



HAL
open science

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

► **To cite this version:**

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. *Journal of the Acoustical Society of America*, 2019, 146 (4), pp.2291-2302. 10.1121/1.5126941 . hal-02315756

HAL Id: hal-02315756

<https://hal.science/hal-02315756v1>

Submitted on 14 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

3 **Laurianne Cabrera,^{1,a)} Huei-Mei Liu,^{2,b)} Lionel Granjon,^{1,a)} Chieh Kao,^{3,c)}**
4 **and Feng-Ming Tsao^{3,d)}**

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

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

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

11 ^{a)}Also at: Integrative Neuroscience and Cognition Center, CNRS, 45 rue des saints pères, 75006,
12 Paris, France.

13 ^{b)}Also at: Institute for Research Excellence in Learning Sciences, National Taiwan Normal
14 University, 162, Section 1, Heping E. Road, Taipei City 106, Taiwan.

15 ^{c)}Also at: Department of Speech-Language-Hearing Sciences, University of Minnesota Twin Cities,
16 115 Shevlin Hall, 164 Pillsbury Drive, Southeast Minneapolis, MN 55455, USA.

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

20
21 Received 8 January 2019; revised 27 August 2019; accepted 3 September 2019;
22 published online 10 October 2019

23 *The Journal of the Acoustical Society of America*

24 <https://doi.org/10.1121/1.5126941>

28 **Abstract**

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

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

47

48 **Keywords:** cochlear implants, lexical tones, consonants, discrimination, word
49 identification

50 I. INTRODUCTION

51 A cochlear implant (CI) is an electronic device implanted in the peripheral auditory
52 system of people with severe-to-profound sensorineural hearing loss. Currently, CIs are
53 fitted to children with congenital hearing loss before one year of age (Lammers *et al.*,
54 2015; Miyamoto *et al.*, 2008; Tomblin *et al.*, 2005). CIs undoubtedly benefit users as
55 speech perception improves on average after implantation (Blamey *et al.*, 2001; Sarant *et al.*,
56 2001; Svirsky *et al.*, 2000; Uhler *et al.*, 2011). However, CI processors do not convey
57 well the fine spectro-temporal information of speech sounds (Shannon, 2012). CI users
58 experience great difficulties in perceiving some phonetic contrasts as well as the pitch of
59 speech (Zeng *et al.*, 2005). How do children using CIs learn the fine phonetic details of
60 their native language in such degraded conditions? The present study focused on the
61 perception of consonants and lexical tones by Mandarin-speaking children using CIs aged
62 between 4 and 7 years.

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

64 Current CI processors convey only the slow amplitude modulations (AM or temporal
65 envelope) of the original external signal via a limited number of electrodes corresponding
66 to relatively broad frequency bands (Shannon, 2012). The temporal fine structure (TFS)
67 of the original signal, which corresponds to the fastest temporal fluctuations and are
68 essential to convey pitch information (Moore, 2008; Rosen, 1992; Smith *et al.*, 2002; Xu
69 and Pfingst, 2003), is not transmitted by CI processors. Instead, it is replaced by a fixed
70 train of pulses in which amplitude is modulated by the original temporal envelope. Thus,
71 fine pitch information is not well transmitted by CIs. Although pitch variations are
72 exhibited in all languages, as they can convey syntactic units (*e.g.*, upward inflection for
73 questions), attract attention, or express emotional information (Collier, 1975;
74 Nooteboom, 1997), they are particularly crucial for manifesting lexical meanings of

75 syllables in tone languages such as Mandarin Chinese. One of the current challenges for
76 CI technology is to better encode this information in order to improve speech perception
77 for listeners using tone languages.

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

95 However, it is important to note that the adult participants of those studies had
96 developed language before their hearing loss. These adult participants with post-lingual
97 hearing loss had already mastered their native language. Studying children who have pre-
98 lingual hearing loss and thus do not have any prior knowledge about their linguistic

99 system will help us to understand how listeners rely on an impoverished speech signal in
100 order to learn a language.

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

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

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

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

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

149 **C. CI-simulation studies and lexical tone perception**

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

165 Recent simulation studies have also shown that reliance on the modulation
166 components of speech is affected by language exposure/experience. Mandarin-speaking
167 adults with NH rely more on the voice-pitch information conveyed by the fine spectro-
168 temporal cues, as they are more affected by its degradation when discriminating lexical
169 tones compared to French-speaking adults with NH (Cabrera *et al.*, 2014). This cross-
170 linguistic difference is also observed at 10 months of age; however, it is not observed at
171 6 months. This observation suggests that the duration of exposure to the native language
172 does influence the use of the acoustic information of the speech signal (Cabrera *et al.*,
173 2015). Interestingly, in those studies, French listeners (infants at 10 months and adults)

174 were shown to be better able to use the remaining temporal envelope to discriminate
175 non-native lexical tones compared to Mandarin listeners. Therefore, the discrimination
176 of lexical tones in such CI-simulation conditions is influenced by language experience.
177 Together, these simulation studies entail two hypotheses regarding the perception of
178 consonants vs. lexical tones by Mandarin-speaking children using CIs. First, like adults
179 with post-lingual hearing loss using CIs, these children may exhibit poorer perception of
180 lexical tones compared to consonants because CI processors convey only the temporal
181 envelope of speech. Second, children with pre-lingual hearing loss may have developed
182 more efficient perceptual strategies (*e.g.*, relying on temporal envelope information) as a
183 result of being exposed to degraded acoustic information while acquiring their native
184 language. If this is the case, we might expect them to perceive lexical tones as well as
185 consonants.

186 **D. Comparison of children’s abilities to discriminate and identify words with** 187 **lexical tone and consonant contrasts**

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

199 (Bouton *et al.*, 2012; Shannon *et al.*, 1995). Based on previous studies with children using
200 CIs, it was predicted that the contrast of tone 1 vs. tone 3 would be easier to perceive
201 because of its larger acoustic distance relative to the tone 1 vs. tone 2 contrast, which
202 shows similar pitch height (Han *et al.*, 2009; Peng *et al.*, 2004). We also expected that
203 perceptual differences between consonants and lexical tones would be task-dependent.
204 A phonetic discrimination task relies more on an acoustic/phonetic level of processing
205 while a word identification task requires access to lexical representations (Wong & Wong,
206 2004).

207 II. METHOD

208 A. Participants

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

220 Another group of 37 children with NH (mean = 5.3 years, sd = 0.6, range= 4.1 to
221 6.7 years; boys N = 20) were tested in quiet rooms at kindergartens in Taiwan. Three
222 additional children were tested in this group but not included because of inability to

223 perform the task (N = 2) or mild hearing loss (N = 1). This group of children had received
224 regular health checks since infancy and exhibited typical developmental history.

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

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

235

236 **B. Stimuli**

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

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

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

261

262 **C. Procedure**

263 Children with NH completed all tasks in approximately 40 min and children with
264 hearing loss were tested within one session of approximately 2 hours including breaks.
265 For both groups, sounds were presented at a level of around 70 dB SPL via loudspeakers,
266 located on each side of the computer monitor in front of the child at approximately 30–
267 40 cm. CI processors were adjusted to each participant’s daily use setting. The
268 experimenter sat next to the child. The phonetic discrimination task was conducted on a
269 laptop computer using E-Prime. In this task, each trial was composed of four spoken
270 pseudo-words while two pictures were displayed on the screen. Children were asked to
271 point at a picture depicting four cows when the sounds were the same and at a picture
272 depicting two cows and two frogs when the sounds were different. Half of the trials were

273 'same' trials (*i.e.*, AA-AA) and the other half were 'change' trials (*i.e.*, AA-BB). Two short
274 familiarization blocks were designed to familiarize the children with the task and the
275 setup. The first familiarization block included up to four trials consisting of either only
276 cow sounds or both cow and frog sounds instead of speech sounds. This was done to make
277 sure that children understood the concepts of 'same' and 'different' (Holt and Lalonde,
278 2012). The second familiarization block included up to eight trials consisting of CVCV
279 pseudowords and same or change trials varying on non-minimal phonetic pairs
280 (consonants and vowels not used in the testing blocks). Feedback, in the form of a positive
281 or negative emoticon, was displayed on the screen after each trial during familiarization.
282 If children could not perform above chance in this second familiarization block within a
283 maximum of three attempts, they were judged to be unable to complete the tasks (two
284 children with CIs were excluded following this criterion). Following familiarization, all
285 children performed a total of four testing blocks, without feedback, comprising 24 trials
286 (12 change and 12 no-change) for each speech contrast (manner, place, tones 1 vs. 2, and
287 tones 1 vs.3) for a total of 96 trials.

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

296 Children were encouraged to point at the 'different' or 'same' picture on the screen
297 to indicate their response. The experimenter recorded the child's response using the

298 computer mouse. In addition, a puzzle piece appeared on the screen after children
299 completed each trial in an effort to maintain interest and let participants track their
300 progress in the task. The correct responses on change and no-change trials were recorded
301 and averaged for each of the four speech contrasts (manner, place, tones 1 vs. 2, and tones
302 1 vs. 3). Participants' accuracy was estimated by a d' score where $d' = Z[p(\text{hit}) - Z[p(\text{false}$
303 $\text{alarms})]$ (Macmillan and Creelman, 1991). If the hit rate was 1 or the false alarm rate was
304 0, then d' was calculated after adjusting the hit rate or false alarm rate by the reciprocal
305 of the number of trials. The maximum d' score in this task was 2.77.

306 Children were also asked to perform a word identification task. For each trial, two
307 pictures of objects were presented on the screen and children were asked in Mandarin to
308 "show the -target object". The experimenter recorded the child's response using the
309 computer mouse. Three training trials, with feedback, were used to familiarize the
310 children with the procedure by presenting words differing in non-minimal phonetic
311 contrasts (that were not used later in the task). This training was repeated a second time
312 if the child was not able to get two correct answers out of three (none of the children were
313 excluded based on this criterion). Once the training was completed, the pictures of two
314 objects whose names differed in minimal pairs from **Table 2** were presented. During the
315 test, no feedback was provided. A total of 12 word pairs was presented four times within
316 two blocks with a break in between. For each word pair (A-B), two trials presented the
317 words produced by speaker 1, and the other two presented the words produced by
318 speaker 2. The presentation side of the target picture (left or right) was counterbalanced
319 for each word pair. Consonant and lexical-tone trials were presented randomly within
320 each block. Correct identification scores were computed for each pair. Half of the children
321 completed the discrimination task first, and the other half completed the identification
322 task first. The order of the blocks was also counterbalanced between children. Finally, a

323 standardized assessment was used to evaluate the receptive vocabulary skills of all
324 children with NH and 11 children using CIs, *i.e.*, the Mandarin version of the Peabody
325 Picture Vocabulary Test (Lu and Liu, 1988).

326 **III. RESULTS**

327 **A. Phonetic discrimination task**

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

329 The performance of the 15 CI users was compared to a sub-group of 15 children
330 with NH who were matched in chronological age [mean age for children with NH = 5.5
331 years, $sd = 0.7$; mean age for children using CIs = 5.9 years, $sd = 1.0$; $t(28) = 1.18, p = .25$]
332 and non-verbal IQ [mean non-verbal IQ for children with NH = 99.0, $sd = 11.2$; mean
333 nonverbal IQ for children using CIs = 98.6, $sd = 14.1$; $t(27) = -0.158, p = .88$; all η^2 reported
334 here are partial]. **Figure 2** shows the d' scores of these two groups for each of the speech
335 contrasts tested. A mixed analysis of variance (ANOVA) with Hearing Status (NH vs. CI)
336 as the between-subject factor and Speech contrast (4) as the within-subject factor was
337 performed on the d' scores. This analysis revealed no main effect of Speech Contrast
338 [$F(3,84) = 0.67, p = .53, \eta^2 = .023$] but a main effect of Hearing Status [$F(1,28) = 7.09, p$
339 $= .013, \eta^2 = .20$]. CI users showed overall lower d' scores than controls with NH. This
340 analysis also revealed a significant interaction between these two factors [$F(3,84) = 5.89,$
341 $p = .003, \eta^2 = .17$]. Pairwise comparisons using Bonferroni corrections indicated that
342 children with NH do not show any difference in their d' scores across the speech contrasts
343 ($ps > .45$), but that CI users had poorer scores for the place of articulation contrast
344 compared to tones 1 vs. 2 ($p = .001$). Moreover, the children with NH showed significantly
345 higher scores than CI users for both manner and place of articulation ($p = .008$ and p
346 $< .001$, respectively), but equivalent scores for tones 1 vs. 2 and 1 vs. 3 ($p = .77, p = .24,$
347 respectively). One-sample t -tests revealed that the averaged d' scores in each condition

348 were above chance level ($= 0$) for the group of CI users ($p = .001, .001, < .001$, and $.001$,
349 respectively for manner, place, tones 1 vs. 2, and tones 1 vs. 3 contrasts) and for the group
350 of children with NH ($ps < .001$).

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

355

356 **2. Exploration of potential factors influencing performance of children using** 357 **CIs**

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

371

372 **3. Generalization of the results for children with NH**

373 Finally, a supplementary analysis was performed for the whole group of children
374 with NH speaking Mandarin. This analysis aimed to assess whether similar results for
375 children with NH would be observed from a larger group of 37 children aged 4–7 years.
376 Their mean scores for manner, place, tones 1 vs. 2, and tones 1 vs. 3 were 1.79 (sd = 0.94),
377 1.53 (sd = 0.82), 1.61 (sd = 0.94), and 1.77 (sd = 0.77), respectively. As the previous
378 analysis with the sub-group of 15 children, a repeated-measures ANOVA performed on
379 the d' scores with Speech contrast as the within-subject factor did not reveal any
380 difference between the four contrasts [$F(3,108) = 2.16, p = .097, \eta^2 = .057$].

381

382 **B. Word identification task**

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

384 The identification scores of the CI users were also compared with those of the
385 same sub-group of children with NH matched in chronological age and non-verbal IQ.
386 **Figure 3** shows the correct identification scores for these two groups as a function of
387 Speech contrast. A mixed ANOVA assessed the effect of Speech contrast (4) as the within-
388 subject factor and Hearing status (2) as the between-subject factor. There were
389 significant effects for both Speech contrast [$F(3,84) = 9.08, p < .001, \eta^2 = .25$] and Hearing
390 status [$F(1,28) = 7.41, p = .011, \eta^2 = .21$], but no Speech contrast vs. Hearing status
391 interaction [$F(3, 84) = 0.75, p = .52, \eta^2 = .026$]. Pairwise comparisons with Bonferroni
392 corrections indicated that, for both the CI users and children with NH, the identification
393 scores for tones 1 vs. 2 contrasts were significantly lower (poorer) than those of manner
394 ($p < .001$), place of articulation ($p = .029$) and tones 1 vs. 3 ($p < .001$). In addition, children
395 with CIs exhibited lower (poorer) word-identification scores than the matched NH group.
396 This is consistent with the fact that CI users showed significantly lower vocabulary scores
397 than controls with NH [raw score range was 16–90 for CI users (mean = 50.5, sd= 33.1)]

398 and 29–99 for children with NH (mean = 59.7, sd = 17.9); $t(28) = 3.44, p = .002, \eta^2 = .30$.
399 One-sample t -tests revealed that, even though some children using CIs showed scores
400 below or chance (0.50), the averaged identification score for the group of CI users was
401 above chance for manner, place, and tones 1 vs. 3, but not for tones 1 vs. 2 contrasts (ps
402 $< .001$ and $= .13$) and were above chance for each contrast for the group of children with
403 NH ($ps < .001$). These analyses thus suggested that, even though CI users had overall
404 poorer word-identification scores than controls with NH, these two groups exhibited the
405 same pattern of response among speech contrasts. Both CI users and children with NH
406 had more difficulties in identifying words differing on tones 1 vs. 2 than on consonants.

407

408 **2. Exploration of potential factors influencing performance of children using** 409 **CIs**

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

415

416 **3. Generalization of the results for children with NH**

417 Finally, the identification scores for the larger cohort of 37 children with NH were
418 also compared between speech contrast. For manner, place, tones 1 vs. 2, and tones 1 vs.
419 3, the mean identification scores were 86 (sd = 10), 79 (sd = 14), 70 (sd = 14), and 81 (sd
420 = 14), respectively. A repeated-measures ANOVA was performed on the percentage of
421 correct identification scores with Speech contrast as the within-subject factor and
422 showed the same main effect of Speech contrast as observed with the sub-group of 15

423 children with NH [$F(3,108) = 10.63, p < .001; \eta^2 = 22.8$]. *Post hoc* tests with Bonferroni
424 corrections revealed that the scores for the contrast of tones 1 vs. 2 were significantly
425 lower than for manner, place, and tones 1 vs. 3 ($p < .001, p = .04$, and $p = .002$, respectively).

426

427 **4. Identification scores and parents' questionnaires**

428 As expected, the contrast with similar pitch heights (*i.e.*, high-level tone 1 and
429 high-rising tone 2) was more difficult than the contrast with distinct pitch heights (*i.e.*,
430 high-level tone 1 and low-dipping tone 3). To further explore the reasons for these low
431 scores with the tones 1 vs. 2 contrast, the parents' responses to the questionnaire
432 regarding whether the words used in the identification task were known by their children
433 were compared to the children's responses for each word pair. According to the parents'
434 questionnaire, the words used in the test of tones 1 vs. 2 were not well-known by both
435 children using CIs and those with NH. On average, parents reported that 72% of the words
436 in the test of tones 1 vs. 2 were not known by the children using CIs (compared to 26%
437 for manner, 38% for place, and 48% for tones 1 vs. 3). For the children with NH, 64% of
438 words were not known in the test of tones 1 vs. 2 (compared to 20% for manner, 33% for
439 place, and 32% for tones 1 vs. 3). The lowest identification scores observed in the tones
440 1 vs. 2 test might be the results of both perceptual difficulties with this tonal contrast and
441 the word pairs selected for this tone test probably being less common to children aged
442 4–7 years. However, it is important to note that the overall proportion of words known
443 per pair as indicated by the parental questionnaire and the actual proportion of children's
444 correct responses did not correlate with each other [$r = -0.056, p = .17$]. For this reason,
445 we preferred not to exclude the trials with words supposed to be not well-known by the
446 children.

447

448 **C. Relationship between phonetic discrimination and word identification**
449 **tasks**

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

458 **IV. GENERAL DISCUSSION**

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

465

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

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

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

479 For lexical-tone perception, a previous study found that tones 1 vs. 3 is the easiest
480 contrast to discriminate for 12-month-old Mandarin-learning infants with NH (Tsao,
481 2008), possibly because of the larger pitch-height differences for this contrast.
482 Nonetheless, in the present study, older children did not differ in their discrimination

483 scores for the two lexical-tone contrasts (tone 1 vs. 2 and tone 1 vs. 3). This finding is
484 consistent with data obtained from a large group of Mandarin-speaking children with NH
485 aged 3–10 years (Zhou *et al.*, 2013). When identifying lexical tones, tone 3 was the most
486 difficult for 3-year-old Mandarin-speaking children with NH when segmental cues were
487 reduced by low-pass filtering (Wong *et al.*, 2005). In the present word-identification task,
488 tones 1 vs. 2 contrasts led to the lowest scores for children with NH. The similar pitch-
489 height between tones 1 and 2 could make this contrast more difficult to process. It is also
490 possible that our choice of stimuli may have accounted for the lower score in this
491 condition, as parent reports suggested that children may have been less familiar with
492 these words.

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

504

505 **B. Perception of lexical tones and consonants using CIs**

506 The goal of this study was to assess the phonetic discrimination and word
507 identification abilities of Mandarin-speaking children using CIs in order to explore

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

517 Nevertheless, the results of the phonetic discrimination task revealed some
518 differences between children with NH and children using CIs. These two groups of
519 children did not perform similarly in detecting changes in consonants and lexical tones.
520 Although children using CIs were able to detect changes in consonants and lexical tones
521 above chance, their discrimination scores for consonants, for both manner and place
522 changes, were significantly lower than their peers with NH. In particular, the CI users
523 showed specific difficulties with the place-of-articulation contrasts. However, they were
524 able to discriminate lexical tones as well as their peers with NH. Thus, this study supports
525 the hypothesis that children using CIs may perform less accurately in discriminating
526 consonant contrasts such as place of articulation, which are mainly conveyed by fine
527 spectro-temporal information, compared to manner, which relies more on envelope
528 information (Rosen, 1992; Bouton *et al.*, 2012). These results also suggest that the CI
529 processors used by participants in this study were sufficient to convey gross pitch
530 information and secondary cues such as duration and amplitude for lexical-tone
531 discrimination (Kuo *et al.*, 2008; Xu *et al.*, 2002). Previous acoustic analysis of the
532 naturally produced Mandarin syllables showed that duration is similar among tones 1, 2,

533 and 3 (Liu, Tsao, & Kuhl, 2007); thus, the secondary duration cue was not considered to
534 be effective for discriminating the present contrasts. The duration of lexical tones might
535 not account for the differences between the NH and CI groups in the present lexical-tone
536 discrimination task. Amplitude cues, however, are likely to play a greater role for CI users,
537 who have better access to temporal cues than to spectral cues (Fu & Zeng, 2000). In the
538 present discrimination task, children using CIs might have benefited more from this
539 secondary amplitude cue for the lexical-tone contrasts than for the consonant contrast of
540 place of articulation, for which amplitude plays a small role (Rosen, 1992).

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

558 tones. Future studies should directly compare perceptual difficulties of children with pre-
559 lingual hearing loss and adults with post-lingual hearing loss using CIs to test this
560 hypothesis.

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

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

588

589 **C. Relationship between hearing experience, verbal skills, and speech** 590 **perception for children using CIs**

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

608 Additionally, consistent with many previous studies, we found that children using CIs
609 exhibited lower vocabulary scores than their peers with NH, suggesting that auditory
610 deprivation in early ages had impeded language acquisition (Lund, 2015; Välimaa *et al.*,
611 2018). Moreover, there is a general trend that the performance variability of the CI group
612 for discrimination and identification of speech sounds is greater (as shown in Figures 1
613 and 2) than that of the NH group, showing that Mandarin-speaking children with similar
614 severity of hearing loss before CI implantation do not perceive consonants and lexical
615 tones in similar ways after CI implantation. Longitudinal and large-scale studies would
616 help to evaluate further the role of hearing experience with CIs on speech and language
617 development.

618

619 **V. CONCLUSIONS**

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

632 highlights the benefits of cochlear implantation in children with severe-to-profound
633 hearing loss, even for those learning a tone language.

634

635 **Acknowledgments**

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

644

645 **Declaration of interests**

646 The authors report no conflicts of interest.

647

648 **REFERENCES**

- 649 Blamey, P. J., Sarant, J. Z., Paatsch, L. E., Barry, J. G., Bow, C. P., Wales, R. J., Wright, M.,
650 et al. (2001). "Relationships among speech perception, production, language, hearing
651 loss, and age in children with impaired hearing," *J. Speech Lang. Hear. Res.*, **44**, 264–
652 285.
- 653 Bouton, S., Serniclaes, W., Bertoncini, J., and Colé, P. (2012). "Perception of speech features
654 by French-speaking children with cochlear implants," *J. Speech Lang. Hear. Res.*
655 *JSLHR*, **55**, 139–153.
- 656 Brown, L., Sherbenou, R. J., and Johnsen, S. K. (2010). *Test of Nonverbal Intelligence:*
657 *TONI-4*, Pro-ed.
- 658 Cabrera, L., Tsao, F.-M., Gnansia, D., Bertoncini, J., and Lorenzi, C. (2014). "The role of
659 spectro-temporal fine structure cues in lexical-tone discrimination for French and
660 Mandarin listeners," *J. Acoust. Soc. Am.*, **136**, 877–882.
- 661 Cabrera, L., Tsao, F.-M., Liu, H.-M., Li, L.-Y., Hu, Y.-H., Lorenzi, C., and Bertoncini, J.
662 (2015). "The perception of speech modulation cues in lexical tones is guided by early
663 language-specific experience," *Frontiers in psychology*, **6**, 1290.
- 664 Ciocca, V., Francis, A. L., Aisha, R., and Wong, L. (2002). "The perception of Cantonese
665 lexical tones by early-deafened cochlear implantees," *J. Acoust. Soc. Am.*, **111**, 2250-
666 2256.
- 667 Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). "Speech recognition in
668 noise as a function of the number of spectral channels: comparison of acoustic hearing
669 and cochlear implants," *J. Acoust. Soc. Am.*, **110**, 1150–1163.
- 670 Fu, Q.-J., and Zeng, F.-G. (2000). "Identification of temporal envelope cues in Chinese tone
671 recognition," *Asia Pac. J. Speech Lang. Hear.*, **5**, 45–58.

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

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

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

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

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

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

687 Holt, R. F., and Lalonde, K. (2012). "Assessing toddlers' speech-sound discrimination," *Int.*
688 *J. Pediatr. Otorhinolaryngol.*, **76**, 680–692.

689 Khouw, E., and Ciocca, V. (2007). "Perceptual correlates of Cantonese tones," *J. Phon.*, **35**,
690 104–117.

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

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

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

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

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

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

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

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

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

713 Macmillan, N. A., and Creelman, C. D. (1991). "Detection theory: A user's guide Cambridge
714 University Press," N. Y.

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

718 Miyamoto, R. T., Hay-McCutcheon, M. J., Kirk, K. I., Houston, D. M., and Bergeson-Dana,
719 T. (2008). "Language skills of profoundly deaf children who received cochlear
720 implants under 12 months of age: a preliminary study," *Acta Otolaryngol. (Stockh.)*,
721 **128**, 373–377.

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

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

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

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

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

735 Shannon, R. V. (2012). "Advances in auditory prostheses," Curr. Opin. Neurol., 25, 61–66.

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

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

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

743 Stilp, C. E., Goupell, M. J., and Kluender, K. R. (2013). "Speech perception in simulated
744 electric hearing exploits information-bearing acoustic change," J. Acoust. Soc. Am.,
745 133, EL136–EL141.

- 746 Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B., and Miyamoto, R. T. (2000).
747 “Language development in profoundly deaf children with cochlear implants,”
748 *Psychol. Sci.*, **11**, 153–158.
- 749 Tomblin, J. B., Barker, B. A., Spencer, L. J., Zhang, X., and Gantz, B. J. (2005). “The effect
750 of age at cochlear implant initial stimulation on expressive language growth in infants
751 and toddlers,” *J. Speech Lang. Hear. Res.*, **48**, 853–867.
- 752 Tsao, F.-M. (2008). “The effect of acoustical similarity on lexical-tone perception of one-
753 year-old Mandarin-learning infants,” *Chin. J. Psychol.*, **50**, 111–124.
- 754 Tsao, F.-M., Liu, H.-M., and Kuhl, P. K. (2006). “Perception of native and non-native
755 affricate-fricative contrasts: cross-language tests on adults and infants,” *J. Acoust.
756 Soc. Am.*, **120**, 2285–2294.
- 757 Uhler, K., Yoshinaga-Itano, C., Gabbard, S. A., Rothpletz, A. M., and Jenkins, H. (2011).
758 “Longitudinal infant speech perception in young cochlear implant users,” *J. Am.
759 Acad. Audiol.*, **22**, 129–142.
- 760 Välimaa, T., Kunnari, S., Laukkanen-Nevala, P., and Lonka, E. (2018). “Early vocabulary
761 development in children with bilateral cochlear implants,” *Int. J. Lang. Commun.
762 Disord.*, **53**, 3–15.
- 763 Vance, T. J. (1976). “An experimental investigation of tone and intonation in Cantonese,”
764 *Phonetica*, **33**, 368–392.
- 765 Whalen, D. H., and Xu, Y. (1992). “Information for Mandarin tones in the amplitude contour
766 and in brief segments,” *Phonetica*, **49**, 25–47.
- 767 Wong, A. O., and Wong, L. L. (2004). “Tone perception of Cantonese-speaking prelingually
768 hearing-impaired children with cochlear implants,” *Otolaryngol. Neck Surg.*, **130**,
769 751–758.

770 Wong, P., Schwartz, R. G., and Jenkins, J. J. (2005). "Perception and production of lexical
771 tones by 3-year-old, Mandarin-speaking children," *J. Speech Lang. Hear. Res.*, **48**,
772 1065–1079.

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

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

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

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

781 Yang, Y. F. (1989). "The vowels and the perception of Chinese tones," *Acta Psychol. Sin.*,
782 **34**, 29–34.

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

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

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

793

794

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			TONES		
Manner contrasts	Discrimination	Word Identification	Tones 1 vs. 3	Discrimination	Word Identification
<i>Unaspirated, Alveolar Affricate vs. Fricative</i> [tʂ] vs. [s]	pa4tʂa1–pa4sa1	走 zou3 (to walk) 叟 sou3 (elder people)	[u]	pa4pu1–pa4pu3	菇 gu1 (mushroom) 鼓 gu3 (drum)
<i>Unaspirated retroflex Fricative vs. affricate</i> [ʂ] vs. [tʂ]	pa4ʂa1–pa4tʂa1	書 shu1 (book) 豬 zhu1 (pig)	[ao]	pa4pao1–pa4pao3	刀 dao1 (knife) 島 dao3 (island)
<i>Unaspirated Velar Fricative vs. Stop</i> [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		
<i>Unaspirated Stop Labial vs. Alveolar</i> [p] vs. [t]	pa4pa1–pa4ta1	繃 beng1 (bandage) 燈 deng1 (lamp)	[u]	pa4pu1–pa4pu2	駒 jyu1 (foal) 桔 jyu2 (mandarin)
<i>Unaspirated Affricate Alveolar vs. Retroflex</i> [tʂ] vs. [tʂ]	pa4tʂa1–pa4tʂa1	支 zhi1 (branch) 髭 zi1 (moustache)	[ao]	pa4pao1–pa4pao2	簞 shao1 (basket) 勺 shao2 (spoon)
<i>Unaspirated Fricative Retroflex vs. Velar</i> [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					Identification scores				
	Vocabulary level	Age at implantation	Duration of CI use	Hearing Threshold Left ear	Hearing Threshold Right ear	Vocabulary level	Age at implantation	Duration of CI use	Hearing Threshold Left ear	Hearing Threshold Right ear
Manner	.406	-.102	.346	-.326	-.100	-.053	-.078	.257	-.241	-.051
	.216	.717	.207	.235	.724	.878	.782	.355	.387	.858
Place	-.169	-.265	.577	.191	-.056	.464	-.084	.375	-.369	.158
	.620	.340	.024*	.495	.844	.150	.767	.168	.176	.575
Tones 1 vs. 2	-.219	.300	.521	-.022	-.080	.435	-.036	.258	-.444	-.270
	.517	.278	.046*	.938	.778	.181	.899	.352	.098	.331
Tones 1 vs. 3	.046	.317	.435	-.373	-.181	.285	.279	.199	-.643	-.036
	.893	.250	.105	.171	.518	.396	.313	.477	.010*	.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.

Correlations between discrimination and identification scores	CI group, N =15				NH group, N = 37			
	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

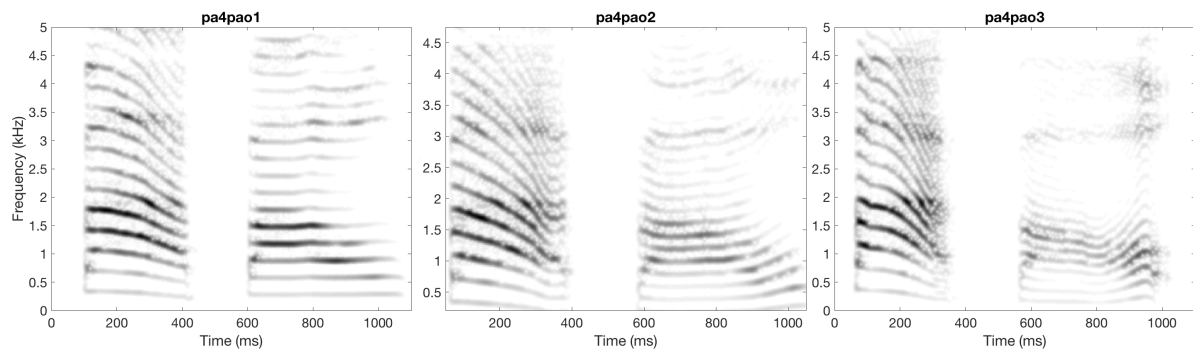


Figure 1. Narrowband spectrograms of the CVCV /pa4pao1/ (left panel), /pa4pao2/ (middle), and /pa4pao3/(right) produced by one speaker. In total, 54 different CVCVs were produced in the tone and consonant conditions by each speaker. The mean duration of those CVCVs was 1.03 s (SD = .05) and 1.05 s (SD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

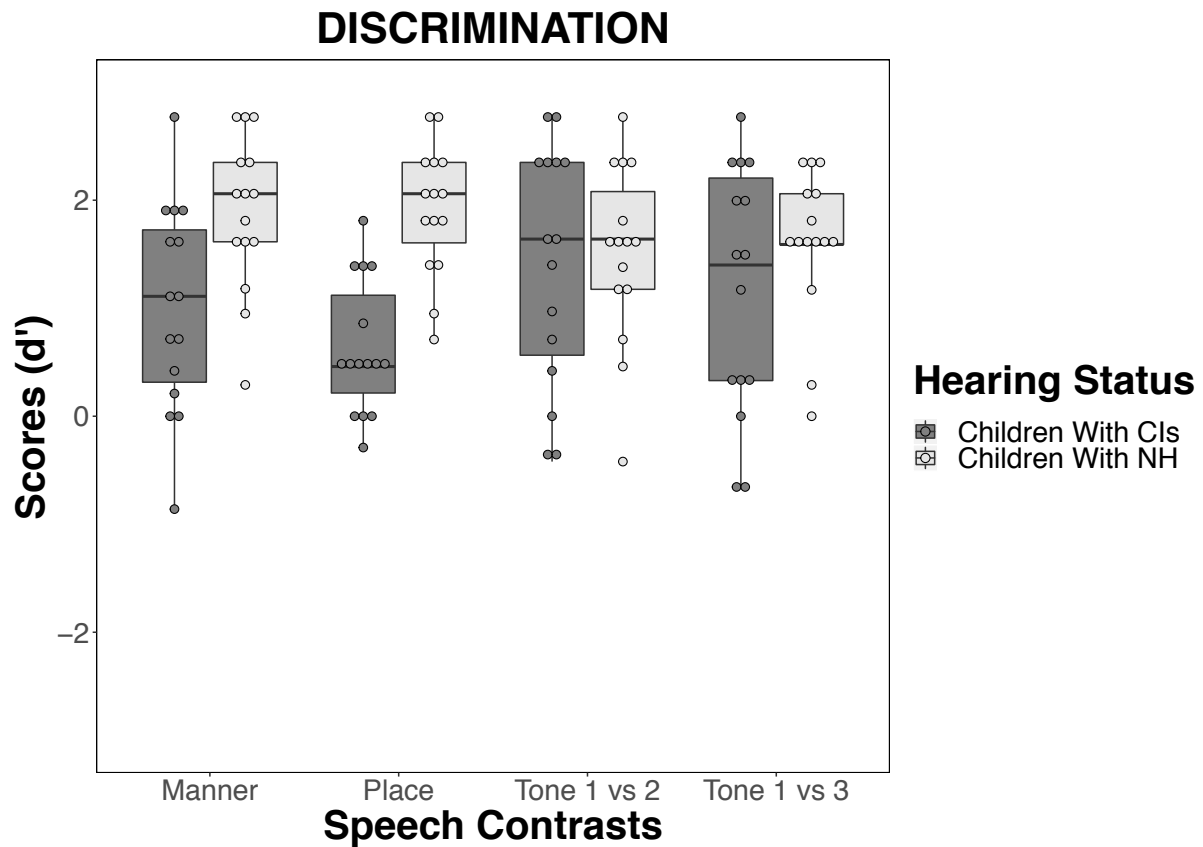


Figure 2. Boxplots for the discrimination scores (d') for the four speech contrasts (manner, place, tones 1 vs. 2, and tones 1 vs. 3) for CI users (dark grey bars) and controls with NH (light grey bars). Each dot represents an individual. The three horizontal lines represent the 25th, 50th, and 75th 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.

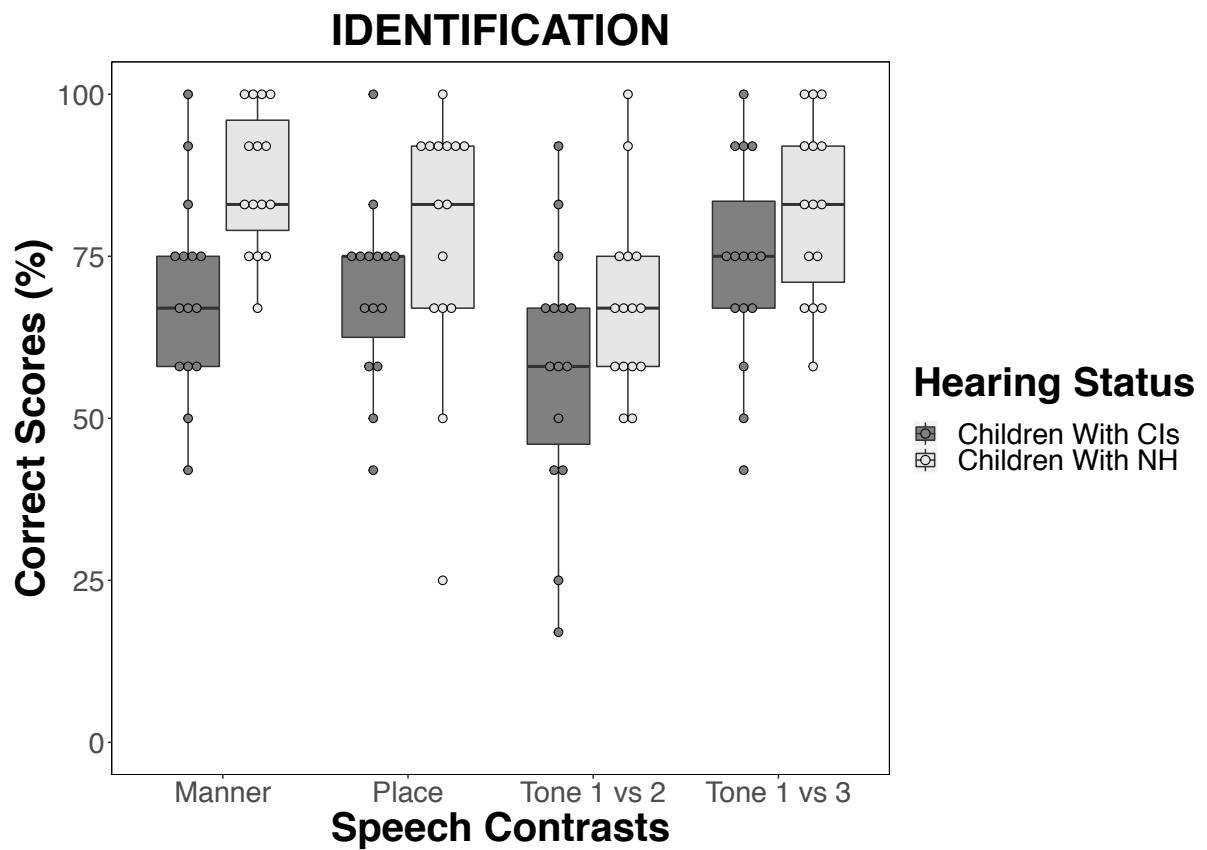


Figure 3. Boxplots for the identification scores (percentage correct) for the four speech contrasts (manner, place, tones 1 vs. 2, and tones 1 vs. 3) for CI users (dark grey bars) and controls with NH (light grey bars). Each dot represents an individual. The three horizontal lines represent the 25th, 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.