Running Title: Infant-directed speech through cochlear implant

1   Cochlear implant-related speech processing may diminish the advantage of exposure to

2   infant-directed speech [a]

3

4   Meisam K. Arjmandi [1,b], Derek Houston [2,c], Yuanyuan Wang [2], Laura C. Dilley [1]

5

6   [1] Department of Communicative Sciences and Disorders, Michigan State University, 1026 Red

7   Cedar Road, East Lansing, Michigan 48824, USA

8

9   [2] Department of Otolaryngology – Head and Neck Surgery, The Ohio State University, 915

10   Olentangy River Road, Columbus, Ohio, 43212, USA

11

12

13

14   [a] Portions of this work were presented at the 177th and 178th Meeting of the Acoustical Society

15   of America.

16   [b] Current affiliation: Department of Otolaryngology – Head and Neck Surgery, Massachusetts Eye

17   and Ear and Harvard Medical School, Boston, Massachusetts, USA. Email:

18   meisam_arjmandi@meei.harvard.edu

19   [c] Also at: Nationwide Children's Hospital, Columbus, OH, USA

Running Title: Infant-directed speech through cochlear implant

ABSTRACT

21 Caregivers modify their speech when talking to infants, a specific type of speech known as infant-

22 directed speech (IDS). This speaking style facilitates language learning compared to adult-directed

23 speech (ADS) in infants with normal hearing (NH). While infants with NH and those with cochlear

24 implants (CIs) prefer listening to IDS over ADS, it is yet unknown how CI speech processing may

25 affect the acoustic distinctiveness between ADS and IDS, as well as the degree of intelligibility of

26 these. This study analyzed speech of seven female adult talkers to model the effects of simulated

27 CI processing on (1) acoustic distinctiveness between ADS and IDS, (2) estimates of intelligibility

28 of caregivers' speech in ADS and IDS, and (3) individual differences in caregivers' ADS-to-IDS

29 modification and estimated speech intelligibility. Results suggest that CI processing is

30 substantially detrimental to the acoustic distinctiveness between ADS and IDS, as well as to the

31 intelligibility benefit derived from ADS-to-IDS modifications. Moreover, the observed variability

32 across individual talkers in acoustic implementation of ADS-to-IDS modification and speech

33 intelligibility was significantly reduced due to CI processing. The findings are discussed in the

34 context of the link between IDS and language learning in infants with CIs.

35 *Keywords*: Infant-directed speech, cochlear implant, acoustic distance, speech intelligibility,

36 individual differences

37

38

39

40

41

Running Title: Infant-directed speech through cochlear implant

42 **I. INTRODUCTION**

43 **A. Infant-directed speech and language learning**

44 Caregivers around the world modify their speaking style from adult-directed speech (ADS) to

45 infant-directed speech (IDS) when talking to infants (Fernald & Kuhl, 1987; Snow, 1977). Greater

46 exposure to IDS and some aspects of its acoustic and linguistic properties facilitates language

47 learning in children with normal hearing (NH) (Cristia & Seidl, 2014; Liu, Kuhl, & Tsao, 2003;

48 Song, Demuth, & Morgan, 2010; Trainor, Austin, & Desjardins, 2000; Weisleder & Fernald,

49 2013). The supportive role of IDS in infants' language learning occurs through several pathways,

50 such as directing and holding infants' attention to speech (Fernald, 1985, 1989; Kitamura,

51 Thanavishuth, Burnham, & Luksaneeyanawin, 2001; Kuhl et al., 1997; Schachner & Hannon,

52 2011; Wang, Bergeson, & Houston, 2017, 2018; Werker, Pegg, & McLeod, 1994), while making

53 speech more clear by providing more salient cues for speech discrimination (Karzon, 1985), word

54 segmentation (Singh & Nestor, 2009), speech segmentation (Thiessen, Hill, & Saffran, 2005), and

55 word learning (Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011). Taken together, these studies

56 demonstrate the supportive role of IDS in infants' early language learning. However, it is not yet

57 clear how cochlear implants may affect acoustic properties of IDS, and its supportive role in

58 language learning. The present study aims to investigate the effects of simulated CI processing on

59 acoustic information pertaining to distinguishing and recognizing words in IDS versus ADS.

60 **B. Infants' attention to IDS**

61 In listening to speech, infants with NH and those with CIs demonstrate increased attention to IDS

62 compared to ADS (The ManyBabies Consortium, 2020; Cooper, Abraham, Berman, & Staska,

63 1997; Fernald, 1985; Wang et al., 2017; Wang, Bergeson, et al., 2018; Werker & McLeod, 1989;

64 Werker et al., 1994). This effect of IDS on attention begins very early as observed in newborns

65 and infants as young as 4 months old (Cooper & Aslin, 1990; Fernald, 1985; Werker & McLeod,

66 1989), demonstrating the capability of infants in using subtle acoustic information to distinguish

67 IDS from ADS while having acquired only minimal linguistic knowledge. Multiple acoustic-

68 phonetic cues have been identified as potentially supporting infants' robust preference for IDS

69 over ADS, such as higher pitch (Cooper & Aslin, 1990; Fernald, 1989), greater pitch fluctuation

70 (Fernald & Simon, 1984), slower speaking rate (Leong, Kalashnikova, Burnham, & Goswami,

71 2014; Narayan & McDermott, 2016; Song et al., 2010), and an expanded vowel space (Burnham

72 et al., 2015) in IDS compared to ADS. These distinctive acoustic-phonetic cues are manifested in

73 spectro-temporal representations of caregivers' speech, which are actively incorporated through

74 infants' highly sensitive auditory systems to recognize IDS as distinct from ADS (Kuhl, 2004;

75 Maye, Werker, & Gerken, 2002; McMurray & Aslin, 2005; Telkemeyer et al., 2009). Although

76 IDS over ADS preference was observed in infants with CIs (Wang et al., 2017), it is still not clear

77 how they are different from their peers with NH in terms of access to spectral information involved

78 in distinguishing IDS from ADS. Infants with CIs may face difficulties in resolving distinctive

79 acoustic information toward performing this task effectively, leading potentially to a disruption in

80 the developmental time course of CI infants' attention to IDS and delayed language development,

81 as compared to that of NH infants. These considerations suggest a benefit of examining how CI-

82 related speech processing may affect acoustic distinctiveness and intelligibility of IDS compared

83 with ADS, and how this acoustic distinctiveness and intelligibility may differ as a function of

84 individual talkers' speech patterns.

85 **C. CI processing and acoustic distinction between IDS and ADS**

86 Infants with CIs have access to a spectro-temporally degraded representation of speech in their

87 language environments, due to limitations of electric stimulation to faithfully transmit this

88   information to the auditory nerve. The limited number of active channels and the broad current

89   spread, that causes the interaction between channels, lead to a degraded representation of spectral

90   and temporal information in speech (Baskent & Gaudrain, 2016; Svirsky, 2017). For instance,

91   Friesen et al. (2001) showed that speech recognition in listeners with NH improved by increasing

92   the number of spectral channels in vocoded speech, highlighting the effect of spectral resolution

93   in recognition of speech in listeners with CIs (Baskent & Gaudrain, 2016; Friesen et al., 2001; Fu,

94   Chinchilla, Nogaki, & Galvin, 2005; Fu & Nogaki, 2005; H. Luo & Poeppel, 2007; Shannon, Zeng,

95   Kamath, Wygonski, & Ekelid, 1995; Svirsky, 2017). Temporal resolution is also reduced in CIs

96   and it is mainly limited to temporal-envelope cues, which disrupts faithful transmission of

97   temporal fine-structure in speech to auditory nerves (Rubinstein, 2004; Svirsky, 2017). This

98   reduced spectro-temporal resolution may negatively affect infants' access to segmental and

99   suprasegmental (i.e., prosodic) features of IDS.

100   Transmission of multiple speech cues are negatively affected by the limited resolution of

101   CIs, including pitch (Mehta & Oxenham, 2017; Qin & Oxenham, 2005; Svirsky, 2017), timbre

102   (Kong, Mullangi, Marozeau, & Epstein, 2011), and melody (Mehta & Oxenham, 2017; Zeng,

103   Tang, & Lu, 2014) in talkers' speech. These cues have been recognized as major attributes that

104   distinguish IDS from ADS (Fernald, 1989; Fernald & Simon, 1984; Wang et al., 2017). For

105   example, caregivers' pitch is identified as an important perceptual cue for children's robust

106   attention in listening to IDS (Fernald, 1985; Piazza, Iordan, & Lew-Williams, 2017) and was also

107   recognized as an important cue in distinguishing IDS from ADS (Fernald & Kuhl, 1987). Higher

108   mean pitch and wider pitch range are recognized as prosodic modifications that signal speech

109   directed to infants as compared to adults (Garnica, 1977; Narayan & McDermott, 2016). Talkers'

110   pitch further contributes to lexical segmentation of speech in listeners with CIs (Spitzer, Liss,

5

111 Spahr, Dorman, & Lansford, 2009), a task critical for word learning. Recognition of emotion from

112 speech is another important aspect in recognizing and processing IDS (Trainor et al., 2000), which

113 relies heavily on perceiving pitch contour and pitch fluctuation in caregivers' speech. These cues

114 are poorly perceived by listeners with CIs. For example, listeners with CIs performed worse in

115 emotion recognition as the number of spectral channels in vocoded speech was decreased (Luo,

116 Fu, & Galvin, 2007). Another instance of degraded representation of speech cues related to IDS is

117 poor perception of melodic pitch that is shown to be negatively affected by reducing the number

118 of spectral channels (Kong, Cruz, Jones, & Zeng, 2004). These findings suggest that children with

119 CIs have probably partial access to spectro-temporal cues in speech that greatly contribute to the

120 distinctiveness of IDS and ADS (e.g., pitch and intonation). This limited access to speech cues that

121 signal IDS may interfere with the connection between recognizing and listening to IDS and

122 developing better language skills. Therefore, it is important to understand how CI may impact the

123 acoustic distinction between IDS and ADS - a major knowledge gap that the present study aimed

124 to address based on acoustic and computational analysis of simulated CI speech.

125 **D. Intelligibility of IDS through CI**

126 Another possible way IDS may foster language learning is by making speech more intelligible.

127 Prior studies showed evidence that IDS may provide acoustic cues that assist infants for parsing

128 linguistic units in speech. Karzon (1985) showed that the exaggerated suprasegmental features of

129 IDS may provide supportive perceptual cues for syllabification of multisyllabic sequences. In

130 another study, infants had better long-term word recognition when the words were presented in

131 IDS compared to ADS (Singh & Nestor, 2009). It was further shown that IDS facilitates word

132 segmentation (Thiessen et al., 2005). Although evidence on intelligibility benefits of IDS over

133 ADS have somehow been contradictory at the level of segmental speech cues (e.g., Kuhl et al.,

134   1997; McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013), there might be an intelligibility

135   benefit associated with other properties of IDS (e.g., prosodic cues), similar to the effects observed

136   in close analogues of a "clear speech" register (Bradlow, Kraus, & Hayes, 2003; S. H. Ferguson

137   & Kewley-Port, 2007). Caregivers' modification of speaking style from ADS to IDS impacts

138   aspects of speech such as pitch contour and speech rate, which have been shown to be strong

139   predictors of speech clarity and intelligibility (Bradlow et al., 2003; Cutler, Dahan, & van

140   Donselaar, 1997; Ferguson & Poore, 2010; Ferguson & Quené, 2014; Spitzer, Liss, & Mattys,

141   2007; Watson & Schlauch, 2008). Different choices of speaking style often impact speech rate,

142   where a slower rate is thought to contribute to enhanced intelligibility of clear speech compared to

143   conversational speech both in typical listeners (Ferguson et al., 2010) and recipients of CIs (Li et

144   al., 2011; Zanto, Hennigan, Östberg, Clapp, & Gazzaley, 2013).

145   Limited spectral resolution of the cochlear implant device affects the quality by which

146   listeners with CIs receive speech cues (Croghan, Duran, & Smith, 2017; Jain & Vipin Ghosh,

147   2018; Peng, Hess, Saffran, Edwards, & Litovsky, 2019; Qin & Oxenham, 2005), which may

148   negatively impact CI users' ability to recognize words and phonemes, especially children (Grieco-

149   calub, Saffran, & Litovsky, 2010; Peng et al., 2019). Therefore, any likely beneficial role of IDS

150   in improvement of speech intelligibility is prone to change when caregivers' speech is perceived

151   through a CI device. Prior studies have shown that performance of listeners with CIs is

152   considerably poorer than listeners with NH in understanding sentences spoken with relatively

153   higher rate (Zanto et al., 2013), which is an acoustic dimension signaling distinction of IDS from

154   ADS (i.e., slower speaking rate in IDS compared to ADS). Variation of fundamental frequency

155   (F_0) is another acoustic dimension of the IDS-vs-ADS distinction that contributes to speech

156   recognition (Spitzer, Liss, Spahr, Dorman, & Lansford, 2009; Spitzer et al., 2007). Findings on the

7

157   effect of these acoustic properties on intelligibility of caregivers' speech is still controversial, and

158   little is known about the effects of CI speech processing on intelligibility of caregivers' speech in

159   ADS and IDS conditions. The present study uses computational modeling approaches to address

160   whether IDS is beneficial for intelligibility improvement and how this effect may be compromised

161   depending on listeners' hearing status (NH vs. CI).

162   **E.   Individual differences in acoustic implementation of IDS and speech intelligibility**

163   Listeners' familiarity with the range of variability in voice of individual talkers is important for

164   robust identification and recognition of individuals' voices (Lavan, Burton, Scott, & McGettigan,

165   2019; Souza, Gehani, Wright, & McCloy, 2013). Caregiver-specific acoustic modification is

166   important for effective caregiver-infant interaction and significantly contributes to infants'

167   recognition of caregivers' voices both before and after birth (Beauchemin et al., 2011; Kisilevsky

168   et al., 2003), as well as for neural development of infants' auditory cortices (Webb, Heller, Benson,

169   & Lahav, 2015). Singh (2008) showed that exposure to IDS with relatively higher variability in

170   vocal affects facilitates word recognition in infants by forming a relatively wider and more

171   generalizable lexical category, highlighting the supportive role of exposure to a wide range of

172   ADS-to-IDS acoustic variation in effective processing of caregivers' speech. The degree of

173   acoustic variability that caregivers introduce to their infants may vary across individuals. The

174   acoustic characteristics of IDS reveal finer details about the unique voice timbre of individual

175   caregivers, which potentially contributes to infant' identification of individual caregivers (Piazza

176   et al., 2017). It is, however, unclear how CI speech processing impacts the range of acoustic

177   variability within and across caregivers due to shifts in caregivers' speaking style (i.e., ADS to

178   IDS), a change which likely eliminates acoustic details and potentially negatively impacts the

179   supportive role of IDS in recognizing caregivers' voices. CI speech processing is also detrimental

180   to the acoustic distance of voices across caregivers, which potentially makes distinguishing talkers

181   from each other more challenging for infants with CIs, compared to those with NH. The present

182   study aimed to investigate these questions by focusing on simulating the limited resolution of CIs

183   using noise-excited envelope vocoder and modeling the intelligibility benefit of ADS-to-IDS

184   modification.

F. Benefits of investigations of acoustic properties of CI-related speech processing

186   Studies aimed at understanding speech perception in individuals with CIs can be characterized as

187   involving one of three general approaches. The first approach examines performance of listeners

188   with NH in response to vocoded speech to simulate hearing in listeners with CIs (Dorman, Loizou,

189   & Rainey, 1997a; Jahn, DiNino, & Arenberg, 2019; Mehta & Oxenham, 2017; Qin & Oxenham,

190   2003, 2005). The second method involves directly study of how listeners with CIs perform in

191   various speech recognition tasks (Brown & Bacon, 2010; Kong, Stickney, & Zeng, 2005; Peng,

192   Tomblin, & Turner, 2008; Peng, Hess, Saffran, Edwards, & Litovsky, 2019). In a third method –

193   the one which we focus on in the present paper – acoustic properties of simulated CI speech and

194   unprocessed analogs are analyzed comparatively. This comparative analysis can be facilitated

195   through the use of various quantitative metrics that emulate speech recognition in CI users as

196   proxies to analyze properties of CI speech and its perception in listeners with CIs (Jain & Vipin

197   Ghosh, 2018; Qin & Oxenham, 2003; Santos, Cosentino, Hazrati, Loizou, & Falk, 2013). Analysis

198   of simulated CI speech and quantitative measures of speech quality and intelligibility tailored to

199   listeners with CIs provide repeatable, automated, inexpensive, and fast tools for gaining

200   preliminary evidence about speech perception in individuals with CIs. As such, a comparative

201   analysis method offers benefits for undertaking efficient investigations for further study, and as

9

202   such provides benefits over various major challenges that accompany the first two categories of

203   studies, such as participant recruitment, time, and cost.

204   **G. Current study**

205   In the present study, we analyzed unprocessed and simulated CI speech of seven female talkers

206   who spoke fifteen utterances both in ADS and IDS. These pairs of ADS and IDS speech were

207   analyzed to examine how the limited frequency resolution of CIs may affect acoustic

208   distinctiveness between ADS and IDS, as well as its effects on the estimated intelligibility of

209   caregivers' speech under modifications of speaking style (i.e., ADS to IDS). ADS and IDS stimuli

210   of the seven female talkers were processed using noise-excited envelope vocoder with different

211   number of channels to simulate the restricted spectral resolution and the amount of distortion in

212   CIs (Dorman, Loizou, Fitzke, & Tu, 1998; Dorman, Loizou, & Rainey, 1997b; Friesen et al., 2001;

213   Fu, Shannon, & Wang, 1998; Loizou, Dorman, & Tu, 1999; Mehta, Lu, & Oxenham, 2020; Mehta

214   & Oxenham, 2017; Qin & Oxenham, 2003, 2005; Rosen, Faulkner, & Wilkinson, 1999; Shannon

215   et al., 1995).

216   To examine the effects of CIs on acoustic distinctiveness between IDS and ADS, we first

217   calculated mel-frequency cepstral coefficients (MFCCs), which have been shown in multiple

218   studies to be effective for characterizing the distinctive qualities of IDS vs. ADS (Inoue,

219   Nakagawa, Kondou, Koga, & Shinohara, 2011; Piazza et al., 2017; Sulpizio et al., 2018). Modeling

220   the frequency response of human auditory system via MFCCs captures acoustic features that are

221   involved in distinguishing IDS and ADS, such as shifts in vocal timbre (Piazza et al., 2017) and

222   acoustic distinctiveness between IDS and ADS (Inoue et al., 2011), including in other languages

223   (e.g., Italian and German) (Sulpizio et al., 2018). MFCCs are also able to effectively reflect

224   caregivers' emotional state and vocal affect (e.g., happiness vs. sadness, Sato & Obuchi, 2007;

10

225   Slaney & McRoberts, 1998), attributes that help explain infants' preference for IDS over ADS

226   (Fernald, 2018; Horowitz, 1983; Mastropieri & Turkewitz, 1999; Moore, Spence, & Katz, 1997;

227   Papoušek, Bornstein, Nuzzo, Papoušek, & Symmes, 1990; Singh, Morgan, & Best, 2002; Walker-

228   Andrews & Grolnick, 1983; Walker-Andrews & Lennon, 1991). Furthermore, MFCCs allowed us

229   to analyze spectral properties of vocoded speech, which was not feasible to do based on calculation

230   of common acoustic properties such as fundamental frequency (F_0), as these cues are partially

231   present in CI-simulated speech (Fuller et al., 2014; Gaudrain & Baskent, 2018). Second, to

232   quantify the acoustic distinctiveness between IDS and ADS, we calculated a *Mahalanobis distance*

233   *(MD)* measure over MFCCs features. MD is a multivariate distance metric that has been widely

234   used to measure the distances between vectors in a variety of multidimensional feature spaces

235   (Arjmandi, Dilley, & Wagner, 2018; Masnan et al., 2015; Xiang, Nie, & Zhang, 2008). Third, to

236   model how CIs may influence intelligibility of caregivers' speech, we calculated the speech-to-

237   reverberation-modulation energy ratio (SRMR) to model the signal-based intelligibility of speech

238   signals delivered to listeners with NH, as well as its CI-tailored version (SRMR-CI) to model

239   intelligibility of speech signals delivered to listeners with CIs (Santos et al., 2013). These measures

240   were examined both within and across speakers to model the impacts of speaking styles (IDS vs.

241   ADS) and listener group (NH vs. CI) on signals' distinctiveness and estimated intelligibility. All

242   these analyses were implemented in Matlab 2019a (The Mathworks Web-Site

243   [http://www.mathworks.com]).

244   The following four specific questions were addressed in this study. First, we asked how

245   simulated CI speech processing affects the acoustic distinctiveness of IDS compared with ADS,

246   particularly as a function of degree of spectral degradation (i.e., number of spectral channels) in

247   CI-simulated speech. We hypothesized that (a) CI-related speech processing significantly degrades

248   the acoustic distinctiveness of IDS compared with ADS, and further that (b) increasing the number

249   of channels would not compensate for degradation imposed by CI-related processing as compared

250   to the unprocessed condition. Second, we asked whether there is signal-based evidence that IDS

251   may be more intelligible than ADS, as gauged by the SRMR metric (to simulate a NH listening

252   condition) or the SRMR-CI metric (to simulate a CI listening condition). We hypothesized that

253   IDS is more intelligible than ADS, but that estimated intelligibility would vary as a function of

254   simulated listening status (NH vs. CI). Third, we asked to what extent the acoustic distinctiveness

255   between IDS and ADS varies across individual caregivers, as well as how such individual

256   differences might be impacted by CI speech processing, focusing on change in the spectral

257   resolution. We predicted individual differences across caregivers in acoustic implementation of

258   the differences between ADS and IDS; we further predicted that CI-related speech processing

259   would decrease the extent of acoustic distinctiveness as a function of speech style, leading to loss

260   of intra- and inter-subject acoustic variability in terms of IDS vs. ADS distinctiveness. Fourth, and

261   finally, we asked how estimated intelligibility *differences* for IDS compared with ADS vary across

262   individual caregivers, and how such intelligibility differ as a function of hearing status (as gauged

263   by SRMR vs. SRMR-CI differences to simulate effects of NH vs. CI listening conditions,

264   respectively). We hypothesized that intelligibility differences for IDS vs. ADS would vary across

265   caregivers, and that these intelligibility differences would be more similar as estimated for CI

266   listeners, compared with NH listeners.

II. METHODS

A. Speech stimuli

269   Seven female adult native talkers of American English ranging in age from 21 to 24 years old

270   spoke fifteen utterances in IDS and ADS (See Appendix in Supplementary Material for the list of

271   stimuli). The utterances were elicited to be also used in separate infant word-learning experiments

272   in which infants were exposed to a novel target word (i.e., "*modi*") in the context of behavioral

273   measures of infant word recognition to assess whether infants learned the novel word from IDS

274   better than ADS. To elicit stimuli, the talkers were instructed to speak utterances as if talking to

275   an infant (IDS condition) or an adult (ADS condition); this procedure was similar to that used in

276   Wang et al., (2017; 2018).

277   Speech stimuli were recorded using an AKG D 542 ST-S microphone in a sound booth and

278   digitized at a sampling rate of 44.1 kHz with 16-bit resolution. The distance between talkers'

279   mouths and the microphone was controlled to assure the quality of the recorded stimuli. Prior to

280   processing as discussed below, the start and end points of recorded utterances were manually

281   identified in Praat software (Boersma & Weenink, 2001) to remove preceding and following silent

282   portions. All participants were fully informed about the purpose and procedure of this study, and

283   they had given informed consent to participate. This study was approved by the Institutional

284   Review Boards of the Ohio State University and Michigan State University.

285   **B. Creation of simulated CI speech**

286   CI-simulated versions of the unprocessed stimuli were created using noise-excited envelope

287   vocoder processing at six levels of spectral degradation, corresponding to 4-, 8-, 12-, 16-, 22- and

288   32-channel noise-vocoded stimuli. The choice of number of channels was made to cover both the

289   actual number of channels in FDA-approved devices (12, 16, 22) and to query ranges of variation

290   ranging from minimal cues used in prior studies (4, 8) out to 32 channels to examine CI vocoder

291   scenarios with higher spectral resolution. The natural stimuli consisting of spoken IDS or ADS

292   were processed in AngelSim[TM] Cochlear Implant and Hearing Loss Simulator (Fu, 2019; Emily

293   Shannon Fu Foundation, www.tigerspeech.com) using 4-, 8-, 12-, 16-, 22-, and 32-channel noise-

294   vocoding CI-simulated stimuli; the noise vocoding method followed the procedure in Shannon et

295   al. (1995). The original stimuli were first band-passed filtered using the Greenwood function

296   (emulating the Greenwood frequency-place map) into N ($N = \{4, 8, 12, 16, 22, 32\}$) adjacent

297   frequency channels ranging from 200 Hz to 8000 Hz. This was implemented in $AngelSim^{TM}$ by

298   setting the absolute lower- and higher-frequency threshold for analysis and carrier filters to 200

299   Hz and 8000 Hz with a filter slope of 24 dB/Oct (Fu, 2019). These frequency ranges are fairly

300   close to the corner frequencies of the Cochlear Nucleus speech processors in CI listeners cochlear

301   implant devices (Crew & Galvin, 2012; Winn & Litovsky, 2015), which emulate the performance

302   of average CI listeners in speech envelope discrimination (Chatterjee & Oberzut, 2011; Chatterjee

303   & Peng, 2008). The same analysis filter and carrier filter of the Greenwood function was used to

304   analyze white noise as a carrier signal (Greenwood, 1990). The $AngelSim$ software used this setup

305   to extract a time-varying amplitude envelope of speech stimuli under each frequency band using

306   half-wave rectification and then modulated independent white-noise carriers. There were 210

307   stimuli per speech style condition (7 talkers x 15 stimuli x 2 speaking styles). Including the noise-

308   vocoded versions of these utterances (at 4, 8, 12, 16, 22, and 32-channels), we analyzed 1470 (210

309   x 7 levels of speech degradation) utterances in the present study.

310   **C. Using MFCC features to characterize acoustic properties of IDS and ADS**

311   We calculated 12 MFCCs for each speech stimulus to characterize its acoustic information (Hunt,

312   Lennig, & Mermeletein, 1980; Imai, 1983; Shaneh & Taheri, 2009). Figure 1 illustrates this step

313   for a sample pair of IDS-ADS stimuli in the process for measuring the acoustic distinctiveness

314   between each ADS-IDS stimulus pair. To calculate MFCCs, each stimulus was re-sampled at 16

315   kHz using a Hamming window of 25 ms applied to each frame, with a frame shift of 10 ms,

316   following the procedure used in Inoue et al. (2011). For each pair of IDS and ADS stimuli ($S_{IDS_{ij}}$

317 and $S_{ADS_{ij}}$ in Figure 1), MFCCs were calculated for all frames of these stimuli ($MFCCs_{IDS_{ij}}$ and

318 $MFCCs_{ADS_{ij}}$). Here, i indicates the index of the speech stimulus, $i = \{1,2,3,\ldots,15\}$, and j is the

319 index for the talker, $j = \{1,2,3,\ldots,7\}$. Unlike previous studies in which a single, time-averaged

320 MFCC vector was calculated to represent acoustic information in IDS and ADS (e.g., Piazza et al.,

321 2017), our approach took advantage of all MFCCs derived from all frames of a speech stimuli to

322 calculate the acoustic distinctiveness between ADS-IDS pairs of stimuli within a multidimensional

323 feature space, thereby preserving the details about spectro-temporal information in speech at the

324 level of frame. Since each analyzed pair of IDS and ADS stimuli contained identical word strings,

325 our analysis was expected to mainly model the acoustic effects of caregivers' speaking styles (IDS

326 vs. ADS).

327

328
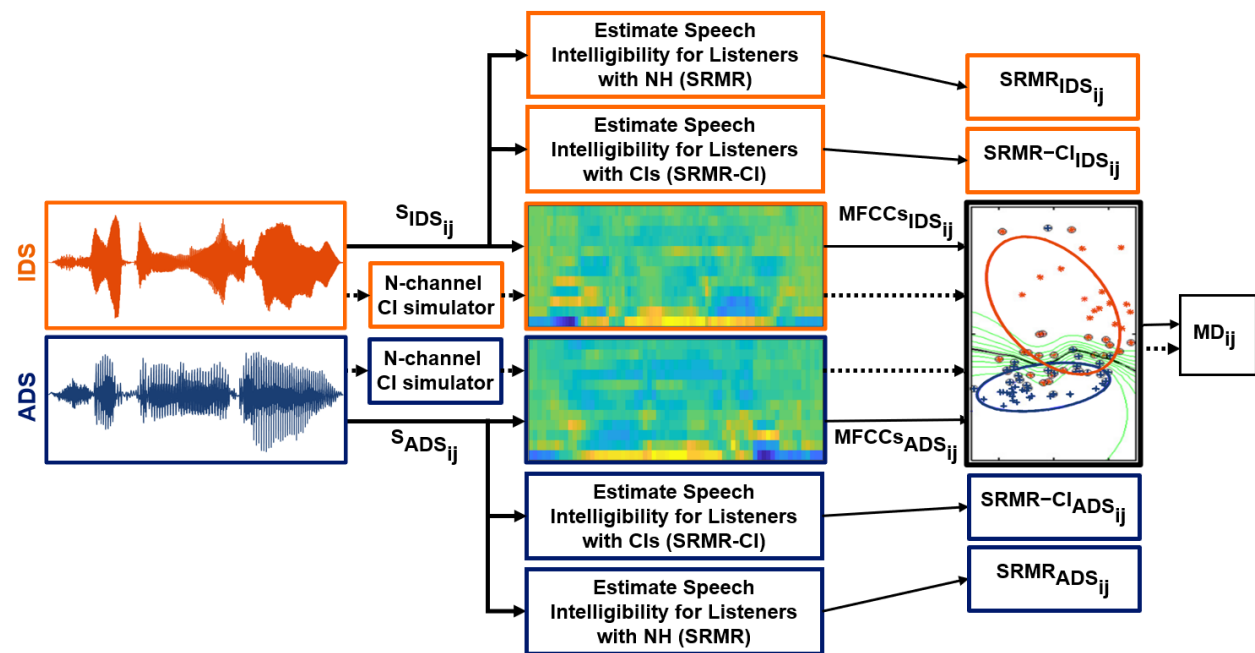
329 **FIG. 1.** Schematic diagram of the approach used in the present study for measuring (1) MD

330 between pairs of MFCCs derived from pairs of IDS-ADS stimuli, and (2) intelligibility of IDS and

331   ADS stimuli as estimated by SRMR value for listeners with NH and SRMR-CI for those with CIs.

332   Note that the dashed line denotes the process for creating and analyzing the noise-vocoded versions

333   of the same pairs of stimuli, while N stands for the number of spectral channels in the noise-excited

334   envelope vocoder. Blocks and lines with blue (dark gray) color indicate paths for processing ADS,

335   while those with orange (light gray) color indicate paths for processing IDS. Note that SRMR and

336   SRMR-CI were calculated only for unprocessed stimuli in order to estimate intelligibility for NH

337   and CI listeners, respectively. In this figure, the waveforms and their corresponding MFCCs are

338   from the utterance "See the modi?" spoken by one of the seven talkers both in IDS and ADS

339   speaking styles. $S_{IDS_{ij}}$ and $S_{ADS_{ij}}$ are the ith pair of IDS and ADS stimuli (i = {1,2,3,…,15}) for

340   talker j (j ={1,2,3,…,7}). MFCCs$_{IDS_{ij}}$ and MFCCs$_{ADS_{ij}}$ are MFCC features derived from $S_{IDS_{ij}}$ and

341    $S_{ADS_{ij}}$ speech stimuli, respectively. The middle two panels show MFCCs obtained from the frames

342   of these IDS and ADS stimuli. MD_{ij} is the MD calculated to measure the acoustic distance between

343   the two matrices for MFCCs$_{IDS_{ij}}$ and MFCCs$_{ADS_{ij}}$. SRMR$_{IDS_{ij}}$ and SRMR$_{ADS_{ij}}$ are the estimated

344   intelligibility for $S_{IDS_{ij}}$ and $S_{ADS_{ij}}$ speech stimuli, respectively, as heard by listeners with NH.

345   SRMR-CI$_{IDS_{ij}}$ and SRMR-CI$_{ADS_{ij}}$ are the estimated speech intelligibility for the same $S_{IDS_{ij}}$ and

346    $S_{ADS_{ij}}$ speech stimuli, respectively, as heard by a listener with CIs.

347   **D. Acoustic distinctiveness quantification using Mahalanobis Distance (MD) measure**

348   After representing acoustic features of pairs of IDS and ADS stimuli by calculating their MFCCs,

349   the acoustic distinctiveness between each pair was measured by calculating MD on the

350   corresponding MFCC matrices (i.e., $MFCCs_{IDS_{ij}}$ and $MFCCs_{ADS_{ij}}$) using a 12-dimensional feature

351   space (i.e., 12 MFCCs) (Maesschalck & Massart, 2000; Masnan et al., 2015; Xiang et al., 2008).

352   MD is a multivariate statistical approach that evaluates distances between two multidimensional

353   feature vectors or matrices that belong to two classes (here IDS vs. ADS) (Arjmandi et al., 2018;

354   Heijden, Ferdinand, Ridder, & Tax, 2005; Maesschalck & Massart, 2000; Masnan et al., 2015).

355   The MD calculation returns the *distance* between means of two classes (here IDS and ADS)

356   relative to the average per-class covariance matrix (Maesschalck & Massart, 2000). Here, acoustic

357   properties of each class were represented by two feature matrices (here, MFCCs). The larger the

358   distance between pairs of $MFCCs_{IDS_{ij}}$ and $MFCCs_{ADS_{ij}}$ feature vectors, the lower the overlap

359   between two classes of IDS and ADS; this corresponds in turn to greater acoustic distance (or

360   distinctiveness) between IDS and ADS stimuli.

361   Each row of the calculated MFCCs matrix for utterance *i* from talker *j* (e.g., $MFCCs_{IDS_{ij}}$ or

362    $MFCCs_{ADS_{ij}}$) corresponds to a frame of that utterance, and 12 MFCCs were presented on the

363   columns of this matrix. Therefore, the dimension of each MFCC matrix was N_F x 12, where N_F

364   refers to the number of frames in that speech stimulus. The acoustic distinctiveness between IDS

365   and ADS stimuli for each female talker *j* was computed by averaging the fifteen MD values

366   obtained from that talker's fifteen IDS-ADS pairs ($MD_j = \frac{1}{N}\sum_{i=1}^{N} MD_{ij}$, $N=15$). This process was

367   separately performed on pairs of IDS and ADS stimuli for seven levels of spectral degradation –

368   *unprocessed*, *32*-, *22*-, *16*-, *12*-, *8*-, and/or *4-channel* noise-vocoded CI-simulated stimuli – to give

369   average MDs used to examine (1) how the acoustic distinctiveness between IDS and ADS stimuli

370   changes as a function of CI processing (and number of spectral channels), (2) how individual

371   caregivers vary in acoustic distinctiveness between IDS and ADS stimuli, and (3) how the limited

372   spectral resolution in CIs may affect this pattern of individual differences across talker.

17

E. Estimation of speech intelligibility using quantitative metrics of speech-to-reverberation-modulation energy ratio (SRMR)

As shown in Figure 1, we estimated the degree of intelligibility of each speech stimulus as it might be experienced by listeners with NH or with CIs. To approximate stimulus intelligibility as associated with NH, we calculated the quantitative metric of *speech-to-reverberation-modulation energy ratio* (SRMR) for IDS and ADS (unprocessed) stimuli, respectively (Santos et al., 2013). This metric has previously been validated as an approximation of intelligibility of speech to listeners with NH in behavioral tasks (Falk et al., 2015; Santos et al., 2013). Further, to approximate intelligibility as might be experienced by listeners with CIs, we calculated an adapted version of the SRMR metric, which was also developed by Falk et al. (2013) to tailor the SRMR metric for CI users; this adapted metric is known as SRMR-CI (Santos et al., 2013).

To calculate SRMR, an (unprocessed) speech stimulus was first filtered by a 23-channel gammatone filterbank in order to emulate cochlear function. Next, a Hilbert transform was applied to output signals from each of the 23 filters in the filterbank to obtain their temporal envelopes. Next, modulation spectral energy for each critical band was calculated by windowing stimulus temporal envelope and computing the discrete Fourier transform. To emulate frequency selectivity in the modulation domain, the modulation frequency bins were grouped into eight overlapping modulation bands with logarithmically-spaced center frequencies between 4 and 128 Hz. Finally, the SRMR value was obtained by calculating the ratio of the average modulation energy in the first four modulation bands (~ 3–20 Hz) to the average modulation energy in the last four modulation bands (~ 20-160 Hz), as explained in Santos et al., (2013).

To calculate SRMR-CI, a 22-channel filterbank was used instead of a 23-channel filterbank in order to emulate the structure of filterbanks in Nucleus CI devices. In addition, the 4-128 Hz

396   range of modulation for filterbank center frequencies was replaced by a 4–64 Hz range to better

397   emulate CI users' performance (Santos et al., 2013). Although these measures have not primarily

398   been developed to estimate the effect of talkers' speaking style on speech intelligibility, the signal

399   processing algorithms used in these metrics estimate the distribution of energy in various

400   frequency bands, as well as other spectral properties that are expected to emulate fairly well

401   performance of NH and CI users in speech recognition. Intelligibility of signals as estimated for

402   different hearing statuses (NH or CI) is inherently modeled through the difference in algorithmic

403   implementation of these metrics (SRMR vs. SRMR-CI, respectively) (Santos et al., 2013).

404   As shown in Figure 1, SRMR and SRMR-CI values were separately calculated ($SRMR_{IDS_{ij}}$,

405   $SRMR_{ADS_{ij}}$, $SRMR\text{-}CI_{IDS_{ij}}$ and $SRMR\text{-}CI_{ADS_{ij}}$) for each pair of stimuli in IDS and ADS conditions

406   ($S_{IDS_{ij}}$ and $S_{ADS_{ij}}$). Intelligibility of speech stimuli produced by talker j in the two conditions (ADS

407   and IDS) was summarized by averaging the relevant SRMR-related values over the 15 stimuli

408   spoken by this talker at each of these two conditions. Therefore, for talker j, four values were

409   calculated: (i) $SRMR_{IDS_j} = \frac{1}{15}\sum_{i=1}^{15} SRMR_{IDS_{ij}}$, (ii) $SRMR_{ADS_j} = \frac{1}{15}\sum_{i=1}^{15} SRMR_{ADS_{ij}}$, (iii)

410   $SRMR\text{-}CI_{IDS_j} = \frac{1}{15}\sum_{i=1}^{15} SRMR\text{-}CI_{IDS_{ij}}$, (iv) $SRMR\text{-}CI_{ADS_j} = \frac{1}{15}\sum_{i=1}^{15} SRMR\text{-}CI_{ADS_{ij}}$. These four

411   values for each talker j provided an approximation of the intelligibility of IDS and ADS stimuli

412   spoken by this talker, as heard by listeners with NH ($SRMR_{IDS_j}$ and $SRMR_{ADS_j}$) and by those with

413   CIs ($SRMR\text{-}CI_{IDS_j}$ and $SRMR\text{-}CI_{ADS_j}$). These values were then used to address (1) how the

414   intelligibility of speech is affected by the two separate factors of caregivers' speaking styles (ADS

415   vs. IDS), as well as by (simulated) group hearing status (NH vs. CI), (2) the extent to which the

416   degree of impact of ADS-to-IDS speaking style modifications on intelligibility varies across

19

417   individual caregivers, and (3) how this variation is affected by (simulated) hearing status (NH vs.

418   CI).

F. Statistical analysis

420   We constructed two Generalized Linear Mixed Models (GLMMs) in Matlab (using the *fitglme*

421   function) (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017; Quené & van den Bergh, 2008)

422   in order to (1) identify whether changes in acoustic distinctiveness (i.e., MD) between IDS and

423   ADS due to CI noise vocoding were statistically significant, and (2) to examine the effect of

424   speaking style (IDS vs. ADS) and simulated hearing status (NH vs. CI), and any potential

425   interaction, on intelligibility of speech (as measured by SRMR and SRMR-CI to simulate NH and

426   CI hearing statuses, respectively). For the first GLMM, the speech degradation level (i.e., 4, 8, 12,

427   16, 22, and 32-channel, and unprocessed) was entered into the model as a fixed predictor, and MDs

428   were defined as the response variable. The individual female talkers and speech stimuli were

429   defined as random effects (intercepts) in the model to account for quality differences that might

430   exist due to talker- and stimulus-specific variation. A post hoc test using Tukey's multiple

431   comparison approach (*multcompare*, MATLAB, Mathworks) was conducted to examine

432   statistically significant mean differences for all possible pairwise comparisons across seven levels

433   of spectral degradation (4, 8, 12, 16, 22, 32-channels, and unprocessed). In the second GLMM,

434   factors of speaking style (IDS or ADS) and simulated hearing status (NH or CI) were entered into

435   the model as independent fixed variables; the estimated speech intelligibility (cf. SRMR or SRMR-

436   CI values) corresponded to the response variable in the model. Talkers and speech stimuli were

437   entered as random-effect intercepts in the model to account for quality and/or information

438   differences that might exist due to talker- and stimulus-specific variation.

439   Measures of central tendency and variability (i.e., mean, standard deviation, and range)

440 were calculated as descriptive statistics to evaluate the degree of variability across caregivers in

441 acoustic implementation of IDS and ADS. We further calculated coefficients of variance (CVs)

442 for MD differences between ADS and IDS conditions for both unprocessed and simulated CI

443 speech within a 22-channel noise-vocoder to examine whether variability across individual

444 caregivers in ADS-IDS acoustic distinctiveness decreased as a function of speech degradation

445 (unprocessed vs. 22-channel simulated CI speech). The same analysis was performed to examine

446 whether changes in speech intelligibility – due to an ADS-to-IDS style shift – were variable across

447 talkers and/or how variability changed as (simulated) listener hearing status changed (from NH to

448 CI). We further calculated Pearson correlation coefficients to examine whether changes in acoustic

449 distinctiveness and speech intelligibility due to changes in speaking style (ADS to IDS) across

450 degradation levels (unprocessed vs. 22-channel vocoded) and hearing status (NH vs. CI) involved

451 a linear transformation.

452 **III. RESULTS**

453 **A. Effects of CI-related spectral degradation on ADS-vs-IDS acoustic distinctiveness**

454 Fig. 2 shows the mean and distribution of MDs obtained as a function of signal degradation that

455 ranged from a no-degradation (i.e., unprocessed) condition to CI-simulated speech for which the

456 number of spectral channels in the noise-vocoder was gradually reduced from 32 to 4. Overall,

457 MDs between IDS and ADS monotonically decreased with decreasing numbers of spectral

458 channels in the noise-vocoder, thereby highlighting the detrimental effects of limited spectral

459 resolution in CI devices for signal fidelity. As illustrated in Fig. 2, the acoustic distinctiveness

460 between IDS and ADS is reduced by CI speech processing systems and further reduced through

461 decreasing the number of spectral channels. Even the highest-fidelity CI-simulated condition (32-

462   channel noise-vocoded speech) showed a substantial drop in acoustic distinctiveness between IDS

463   and ADS, relative to the unprocessed condition. The average MD across the six noise-vocoded

464   conditions showed a decline of approximately 67% relative to MD at the unprocessed condition,

465   indicating that a large portion of acoustic information involved in conveying the distinction

466   between IDS and ADS was lost due to CI-related speech processing.

467   Table I represents the results of the GLMM analysis to test statistical significance of the

468   effect of CI-related speech processing on MD. As this table suggests, the detrimental effect of CI-

469   related processing, as captured in the decrease in the number of spectral channels on the

470   distinctiveness between IDS and ADS was statistically significant (β = -0.886, t = -16.54, $p <$

471   0.0001). Post hoc testing using a Tukey's multiple comparison approach revealed that the

472   simulated effect of limited spectral resolution in CI processing on the decrease in MDs was

473   statistically significant for comparisons of all pairs of speech degradation conditions except three

474   (32 vs. 22 channels, 22 vs. 16 channels, and 16 vs. 12 channels).

475
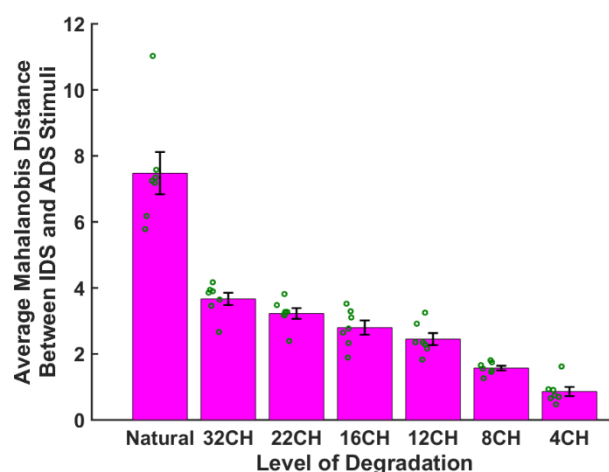
476   **FIG. 2.** The mean (bar) and ±1 standard error (vertical error bar in black) of Mahalanobis distance

477   (MD) across groups of IDS and ADS stimuli at seven levels of spectro-temporal degradation,

478   ranging from no-degradation (natural/unprocessed) to 4-channel noise-vocoded stimuli. Green

479   (dark gray) circles show the mean MDs for each talker derived by averaging MDs over 15 pairs of

480   IDS-ADS stimuli for that talker.

481   **TABLE I.** Statistical GLMM for modeling the effect of level of speech degradation on acoustic

482   distinctiveness between IDS and ADS as quantified by MD.

	β Estimate	St. Error	t	Pr(>|t|)
(intercept)	6.694	0.58	11.45	0.000[a]
Level of degradation	-0.886	0.053	-16.54	0.000[a]

483   [a] The p-value was statistically significant.

484   **B. Effects of CI-related speech processing on intelligibility of ADS and IDS**

485   Fig. 3 shows average SRMR or SRMR-CI values estimating intelligibility of ADS and IDS stimuli

486   for NH and CI listeners. The bar graphs show the group data, corresponding to the means averaged

487   across talkers. The figure suggests that speech spoken in IDS is more intelligible than matched

488   utterances in ADS, regardless of whether intelligibility is estimated for listeners with NH or those

489   with CIs. On average, the stimuli spoken in IDS were ~15% more intelligible than those spoken

490   in ADS for listeners with normal hearing, as measured by the SRMR metric. This average

491   improvement in intelligibility of IDS stimuli over ADS stimuli was ~5.7% for listeners with CIs,

492   as estimated by SRMR-CI, suggesting that the supportive effects of IDS for speech intelligibility

493   improvement is decreased through CIs.

494   The results from the GLMM analysis for examining the effect of talkers' speaking style

495   (IDS vs. ADS), listeners' group (NH vs. CI), and their interaction on intelligibility of speech (Table

496   II) revealed a significant effect of speaking style ($\beta = 3.05$, $SE = 0.74$, $t = 4.14$, $p < 0.0001$). This

497   analysis provides new evidence that IDS may likely improve intelligibility of caregivers' speech

498   to both listeners with NH and those with CIs. These results further showed a significant, negative

499   effect of CI-related processing on intelligibility of caregivers' speech ($\beta = -2.23$, $SE = 0.74$, $t = -$

500   3.02, $p < 0.0001$). In addition, a significant interactive effect between speaking style and listeners

501   group was revealed by the GLMM model ($\beta = -1.337$, $SE = 0.466$, $t = -2.865$, $p < 0.001$), reflecting

502   the fact that hearing speech through CIs resulted in a smaller intelligibility gain from IDS relative

503   to ADS than was observed in the NH condition.

504
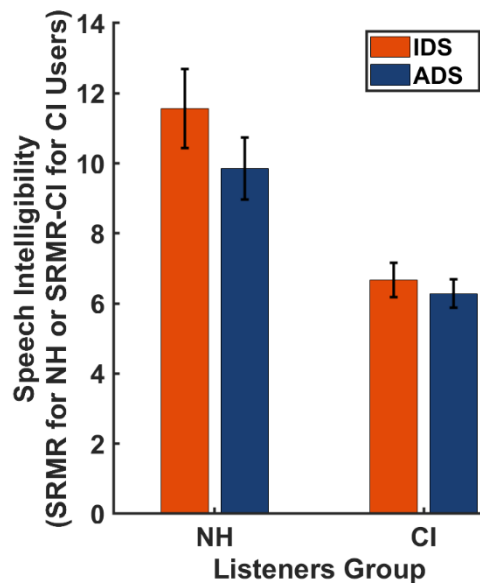
505   **FIG. 3.** Estimated intelligibility of speech stimuli spoken in IDS (orange; light gray) and in ADS

506   (blue; dark gray) styles (simulated) groups of listeners with different hearing statuses: NH

507   (estimated by SRMR) and those with CIs (estimated by SRMR-CI). The bar graphs represent

508   average values of SRMR or SRMR-CI over the seven talkers.

509   **TABLE II.** Statistical GLMM for the effects of speaking style, level of speech degradation, and

510   their interaction on intelligibility of speech, as measured by SRMR or SRMR-CI.

	β Estimate	*St. Error*	*t*	*Pr(>|t|)*
(intercept)	10.365	1.339	7.74	<0.00001 [a]
Speaking Style	3.054	0.737	4.14	<0.00001 [a]
Level of degradation	-2.23	0.737	-3.024	<0.0001 [a]
Speaking Style:Level of degradation	-1.336	0.466	-2.865	<0.0001 [a]

511   [a] The p-value was statistically significant.

512 **C. Individual differences across caregivers in ADS-vs-IDS acoustic distinctiveness and the**

513 **effect of simulated CI processing**

514 To examine individual differences across talkers in acoustic distinctiveness of IDS from ADS, we

515 focused on MD data from the individual talkers in two conditions (unprocessed and 22-channel

516 CI-simulated noise-vocoded speech), as shown in Fig. 4A (left vs. right bars, respectively). Two

517 general observations are noticeable in this data. First, caregivers vary considerably in terms of the

518 amount of acoustic variability created by their ADS-to-IDS modification (mean = 7.47, SD = 1.70,

519 range = 5.23). For example, the ADS-to-IDS shift in MD for talker 2 exposes a child with NH (cf.

520 natural, unprocessed condition) to a much wider range of acoustic information compared to that

521 for talker 3. This, in turn, suggests that a child of talker 2 would be more advantaged in learning

522 acoustic patterns of her caregivers' speech compared to a child of talker 3.

523 The second observation is that the magnitude of this acoustic variability significantly

524 reduced for CI-simulated signals (mean = 3.22, SD = 0.43, range = 1.4). An analysis of coefficient

525 of variation (CV) revealed that the standard deviation of MD for unprocessed (natural) speech

526 stimuli was 22.7% of its mean, whereas this value was 13.3% for CI-simulated speech stimuli for

527 the 22-channel noise-vocoder. Comparing these two CV values shows that the magnitude of

528 variation in ADS-IDS acoustic distinctiveness (i.e., MD) reduces approximately by half due to CI-

529 related speech processing, highlighting a major loss of acoustic information useful for recognition

530 of caregivers themselves and for distinguishing them from other caregivers.

531 If CI-simulated processing involves a *linear* transformation of acoustic distinctiveness

532 across speech styles for each talker, then we should observe MD differences to be significantly

533 correlated across natural (i.e., unprocessed) and 22-channel noise-vocoded conditions. The

534 Pearson's correlation coefficient was $r(5) = 0.07$ ($p = 0.87$), suggesting that CI-simulated speech

535   processing did not impact talkers' speech proportionately in terms of the acoustic distinctiveness

536   between their IDS and ADS. Instead, CI-related speech processing reduced the IDS-ADS acoustic

537   distance for some talkers (e.g., talker 6) more than others (e.g., talker 4).

538
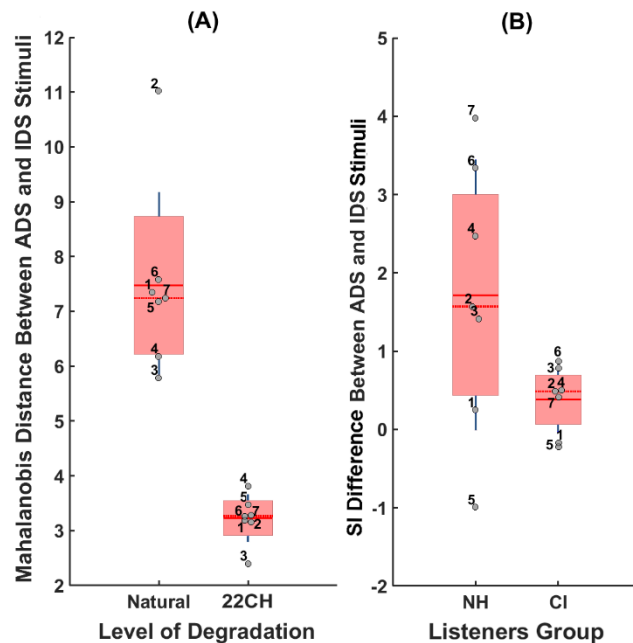
539   **FIG. 4.** Variability across seven female talkers in (A) acoustic distinctiveness between their IDS

540   and ADS for unprocessed stimuli (Natural) and simulated CI speech within a 22-channel noise

541   vocoder (22CH), and (B) the change in talkers' speech intelligibility (SI) due to a change in their

542   speaking style (ADS to IDS), as heard for two (simulated) listener groups with either NH

543   (estimated by SRMR) or CIs (estimated by SRMR-CI). The data points (gray circles) are laid over

544   a 1.96 standard error of the mean (95% confidence interval) in red (rectangle area with light gray)

545   and 1 standard deviation shown by blue lines (vertical dark gray lines). The solid and dotted red

546   lines (horizontal solid and dotted dark gray lines) show the mean and median, respectively.

D. Individual differences across caregivers in IDS and ADS intelligibility and the effect of

hearing status

To examine how individual talkers vary in the effects of their ADS-to-IDS modifications

on intelligibility of speech to listeners with NH and those with CIs, we focused on patterns of

SRMR and SRMR-CI metrics from the individual talkers. For (simulated) listeners with NH,

SRMRs of pairs of IDS and ADS stimuli were subtracted for each talker and averaged over 15

pairs of ADS-IDS stimuli in order to measure the amount of change in her speech intelligibility

due to ADS-to-IDS modification (for talker j, $\Delta SRMR_{(IDS\text{-}ADS)_j} = \frac{1}{15}\sum_{i=1}^{15} (SRMR_{IDS_{ij}} - SRMR_{ADS_{ij}})$).

Likewise, for (simulated) listeners with CI, SRMR-CIs of pairs of IDS and ADS stimuli were

subtracted for each talker and averaged over 15 pairs of ADS-IDS stimuli in order to quantify the

amount of change in her speech intelligibility due to ADS-to-IDS modification (for talker j,

$\Delta SRMR\text{-}CI_{(IDS\text{-}ADS)_j} = \frac{1}{15}\sum_{i=1}^{15} (SRMR\text{-}CI_{IDS_{ij}} - SRMR\text{-}CI_{ADS_{ij}})$). Thus, a positive value of this

difference measure represents IDS being, on average, more intelligible than ADS for a talker,

showing an intelligibility benefit for IDS over ADS for that talker.

Figure 4B (right panel) shows the results for changes in intelligibility of speech of talkers

due to modifications to speaking style. The bar plot on the left in Figure 4B (labeled as NH on the

x-axis) presents the difference across seven talkers in IDS-ADS intelligibility as estimated for

listeners with NH. The data in this scatter plot shows how much speech of a talker (e.g., talker 6)

would be heard as more or less intelligible as the talker speaks the utterances or words in IDS

compared to ADS.

Two patterns are observable in Figure 4B. First, the change in intelligibility of talkers'

speech due to the speaking style modification (ADS to IDS) was fairly variable across talkers for

27

569   simulated listeners with NH (mean = 1.71, *SD* = 1.73, range = 4.97), such that this change was

570   associated with relatively more intelligible speech for some talkers compared to others. As this

571   plot suggests, when talkers change their speaking style from ADS to IDS, not only does this not

572   lead to an equal amount of change in intelligibility of speech across talkers, but for one talker the

573   direction of this effect is even slightly negative (talker 5). For the majority of talkers (6 out of 7

574   talkers), IDS was more intelligible than ADS (as shown by a positive difference value).

575   Second, relative to the size of the ADS-to-IDS shift for simulated NH listeners, the size of

576   this ADS-to-IDS shift was overall considerably reduced for simulated listeners with CIs (mean =

577   0.38, SD = 0.43, range = 1.10). The bar plot on the right panel of figure 4B (labeled as CI on the

578   x-axis) presents the dispersion of change in speech intelligibility for each talker due to changing

579   speaking style from ADS to IDS, as estimated for listeners with CIs. Notably, CI-related

580   processing decreases the degree of variability in speech intelligibility, compared with that for NH,

581   as expected. An analysis of CV showed that although the mean of change in speech intelligibility

582   due to ADS-to-IDS modifications was overall smaller for the simulated CI condition compared to

583   the simulated NH condition, while the magnitude of variability was almost the same for two groups

584   of listeners (CV $\cong$ 11%). Highlighting non-linearity of effects of CI processing on speech across

585   different talkers, there was no significant correlation across the seven talkers for ADS-to-IDS

586   intelligibility differences in simulated CI vs. NH listening conditions ($r(5) = 0.74$, $p = 0.58$).

587   **IV. DISCUSSIONS**

588   The present study investigated the simulated effects of CI speech processing on acoustic

589   distinctiveness and intelligibility of speech signals as a function of IDS and ADS speaking styles.

590   Results from the present study supported the hypothesis that the limited spectral resolution in CIs

591   significantly degrades acoustic information involved in distinguishing IDS from ADS. This

592   significant loss of acoustic information related to the distinction between IDS and ADS may

593   negatively impact infants' recognizing IDS as distinct from ADS, leading possibly to less attention

594   to caregivers' speech, with potential consequences for children with CIs developing relatively

595   poorer language skills compared to those with NH.

596   These results are in line with prior findings demonstrating difficulties by listeners with CIs

597   in distinguishing among various speaking styles (Tamati, Janse, & Başkent, 2019). Distinguishing

598   caregivers' speaking styles is tied to advances in language development (Karzon, 1985; Singh &

599   Nestor, 2009; Thiessen et al., 2005; Wang et al., 2017), but is made challenging when speech is

600   degraded by CIs and/or other undesirable sources, e.g., noise and reverberation (Fetterman &

601   Domico, 2002; Hazrati & Loizou, 2012; Zheng, Koehnke, & Besing, 2011). The significant

602   elimination of IDS-related acoustic information suggested by the present study indicates that

603   infants with CIs likely do not have access to as wide a range of spectro-temporal information in

604   caregivers' speech to foster recognizing when speech is directed to them. As a result, these children

605   may experience considerable difficulties in recognizing IDS as distinct from ADS, particularly in

606   complex linguistic environments. It should be noted that, in the absence of fine-grained spectral

607   cues, infants with CIs may utilize other well-coded cues through CI devices such as caregivers'

608   speaking rate to detect whether speech is directed to them.

609   Although Wang et al., (2017) showed that IDS enhanced attention to speech in infants with

610   CIs compared to ADS, results from the present study suggest that infants' capability to recognize

611   IDS from ADS and to prefer attending to IDS over ADS is not probably comparable to that of

612   children with NH and is expected to be largely reduced. Furthermore, the fact that infants with CIs

613   showed this preference after 12 months of CI experience in Wang et al.'s (2017) study highlights

614   the possibility that the time course of developing certain capabilities for differentiating IDS and

615 ADS would be longer than peers with NH and would depend on the amount of experience with

616 CIs.

617   Our results also suggest that infants with CIs must develop a different cue-weighting

618 system for recognition of IDS from ADS compared to infants with NH, where the type and

619 magnitude of the relevant acoustic cues would be different from what infants with NH develop.

620 For example, in the absence of prominent acoustic cues of IDS such as F_0 (Fernald & Mazzie,

621 1991; Kuhl & Meltzoff, 1999; Mehler, 1981), which is poorly perceived through CI devices

622 (Mehta et al., 2020; Mehta & Oxenham, 2017; Qin & Oxenham, 2003), infants with CIs may rely

623 more on suprasegmental cues such as speech rate (Peng et al., 2017) and/or periodicity in the

624 temporal envelope (Fu, Chinchilla, & Galvin, 2004; Kong et al., 2004). However, perception of

625 pitch through periodicity in the temporal envelope is mostly limited to F_0s below around 300 Hz

626 (Carlyon, Deeks, & McKay, 2010; Kong & Carlyon, 2010), which is generally below the range of

627 F_0 variation in IDS. This lack of access to the entire spectro-temporal cue range in IDS puts infants

628 with CIs at high risk for missing IDS-related communicative events and thus for developing sub-

629 optimal language skills.

630   Our results corroborated our hypothesis that caregivers likely expose infants to more

631 intelligible speech when speaking in IDS compared to ADS, providing the first evidence on the

632 intelligibility benefit of IDS over ADS. This was shown using a novel application of a recently-

633 developed metric of intelligibility, which models intelligibility of speech for listeners with NH

634 (i.e., SRMR) and those with CIs (i.e., SRMR-CI) (Falk et al., 2015; Santos et al., 2013). This

635 positive effect of IDS on speech intelligibility might be because caregivers provide more clear

636 speech by speaking louder, slower, and/or in a hyperarticulated fashion when using IDS (Hazan et

637 al., 2018; Krause & Braida, 2002; Li et al., 2011; Liu, Del Rio, Bradlow, & Zeng, 2004). These

638   specific speaking patterns likely lead to an intelligibility benefit compared to ADS (Janse,

639   Nooteboom, & Quené, 2007; Liu et al., 2004). Greater intelligibility for IDS over ADS suggests

640   that use of IDS during caregiver-infant spoken communication assists infants in better

641   understanding speech, conceivably supporting infants' word learning process through a direct link

642   between exposure to IDS and improved speech intelligibility. However, our results also revealed

643   an interactive effect, showing that the size of this positive effect of IDS on caregivers' speech

644   intelligibility is smaller for infants with CIs compared to their peers with NH. This, in turn,

645   suggests that the limited spectral resolution of CI devices not only disrupts the communication of

646   IDS-ADS-specific acoustic information, but also decreases the intelligibility of caregivers' IDS,

647   indicating that infants with CIs are not expected to benefit from exposure to IDS as much as their

648   NH peers. These results are consistent with prior findings demonstrating greater difficulties that

649   adult listeners with CIs incur in processing speech and accomplishing lexical access, compared to

650   listeners with NH (McMurray, Farris-Trimble, & Rigler, 2017; Nagels, Bastiaanse, Başkent, &

651   Wagner, 2020). Although prior studies demonstrated a supportive role of IDS on infants' language

652   outcomes, to our knowledge, this is the first study that provides evidence supporting detrimental

653   effects of the CI speech processing on the supportive role of IDS in speech understanding.

654   Prior studies on speech recognition in listeners with CIs have highlighted the importance

655   of further examining patterns of individual differences, in addition to group differences (Dilley et

656   al., 2020; Nagels et al., 2020; Peng et al., 2019; Spencer, 2004; Szagun & Schramm, 2016). Our

657   investigation of patterns of individual differences across caregivers in acoustic distinctiveness

658   between IDS and ADS and the simulated effect of limited spectral resolution in CI on these patterns

659   highlighted multiple important observations. The first observation relates to the amount of acoustic

660   variability produced by each talker due to changing speaking style from ADS to IDS. The amount

661   of resultant acoustic variability varied across talkers, indicating that infants would experience

662   different language environments in terms of qualities of their caregivers' speech. This is expected

663   to result in some caregivers' exposing infants to relatively larger ranges of acoustic information,

664   as compared to others, which would foster their infants' speech processing and language

665   acquisition by assisting infants in more robust recognition of their caregivers' voices (Beauchemin

666   et al., 2011; Kisilevsky et al., 2003), as well as better development of auditory cortical processing

667   for language development (Webb et al., 2015). Similar to the recent findings by Dilley et al. (2020),

668   these results also suggest that IDS is not always readily distinguishable from ADS, due to the fact

669   that caregivers vary in implementation of IDS. As such, these results confirm that conditions for

670   recognition of IDS by infants is not always optimal (Piazza et al., 2017) and might differentially

671   affect language outcomes for infants with CIs (Dilley et al., 2020).

672   More importantly, the amount of ADS-IDS acoustic difference was considerably reduced

673   for each caregiver as her speech passed through a CI simulator, indicating that infants with CIs

674   probably have access to a much narrower range of acoustic variability in the voices of their

675   caregivers, as compared to their peers with NH. This large decline in the degree of acoustic

676   variability in caregivers' voices may negatively impact infants' robust identification of their

677   caregivers' voices, thereby preventing their gaining the maximal benefit from language input that

678   may be available in their linguistic environments. In fact, encoding and learning these cues is

679   crucial as they both significantly contribute to infants' familiarity with ranges of acoustic variation

680   in talkers' speech and their robust performance in understanding speech, despite of multiple

681   sources of variability, such as talker variation (Allen, Miller, & DeSteno, 2009; Eskenazi, 1993),

682   language context (Mattys, 2000; McMurray & Aslin, 2005; Miller, 1994), speech rate (Sommers,

683   Nygaard, & Pisoni, 1992), and background noise and/or reverberation (Hawkins, 2004). Thus,

684   infants with CIs are probably at risk for partial learning of subtle voice cues specific to their

685   caregivers, as reflected in their fine-structure spectro-temporal information.

686   Acoustic implementation of ADS-to-IDS modification varies across individual caregivers,

687   which may differentially impact infants' understanding of speech and thus their language

688   outcomes (Dilley et al., 2020; Hoff, 2006; Weisleder & Fernald, 2013). Results from the present

689   study showed that individual talkers varied in the effect of their speaking style modification on

690   speech intelligibility, suggesting that modification of speaking style from ADS to IDS does not

691   always cause an equal degree of improvement in intelligibility of caregivers' speech. Notably, our

692   results showed that this variability across caregivers in the impact of their speaking style

693   modification on speech intelligibility was reduced due to limited spectral resolution in CIs. In the

694   present study, we only studied variability across caregivers in acoustic information and

695   intelligibility of their IDS compared to ADS, whereas caregivers of infants with CIs may vary in

696   other aspects of spoken communication such as gestural and proprioceptive behaviors, which very

697   likely change the degree of intelligibility by which infants eventually perceive caregivers' speech

698   (Kirk et al., 2007; Kirk & Pisoni, 2002; Lachs, Pisoni, & Kirk, 2001).

699   These findings can be further interpreted in the context of spectro-temporal information

700   available to infants' auditory systems, which is very sensitive to subtle changes in speech acoustics

701   (Jusczyk, Hohne, & Bauman, 1999; Kuhl, 2004). When infants' accessibility to fine-grained

702   spectro-temporal structures in speech is largely compromised because of limited spectral

703   resolution in CIs, they may not be able to readily recognize and attend to rich IDS. In the absence

704   of this fine-structure spectral information at the output of CI electrodes, infants must rely on

705   course-grained cues in the speech envelope, such as temporal envelope periodicity (Green,

706   Faulkner, & Rosen, 2002; Moore, 2003) or cues from other sensory modalities (e.g., visual and

707   tactile, Green, Nip, Wilson, Mefferd, & Yunusova, 2010; Rohlfing, Fritsch, Wrede, & Jungmann,

708   2006) in order to recognize IDS from ADS. This increases the cognitive load in processing speech

709   particularly in complex auditory environments and negatively contributes to the observed poor

710   language outcomes in some children (Davidson, Geers, & Nicholas, 2014; Dunn et al., 2014;

711   Geers, Strube, Tobey, & Moog, 2011; Geers, Nicholas, Tobey, & Davidsonb, 2015; Houston,

712   Pisoni, Kirk, Ying, & Miyamoto, 2003; Houston, Stewart, Moberly, Hollich, & Miyamoto, 2012;

713   Miyamoto, Houston, Kirk, Perdew, & Svirsky, 2003; Niparko et al., 2010; Pisoni et al., 2007;

714   Pisoni, Kronenberger, Harris, & Moberly, 2018).

715   The present study used a within-talker manipulation of speech style to investigate how

716   limited spectral resolution in CIs may affect processing of speech style shifts which are tied to

717   language development, namely IDS vs. ADS signals. While the study involved a sizeable number

718   of individual utterances with control of phonetic properties within talkers, the amount of speech

719   collected from each of the seven talkers was relatively small. Further work will be needed to test

720   how these results extend to a larger sample of talkers with utterances with more varied segmental

721   and lexical composition. Furthermore, studying natural IDS and ADS would be ideal; however, it

722   is difficult to control the content and quality of the speech collected from naturalistic environments.

723   Importantly, studies based on simulated CI speech and computational models of human hearing

724   provide valuable evidence for understanding speech perception in listeners with normal and

725   impaired hearing (Litvak, Delgutte, & Eddington, 2001; Mehta et al., 2020; Rubinstein, Wilson,

726   Finley, & Abbas, 1999; Throckmorton & Collins, 2002). Findings from these studies should be

727   viewed as evidence for general trends, rather than the actual performance of listeners with CIs,

728   subject to further investigation. Although we simulated limitation of CIs in spectral resolution by

729   changing the number of channels in noise-vocoded speech, the effect of other aspects of CI

34

730   processing on spectral resolution, such as channel interactions (i.e., steepness of spectral slope)

731   (Crew & Galvin, 2012; Mehta et al., 2020), requires further investigation. Additionally, metrics of

732   speech intelligibility of SRMR and SRMR-CI have been validated for adult listeners with CIs and

733   are not direct estimates of speech intelligibility, which is very likely different from what infants

734   with CIs experience in terms of recognition of utterances. Considering these limitations, the results

735   should not be taken as the final determination of how children with CIs perform in recognition

736   between IDS and ADS and how ADS-to-IDS modification impacts the degree of intelligibility of

737   caregivers' speech to children with CIs.

738   Despite these limitations, the simulated results from the present study imply that, compared

739   to infants with NH, infants with CIs could be disadvantaged in perceiving IDS and acquiring

740   spoken language due to multiple factors. First, partial transmission of spectro-temporal cues

741   because of limited spectral resolution likely decreases the supportive role of IDS in infants'

742   language learning by disrupting the link between attending to IDS and speech comprehension. In

743   fact, degraded representation of spectral information that contributes to acoustic distinction

744   between IDS and ADS in CIs may have detrimental effects on infants' ability to pay attention to

745   caregivers' speech, something that is a fundamental cognitive skill for spoken language acquisition

746   (Bergeson, 2014; Glenn, Cunningham, & Joyce, 1981; Houston et al., 2003; Rottmann & Zobrist,

747   2004; Wang, Shafto, & Houston, 2018). It is worth mentioning that infants' language learning

748   involves incorporating information from multiple interwind communication dimensions (i.e.,

749   visual, social, tactile, and emotional), which creates a very rich channel for learning language

750   through infant-directed speech (Gogate, Bahrick, & Watson, 2000; Nomikou & Rohlfing, 2011),

751   even in the absence of major spectro-temporal cues such as talker's F_0 in the output of CI

752   electrodes. Our findings further suggested that CIs diminish the benefit of exposure to more

753 intelligible IDS, as compared to ADS, which may negatively impact infants' abilities to process

754 caregivers' speech, possibly with further significant downstream consequences for later language

755 skills. Last but not least, our simulated results imply that, compared to infants with NH, infants

756 with CIs have access to a relatively narrower range of acoustic information (corresponding to

757 smaller acoustic variability) in their caregivers' speech, which probably leads to experiencing

758 greater difficulties in robust identification and recognition of their caregivers' voices (Beauchemin

759 et al., 2011; Kisilevsky et al., 2003; Lavan et al., 2019), as well as developing poorer word

760 recognition skills (Singh, 2008) and an impaired auditory system for language processing (Webb

761 et al., 2015). Despite the reduced ADS-vs-IDS distinctiveness, intelligibility, and variability of

762 vocoded IDS, it is possible that children with CIs may have developed certain coping/adapting

763 strategies to mitigate the degraded speech input. For example, it is possible that children with CIs

764 may have higher sensitivity (lower threshold) to the acoustic cues than children with NH.

765 In summary, the current study used computational and signal processing approaches in

766 order to provide new evidence for how CI-related speech processing may impact recognition of

767 IDS from ADS in children with CIs, as well as how these style differences may affect intelligibility

768 benefits derived by style shifts from ADS to IDS. These findings provide solid grounding for

769 developing new perceptual studies to test abilities of infants with CIs to recognize their caregivers'

770 speaking style (ADS vs. IDS) and to recognize intelligible words in caregivers' speech. Focusing

771 on computational metrics as undertaken here provides an important complement to costly,

772 complex, labor-intensive perceptual studies. The results provided preliminary evidence for how

773 CI-related speech processing may alter the pathway from exposure to IDS to processing speech

774 and, by extension, the acquisition of language by children with CIs compared to that of their NH

775 peers in two ways: (1) making it probably harder for children with CIs to recognize IDS from

776   ADS, (2) decreasing the ADS-to-IDS intelligibility benefit. The most direct and immediate

777   implication of these findings is the imperative need to improve signal processing in CI devices to

778   assure the faithful transmission of acoustic cues relevant to identification and recognition of IDS

779   from ADS. Until then, the major clinical implication of these findings is that the maximum benefit

780   from exposure to IDS for language learning in infants with CIs requires caregivers' active use of

781   multimodal (i.e., gesture, tactile, visual, social, and emotional) communicative behaviors in order

782   to compensate for the degraded representation of acoustic information relevant to IDS and to

783   support its robust perception.

ACKNOWLEDGMENTS

788   **REFERENCES**

789   Allen, J. S., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice-onset-time:

790   Contextual influences. *The Journal of the Acoustical Society of America*, *125*(6), 3974–3982.

791   Arjmandi, M., Dilley, L. C., & Wagner, S. E. (2018). Investigation of acoustic dimension use in

792   dialect production: machine learning of sonorant sounds for modeling acoustic cues of

793   African American dialect. *11th International Conference on Voice Physiology and*

794   *Biomechanics*, 12–13. East Lansing, USA.

795   Baskent, D., & Gaudrain, E. (2016). Perception and Psychoacoustics of Speech in Cochlear

796   Implant Users. *Scientific Foundations of Audiology. Perspectives from Physics, Biology,*

797   *Modelling,   and   Medicine*,   185–320.   Retrieved   from

Running Title: Infant-directed speech through cochlear implant

798   https://books.google.de/books?hl=de&lr=&id=EtAyDAAAQBAJ&oi=fnd&pg=PA285&dq

799   =Scientific+Foundations+of+Audiology&ots=cfEfTicv7h&sig=1cTQmXsc_FR7oNQiwW

800   YgpklOkN0

801 Beauchemin, M., González-Frankenberger, B., Tremblay, J., Vannasing, P., Martínez-Montes, E.,

802   Belin, P., … Lassonde, M. (2011). Mother and stranger: An electrophysiological study of

803   voice processing in newborns. *Cerebral Cortex*, *21*(8), 1705–1711.

804 Bergeson, T. R. (2014). Hearing versus Listening: Attention to Speech and Its Role in Language

805   Acquisition in Deaf Infants with Cochlear Implants. *Lingua*, 10–25.

806 Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot

807   International*, *5:9/10*, 341–345.

808 Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning

809   disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing

810   Research*, *46*(1), 80–97.

811 Brown, C. A., & Bacon, S. P. (2010). Fundamental frequency and speech intelligibility in

812   background noise. *Hearing Research*, *266*(1–2), 52–59.

813 Burnham, E. B., Wieland, E. A., Kondaurova, M. V., McAuley, J. D., Bergeson, T. R., & Dilley,

814   L. C. (2015). Phonetic Modification of Vowel Space in Storybook Speech to Infants up to 2

815   Years of Age. *Journal of Speech, Language, and Hearing Research*, *58*(2), 241–253.

816 Carlyon, R. P., Deeks, J. M., & McKay, C. M. (2010). The upper limit of temporal pitch for

817   cochlear-implant listeners: Stimulus duration, conditioner pulses, and the number of

818   electrodes stimulated. *The Journal of the Acoustical Society of America*, *127*(3), 1469–1478.

819   Chatterjee, M., & Oberzut, C. (2011). Detection and rate discrimination of amplitude modulation

820   in electrical hearing. *The Journal of the Acoustical Society of America*, *130*(3), 1567–1580.

821   Chatterjee, M., & Peng, S. C. (2008). Processing F0 with cochlear implants: Modulation frequency

822   discrimination and speech intonation recognition. *Hearing Research*, *235*(1–2), 143–156.

823   Consortium, M. (2020). Quantifying sources of variability in infancy research using the infant-

824   directed speech preference. *Advances in Methods and Practices in Psychological Science*,

825   *3*(1), 24–52. https://doi.org/10.17605/OSF.IO/S98AB

826   Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants'

827   preference for motherese. *Infant Behavior and Development*, *20*(4), 477–488.

828   Cooper, R. P., & Aslin, R. N. (1990). Preference for Infant-Directed Speech in the First Month

829   after Birth. *Child Development*, *61*(5), 1584–1595.

830   Crew, J. D., & Galvin, J. J. (2012). Channel interaction limits melodic pitch perception in

831   simulated cochlear implants. *The Journal of the Acoustical Society of America*, *132*(5),

832   EL429–EL435.

833   Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal*

834   *of Child Language*, *41*(4), 935.

835   Croghan, N. B. H., Duran, S. I., & Smith, Z. M. (2017). Re-examining the relationship between

836   number of cochlear implant channels and maximal speech intelligibility. *The Journal of the*

837   *Acoustical Society of America*, *142*(6), EL537–EL543.

838   Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the Comprehension of Spoken

839   Language: A Literature Review. *Language and Speech*, *40*, 141–201.

39

840   Davidson, L. S., Geers, A. E., & Nicholas, J. G. (2014). The effects of audibility and novel word

841   learning ability on vocabulary level in children with cochlear implants. *Cochlear Implants*

842   *International*, *15*(4), 211–221.

843   Dilley, L., Lehet, M., Wieland, E. A., Arjmandi, M. K., Houston, D., Kondaurova, M., …

844   Bergeson, T. (2020). Individual differences in mothers' spontaneous infant-directed speech

845   predict language attainment in children with cochlear implants. *Journal of Speech, Language,*

846   *and Hearing Research*, 1–15.

847   Dorman, M. F., Loizou, P. C., Fitzke, J., & Tu, Z. (1998). The recognition of sentences in noise

848   by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–

849   20 channels. *The Journal of the Acoustical Society of America*, *104*(6), 3583–3585.

850   Dorman, M. F., Loizou, P. C., & Rainey, D. (1997a). Simulating the effect of cochlear-implant

851   electrode insertion depth on speech understanding. *The Journal of the Acoustical Society of*

852   *America*, *102*(5), 2993–2996.

853   Dorman, M. F., Loizou, P. C., & Rainey, D. (1997b). Speech intelligibility as a function of the

854   number of channels of stimulation for signal processors using sine-wave and noise-band

855   outputs. *The Journal of the Acoustical Society of America*, *102*(4), 2403–2411.

856   Dunn, C. C., Walker, E. A., Oleson, J., Kenworthy, M., Voorst, T. Van, Tomblin, J. B., … Gantz,

857   B. J. (2014). Longitudinal speech perception and language performance in pediatric cochlear

858   implant users: The effect of age at implantation. *Ear and Hearing*, *35*(2), 148–160.

859   Eskenazi, M. (1993). Trends in speaking styles research. *Third European Conference on Speech*

860   *Communication and Technology*, 501–509.

861   Falk, T. H., Parsa, V., Santos, J. F., Arehart, K., Hazrati, O., Falk, T. H., … Scollie, S. (2015).

862   Objective Quality Prediction for Users of and Intelligibility Assistive Listening Devices.

863   *IEEE Signal Processing Magazine*, *32*(2), 114–124.

864   Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech:

865   Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*,

866   *50*(5), 1241–1255.

867   Ferguson, S. H., Poore, M. A., Shrivastav, R., Kendrick, A., McGinnis, M., & Perigoe, C. (2010).

868   Acoustic Correlates of Reported Clear Speech Strategies. *Journal of the Academy of*

869   *Rehabilitative Audiology*, *43*, 45–64.

870   Ferguson, S. H., & Quené, H. (2014). Acoustic correlates of vowel intelligibility in clear and

871   conversational speech for young normal-hearing and elderly hearing-impaired listeners. *The*

872   *Journal of the Acoustical Society of America*, *135*(6), 3570–3584.

873   Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and*

874   *Development*, *8*(2), 181–195.

875   Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the

876   melody the message? *Child Development*, *60*(6), 1497–1510.

877   Fernald, A. (1993). Approval and Disapproval : Infant Responsiveness to Vocal Affect in Familiar

878   and Unfamiliar Languages. *Child Development*, *64*(3), 657–674.

879   Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults.

880   *Developmental Psychology*, *27*(2), 209–221.

881   Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns.

882   *Developmental Psychology*, *20*(1), 104–113.

883   Fernald, & Kuhl. (1987). Acoustic determinants of infant preference for motherse speech. *Infant*

884   *Behaviour and Development*, *10*, 279–293.

885   Fetterman, B. L., & Domico, E. H. (2002). Speech recognition in background noise of cochlear

886   implant patients. *Otolaryngology - Head and Neck Surgery*, *126*(3), 257–263.

887   Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a

888   function of the number of spectral channels: Comparison of acoustic hearing and cochlear

889   implants. *The Journal of the Acoustical Society of America*, *110*(2), 1150–1163.

890   Fu, Q.-J. (2019). *AngelSim: Cochlear implant and hearing loss simulator*. Retrieved from

891   http://www.tigerspeech.com/angelsim/angelsim_about.html

892   Fu, Q.-J., Chinchilla, S., & Galvin, J. J. (2004). The role of spectral and temporal cues in voice

893   gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the*

894   *Association for Research in Otolaryngology*, *5*(3), 253–260.

895   Fu, Q.-J., Chinchilla, S., Nogaki, G., & Galvin, J. J. (2005). Voice gender identification by cochlear

896   implant users: The role of spectral and temporal resolution. *The Journal of the Acoustical*

897   *Society of America*, *118*(3), 1711–1718.

898   Fu, Q.-J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: The role of spectral

899   resolution and smearing. *Journal of the Association for Research in Otolaryngology*, *6*(1),

900   19–27.

901   Fu, Q.-J., Shannon, R. V., & Wang, X. (1998). Effects of noise and spectral resolution on vowel

902   and consonant recognition: Acoustic and electric hearing. *The Journal of the Acoustical*

903   *Society of America*, *104*(6), 3586–3596.

904   Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R. H., & Başkent, D. (2014).

905   Gender Categorization in Cochlear Implant Users. *JARO - Journal of the Association for*

906   *Research in Otolaryngology*, *15*(6), 1037–1048. https://doi.org/10.1007/s10162-014-0483-7

907   Garnica, O. (1977). On some prosodic and paralinguistic features of speech to young children. In

908   C. S. and C. Ferguson (Ed.), *Talking to Children: Language Input and Acquisition* (pp. 271–

909   285). Cambridge, UK: Cambridge University Press.

910   Gaudrain, E., & Baskent, D. (2018). Discrimination of voice pitch and vocal-tract length in

911   cochlear   implant   users.   *Ear   and   Hearing*,   *39*(2),   226–237.

912   https://doi.org/10.1097/AUD.0000000000000480

913   Geers, A. E., Strube, M. J., Tobey, E. A., & Moog, J. S. (2011). Epilogue: factors contributing to

914   long-term outcomes of cochlear implantation in early childhood. *Ear and Hearing*, *32*(1

915   Suppl), 84S.

916   Geers, Ann E., Nicholas, J., Tobey, E., & Davidsonb, L. (2016). Persistent Language Delay Versus

917   Late Language Emergence in Children With Early Cochlear Implantation. *Journal of Speech,*

918   *Language, and Hearing Research*, *59*(1), 155–170.

919   Glenn, S. M., Cunningham, C. C., & Joyce, P. F. (1981). A Study of Auditory Preferences in

920   Nonhandicapped Infants and Infants with down ' s Syndrome. *Child Development*, 1303–

921   1307.

922   Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role

923   of temporal synchrony between verbal labels and gestures. *Child Development*, *71*(4), 878–

924   894.

925   Green, J. R., Nip, I. S. B., Wilson, E. M., Mefferd, A. S., & Yunusova, Y. (2010). Lip movement

926   exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing*

927   *Research*, *53*(6), 1529–1542.

928   Green, T., Faulkner, A., & Rosen, S. (2002). Spectral and temporal cues to pitch in noise-excited

929   vocoder simulations of continuous-interleaved-sampling cochlear implants. *The Journal of*

930   *the Acoustical Society of America*, *112*(5), 2155–2164.

931   Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years

932   later. *Journal of the Acoustical Society of America*, *87*(6), 2592–2605.

933   Grieco-calub, T. M., Saffran, J. R., & Litovsky, R. Y. (2010). Spoken Word Recognition in

934   Toddlers Who Use Cochlear Implants. *Journal of Speech Language and Hearing Research*,

935   *52*(6), 1390–1400.

936   Hawkins, S. (2004). Puzzles and patterns in 50 years of research on speech perception. *From Sound*

937   *to Sense: 50+ Years of Discoveries in Speech Communication*, 223–246.

938   Hazan, V., Tuomainen, O., Kim, J., Davis, C., Sheffield, B., & Brungart, D. (2018). Clear speech

939   adaptations in spontaneous speech produced by young and older adults. *The Journal of the*

940   *Acoustical Society of America*, *144*(3), 1331–1346.

941   Hazrati, O., & Loizou, P. C. (2012). The combined effects of reverberation and noise on speech

942   intelligibility by cochlear implant listeners. *International Journal of Audiology*, *51*(6), 437–

943   443.

944   Heijden, V. Der, Ferdinand, R. P. D., Ridder, D. De, & Tax, D. M. (2005). *Classification,*

945   *Parameter Estimation and State Estimation An Engineering Approach Using MATLAB*. John

946   Wiley & Sons.

947   Hoff, E. (2006). How social contexts support and shape language development. *Developmental*

948   *Review*, *26*(1), 55–88.

949   Horowitz, F. D. (1983). The Effects of Intonation on Infant Attention: The Role of the Rising

950   Intonation Contour. *Journal of Child Language*, *10*(3), 521–534.

951   Houston, D. M., Pisoni, D. B., Kirk, K. I., Ying, E. A., & Miyamoto, R. T. (2003). Speech

952   perception skills of deaf infants following cochlear implantation: A first report. *International*

953   *Journal of Pediatric Otorhinolaryngology*, *67*(5), 479–495.

954   Houston, D. M., Stewart, J., Moberly, A., Hollich, G., & Miyamoto, R. T. (2012). Word learning

955   in deaf children with cochlear implants: Effects of early auditory experience. *Developmental*

956   *Science*, *15*(3), 448–461.

957   Hunt, N. J., Lennig, N., & Mermeletein, P. (1980). Experiments in syllable-based recognition of

958   continuous speech. *IEEE International Conference on Acoustics, Speech, and Signal*

959   *Processing*, (5), 880–883.

960   Imai, S. (1983). Cepstral analysis synthesis on the mel frequency scale. *IEEE International*

961   *Conference on Acoustics, Speech, and Signal Processing*, 93–96.

962   Inoue, T., Nakagawa, R., Kondou, M., Koga, T., & Shinohara, K. (2011). Discrimination between

963   mothers' infant- and adult-directed speech using hidden Markov models. *Neuroscience*

964   *Research*, *70*(1), 62–70.

965   Jahn, K. N., DiNino, M., & Arenberg, J. G. (2019). Reducing Simulated Channel Interaction

966   Reveals Differences in Phoneme Identification Between Children and Adults With Normal

967   Hearing. *Ear and Hearing*, *40*(2), 295–311.

968 Jain, S., & Vipin Ghosh, P. G. (2018). Acoustic simulation of cochlear implant hearing: Effect of

969   manipulating various acoustic parameters on intelligibility of speech. *Cochlear Implants*

970   *International*, *19*(1), 46–53.

971 Janse, E., Nooteboom, S. G., & Quené, H. (2007). Coping with gradient forms of /t/-deletion and

972   lexical ambiguity in spoken word recognition. In *Language and Cognitive Processes* (Vol.

973   22).

974 Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for

975   word segmentation. *Perception and Psychophysics*, *61*(8), 1465–1476.

976 Karzon, R. G. (1985). Discrimination of polysyllabic sequences by one- to four-month-old infants.

977   *Journal of Experimental Child Psychology*, *39*(2), 326–342.

978 Kirk, K. I., Hay-McCutcheon, M. J., Holt, R. F., Gao, S., Qi, R., & Gerlain, B. L. (2007).

979   Audiovisual spoken word recognition by children with cochlear implants. *Audiological*

980   *Medicine*, *5*(4), 250–261.

981 Kirk, K. I., & Pisoni, D. B. (2002). Audiovisual integration of speech by children and adults with

982   cochlear implants. *International Conference on Spoken Language Processing*, 1689.

983 Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., … Wang, Z. (2003). Effects

984   of experience on fetal voice recognition. *Psychological Science*, *14*(3), 220–224.

985 Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2001). Universality and

986   specificity in infant-directed speech: Pitch modifications as a function of infant age and sex

987   in a tonal and non-tonal language. *Infant Behavior and Development*, *24*(4), 372–392.

988   Kong, Y.-Y., & Carlyon, R. P. (2010). Temporal pitch perception at high rates in cochlear

989   implants. *The Journal of the Acoustical Society of America*, *127*(5), 3114–3123.

990   Kong, Y. Y., Mullangi, A., Marozeau, J., & Epstein, M. (2011). Temporal and Spectral Cues for

991   Musical Timbre Perception in Electric Hearing. *Journal of Speech, Language, and Hearing*

992   *Research*, *54*, 981–994. https://doi.org/10.1044/1092-4388(2010/10-0196).Temporal

993   Kong, Ying Yee, Cruz, R., Jones, J. A., & Zeng, F. G. (2004). Music Perception with Temporal

994   Cues in Acoustic and Electric Hearing. *Ear and Hearing*, *25*(2), 173–185.

995   Kong, Ying Yee, Stickney, G. S., & Zeng, F.-G. (2005). Speech and melody recognition in

996   binaurally combined acoustic and electric hearing. *The Journal of the Acoustical Society of*

997   *America*, *117*(3), 1351–1361.

998   Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects

999   of speaking rate and speaking mode on intelligibility. *The Journal of the Acoustical Society*

1000   *of America*, *112*(5), 2165–2172.

1001   Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews*

1002   *Neuroscience*, *5*(11), 831–843.

1003   Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina,

1004   V. L., … Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed

1005   to infants. *Science*, *277*(5326), 684–686.

1006   Kuhl, P. K., & Meltzoff, A. N. (1999). The intermodal representation of speech in newborns.

1007   *Developmental Science*, *2*(1), 42–46.

47

1008   Lachs, L., Pisoni, D. B., & Kirk, K. I. (2001). Use of audiovisual information in speech perception

1009   by prelingually deaf children with cochlear implants: A first report. *Ear and Hearing*, *22*(3),

1010   236–251.

1011   Lavan, N., Burton, A. M., Scott, S. K., & McGettigan, C. (2019). Flexible voices: Identity

1012   perception from variable vocal signals. *Psychonomic Bulletin and Review*, *26*(1), 90–102.

1013   Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2014). Infant-directed speech

1014   enhances temporal rhythmic structure in the envelope. *Proceedings of the Annual Conference*

1015   *of the International Speech Communication Association, INTERSPEECH*, (September),

1016   2563–2567.

1017   Li, Y., Zhang, G., Kang, H., Liu, S., Han, D., & Fu, Q.-J. (2011). Effects of speaking style on

1018   speech intelligibility for Mandarin-speaking cochlear implant users. *The Journal of the*

1019   *Acoustical Society of America*, *129*(6), EL242–EL247.

1020   Litvak, L., Delgutte, B., & Eddington, D. (2001). Auditory nerve fiber responses to electric

1021   stimulation: Modulated and unmodulated pulse trains. *The Journal of the Acoustical Society*

1022   *of America*, *110*(1), 368–379.

1023   Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity

1024   and infants' speech discrimination skills. *Developmental Science*, *6*(3), 1–10.

1025   Liu, S., Del Rio, E., Bradlow, A. R., & Zeng, F.-G. (2004). Clear speech perception in acoustic

1026   and electric hearing. *The Journal of the Acoustical Society of America*, *116*(4), 2374–2383.

1027   Loizou, P. C., Dorman, M., & Tu, Z. (1999). On the number of channels needed to understand

1028   speech. *The Journal of the Acoustical Society of America*, *106*(4), 2097–2103.

1029   Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate

1030   Speech in Human Auditory Cortex. *Neuron*, *54*(6), 1001–1010.

1031   Luo, X., Fu, Q.-J., & Galvin, J. J. (2007). Vocal Emotion Recognition by Normal-Hearing

1032   Listeners and Cochlear Implant Users. *Trends in Amplification*, *11*(4), 301–315.

1033   Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant-

1034   and adult-directed speech. *Language Learning and Development*, *7*(3), 185–201.

1035   Maesschalck, R. D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and*

1036   *Intelligent Laboratory Systems*, *50*(1), 1–18.

1037   Masnan, M. J., Mahat, N. I., Shakaff, A. Y. M., Abdullah, A. H., Zakaria, N. Z. I., Yusuf, N., …

1038   Aziz, A. H. A. (2015). Understanding Mahalanobis distance criterion for feature selection.

1039   *AIP Conference Proceedings*, *1660*, 050075. AIP Publishing LLC.

1040   Mastropieri, D., & Turkewitz, G. (1999). Prenatal experience and neonatal responsiveness to vocal

1041   expressions of emotion. *Developmental Psychobiology*, *35*(3), 204–214.

1042   Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception and*

1043   *Psychophysics*, *62*(2), 253–265.

1044   Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error

1045   and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

1046   Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can

1047   affect phonetic discrimination. *Cognition*, *82*(3), 101–111.

1048   McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech

1049   perception. *Cognition*, *95*(2), B15–B26.

1050 McMurray, B., Farris-Trimble, A., & Rigler, H. (2017). Waiting for lexical access: Cochlear

1051   implants or severely degraded input lead listeners to process speech less incrementally.

1052   *Cognition*, *169*(August), 147–164.

1053 McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech

1054   and the development of speech perception: Enhancing development or an unintended

1055   consequence? *Cognition*, *85*(0 1), 1–27.

1056 Mehler, J. (1981). The role of syllables in speech processing: infant and adult data. *Philosophical*

1057   *Transactions of the Royal Society of London. B, Biological Sciences*, *295*(1077), 333–352.

1058 Mehta, A. H., Lu, H., & Oxenham, A. J. (2020). The Perception of Multiple Simultaneous Pitches

1059   as a Function of Number of Spectral Channels and Spectral Spread in a Noise-Excited

1060   Envelope Vocoder. *JARO - Journal of the Association for Research in Otolaryngology*, *21*(1),

1061   61–72.

1062 Mehta, A. H., & Oxenham, A. J. (2017). Vocoder Simulations Explain Complex Pitch Perception

1063   Limitations Experienced by Cochlear Implant Users. *JARO - Journal of the Association for*

1064   *Research in Otolaryngology*, *18*(6), 789–802.

1065 Miller, J. L. (1994). On the internal structure of phonetic categories: a progress report. *Cognition*,

1066   *50*(1–3), 271–285.

1067 Miyamoto, R. T., Houston, D. M., Kirk, K. I., Perdew, A. E., & Svirsky, M. A. (2003). Language

1068   development in deaf infants following cochlear implantation. *Acta Oto-Laryngologica*,

1069   *123*(2), 241–244.

1070 Moore, B. C. J. (2003). Coding of sounds in the auditory system and its relevance to signal

1071   processing and coding in cochlear implants. *Otology and Neurotology*, *24*(2), 243–254.

1072   Moore, D. S., Spence, M. J., & Katz, G. S. (1997). Six-month-olds' categorization of natural

1073   infant-directed utterances. *Developmental Psychology*, *33*(6), 980–989.

1074   Nagels, L., Bastiaanse, R., Başkent, D., & Wagner, A. (2020). Individual differences in lexical

1075   access among cochlear implant users. *Journal of Speech, Language, and Hearing Research*,

1076   *63*(1), 286–304.

1077   Narayan, C. R., & McDermott, L. C. (2016). Speech rate and pitch characteristics of infant-directed

1078   speech: Longitudinal and cross-linguistic observations. *The Journal of the Acoustical Society*

1079   *of America*, *139*(3), 1272–1281.

1080   Niparko, J. K., Tobey, E. A., Thal, D. J., Eisenberg, L. S., Wang, N.-Y., Quittner, A. L., & Fink,

1081   N. E. (2010). Spoken language development in children following cochlear implantation.

1082   *Jama*, *303*(15), 1498–1506.

1083   Nomikou, I., & Rohlfing, K. J. (2011). Language does something: Body action and language in

1084   maternal input to three-month-olds. *IEEE Transactions on Autonomous Mental Development*,

1085   *3*(2), 113–128.

1086   Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses

1087   to prototypical melodic contours in parental speech. *Infant Behavior and Development*, *13*(4),

1088   539–545.

1089   Peng, S. C., Lu, H. P., Lu, N., Lin, Y. S., Deroche, M. L. D., & Chatterjee, M. (2017). Processing

1090   of acoustic cues in lexical-tone identification by pediatric cochlear-implant recipients.

1091   *Journal of Speech, Language, and Hearing Research*, *60*(5), 1223–1235.

51

1092   Peng, S. C., Tomblin, J. B., & Turner, C. (2008). Production and Perception of Speech Intonation

1093   in Pediatric Cochlear Implant Recipients and Individuals with Normal Hearing. *Ear and*

1094   *Hearing*, *29*(3), 336–351.

1095   Peng, Z., Hess, C., Saffran, J. R., Edwards, J. R., & Litovsky, R. Y. (2019). Assessing Fine-Grained

1096   Speech Discrimination in Young Children With Bilateral Cochlear Implants. *Otology &*

1097   *Neurotology*, *40*(3), e191–e197.

1098   Peter, J. W., & Robert S, S. (2008). The effect of fundamental frequency on the intelligibility of

1099   speech with flattened intonation contours. *American Journal of Speech-Language Pathology*,

1100   *17*(4), 348–355.

1101   Piazza, E. A., Iordan, M. C., & Lew-Williams, C. (2017). Mothers Consistently Alter Their Unique

1102   Vocal Fingerprints When Communicating with Infants. *Current Biology*, *27*(20), 3162-

1103   3167.e3.

1104   Pisoni, D. B., Conway, C. M., Kronenberger, W., Horn, D. L., Karpicke, J., & Henning, S. (2007).

1105   Efficacy and effectiveness of cochlear implants in deaf children. *Research on Spoken*

1106   *Language Processing*, *28*(28), 3–46.

1107   Pisoni, D. B., Kronenberger, W. G., Harris, M. S., & Moberly, A. C. (2018). Three challenges for

1108   future research on cochlear implants. *World Journal of Otorhinolaryngology - Head and Neck*

1109   *Surgery*, *3*(4), 240–254.

1110   Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech

1111   reception in fluctuating maskers. *The Journal of the Acoustical Society of America*, *114*(1),

1112   446–454.

1113   Qin, M. K., & Oxenham, A. J. (2005). Effects of Envelope-Vocoder Processing on F0

1114   Discrimination. *Ear & Hearing*, *26*(5), 451–460.

1115   Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random

1116   effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425.

1117   Rohlfing, K. J., Fritsch, J., Wrede, B., & Jungmann, T. (2006). How can multimodal cues from

1118   child-directed interaction reduce learning complexity in robots? *Advanced Robotics*, *20*(10),

1119   1183–1199.

1120   Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral

1121   shifts of speech: Implications for cochlear implants. *The Journal of the Acoustical Society of*

1122   *America*, *106*(6), 3629–3636.

1123   Rottmann, N., & Zobrist, P. (2004). Tuned to the signal: The privileged status of speech for young

1124   infants. *Developmental Science*, *7*(3), 270–276.

1125   Rubinstein, J. T., Wilson, B. S., Finley, C. C., & Abbas, P. J. (1999). Pseudospontaneous activity:

1126   Stochastic independence of auditory nerve fibers with electrical stimulation. *Hearing*

1127   *Research*, *127*(1–2), 108–118.

1128   Rubinstein, Jay T. (2004). How cochlear implants encode speech. *Current Opinion in*

1129   *Otolaryngology and Head and Neck Surgery*, *12*(5), 444–448.

1130   Santos, J. F., Cosentino, S., Hazrati, O., Loizou, P. C., & Falk, T. H. (2013). Objective speech

1131   intelligibility measurement for cochlear implant users in complex listening environments.

1132   *Speech Communication*, *55*(7–8), 815–824.

1133   Sato, N., & Obuchi, Y. (2007). Emotion Recognition using Mel-Frequency Cepstral Coefficients.

1134   *Information and Media Technologies*, *2*(3), 835–848.

1135   Schachner, A., & Hannon, E. E. (2011). Infant-Directed Speech Drives Social Preferences in 5-

1136   Month-Old Infants. *Developmental Psychology*, *47*(1), 19–25.

1137   Shaneh, M., & Taheri, A. (2009). Voice command recognition system based on MFCC and VQ

1138   algorithms. *World Academy of Science, Engineering and Technology*, *57*, 534–538.

1139   Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition

1140   with primarily temporal cues. *Science*, *270*(5234), 303–304.

1141   Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*,

1142   *106*(2), 833–870.

1143   Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy

1144   talk? *Infancy*, *3*(3), 365–394.

1145   Singh, L., & Nestor, S. (2009). Influences of Infant-Directed Speech on Early Word Recognition.

1146   *Infancy*, *14*(6), 654–666.

1147   Slaney, M., & Mcroberts, G. (1998). BABY EARS: A recognition system for affective

1148   vocalizations. *IEEE International Conference on Acoustics, Speech and Signal Processing*,

1149   985–988.

1150   Snow, C. E. (1977). Mothers' speech research: From input to interaction. *Talking to Children:*

1151   *Language Input and Acquisition*, *3149*.

1152   Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1992). Stimulus variability and the perception

1153   of spoken words: Effects of variations in speaking rate and overall amplitude. *International*

1154   *Conference on Spoken Language Processing*, (October), 217–220.

1155   Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed

1156   speech on infant word recognition. *The Journal of the Acoustical Society of America*, *128*(1),

1157   389–400.

1158   Souza, P., Gehani, N., Wright, R., & McCloy, D. (2013). The advantage of knowing the talker.

1159   *Journal of the American Academy of Audiology*, *24*(8), 689–700.

1160   Spencer, P. E. (2004). Individual Differences in Language Performance after Cochlear

1161   Implantation at One to Three Years of Age: Child, Family, and Linguistic Factors. *Journal of*

1162   *Deaf Studies and Deaf Education*, *9*(4), 395–412.

1163   Spitzer, S. M., Liss, J. M., & Mattys, S. L. (2007). Acoustic cues to lexical segmentation: A study

1164   of resynthesized speech. *The Journal of the Acoustical Society of America*, *122*(6), 3678–

1165   3687.

1166   Spitzer, S. M., Liss, J., Spahr, T., Dorman, M., & Lansford, K. (2009). The use of fundamental

1167   frequency for lexical segmentation in listeners with cochlear implants. *The Journal of the*

1168   *Acoustical Society of America*, *125*(6), EL236–EL241.

1169   Sulpizio, S., Kuroda, K., Dalsasso, M., Asakawa, T., Bornstein, M. H., Doi, H., … Shinohara, K.

1170   (2018). Discriminating between Mothers' Infant- and Adult-Directed Speech: Cross-

1171   Linguistic Generalizability from Japanese to Italian and German. *Neurosci Res.*, *133*, 21-27.

1172   Svirsky, M. A. (2017). Cochlear implants and electronic hearing. *Physics Today*, *70*(8), 53–58.

1173   Szagun, G., & Schramm, S. A. (2016). Sources of variability in language development of children

1174   with cochlear implants: Age at implantation, parental language, and early features of

1175   children's language construction. *Journal of Child Language*, *43*(3), 505–536.

1176   Tamati, T. N., Janse, E., & Başkent, D. (2019). Perceptual Discrimination of Speaking Style under

1177   Cochlear Implant Simulation. *Ear and Hearing*, *40*(1), 63–76.

1178   Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., … Wartenburger,

1179   I. (2009). Sensitivity of Newborn Auditory Cortex to the Temporal Structure of Sounds.

1180   *Journal of Neuroscience*, *29*(47), 14726–14733.

1181   Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant directed speech facilitates word

1182   segmentation. *Infancy*, *7*(1), 53–71.

1183   Throckmorton, C. S., & Collins, L. M. (2002). The effect of channel interactions on speech

1184   recognition in cochlear implant subjects: Predictions from an acoustic model. *The Journal of*

1185   *the Acoustical Society of America*, *112*(1), 285–296.

1186   Trainor, L. J., Austin, C. M., & Desjardins, N. (2000). Is infant-directed speech prosody a result

1187   of the vocal expression of emotion? *Psychological Science*, *11*(3), 188–195.

1188   Walker-Andrews A. S. & Grolnick, W. (1983). Discrimination of vocal expression by young

1189   infants. *Infant Behavior and Development*, *6*, 491–498.

1190   Walker-Andrews, A. S., & Lennon, E. (1991). Infants' discrimination of vocal expressions:

1191   Contributions of auditory and visual information. *Infant Behavior and Development*, *14*(2),

1192   131–142.

1193   Wang, Y., Bergeson, T. R., & Houston, D. M. (2017). Infant-Directed Speech Enhances Attention

1194   to Speech in Deaf Infants With Cochlear Implants. *Journal of Speech Language and Hearing*

1195   *Research*, *60*(11), 3321.

1196   Wang, Y., Bergeson, T. R., & Houston, D. M. (2018). Preference for Infant-Directed Speech in

1197   Infants With Hearing Aids: Effects of Early Auditory Experience. *Journal of Speech,*

1198   *Language & Hearing Research*, *61*(9), 2431–2439.

1199   Wang, Y., Shafto, C. L., & Houston, D. M. (2018). Attention to speech and spoken language

1200   development in deaf children with cochlear implants: a 10-year longitudinal study.

1201   *Developmental Science*, (July 2017), e12677.

1202   Webb, A. R., Heller, H. T., Benson, C. B., & Lahav, A. (2015). Mother's voice and heartbeat

1203   sounds elicit auditory plasticity in the human brain before full gestation. *Proceedings of the*

1204   *National Academy of Sciences of the United States of America*, *112*(10), 3152–3157.

1205   Weisleder, A., & Fernald, A. (2013). Talking to Children Matters: Early Language Experience

1206   Strengthens Processing and Builds Vocabulary. *Psychological Science*, *24*(11), 2143–2152.

1207   Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed

1208   talk: a developmental study of attentional and affective responsiveness. *Canadian Journal of*

1209   *Psychology*, *43*(2), 230–246.

1210   Werker, J. F., Pegg, J., & McLeod, P. J. (1994). A cross-language investigation of infant preference

1211   for infant-directed communication. *Infant Behavior and Development*, *17*(3), 323–333.

1212   Winn, M. B., & Litovsky, R. Y. (2015). Using speech sounds to test functional spectral resolution

1213   in listeners with cochlear implants. *The Journal of the Acoustical Society of America*, *137*(3),

1214   1430–1442.

1215   Xiang, S., Nie, F., & Zhang, C. (2008). Learning a Mahalanobis distance metric for data clustering

1216   and classification. *Pattern Recognition*, *41*(12), 3600–3612.

1217   Zanto, T. P., Hennigan, K., Östberg, M., Clapp, W. C., & Gazzaley, A. (2013). Effect of speaking

1218   rate on recognition of synthetic and natural speech by normal-hearing and cochlear implant

1219   listeners. *Ear and Hearing*, *34*(3), 313.

1220   Zeng, F. G., Tang, Q., & Lu, T. (2014). Abnormal pitch perception produced by cochlear implant

1221   stimulation. *PLoS ONE*, *9*(2).

1222   Zheng, Y., Koehnke, J., & Besing, J. (2011). Effects of Noise and Reverberation on Virtual Sound

1223   Localization for Listeners With Bilateral Cochlear Implants. *American Journal of Audiology*,

1224   *32*(5), 569–572.

1225