	.022	.055	.054	*.325*	*.062*	*.303*	*.490*
Place	.642*	.348	.347	.534*	.146	.154	.117	.027
	.010	.204	.205	.040	*.387*	*.363*	*.492*	*.872*
Tones 1 vs. 2	.715**	.292	.388	.667*	.210	.181	.533**	.277
	.003	.292	.153	.007	*.212*	*.285*	*.001*	*.097*
Tones 1 vs. 3	.656*	.232	.444	.721**	.115	.248	.138	.051
	.008	.406	.097	.002	*.497*	*.138*	*.416*	*.765*

Note: *: Correlation is significant at the .05 level (two-tailed).

**: Correlation is significant at the .003 level corrected for multiple comparisons (two-tailed).

Table 4. Correlation coefficients (first row) and *p* values (second row) between speech perception scores in the phonetic discrimination and word identification tasks for the groups of children using CIs (Pearson coefficients) and children with NH (Spearman coefficients).

Figures

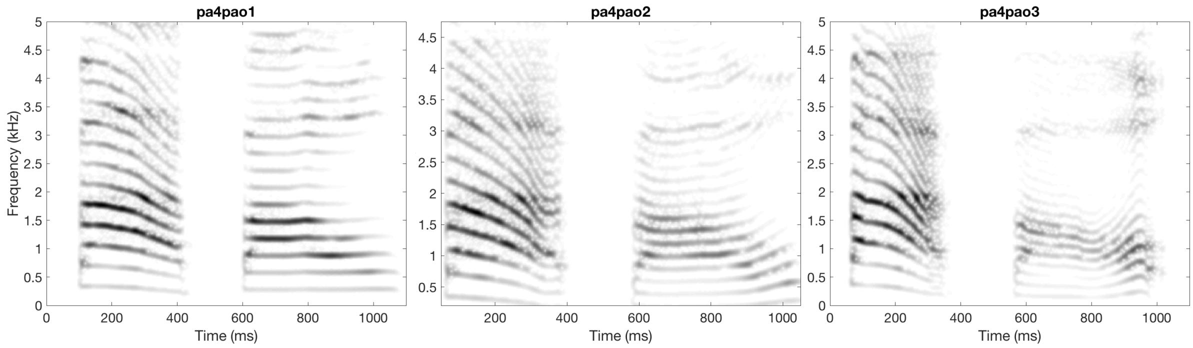

Figure 1. Narrowband spectrograms of the CVCV /pa4pao1/ (left panel), /pa4pao2/ (middle), and /pa4pao3/(right) produced by one speaker. In total, 54 different CVCVs were produced in the tone and consonant conditions by each speaker. The mean duration of those CVCVs was 1.03 s (SD = .05) and 1.05 s (SD = 0.04) for speaker 1 and 2, respectively. The mean F0 of each speaker was also different (287 and 245 Hz).

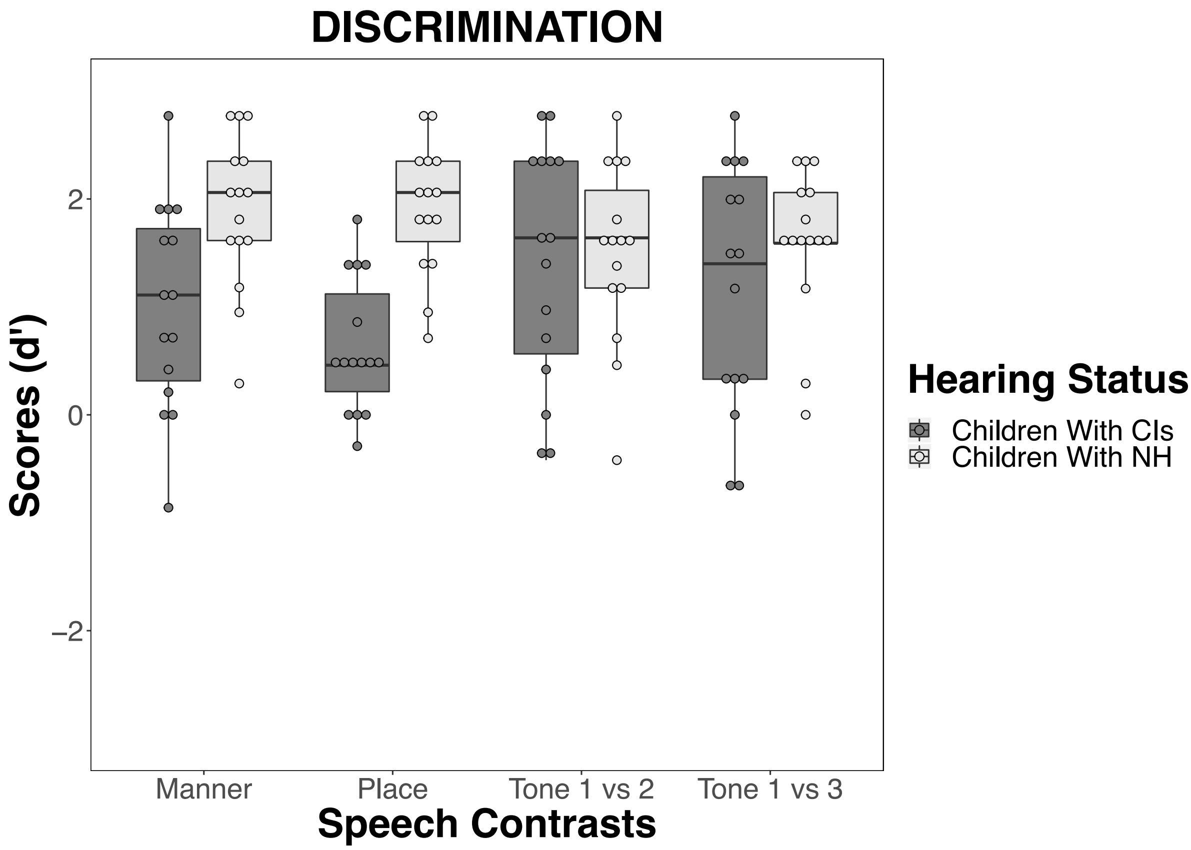

Figure 2. Boxplots for the discrimination scores (*d*') for the four speech contrasts (manner, place, tones 1 vs. 2, and tones 1 vs. 3) for CI users (dark grey bars) and controls with NH (light grey bars). Each dot represents an individual. The three horizontal lines represent the 25th, 50th, and 75th percentiles, respectively, and approximately 95% of the data are expected to lie between the vertical bars. The points that do not fall on the vertical bars represent the outliers.

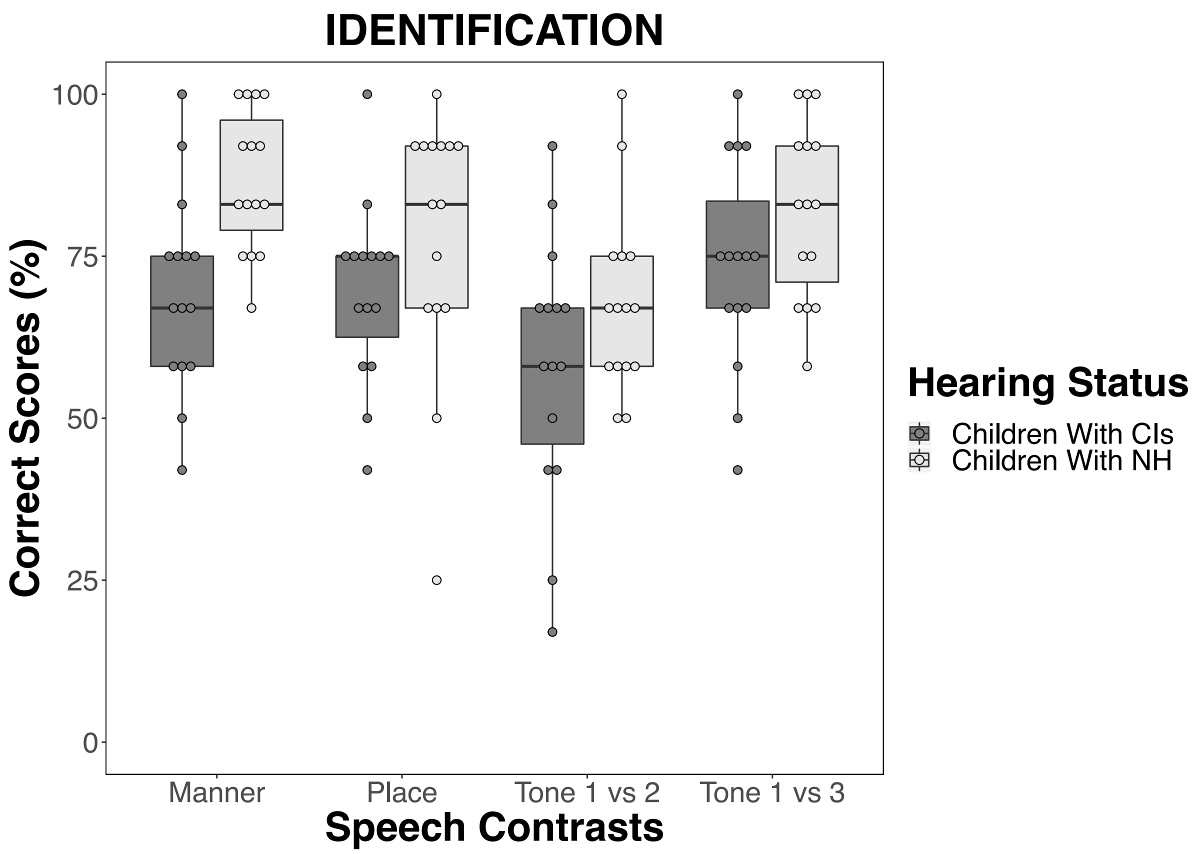

Figure 3. Boxplots for the identification scores (percentage correct) for the four speech contrasts (manner, place, tones 1 vs. 2, and tones 1 vs. 3) for CI users (dark grey bars) and controls with NH (light grey bars). Each dot represents an individual. The three horizontal lines represent the 25th, 50th, and 75th percentiles, respectively, and approximately 95% of the data are expected to lie between the vertical bars. The points that do not fall on the vertical bars represent the outliers.