

Running title: Voice quality through cochlear implant

1 **Effects of simulated cochlear-implant processing on voice quality distinction: Evidence from analysis**
2 **of disordered voices**^{a)}

3

4 Meisam K. Arjmandi^{1,b)}, Hamzeh Ghasemzadeh^{1,c)}, Laura C. Dilley¹

5

6 ¹ Department of Communicative Sciences and Disorders, Michigan State University, 1026 Red
7 Cedar Road, East Lansing, Michigan 48824, USA

8

9

10 ^{a)} Portions of this work were presented at the 177th Meeting of the Acoustical Society of America,
11 Louisville, Kentucky, USA, May 2019.

12

13 ^{b)} Current affiliation: Department of Otolaryngology – Head and Neck Surgery, Massachusetts Eye
14 and Ear and Harvard Medical School, Boston, Massachusetts, USA. Email:
15 meisam_arjmandi@meei.harvard.edu

16

17 ^{c)} Also affiliated with: The Department of Computational Mathematics Science and Engineering,
18 Michigan State University, East Lansing, Michigan, USA.

19

20

21

Running title: Voice quality through cochlear implant

22 **ABSTRACT**

23 The ability to discern variations in voice quality from speech is important for effective talker
24 identification and robust speech processing; yet, little is known about how faithfully acoustic
25 information relevant to variations in talkers' voice quality is transmitted through a cochlear implant
26 (CI) device. The present study analyzed unprocessed and CI-simulated versions of sustained /a/
27 vowel sounds from two groups of individuals with normal and disordered voice qualities in order
28 to explore the effects of CI speech processing on acoustic information relevant for the distinction
29 of voice quality. The CI-simulated voices were created by processing the vowel sounds along with
30 4-, 8-, 12-, 16-, 22-, and 32-channel noise-vocoders. The variations in voice quality for each voice
31 sound was characterized by calculating mel-frequency cepstral coefficients (MFCCs). The effects
32 of simulated CI speech processing on the acoustic distinctiveness between normal and disordered
33 voices were then measured by calculating the Mahalanobis distance (MD) metric, as well as
34 accuracy of support vector machines (SVMs) applied to MFCC features. The results showed that
35 CI speech processing, as simulated by noise vocoding, is highly detrimental to the acoustic
36 information involved in conveying voice quality distinctions. This supports the view that listeners
37 with CIs will likely experience difficulties in perceiving voice quality variations due to the reduced
38 spectral resolution, shedding light on challenges listeners with CIs may face for effective
39 recognition and processing of talkers' voices.

40 *Keywords:* Voice quality, cochlear implant, acoustic distinctiveness

41

42

43

Running title: Voice quality through cochlear implant

44 **1. INTRODUCTION**

45 The spectro-temporal features in speech contains rich acoustic information from which listeners
46 can learn and retrieve a variety of linguistic and indexical cues that are important for robust speech
47 processing. Voice quality is an important aspect of speech that reflects the configuration and
48 function of individual talkers' vocal apparatus, contributing important information to speech
49 understanding (Abercrombie, 1967; Podesva, 2007). For example, voice quality variations may
50 provide perceptually salient grammatical and phonological cues for language comprehension
51 (Cameron, 2001; Dicano, 2009; Dolar, 2006; Garellek & Keating, 2011; Gordon, 2001; Gordon
52 & Ladefoged, 2001; Henton, 1986; Ogden, 2001), not to mention indexical information such as
53 gender, age, and affective state (Abberton & Fourcin, 1978; Laver, 1968; Scott & McGettigan,
54 2015). Voice quality also contributes to speech understanding through constructing stance in
55 communicative interaction (Aubergé & Cathiard, 2003; Guzman, Correa, Muñoz, & Mayerhoff,
56 2013; Podesva, 2007; Sicoli, 2010; Tsai et al., 2010; Zimman, 2012).

57 While listeners with normal hearing (NH) have access to voice quality-related indexical
58 and sociolinguistic information, little is known about how listeners with cochlear implants (CIs)
59 may be disadvantaged in accessing this information due to the reduced spectral resolution.
60 Specifically, it is not clear how the limited spectral resolution in CIs may impact the faithful
61 transmission of acoustic information relevant to the distinction of voice quality. The present work
62 studied voices produced by individuals with normal and disordered voice qualities to examine
63 effects of simulated CI processing on the voice quality distinction.

64 Acoustic information relevant to voice quality signals a range of attributes, thereby
65 facilitating robust speech perception. This acoustic information allows listeners to associate the
66 variations in talkers' voice into perceptual attributes of voice quality, such as breathiness,

Running title: Voice quality through cochlear implant

67 roughness, and strain (Childers & Lee, 1991; Eskenazi, Childers, & Hicks, 1990; Klatt & Klatt,
68 1990), which can provide acoustic cues for phonemic contrasts (Dicanio, 2009; Garellek &
69 Keating, 2011; Gordon, 2001; Gordon & Ladefoged, 2001). Voicing behaviors like creaky voice
70 signal phrase-final position (Henton, 1986; Ogden, 2001) and convey linguistic information at
71 segmental and prosodic levels (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996; Dilley, Arjmandi,
72 Ireland, Heffner, & Pitt, 2016; Redi & Shattuck-Hufnagel, 2001). These findings support the
73 premise that access to the acoustic information about voice quality facilitates robust perception of
74 speech.

75 Variations in voice quality also assist listeners in identification of individual talkers'
76 voices. Multiple traits, such as a talker's gender, age, dialect, and social group can be often readily
77 incorporated by listeners with NH to recognize talkers' voice. Voice quality can signal gender
78 (e.g., Gussenhoven, 2004; Ohala, 1983; Puts, Hodges, Cárdenas, & Gaulin, 2007), race (Alim,
79 2004; Irwin, 1977; Moisik, 2013; Thomas & Reaser, 2004), dialect (Purnell, Idsardi, & Baugh,
80 1999), and/or social group (Esling, 1978; Sicoli, 2007; Stuart-Smith, 1999)- all of which convey
81 information about talker identity. For instance, studies on African American English demonstrated
82 a connection between talkers' voice quality and their dialect and race (Arjmandi, Dilley, &
83 Wagner, 2018; Irwin, 1977; Thomas & Reaser, 2004). Non-modal voice qualities were found to
84 be frequently used by African American talkers, leading to a relatively harsh voice quality (Alim,
85 2004; Britt, 2011). Therefore, perception of acoustic information relevant to voice quality facilitates
86 the identification of talkers as an important step in talker normalization toward robust, talker-
87 independent speech perception (Johnson, 2005; Kleinschmidt & Jaeger, 2015; Smith & Patterson,
88 2005).

Running title: Voice quality through cochlear implant

89 Voice quality contributes to learning about several other types of information about talkers
90 such as their physical, psychological and mental health (e.g., voice disorders, anxiety level, mood;
91 e.g., Eskenazi et al., 1990; Kreiman, Vanlancker-Sidtis, & Gerratt, 2005). Vocal fold disorders
92 (e.g., vocal fold polyps, nodules, etc.) represent an instance where voice quality is abnormally
93 altered (Arjmandi & Pooyan, 2012; Arjmandi, Pooyan, Mikaili, Vali, & Moqarehzadeh, 2011;
94 Ghasemzadeh & Arjmandi, 2019; Umaphathy, Krishnan, Parsa, & Jamieson, 2005). These
95 abnormalities (physiological, neurological, and/or functional) are associated with the variations in
96 talkers' voice quality and are recognized as disordered voice quality, as compared to normal voice
97 quality (Blood, Mahan, & Hyman, 1979; Oates, 2009). Abnormalities associated with voice
98 disorders lead to perturbations of vocal fold vibratory patterns in some or all glottal cycles, which
99 can impact voice spectra in multiple frequency regions (Eskenazi et al., 1990; Hammarberg,
100 Fritzell, Gaufin, Sundberg, & Wedin, 1980; Hartl, Hans, Vaissière, Riquet, & Brasnu, 2001; Krom,
101 1995; Naranjo, Lara, Rodríguez, & García, 1994; Wolfe, Cornell, & Palmer, 1991). Examples of
102 acoustic manifestations of disordered voices include changes in the low and high frequency energy
103 profiles of talkers' voices (e.g., increased spectral energy within the mid- and high-frequency
104 regions) (Arjmandi & Pooyan, 2012; Ball & Code, 2008; Ghasemzadeh & Arjmandi, 2019;
105 Kitzing & Åkerlund, 1993; Naranjo, Lara, Rodríguez, & García, 1994), as well as decreased
106 steepness of the low-frequency spectral slope (Arjmandi & Pooyan, 2012; Behroozmand &
107 Almasganj, 2007; Hartl et al., 2001; Klatt & Klatt, 1990; Guus de Krom, 1995). Taken together,
108 these studies show that disordered voice quality impacts spectral characteristics of voice in
109 multiple sub-bands in ways that are distinctive from spectral profile of normal voice.

110 Unlike listeners with NH who have access to acoustic information signaling voice quality,
111 access to spectral profiles of voice through CIs is limited by factors such as the number of spectral

Running title: Voice quality through cochlear implant

112 channels (cf. number of electrodes) and mismatches between the frequencies in acoustic content
113 and frequencies associated with auditory nerves at the place of electrode stimulation on the cochlea
114 (e.g., Fu, Chinchilla, Nogaki, & Galvin, 2005; Svirsky, 2017). As a result, listeners with CIs
115 receive partial, degraded representations of acoustic information in talkers' voices. An important
116 portion of voice quality-related acoustic information is associated with the change in vocal folds
117 that appears in mid- and high-frequency regions of the voice spectrum (Arjmandi & Pooyan, 2012;
118 de Krom, 1993; Fukazawa, el-Assuooty, & Honjo, 1988; Hartl et al., 2001; O'Leidhin & Murphy,
119 2005; Yumoto, Gould, & Baer, 1982). This acoustic information includes, but is not limited to,
120 elevated energy levels in these frequency regions and deformation of formant frequencies. These
121 spectral regions are either completely filtered or partially transmitted through CIs (Goupell, 2015;
122 Svirsky, 2017). Low-frequency harmonics are also affected through CIs. Acoustic cues such as
123 fundamental frequency (F_0) and $H1-H2$ (first and second harmonic amplitude differences) and $H1-$
124 $A1$ (first harmonic and first formant amplitude differences) are low-frequency harmonic
125 components that moderately reflect changes in the quality of talkers' voice (e.g., breathy, strained,
126 rough, etc.), where these can potentially contribute to recognition of talkers' gender and individual
127 identity (Fu, Chinchilla, & Galvin, 2004; Gelfer & Bennett, 2013). We currently have limited
128 evidence about how the reduced spectral resolution in CIs may degrade acoustic information
129 relevant to detecting voice quality variation.

130 In the present study, we used voice samples from two groups of talkers, one with normal
131 voice quality and one with disordered voice quality, in order to examine effects of simulated CI
132 processing of such voices on acoustic information available for distinguishing between these two
133 classes of voice qualities. Noise-vocoding was used to simulate the limited spectral resolution of
134 CIs in representation of the acoustic information through varying the number of spectral channels

Running title: Voice quality through cochlear implant

135 in the noise vocoder, thereby simulating varying degrees of spectral resolution while keeping other
136 parameters of the simulation constant (e.g., the slopes of the synthesis bands). Acoustic simulation
137 of CI speech is a useful and precious approach to simulate what acoustic information listeners with
138 CIs have access to when hearing speech (Do, 2012; Do, Pastor, & Goalic, 2012; Santos, Cosentino,
139 Hazrati, Loizou, & Falk, 2013). In our analysis, we first investigated the variations in the
140 magnitude spectrum of normal and disordered voices to understand how CIs may impact reception
141 of information pertinent to distinguishing between these two classes of voice qualities. We further
142 quantified the acoustic distance between normal and disordered voices as a function of the number
143 of spectral channels, in order to examine effect of CI-related spectral resolution (number of
144 channels in the noise vocoder) on availability of acoustic information related to voice quality
145 distinctions.

146 **2. MATERIALS AND METHODS**

147 **2.1. Voice samples**

148 The voice samples in this study were sustained /a/ vowel sounds from the voice disorders database
149 model 4337, version 1.03 (Kay Elemetrics Corporation, Lincoln Park, NJ), developed by
150 Massachusetts Eye and Ear Infirmary (MEEI), Voice and Speech Lab. Two groups of participants
151 who were diagnosed as having either normal or disordered voices were asked to sustain the vowel
152 /a/. Participants' voices were recorded at a sampling frequency of 44.1 kHz with 16-bit resolution.
153 All the analyzed voice segments were 1-second long, extracted from the middle of each excerpt to
154 deal with the length difference between normal and disordered voice samples, as well as the
155 transient patterns during the onset and offset of the phonations. The vowel sounds from 293
156 individuals were analyzed. 53 talkers had normal voice quality (21 males) and the remaining 240
157 talkers (96 males) were diagnosed with one or multiple voice disorders (e.g., vocal fold nodules,

Running title: Voice quality through cochlear implant

158 vocal fold paralysis, vocal fold cysts) that had resulted from abnormal physiological, neurological,
159 and/or functional changes affecting vocal fold health.

160 **2.2. Creation of noise-vocoded voice samples**

161 CI-simulated versions of unprocessed voice samples were created using a noise-excited envelope
162 vocoder, the *AngelSim*TM Cochlear Implant and Hearing Loss Simulator (Fu, 2019; Emily
163 Shannon Fu Foundation, www.tigerspeech.com). The CI-simulated vocoding simulation process
164 used in prior studies was followed to create CI-simulated voices (Shannon, Zeng, Kamath,
165 Wygonski, & Ekelid, 1995). This process involved dividing each voice spectrum into a variable
166 number of logarithmically-spaced frequency bands between absolute lower and higher frequencies
167 of 200 Hz and 7000 Hz (24 dB/octave analysis filter slopes), corresponding to the frequency-place
168 map simulated by the Greenwood function (Greenwood, 1990). These frequency limits
169 approximate the corner frequencies in Cochlear Nucleus speech processors (Crew & Galvin, 2012;
170 Winn & Litovsky, 2015). The amplitude envelope of each signal, obtained from filtering the voice
171 spectrum under each sub-band, was captured using half-wave rectification and a low-pass filter
172 with a cut-off frequency of 160 Hz and filter slope of 24 dB/oct; this simulated the performance
173 of the average CI listener in envelope discrimination (Chatterjee & Oberzut, 2011; Chatterjee &
174 Peng, 2008). The extracted amplitude envelopes were then used to modulate band-pass filtered
175 white-noise carrier signals, which were created using a filter identical to that implemented for the
176 analysis filter. The final noise-vocoded version of each voice stimulus was created by summing
177 amplitude-modulated signals. This process replaces fine spectro-temporal structures in voice
178 signal with noise while preserving most of the coarse-grained temporal structures. The quality of
179 CI-simulated voice depends on the number of spectral channels in the vocoder. The noise-excited
180 envelope vocoder was used in *AngelSim* software to process unprocessed voices and create their

Running title: Voice quality through cochlear implant

181 noise-vocoded versions with 4-, 8-, 12-, 16-, 22-, and 32-channel. Therefore, the simulated
182 cochlear-implant voices were created at six levels of spectral degradation (4-, 8-, 12-, 16-, 22-, and
183 32-channels). The choice of the number of spectral channels was made to simulate a wide range
184 of spectral degradation and their corresponding perceived difficulty in speech processed through
185 CIs (Shannon, Fu, & Galvin, 2004), as well as to cover the current set-up of between 12-24 active
186 channels in cochlear implant devices. Considering the assumption made about the relationship
187 between electrical spread in the cochlea and acoustical filter slope (Bingabr, Espinoza-Varas, &
188 Loizou, 2008; Oxenham & Kreft, 2014), the selected filter slope of 24 dB/Oct is in the highest
189 range of steepness provided by current CI technology (filter slope varies between 8 and 24 dB/oct),
190 corresponding to the minimum channel interaction available in the current CI devices.

191 **2.3. Analysis of voice spectra**

192 We first investigated the average spectra of the two groups of voices (unprocessed normal voices
193 vs. unprocessed disordered voices) to understand how the noise-vocoder and the number of
194 spectral channels would likely affect the distinctive features of these two classes of voice qualities
195 across frequency regions. The average spectrum of voice signals derived over the spectrum of all
196 voice samples from each class of voice quality (i.e., normal or disordered) were estimated using
197 12th-order linear predictive coding (LPC) (Rabiner & Schafer, 1978). Variations in characteristics
198 of average voice spectra were investigated under seven levels of spectral degradation
199 (unprocessed, 32, 22, 16, 12, 8, and 4-channels noise-vocoder) to simulate how differing fidelity
200 of CI processing may affect the acoustic information available to signal the distinction between
201 normal vs. disordered voice qualities.

Running title: Voice quality through cochlear implant

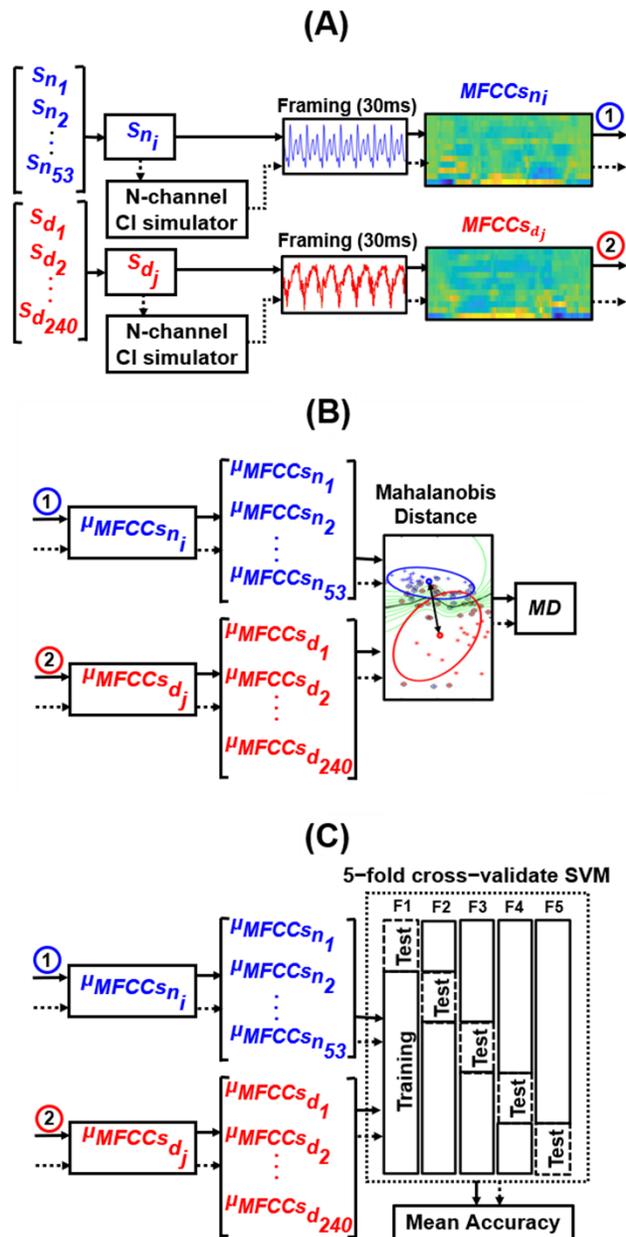
202 **2.4. Using MFCC features to characterize acoustic information**

203 Mel-frequency cepstral coefficients (MFCCs) were used to characterize variations in spectral
204 profile of unprocessed normal vs. disordered voices, as well as their CI-simulated counterparts.
205 MFCC features approximate the filtering structure and frequency resolution of the human auditory
206 system (Fant, 1973; Hunt, Lennig, & Mermelestein, 1980; Davis and Mermelstein, 1980; Shaneh
207 & Taheri, 2009; Stevens, Volkman, & Newman, 1937). MFCC features have been shown in prior
208 studies to characterize well canonical features distinguishing normal and disordered voices (Ali,
209 Alsulaiman, Muhammad, Elamvazuthi, & Mesallam, 2013; Dibazar, Narayanad, & Berger, 2002;
210 Firdos & Umarani, 2016; Godino-Llorente, Gomez-Vilda, & Blanco-Velasco, 2006; Panek,
211 Skalski, Gajda, & Tadeusiewicz, 2015). Furthermore, features that are developed based on speech
212 production behaviors such as F_0 and the first and second harmonics ($H1$ and $H2$) are either heavily
213 degraded or are absent in the CI-simulated voices. Using MFCC features also permitted us to deal
214 robustly with these methodological challenges on the way to characterizing changes in spectral
215 properties of talkers' voice spectra across normal vs. disordered voice quality.

216 Fig. 1A shows the schematic of the approach used for calculating MFCCs for samples of
217 voices with normal and disordered qualities. To calculate MFCCs, each /a/ vowel sound (S_{n_i} or
218 S_{d_j} in Fig. 1A) was first segmented into frames of 30 ms with a frame shift of 15 ms. Here, i and j
219 indicate the index of voice stimuli for normal and disordered voices, respectively ($i = \{1,2,3,$
220 $\dots,53\}$ and $j = \{1,2,3, \dots,240\}$). A Hamming window was then applied to each frame to decrease
221 the effect of sidelobes for better frequency-selective analysis (Rabiner & Schafer, 1978). The
222 power spectrum of each frame was calculated based on Fast Fourier transform (FFT) analysis.
223 Then, 32 mel-filterbanks were generated and applied to voice power spectra. MFCCs were derived
224 by calculating the DCT of the logarithm of all filterbank energies (Rabiner & Schafer, 1978).

Running title: Voice quality through cochlear implant

225 Eventually, the first twelve components were preserved as MFCC features for each frame of a
 226 voice signal ($MFCCs_{n_i}$ or $MFCCs_{d_j}$ in Fig. 1). Each voice signal was represented by an MFCC
 227 matrix of size $F \times 12$, where F indicates the number of frames in a voice signal. The same procedure
 228 was performed to calculate MFCC matrices for CI-simulated versions of the unprocessed normal
 229 and disordered voices, as indicated by dashed lines in Fig. 1A.



Running title: Voice quality through cochlear implant

231 **Fig. 1.** (Color online) Schematic diagram of methods used in the current study to (A) characterize
232 acoustic properties of normal and disordered voice stimuli based on MFCCs features; (B) evaluate
233 the acoustic distance between voices with normal and disordered qualities based on calculating
234 *Mahalanobis distance* between MFCCs; and (C) evaluate the acoustic distance between normal
235 vs. disordered voice qualities based on the classification accuracy derived from applying 5-fold
236 validation to a support vector machine (SVM) classifier. The dashed lines refer to the process for
237 creating and analyzing the CI-simulated versions of the voice stimuli. N in the “*N-channel*
238 *Simulator*” block stands for the number of spectral channels in the CI-simulated noise vocoder.
239 Components, blocks, and lines with blue (dark gray) color show the paths for processing voices
240 with normal quality, whereas the components with red (light gray) color show the paths for
241 processing voices with disordered quality.

242 **2.5. Acoustic distance quantification using Mahalanobis distance**

243 To examine the acoustic distance between two classes of normal and disordered voices as a
244 function of the level of spectral degradation (imposed by the variable number of spectral channels
245 in the CI-simulated voices), we calculated Mahalanobis distance (MD) metrics on MFCC features.
246 MD is a distance measure, which calculates the distance between two or more classes at a
247 multidimensional feature space (Arjmandi et al., 2018; Maesschalck & Massart, 2000; Masnan et
248 al., 2015; Xiang, Nie, & Zhang, 2008). MD is analogous to a multidimensional d' , as used in signal
249 detection theory (Macmillan & Creelman, 2004). This multivariate statistical approach uses two
250 feature matrices (or vectors) from two separate classes to evaluate the extent to which the two
251 classes can be distinguished after sphering the distance matrix between the two classes using the
252 average covariance matrix of the per-class centered data (Maesschalck & Massart, 2000; Masnan
253 et al., 2015). Hence, a relatively greater MD value for a condition (e.g., unprocessed) means a

Running title: Voice quality through cochlear implant

254 relatively larger distance between the two classes of normal and disordered voice qualities with a
255 relatively lower between-class overlap for that condition compared to other conditions (e.g., 32-
256 channel noise-vocoded voices). Therefore, a relatively larger distance indicates the presence of
257 more discriminative acoustic information relevant to the distinction between two classes of voice
258 qualities for a condition compared to other conditions.

259 As shown in Fig. 1B, the acoustic properties of each voice stimulus (S_{n_i} or S_{d_j}) were
260 characterized by a single, time-averaged MFCC vector ($\mu_{MFCCs_{n_i}}$ or $\mu_{MFCCs_{d_j}}$), derived by averaging
261 MFCCs across frames of each voice stimulus. These time-averaged MFCC features have been
262 shown to successfully represent the unique spectral characteristics of a speech sound (Mckinney
263 & Breebaart, 2003; Davis and Mermelstein, 1980; Terasawa, Slaney, & Berger, 2005). Therefore,
264 53 normal voices were presented by fifty-three 12-dimensional average MFCC vectors,
265 constructing a 53×12 feature matrix. Likewise, 240 disordered voices were characterized by 240
266 12-dimensional average MFCC vectors, leading to a 240×12 feature matrix. The acoustic distance
267 between normal and disordered voices was then measured by calculating the MD between these
268 two matrices of MFCCs (Fig. 1B). The same procedure was followed for simulated versions of the
269 unprocessed normal and the unprocessed disordered voice stimuli (dashed lines in Fig. 1B),
270 leading to seven values of MD corresponding to seven levels of spectral degradation from the
271 unprocessed to 4-channel noise-vocoded voices. The calculated MDs of seven levels of spectral
272 degradation were comparatively examined to identify the extent to which the spectral information
273 involved in distinguishing voices with normal and disordered qualities were affected by CI noise
274 vocoding, as well as the number of channels in the CI noise-vocoding process.

Running title: Voice quality through cochlear implant

275 **2.6. Acoustic distance quantification using support vector machines (SVMs)**

276 We further used SVMs classifier to measure the acoustic distance between two classes of voice
277 qualities. SVMs have been frequently used as a successful classification method for various
278 classification purposes, including classification of normal and disordered voice qualities (Akbari
279 & Arjmandi, 2015; Arjmandi et al., 2011; Ghasemzadeh, Khass, Arjmandi, & Pooyan, 2015;
280 Ghasemzadeh & Arjmandi, 2019; Arjmandi, Pooyan, Mohammadnejad, & Vali, 2010; Umapathy,
281 Rachel, & Thulasi, 2018). One application of SVM classifiers is to evaluate features that
282 maximally distinguish between two classes, where the classification accuracy of SVM classifiers
283 is a criterion for feature evaluation (Heijden, Ferdinand, Ridder, & Tax, 2005). We took advantage
284 of this property of SVMs to examine effects of CI speech processing and number of channels on
285 the acoustic distinctiveness between normal and disordered voices, providing a complementary
286 analysis to MD. Higher classification accuracy between two classes indicates that there was more
287 distinctive acoustic information with respect to class separation.

288 As illustrated in Figure 1C, a 5-fold cross-validation analysis was performed to train and
289 then test an SVM classifier on its classification accuracy in distinguishing between normal and
290 disordered voice qualities at seven levels of spectral degradation (unprocessed, 32-, 22-, 16-, 12-,
291 8-, and 4-channel). Two feature matrices of 53x12 and 240x12 MFCC features from normal and
292 disordered classes were entered into the SVM classifier for training and testing phases, as executed
293 through the 5-fold cross-validation procedure (Kohavi, 1995; Reilly, Moran, & Lacy, 2004). The
294 output of the SVM classifier was the mean SVM classification accuracy over classification
295 accuracies, obtained from five repetitions of cross-validation. The average classification
296 accuracies at six levels of spectral degradation were examined with reference to that of the
297 unprocessed condition (as baseline performance) to understand the extent of degradation imposed

Running title: Voice quality through cochlear implant

298 by CI noise vocoding on acoustic information involved in voice quality distinction. The radial
299 basis function (RBF) kernel was used in SVM classifier. The parameters of RBF kernel and the
300 regularization parameter (ξ) of SVM were set to their default values in *Matlab*.

301 **3. RESULTS**

302 In three steps, we examined the effects of simulated cochlear-implant processing on spectral
303 information relevant to voice quality distinction. We first examined spectra of normal and
304 disordered voice samples under the seven spectral degradation conditions. Fig. 2 shows the average
305 magnitude spectra of the two groups of voices with normal (blue or dark gray) and disordered (red
306 or light gray) qualities across all voice samples. The standard deviations of magnitude spectra are
307 also shown as a shaded area across the average lines. The average voice spectra are selectively
308 shown for unprocessed (panel A) and simulated cochlear implant voices with 16- (panel B) and 4-
309 channel (panel C) spectral resolution. These spectra are computed by averaging individual
310 frequency spectrum over all voice samples from a class of voice quality. Overall, these plots
311 demonstrate the detrimental effect of CI processing on spectral information that could signal
312 differences in talkers' voice quality. The patterns of variation in average spectral energy of
313 disordered voices compared to normal voices at different frequency sub-bands reflect voice quality
314 variations, caused by a wide range of physiological, neurological, and/or functional voice
315 disorders.

316 The difference between the average magnitude spectrum for normal voices (the blue or
317 dark gray line) and that of disordered voices (the red or light gray line) in Fig. 1A reveals
318 distinctive patterns of spectral energy within low-, mid-, and high-frequency regions. For the
319 unprocessed condition, a peak in the frequency regions between 1 and 2 kHz distinguishes average
320 spectrum of voices with normal quality from the average spectrum of disordered voices. These

Running title: Voice quality through cochlear implant

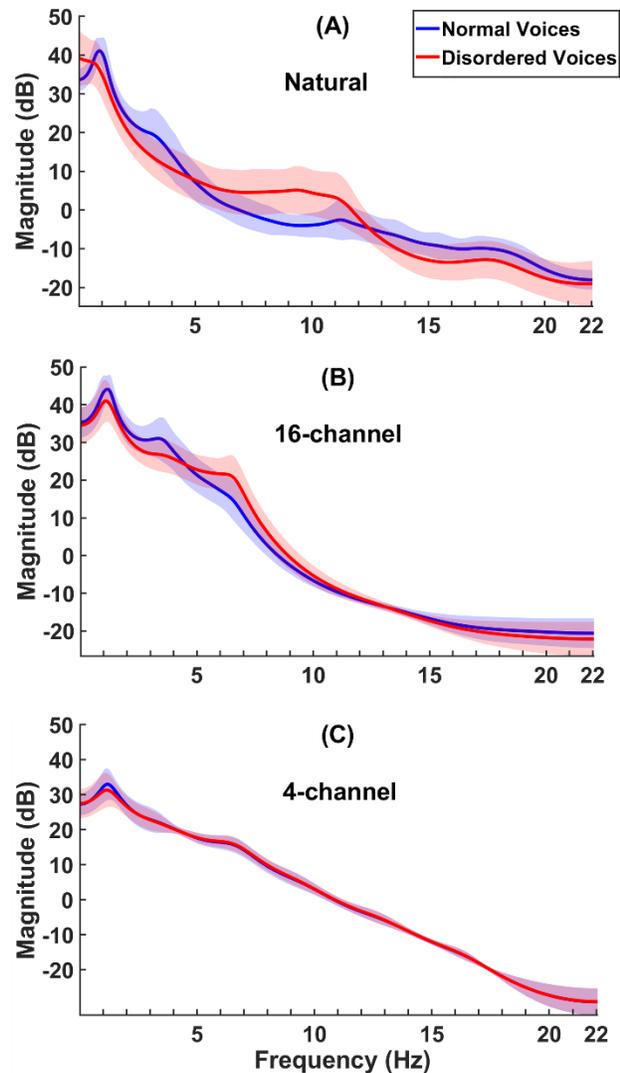
321 degraded low-frequency patterns in the disordered voice spectrum are considered to be signs of
322 partial closure of the vocal folds (Hartl et al., 2001; Kitzing & Åkerlund, 1993; de Krom, 1995).
323 These differences in the spectral level in frequency bands covering the first formant (*e.g.*,
324 breakdown in formant structure) may be associated with breathy voice quality reported in some
325 voice disorders (Kitzing & Åkerlund, 1993; Krom, 1995; Rontal, Rontal, & Rolnick, 1975;
326 Thomas, 2008; Wolfe & Bacon, 1971). The relative reduction in low-numbered harmonic
327 components is also visible in low-frequency regions, which is due probably to irregular vibratory
328 patterns of vocal folds and hoarse voices in disordered voice quality (Fex, Fex, Shiromoto, &
329 Hirano, 1994; Roy & Leeper, 1993; Thomas, 2008; Yanagihara, 1967). A relatively higher level
330 of energy in mid-frequency bands (~4.7 KHz-12.4 kHz) is evident in the average spectrum of
331 samples with disordered voice quality compared to those from the normal group, which potentially
332 signals the presence of high degree of breathiness in disordered voice samples and an increase in
333 the level of the turbulence noise components in the vocal excitation signal (Askenfelt &
334 Hammarberg, 1986; Fukazawa et al., 1988; Hanson, 1997; O'Leidhin & Murphy, 2005). The
335 presence of a wide-band noise in this frequency region (*i.e.*, between ~ 5 kHz and ~12 kHz) in the
336 average spectrum of disordered voices may also be attributed to a rough voice quality in disordered
337 voices (de Krom, 1995).

338 Comparing the average spectrum of two groups of voices with normal and disordered
339 qualities among three levels of spectral degradation in Fig. 2 (unprocessed in panel (A), 16-channel
340 in panel (B), and 4-channel in panel (C)) demonstrates that CI noise vocoding substantially
341 degrades acoustic information involved in voice quality distinction. In general, the noise vocoding
342 process- simulating CI speech processing - caused a major loss of acoustic information that might
343 be used by listeners to distinguish voice quality at low-, mid-, and high-frequency ranges of voice

Running title: Voice quality through cochlear implant

344 spectra. The detrimental effect of noise-vocoding increased as the number of channels decreased;
345 this pattern was observed to the extent that spectra of two classes of voice qualities become almost
346 visually indistinguishable at 4-channel CI-simulated voices. This reduction in acoustic distance
347 between groups of voices with normal and disordered qualities highlights the detrimental effect of
348 low spectral resolution in CI speech processing on discarding voice quality-related acoustic
349 information. As the spectra of normal and disordered voices for 16- and 4-channel noise-vocoded
350 voices suggest, a large portion of spectral information at low frequencies is expected to be
351 discarded due to CI processing. This spectral region is particularly important for the perception of
352 voice quality variations as it is where the low-numbered harmonics are located. CI listeners do not
353 have access to the frequency components in these low-frequency regions (Bernstein & Oxenham,
354 2003; Smurzynski, 1990), which may negatively impact their ability to perceive variations in voice
355 quality. As the number of spectral channels decreases, the noise-like distinctive patterns in the
356 mid-frequency range (between 5 and 12 kHz) disappear, suggesting that listeners with CIs do not
357 have access to acoustic cues relevant to variations in voice quality within these spectral regions.

Running title: Voice quality through cochlear implant

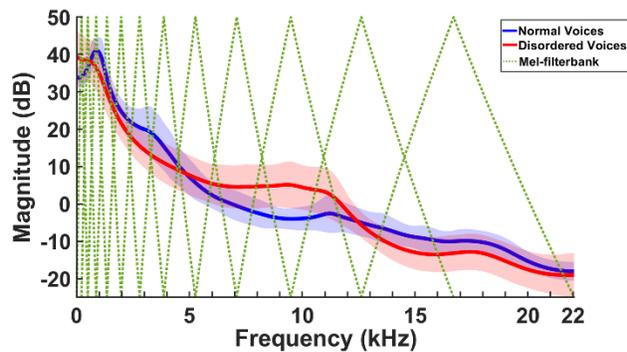


358

359 **Fig. 2.** (Color online) Average magnitude spectra of voice stimuli with normal (blue or dark gray
360 lines) and disordered (red or light gray lines) qualities for (A) unprocessed voices, (B) simulated
361 cochlear-implant voices with 16-channel, and (C) 4-channel in the CI noise-vocoder. The standard
362 deviations of the magnitude spectra are also shown as the blue (dark gray) shaded area across the
363 average blue line (dark gray) for the group of normal voices and as the red (light gray) shaded area
364 across the average red line (light gray) for the group of disordered voices.

Running title: Voice quality through cochlear implant

365 We further quantified the acoustic distance between voices with normal and disordered
366 qualities at seven levels of spectral degradation to examine the effects of simulated CI processing
367 on acoustic information distinctive of voice quality. Fig. 3 illustrates an example of the filterbank
368 of mel-spaced triangular filters through which voice spectrum of each normal or disordered voice
369 was passed to characterize the variations in their spectral energy at different frequency sub-bands.



370
371 **Fig. 3.** (Color online) The filterbank of mel-spaced triangular filters (green or dark gray dotted
372 lines) superimposed on average magnitude spectra of voices with normal (blue or dark gray line)
373 and disordered (red or light gray line) qualities. In this example, mel-filterbak contains 12 filters,
374 which starts at 0 Hz and expands to 22.05 kHz, corresponding to half of the sampling frequency
375 (44.1 kHz). The actual filterbank in the calculation of MFCC features was comprised of 32 mel
376 filters.

377 Fig. 4 shows the calculated MD between MFCCs of normal and disordered voices as a
378 function of different levels of spectral degradation, corresponding to change in the number of
379 spectral channels in CI-simulated voices. This figure illustrates that the acoustic distance between
380 voices with normal and disordered qualities decreased largely as a function of noise vocoding,
381 supporting the contention that CI-related processing is detrimental to the reception of voice
382 quality-related information. On average, there was an approximately 33% decline in MD due to

Running title: Voice quality through cochlear implant

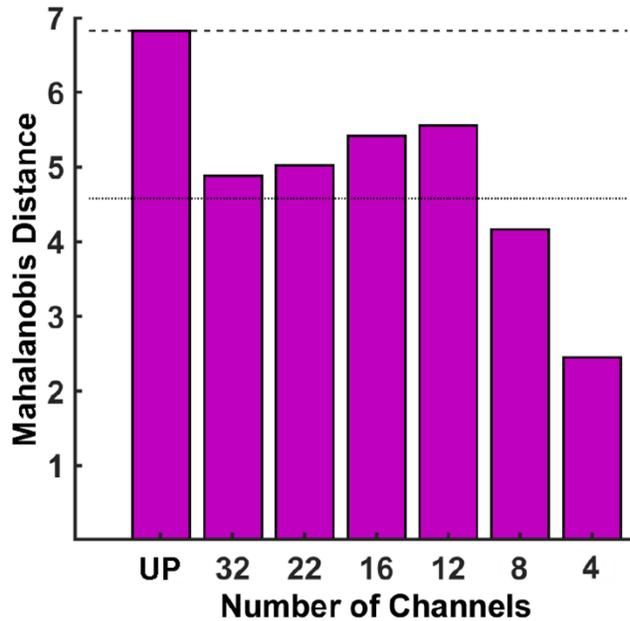
383 noise-vocoding process when comparing the MD at the baseline (i.e., unprocessed condition) (the
384 top dashed line in Fig. 4) with the average MD derived across six levels of spectral degradation
385 (the middle dotted line in Fig. 4). This large decline in acoustic distance between normal and
386 disordered voice qualities suggests that the CI potentially discards an important portion of acoustic
387 information responsible for signaling variation in talkers' voice qualities.

388 An interesting pattern was the increase in MD as the number of spectral channels decreased
389 from 32 channels to 22, 16, and 12 channels. Our visual investigation of normal and disordered
390 voice spectra showed that noise vocoding interestingly resulted in more distinctive patterns of
391 spectral energy between normal and disordered voice qualities as the number of spectral channels
392 changed from 32 to 12 channels. This pattern was particularly noticeable in low frequency regions
393 where mel-filterbank is more sensitive to variations in voice spectrum because of its narrower
394 filters, as compared with high-frequency regions with wider filters (see Fig. 3). This unexpected
395 pattern suggests that increasing spectral channels in noise-vocoder does not necessarily mitigate
396 the information loss relevant to voice quality distinction. The range of decrease in acoustic distance
397 (i.e., MD) due to noise vocoding relative to acoustic distance in the unprocessed condition was
398 between ~64% for 4-channel CI-simulated voices and ~19% for 12-channel CI-simulated voice,
399 which was still relatively high.

400

401

Running title: Voice quality through cochlear implant



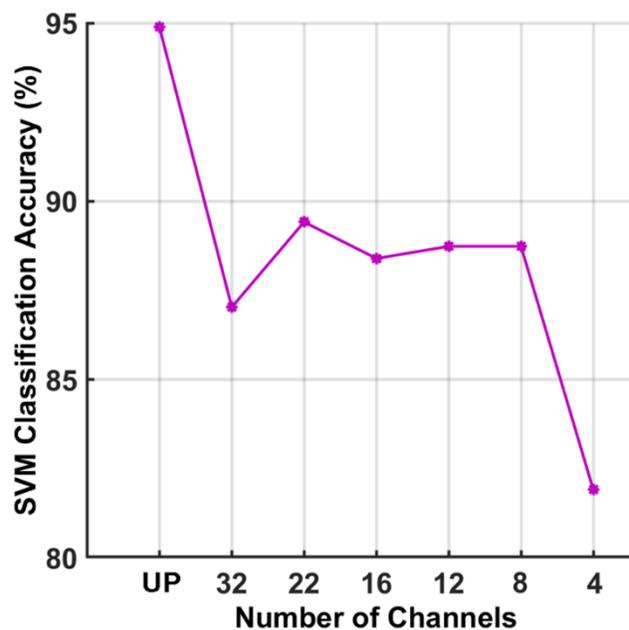
402

403 **Fig. 4.** (Color online) *Mahalanobis distance* between two groups of voices with normal and
404 disordered qualities as a function of spectral degradation in CI-simulated voices (i.e., number of
405 spectral channels in noise-vocoder). The top, horizontal dashed line shows the MD derived from
406 unprocessed voices and the dotted line in the middle shows the average of MDs across six levels
407 of spectral degradation. The unprocessed condition is labeled as “UP”.

408 Fig. 5 shows the results of normal-vs-disordered SVM classification accuracy for seven
409 levels of spectral resolution from the unprocessed condition to the highly spectrally-degraded CI-
410 simulated voices created by 4-channel noise-vocoder. Results from SVM classification supports
411 the general trend displayed by MD on the effect of CI speech processing on acoustic information
412 involved in normal-vs-disordered voice distinction. However, there was an approximately 8%
413 decline in the accuracy of SVM in classification of normal and disordered voice qualities between
414 the unprocessed condition and the average accuracy obtained across six levels of spectral
415 degradation. Classification accuracy in Fig. 5 shows three categories of performance between 80-
416 85%, 85-90%, and 90-95%. These simulated results suggest that the performance of the current CI

Running title: Voice quality through cochlear implant

417 technology falls within the second category in terms of the effect of number of spectral channels
418 on voice quality distinction (12 channels in MED-EL devices, 16 channels in Advanced Bionics
419 devices, and 22 channels in Cochlear). It is noticeable that the performance in this region varies
420 between 87% and 90% classification accuracy for 12, 16, and 22 channels, which is still at least
421 5% below the unprocessed condition. We speculate that this difference between SVM and
422 Mahalanobis distance in measuring voice quality-related acoustic distinction is because of the
423 exposure phenomenon simulated by SVM as being trained on a subset of data in a supervised
424 fashion. Another explanation could be related to the calculation of MD, which assumes that
425 features have multivariate normal distribution, which might not be necessarily valid for the MFCC
426 features in this study.



427
428 **Fig. 5.** (Color online) The accuracy of SVM in classification between two groups of normal and
429 disordered voices at seven levels of spectral degradation, corresponding to change in the number
430 of noise-vocoder frequency channels (i.e., unprocessed (UP), 32-, 22-, 16-, 12-, 8-, and 4-channels
431 noise-vocoder).

Running title: Voice quality through cochlear implant

432 **4. DISCUSSION AND CONCLUSIONS**

433 This study investigated how CI processing affects acoustic information involved in signaling
434 variations in talkers' voice quality. We analyzed /a/ vowels spoken by talkers with normal or
435 disordered voice qualities to examine to effects of CI speech processing on acoustic information
436 available for distinguishing voice quality. CI speech processing was simulated by noise-excited
437 envelope vocoders. The results showed large decreases in acoustic distance between normal vs.
438 disordered voices, shedding light on possible difficulties that listeners with CIs may experience in
439 perception of talkers' voice quality, which in turn may reduce their ability to identify talkers' voice.

440 Our investigation of /a/ vowel spectra within different frequency sub-bands across normal
441 vs. disordered voice quality showed that simulated CI processing, based on a noise-vocoder, has a
442 detrimental impact on acoustic information signaling changes in talkers' voice quality. The CI
443 processing considerably degraded spectral information in low- (<2 kHz), mid- (~4-12 kHz), and
444 high-frequency regions (>12 kHz) of the voice spectrum that could contribute to voice quality
445 distinction (Arjmandi & Pooyan, 2012; Ball & Code, 2008; Behroozmand & Almasganj, 2007).
446 The discriminative spectral information under these frequency regions signals various degrees of
447 distinctive acoustic information that listeners may utilize to perceive fine variations in talkers'
448 voice quality (Eskenazi et al., 1990; Hillenbrand, Cleveland, & Erickson, 1994; Moisik, 2013;
449 Park et al., 2016; Podesva, 2007; Sicoli, 2010). These patterns of loss in voice quality-related
450 information due to CI processing suggest that listeners with CIs potentially do not receive a large
451 portion of acoustic information, due to the partial transmission of fine-grained spectral structures,
452 that could signal changes in talkers' voice quality.

453 We further measured the spectral distance between voices with normal and disordered
454 qualities by first characterizing their vowels' spectral variations using MFCC features and then

Running title: Voice quality through cochlear implant

455 calculating the distance between MFCC features using MD. The MD between normal and
456 disordered vowel sounds were examined at different levels of spectral degradation (i.e.,
457 unprocessed, 32, 22, 16, 12, 8, and 4-channel CI noise-vocoder processing) to identify how
458 simulated CI speech processing affects the acoustic distance between voices with normal and
459 disordered qualities. We further examined this effect as a function of the number of spectral
460 channels in the noise-vocoder. The results showed a large decrease in acoustic distance between
461 normal and disordered voice qualities because of CI speech processing, highlighting the loss of
462 acoustic information related to talkers' voice quality in CI processed speech. Therefore, listeners
463 with CIs potentially face difficulties compared to listeners with NH in incorporating talkers' voice
464 quality information to construct the corresponding mental representation for identification and
465 recognition of talkers' voice and processing their speech.

466 The results of SVM accuracy in classification between normal and disordered voice
467 qualities corroborated the detrimental effect of the simulated CI processing on acoustic
468 information involved in voice quality distinction. However, the noise vocoding and the number of
469 channels in the noise-vocoder resulted in lower degrees of drop based on SVM classification
470 compared to the amount of decline quantified by MD. Our interpretation of this difference between
471 SVM and MD is that the training phase in SVM somehow simulated the effect of exposure/learning
472 in making sense of the degraded voice signals for voice quality distinction. The results of SVM
473 classification, as executed through 5-fold cross-validation procedure, suggested a good
474 performance of higher than 80%, even in the highly degraded condition of 4 spectral channels. The
475 average drop of ~ 7% in normal-vs-disordered voice classification between unprocessed and CI-
476 simulated conditions further highlights the detrimental effect of CI speech processing to acoustic
477 information relevant to the recognition of talkers' voice.

Running title: Voice quality through cochlear implant

478 These results can be interpreted in the context of perception of voice cues involved in talker
479 recognition, as well as speech processing. The observed lack of faithful transmission of acoustic
480 information, that is more or less related to various perceptual attributes of voice quality (e.g.,
481 breathiness, harshness, and strain), suggests that CI listeners may not benefit from voice quality-
482 related acoustic variations as much as their peers with NH in processing segmental and
483 suprasegmental information for speech comprehension (Dicanio, 2009; Dilley et al., 1996; Dilley
484 et al., 2016; Garellek & Keating, 2011; Gordon, 2001; Gordon & Ladefoged, 2001; Henton, 1986;
485 Ogden, 2001; Redi & Shattuck-Hufnagel, 2001), as well as in recognition of talkers' gender
486 (Gussenhoven, 2004; Ohala, 1983; Puts, Hodges, Cárdenas, & Gaulin, 2007), race (Alim, 2004;
487 Irwin, 1977; Moisik, 2013; Thomas & Reaser, 2004) and social and cultural class (Esling, 1978;
488 Rilliard et al., 2009; Sicoli, 2007; Stross, 2013; Stuart-Smith, 1999). Our investigation of normal
489 and disordered voice qualities suggests that CI processing substantially degrades spectral
490 properties signaling voice quality variations (Dicanio, 2009; Garellek & Keating, 2011), which
491 probably negatively impact CI listeners' access and learning talker-specific information as an
492 important skill for robust speech recognition (Johnson, 2005; Kleinschmidt & Jaeger, 2015; Pisoni,
493 1992). Results from prior studies demonstrated that listeners with CIs do not have access to low-
494 numbered harmonics for robust perception of F_0 , leading to poor performance in talker
495 identification and discrimination (Gaudrain & Baskent, 2018), prosody perception elicited by
496 dynamic pitch (Deroche, Kulkarni, Christensen, Limb, & Chatterjee, 2016), and speech
497 recognition in complex listening conditions such as multi-talker situations (Rosen, Souza,
498 Ekelund, & Majeed, 2013; Stickney, Assmann, Chang, & Zeng, 2007; Stickney, Zeng, Litovsky,
499 & Assmann, 2004). In addition, listeners with NH may incorporate other cues such as vocal-tract
500 length (VTL) and formant frequencies in constructing talkers' voice quality to distinguish between

Running title: Voice quality through cochlear implant

501 talkers, cues that are poorly perceived by listeners with CIs (Gaudrain & Baskent, 2018). Our
502 results provide further evidence in explaining the poor performance of listeners with CIs in
503 perception and effective use of talkers' voice cues (Başkent, Luckmann, Ceha, Gaudrain, &
504 Tamati, 2018; Gaudrain & Baskent, 2018; Mehta, Lu, & Oxenham, 2020; Mehta & Oxenham,
505 2017; Moore & Carlyon, 2005; Stickney et al., 2007) by showing that an important portion of this
506 acoustic information is discarded by cochlear implant speech processing.

507 There are multiple limitations in the current CI devices including the number of spectral
508 channels (electrodes), which restricts spectral and temporal resolution in CI devices in
509 representation of speech. Our results highlight the need for improving CI speech processing
510 strategies to assure that acoustic cues related to voice quality are faithfully transferred through CIs.
511 Developing more effective strategies requires researchers to evaluate the mechanisms underlying
512 encoding spectral and temporal cues responsible for representing voice quality measures.
513 Therefore, further studies are required to understand how listeners with CIs perceive acoustic cues
514 related to voice quality variations and how possible loss of information at this level may impact
515 their ability to identify talkers and process their speech. Tamati et al., (2017) found that listeners
516 with CIs perform poorer than their NH peers in speech recognition when there is large talker
517 variability. They also showed that CI users experience difficulties in the recognition of talkers'
518 voices and accents while their performance were also largely variable compared to listeners with
519 NH. Results from the present study provide further evidence that listeners with CIs may not have
520 access to voice quality cues for robust identification of talkers. This lack of access to the voice
521 quality cues may negatively impact CI listeners' ability to overcome talker variability for
522 successful speech perception.

Running title: Voice quality through cochlear implant

523 Despite the limitations of CIs in robust and reliable transformation of speech, Vongpaisal
524 et al. (2010) showed that children with CIs are able to develop models of talker identity, which
525 may reflect the important role of neural plasticity and more powerful speech processing at higher
526 cortical levels for auditory processing and language development. As simulated by SVM, there
527 might be a large effect of exposure or training that can improve the performance of CI listeners in
528 distinction between various voice qualities. In fact, this phenomenon can be logically expanded to
529 how CIs listeners may use the information in the voice delivered through CI at higher levels of
530 speech processing and language learning to compensate for the lack of various acoustic cues such
531 as those related to the perception of talkers' voice quality (Moore & Shannon, 2009). Speech
532 recognition of children with CIs significantly improved as they had more experience in listening
533 to speech through a CI device (Brown et al., 2004; Fryauf-Bertschy, Tyler, Kelsay, Gantz, &
534 Woodworth, 1997; Miyamoto, Osberger, & Kessler, 1996; Tyler et al., 2000). Another factor that
535 is not modeled in our study is the effect of linguistic and contextual cues in continuous speech that
536 listeners with CIs can incorporate to infer talkers' voice quality for talker recognition and language
537 processing. The significant effect of these cues on sentence recognition was shown in listeners
538 with CIs (Geers, 2002; Meyer & Svirsky, 2000). Despite these contextual effects, having access
539 to acoustic information relevant to talkers' voice quality is still critical for speech processing
540 (Başkent et al., 2018; Gaudrain & Baskent, 2018), particularly in complex listening conditions
541 such as speech recognition in multi-talker scenarios (Rosen et al., 2013; Stickney et al., 2007,
542 2004).

543 The present study had some limitations that should be considered while interpreting the
544 findings. Although studies based on CI-simulated speech are valuable in general, these findings
545 should be viewed as a general trend rather than the actual performance of listeners with CIs in

Running title: Voice quality through cochlear implant

546 perception of talkers' voice quality. One factor is the place-frequency mismatch due to the
547 configuration of the electrode array in the cochlea that its effect is not examined in this study.
548 Furthermore, characterization of voice quality based on MFCCs features might not completely
549 reflect the normal hearing system in perception of voice quality variations as shown by recent
550 studies (Anand, Kopf, Shrivastav, & Eddins, 2019; Eddins, Anand, Lang, & Shrivastav, 2020). It
551 is also worth mentioning that listeners may incorporate segmental and suprasegmental cues at word
552 and/or sentence levels for recognition of talkers' voice quality rather than merely relying on
553 spectral variations of vowel sounds. Regardless of these limitations, the present study provided
554 new evidence showing that acoustic information involved in distinguishing talkers' voice quality
555 is substantially degraded in CI-simulated voices. Our results suggest that listeners who use CIs
556 may have great difficulties incorporating voice quality cues for talkers' voice recognition. The
557 poor spectral resolution provided by cochlear implant device to CI listeners may negatively impact
558 acoustic cues involved in voice quality transmission, leading possibility to subsequent poor
559 perception of talkers' voice quality in listeners with CIs. Future perceptual studies will determine
560 which specific acoustic cues relevant to talkers' voice quality are not faithfully transmitted through
561 cochlear implants. The findings from the current study underscore the need in two directions: (a)
562 the need for examining the current signal processing strategies in CIs for their fidelity in passing
563 voice quality cues and developing more advanced speech processing strategies in CI device to
564 assure faithful transmission of these cues, and (b) the need for active use of multimodal (i.e.,
565 gesture, tactile, and visual) communicative behaviors to provide supportive cues for listeners with
566 CIs in recognition of talkers' voice.

Running title: Voice quality through cochlear implant

567

Acknowledgments

568 This work was partially supported by Dissertation Completion Fellowship, awarded by Michigan
569 State University to Meisam K. Arjmandi.

570

REFERENCES

571 Abberton, E., & Fourcin, A. J. (1978). Intonation and speaker identification. *Language and*
572 *Speech*, 21(4), 305–318.

573 Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh, Scotland: Edinburgh
574 University Press.

575 Akbari, A., & Arjmandi, M. K. (2015). Employing linear prediction residual signal of wavelet sub-
576 bands in automatic detection of laryngeal pathology. *Biomedical Signal Processing and*
577 *Control*, 18, 293–302.

578 Ali, Z., Alsulaiman, M., Muhammad, G., Elamvazuthi, I., & Mesallam, T. A. (2013). Vocal fold
579 disorder detection based on continuous speech by using MFCC and GMM. *2013 7th IEEE*
580 *GCC Conference and Exhibition, GCC 2013*, 292–297.

581 Alim, S. H. (2004). You Know My Steez: An Ethnographic and Sociolinguistic Study of
582 Styleshifting in a Black American Speech Community. *Journal of Linguistic Anthropology*,
583 17(1), 149–151.

584 Anand, S., Kopf, L. M., Shrivastav, R., & Eddins, D. A. (2019). Objective Indices of Perceived
585 Vocal Strain. *Journal of Voice*, 33(6), 838–845.

586 Arjmandi, M., Dilley, L. C., & Wagner, S. E. (2018). Investigation of acoustic dimension use in
587 dialect production: machine learning of sonorant sounds for modeling acoustic cues of

Running title: Voice quality through cochlear implant

588 African American dialect. *11th International Conference on Voice Physiology and*
589 *Biomechanics*, 12–13. East Lansing, USA.

590 Arjmandi, M. K., & Pooyan, M. (2012). An optimum algorithm in pathological voice quality
591 assessment using wavelet-packet-based features, linear discriminant analysis and support
592 vector machine. *Biomedical Signal Processing and Control*, 7(1), 3–19.

593 Arjmandi, M. K., Pooyan, M., Mikaili, M., Vali, M., & Moqarehzadeh, A. (2011). Identification
594 of Voice Disorders Using Long-Time Features and Support Vector Machine With Different
595 Feature Reduction Methods. *Journal of Voice*, 25(6), e275–e289.

596 Arjmandi, M. K., Pooyan, M., Mohammadnejad, H., & Vali, M. (2010). Voice disorders
597 identification based on different feature reduction methodologies and support vector machine.
598 *2010 18th Iranian Conference on Electrical Engineering*, 45–49.

599 Askenfelt, A. G., & Hammarberg, B. (1986). (1986). Speech waveform perturbation analysis: a
600 perceptual-acoustical comparison of seven measures. *Journal of Speech, Language, and*
601 *Hearing Research*, 29(1), 50–64.

602 Aubergé, V., & Cathiard, M. (2003). Can we hear the prosody of smile? *Speech Communication*,
603 40(1–2), 87–97.

604 Ball, M. J., & Code, C. (2008). *Instrumental Clinical Phonetics*. Whurr Publishers.

605 Başkent, D., Luckmann, A., Ceha, J., Gaudrain, E., & Tamati, T. N. (2018). The discrimination of
606 voice cues in simulations of bimodal electro-acoustic cochlear-implant hearing. *The Journal*
607 *of the Acoustical Society of America*, 143(4), EL292–EL297.

608 Behroozmand, R., & Almasganj, F. (2007). Optimal selection of wavelet-packet-based features

Running title: Voice quality through cochlear implant

- 609 using genetic algorithm in pathological assessment of patients' speech signal with unilateral
610 vocal fold paralysis. *Computers in Biology and Medicine*, 37(4), 474–485.
- 611 Bernstein, J. G., & Oxenham, A. J. (2003). Pitch discrimination of diotic and dichotic tone
612 complexes: Harmonic resolvability or harmonic number? *The Journal of the Acoustical*
613 *Society of America*, 113(6), 3323.
- 614 Bingabr, M., Espinoza-Varas, B., & Loizou, P. C. (2008). Simulating the effect of spread of
615 excitation in cochlear implants. *Hearing Research*, 241(1–2), 73–79.
- 616 Blood, G. W., Mahan, B. W., & Hyman, M. (1979). Judging personality and appearance from
617 voice disorders. *Journal of Communication Disorders*, 12(1), 63–67.
- 618 Britt, E. (2011). Can the church say amen: Strategic uses of black preaching style at the State of
619 the Black Union. *Language in Society*, 40(2), 211–233.
- 620 Brown, J., Geers, A., Herrmann, B., Kirk, I., Tomblin, J. B., & Waltzman, S. (2004). Cochlear
621 Implants. *Asha Supplement*, 1–39.
- 622 Cameron, D. (2001). Designer voices. *Critical Quarterly*, 43(4), 81–85.
- 623 Chatterjee, M., & Oberzut, C. (2011). Detection and rate discrimination of amplitude modulation
624 in electrical hearing. *The Journal of the Acoustical Society of America*, 130(3), 1567–1580.
- 625 Chatterjee, M., & Peng, S. C. (2008). Processing F0 with cochlear implants: Modulation frequency
626 discrimination and speech intonation recognition. *Hearing Research*, 235(1–2), 143–156.
- 627 Childers, D., & Lee, C. K. (1991). Vocal quality factors: analysis, synthesis, and perception.
628 *Journal of the Acoustical Society of America*, 90(5), 2394–2410.
- 629 Crew, J. D., & Galvin, J. J. (2012). Channel interaction limits melodic pitch perception in

Running title: Voice quality through cochlear implant

630 simulated cochlear implants. *The Journal of the Acoustical Society of America*, 132(5),
631 EL429–EL435.

632 de Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in
633 speech signals. *Journal of Speech, Language, and Hearing Research*, 36(2), 254–266.

634 Deroche, M. L. D., Kulkarni, A. M., Christensen, J. A., Limb, C. J., & Chatterjee, M. (2016).
635 Deficits in the sensitivity to pitch sweeps by school-aged children wearing cochlear implants.
636 *Frontiers in Neuroscience*, 10(MAR), 1–15.

637 Dibazar, A. A., Narayanad, S., & Berger, T. W. (2002). Feature Analysis for Automatic Detection
638 of Pathological Speech. *Proceedings of EMBS*, 182–183.

639 Dicanoio, C. T. (2009). The phonetics of register in Takhian Thong Chong. *Journal of the*
640 *International Phonetic Association*, 39(2), 162–188.

641 Dilley, L. C., Arjmandi, M. K., Ireland, Z., Heffner, C., & Pitt, M. (2016). Glottalization, reduction,
642 and acoustic variability in function words in American English. *The Journal of the Acoustical*
643 *Society of America*, 140(4), 3114–3114.

644 Dilley, L. C., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels
645 as a function of prosodic structure. *Journal of Phonetics*, 24(4), 423–444.

646 Do, C.-T. (2012). Acoustic Simulations of Cochlear Implants in Human and Machine Hearing
647 Research. *Cochlear Implant Research Updates*, 117.

648 Do, C.-T., Pastor, D., & Goalic, A. (2012). A novel framework for noise robust ASR using cochlear
649 implant-like spectrally reduced speech. *Speech Communication*, 54(1), 119–133.

650 Dolar, M. (2006). *A voice and nothing more*. MIT Press.

Running title: Voice quality through cochlear implant

- 651 Eddins, D. A., Anand, S., Lang, A., & Shrivastav, R. (2020). Developing Clinically Relevant
652 Scales of Breathy and Rough Voice Quality. *Journal of Voice*.
- 653 Eskenazi, L., Childers, D. G., & Hicks, D. M. (1990). Acoustic correlates of vocal quality. *Journal*
654 *of Speech and Hearing Research*, 33(2), 298–306.
- 655 Esling, J. (1978). The identification of features of voice quality in social groups. *Journal of the*
656 *International Phonetic Association*, 8(1–2), 18–23.
- 657 Fant, G. (1973). Acoustic description and classification of phonetic units. *Speech Sounds and*
658 *Features*, 32–83.
- 659 Fex, B., Fex, S., Shiromoto, O., & Hirano, M. (1994). Acoustic analysis of functional dysphonia:
660 Before and after voice therapy (accent method). *Journal of Voice*, 8(2), 163–167.
- 661 Firdos, S., & Umarani, K. (2016). Disordered Voice Classification using SVM and Feature
662 Selection using GA. *2016 Second International Conference on Cognitive Computing and*
663 *Information Processing (CCIP)*, 1–6.
- 664 Fryauf-Bertschy, H., Tyler, R. S., Kelsay, D. M., Gantz, B. J., & Woodworth, G. G. (1997).
665 Cochlear implant use by prelingually deafened children: the influences of age at implant and
666 length of device use. *Journal of Speech, Language, and Hearing Research*, 40(1), 183–199.
- 667 Fu, Q.-J. (2019). *AngelSim: Cochlear implant and hearing loss simulator*. Retrieved from
668 http://www.tigerspeech.com/angelsim/angelsim_about.html
- 669 Fu, Q.-J., Chinchilla, S., & Galvin, J. J. (2004). The role of spectral and temporal cues in voice
670 gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the*
671 *Association for Research in Otolaryngology*, 5(3), 253–260.

Running title: Voice quality through cochlear implant

- 672 Fu, Q.-J., Chinchilla, S., Nogaki, G., & Galvin, J. J. (2005). Voice gender identification by cochlear
673 implant users: The role of spectral and temporal resolution. *The Journal of the Acoustical*
674 *Society of America*, 118(3), 1711–1718.
- 675 Fukazawa, T., el-Assuooty, a, & Honjo, I. (1988). A new index for evaluation of the turbulent
676 noise in pathological voice. *The Journal of the Acoustical Society of America*, 83(3), 1189–
677 1193.
- 678 Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions
679 in Jalapa Mazatec. *Journal of the International Phonetic Association*, 41(2), 185–205.
- 680 Gaudrain, E., & Baskent, D. (2018). Discrimination of voice pitch and vocal-tract length in
681 cochlear implant users. *Ear and Hearing*, 39(2), 226–237.
- 682 Geers, A. E. (2002). Factors Affecting the Development of Speech, Language, and Literacy in
683 Children With Early Cochlear Implantation. *Language Speech and Hearing Services in*
684 *Schools*, 33(3), 172.
- 685 Gelfer, M. P., & Bennett, Q. E. (2013). Speaking fundamental frequency and vowel formant
686 frequencies: Effects on perception of gender. *Journal of Voice*, 27(5), 556–566.
- 687 Ghasemzadeh, H., Khass, M. T., Arjmandi, M. K., & Pooyan, M. (2015). Detection of vocal
688 disorders based on phase space parameters and Lyapunov spectrum. *Biomedical Signal*
689 *Processing and Control*, 22, 135–145.
- 690 Ghasemzadeh, H., & Arjmandi, M. K. (2019). Toward Optimum Quantification of Pathology-
691 induced Noises: An Investigation of Information Missed by Human Auditory System.
692 *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, 519–528.

Running title: Voice quality through cochlear implant

- 693 Godino-Llorente, J. I., Gomez-Vilda, P., & Blanco-Velasco, M. (2006). Dimensionality reduction
694 of a pathological voice quality assessment system based on Gaussian mixture models and
695 short-term cepstral parameters. *IEEE Transactions on Biomedical Engineering*, 53(10),
696 1943–1953.
- 697 Gordon, M. (2001). Linguistic aspects of voice quality with special reference to Athabaskan.
698 *Proceedings of the Athabaskan Languages Conference*, 163–178.
- 699 Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of*
700 *Phonetics*, 29(4), 383–406.
- 701 Goupell Matthew J. (2015). Pushing the Envelope of Auditory Research with Cochlear Implants.
702 *Acoustics Today*, 11(2), 26–33.
- 703 Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years
704 later. *Journal of the Acoustical Society of America*, 87(6), 2592–2605.
- 705 Gussenhoven, C. (2004). The phonology of tone and intonation. *The Phonology of Tone and*
706 *Intonation*, 1–355.
- 707 Guzman, M., Correa, S., Muñoz, D., & Mayerhoff, R. (2013). Influence on spectral energy
708 distribution of emotional expression. *Journal of Voice*, 27(1), 129.e1-129.e10.
- 709 Hammarberg, B., Fritzell, B., Gaufin, J., Sundberg, J., & Wedin, L. (1980). Perceptual and acoustic
710 correlates of abnormal voice qualities. *Acta Oto-Laryngologica*, 90(1–6), 441–451.
- 711 Hanson, H. M. (1997). Glottal characteristics of female speakers: acoustic correlates. *The Journal*
712 *of the Acoustical Society of America*, 101(1), 466–481.
- 713 Hartl, D. M., Hans, S., Vaissière, J., Riquet, M., & Brasnu, D. F. (2001). Objective voice quality

Running title: Voice quality through cochlear implant

- 714 analysis before and after onset of unilateral vocal fold paralysis. *Journal of Voice*, 15(3), 351–
715 361.
- 716 Heijden, V. Der, Ferdinand, R. P. D., Ridder, D. De, & Tax, D. M. (2005). *Classification,*
717 *Parameter Estimation and State Estimation An Engineering Approach Using MATLAB*. John
718 Wiley & Sons.
- 719 Henton, C. G. (1986). Creak as a sociophonetic marker. *The Journal of the Acoustical Society of*
720 *America*, 80(S1), S50–S50.
- 721 Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal
722 quality. *Journal of Speech, Language, and Hearing Research*, 37(4), 769–778.
- 723 Houston, M., Beer, Bergeson, R., Chin, B., Pisoni, B., & Miyamoto, T. (2012). The ear is
724 connected to the brain: some new directions in the study of children with cochlear implants
725 at Indiana University. *Journal of American Academy of Audiology*, 23(6), 446–463.
- 726 Hunt, N. J., Lennig, N., & Mermelestein, P. (1980). Experiments in syllable-based recognition of
727 continuous speech. *IEEE International Conference on Acoustics, Speech, and Signal*
728 *Processing*, (5), 880–883.
- 729 Irwin, R. B. (1977). Judgments of vocal quality, speech fluency, and confidence of southern black
730 and white speakers. *Language and Speech*, 20(3), 261–266.
- 731 John J. Ohala. (1983). Cross-Language Use of Pitch: An Ethological View. *Phonetica*, Vol. 40,
732 pp. 1–18.
- 733 Johnson, K. (2005). Speaker Normalization in Speech Perception. In *The Handbook of Speech*
734 *Perception* (pp. 363–389).

Running title: Voice quality through cochlear implant

- 735 Kitzing, P., & Åkerlund, L. (1993). Long-time average spectrograms of dysphonic voices before
736 and after therapy. *Folia Phoniatrica et Logopaedica*, 45(2), 53–61.
- 737 Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations
738 among female and male talkers. In *The Journal of the Acoustical Society of America* (Vol.
739 87).
- 740 Kleinschmidt, D., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar,
741 generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203.
- 742 Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model
743 Selection. *Appears in the International Joint Conference on Artificial Intelligence (IJCAI)*, 1–
744 7.
- 745 Kreiman, J., Vanlancker-Sidtis, D., & Gerratt, B. R. (2005). Perception of voice quality. In
746 *Handbook of speech perception* (pp. 338–362). Malden, MA: Blackwell.
- 747 Krom, Guus de. (1995). Some Spectral Correlates of Pathological Breathy and Rough Voice
748 Quality for Different Types of Vowel Fragments. *Journal of Speech Language and Hearing*
749 *Research*, 38(4), 794.
- 750 Laver, J. D. M. (1968). Voice Quality and Indexical Information. *International Journal of*
751 *Language & Communication Disorders*, 3(1), 43–54.
- 752 Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.
- 753 Maesschalck, R. D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and*
754 *Intelligent Laboratory Systems*, 50(1), 1–18.
- 755 Masnan, M. J., Mahat, N. I., Shakaff, A. Y. M., Abdullah, A. H., Zakaria, N. Z. I., Yusuf, N., ...

Running title: Voice quality through cochlear implant

- 756 Aziz, A. H. A. (2015). Understanding Mahalanobis distance criterion for feature selection.
757 *AIP Conference Proceedings*, 1660, 050075. AIP Publishing LLC.
- 758 Mckinney, M. F., & Breebaart, J. (2003). Features for Audio and Music Classification. *Proc*
759 *ISMIR*, 4, 151–158.
- 760 Mehta, A. H., Lu, H., & Oxenham, A. J. (2020). The Perception of Multiple Simultaneous Pitches
761 as a Function of Number of Spectral Channels and Spectral Spread in a Noise-Excited
762 Envelope Vocoder. *JARO - Journal of the Association for Research in Otolaryngology*, 21(1),
763 61–72.
- 764 Mehta, A. H., & Oxenham, A. J. (2017). Vocoder Simulations Explain Complex Pitch Perception
765 Limitations Experienced by Cochlear Implant Users. *JARO - Journal of the Association for*
766 *Research in Otolaryngology*, 18(6), 789–802.
- 767 Meyer, T. A., & Svirsky, M. A. (2000). Speech perception by children with the CLARION (CIS)
768 or nucleus 22 (speak) cochlear implant or hearing aids. *Annals of Otology, Rhinology and*
769 *Laryngology*, 109(12 II SUPPL.), 49–51.
- 770 Miyamoto, R. T., Osberger, M. J., & Kessler, K. (1996). Cochlear implant in aural re (habilitation)
771 of adults and children. *J Otolaryngology Head and Neck Surgery*, 116, 1142–1152.
- 772 Moisik, S. R. (2013). Harsh voice quality and its association with blackness in popular American
773 media. *Phonetica*, 69(4), 193–215.
- 774 Moore, B. C. J., & Carlyon, R. P. (2005). Perception of Pitch by People with Cochlear Hearing
775 Loss and by Cochlear Implant Users. In *Pitch* (pp. 234–277). [https://doi.org/10.1007/0-387-](https://doi.org/10.1007/0-387-28958-5_7)
776 [28958-5_7](https://doi.org/10.1007/0-387-28958-5_7)

Running title: Voice quality through cochlear implant

- 777 Moore, D. R., & Shannon, R. V. (2009). Beyond cochlear implants: Awakening the deafened brain.
778 *Nature Neuroscience*, *12*(6), 686–691.
- 779 Naranjo, N. V., Lara, E. M., Rodríguez, I. M., & García, G. C. (1994). High-frequency components
780 of normal and dysphonic voices. *Journal of Voice*, *8*(2), 157–162.
- 781 O’Leidhin, E., & Murphy, P. (2005). Analysis of spectral measures for voiced speech with varying
782 noise and perturbation levels. *ICASSP, IEEE International Conference on Acoustics, Speech*
783 *and Signal Processing - Proceedings, I*, 869–872.
- 784 Oates, J. (2009). Auditory-perceptual evaluation of disordered voice quality. *Folia Phoniatica et*
785 *Logopaedica*, *61*(1), 49–56.
- 786 Ogden, R. (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of*
787 *the International Phonetic Association*, *31*(1), 139–152.
- 788 Oxenham, A. J., & Kreft, H. A. (2014). Speech perception in tones and noise via cochlear implants
789 reveals influence of spectral resolution on temporal processing. *Trends in Hearing*, *18*, 1–14.
- 790 Panek, D., Skalski, A., Gajda, J., & Tadeusiewicz, R. (2015). Acoustic Analysis Assessment in
791 Speech Pathology Detection. *International Journal of Applied Mathematics and Computer*
792 *Science*, *25*(3), 631–643.
- 793 Park, S. J., Sigouin, C., Kreiman, J., Keating, P., Guo, J., Yeung, G., ... Alwan, A. (2016). Speaker
794 identity and voice quality: Modeling human responses and automatic speaker recognition.
795 *Proceedings of the Annual Conference of the International Speech Communication*
796 *Association, INTERSPEECH, 08-12-Sept*, 1044–1048.
- 797 Pisoni, D. B. (1992). Talker normalization in speech perception. *Speech Perception, Production*

Running title: Voice quality through cochlear implant

- 798 *and Linguistic Structure*, (1974), 143–151.
- 799 Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a
800 persona. *Journal of Sociolinguistics*, 11(4), 478–504.
- 801 Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and Phonetic Experiments on American
802 English Dialect Identification. *Journal of Language and Social Psychology*, 18(10), 10–30.
- 803 Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. C. (2007). Men’s voices as dominance
804 signals: vocal fundamental and formant frequencies influence dominance attributions among
805 men. *Evolution and Human Behavior*, 28(5), 340–344.
- 806 Rabiner, L. R., & Schafer, R. W. (1978). *Digital Processing of Speech Signals*. New Jersey:
807 Prentice-Hal.
- 808 Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal
809 speakers. *Journal of Phonetics*, 29(4), 407–429.
- 810 Reilly, R. B., Moran, R., & Lacy, P. (2004). Voice pathology assessment based on a dialogue
811 system and speech analysis. *Proceedings of the AAAI Fall Symposium on Dialogue Systems*
812 *for Health Communication, Washington DC*, 104–109.
- 813 Rilliard, A., Shochi, T., Martin, J. C., Erickson, D., & Aubergé, V. (2009). Multimodal Indices to
814 Japanese and French Prosodically Expressed Social Affects. *Language and Speech*, 52(2–3),
815 223–243.
- 816 Rontal, E., Rontal, M., & Rolnick, M. I. (1975). Objective evaluation of vocal pathology using
817 voice spectrography. *Annals of Otology, Rhinology & Laryngology*, 84(5), 662–671.
- 818 Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background

Running title: Voice quality through cochlear implant

- 819 of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical*
820 *Society of America*, 133(4), 2431–2443.
- 821 Roy, N., & Leeper, H. A. (1993). Effects of the manual laryngeal musculoskeletal tension
822 reduction technique as a treatment for functional voice disorders: Perceptual and acoustic
823 measures. *Journal of Voice*, 7(3), 242–249.
- 824 S. Davis and P. Mermelstein. (1980). Comparison of Parametric Representations for Monosyllabic
825 Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics,*
826 *Speech, and Signal Processing*, 28(4), 357–366.
- 827 Santos, J. F., Cosentino, S., Hazrati, O., Loizou, P. C., & Falk, T. H. (2013). Objective speech
828 intelligibility measurement for cochlear implant users in complex listening environments.
829 *Speech Communication*, 55(7–8), 815–824.
- 830 Scott, S., & McGettigan, C. (2015). The voice: From identity to interactions. In *APA Handbook of*
831 *Nonverbal Communication*.
- 832 Shaneh, M., & Taheri, A. (2009). Voice command recognition system based on MFCC and VQ
833 algorithms. *World Academy of Science, Engineering and Technology*, 57, 534–538.
- 834 Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition
835 with primarily temporal cues. *Science*, 270(5234), 303–304.
- 836 Shannon, Robert V., Fu, Q. J., & Galvin, J. (2004). The number of spectral channels required for
837 speech recognition depends on the difficulty of the listening situation. *Acta Oto-*
838 *Laryngologica, Supplement*, 124(552), 50–54.
- 839 Sicoli, M. A. (2007). *Tono: A linguistic ethnography of tone and voice in a Zapotec region*

Running title: Voice quality through cochlear implant

840 (Doctoral dissertation).

841 Sicoli, Mark A. (2010). Shifting voices with participant roles: Voice qualities and speech registers
842 in Mesoamerica. *Language in Society*, 39(4), 521–553.

843 Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract
844 length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of*
845 *America*, 118(5), 3177–3186.

846 Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many
847 harmonics. *Journal of the Acoustical Society of America*, 87(1), 304–310.

848 Stevens, S. S., Volkman, J., & Newman, E. B. (1937). A scale for the measurement of the
849 psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8(3), 185–
850 190.

851 Stickney, G. S., Assmann, P. F., Chang, J., & Zeng, F.-G. (2007). Effects of cochlear implant
852 processing and fundamental frequency on the intelligibility of competing sentences. *The*
853 *Journal of the Acoustical Society of America*, 122(2), 1069–1078.

854 Stickney, G. S., Zeng, F.-G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech
855 recognition with speech maskers. *The Journal of the Acoustical Society of America*, 116(2),
856 1081–1091.

857 Stross, B. (2013). Falsetto voice and observational logic: Motivated meanings. *Language in*
858 *Society*, 42(2), 139–162.

859 Stuart-Smith, J. (1999). Glasgow: Accent and voice quality. *Urban Voices: Accent Studies in the*
860 *British Isles*, 203–222.

Running title: Voice quality through cochlear implant

- 861 Svirsky, M. A. (2017). Cochlear implants and electronic hearing. *Physics Today*, 70(8), 53–58.
- 862 Tamati, T. N., Janse, E., Pisoni, D. B., & Baskent, D. (2017). Talker variability in real-life speech
863 recognition by cochlear implant users. *The Journal of the Acoustical Society of America*,
864 141(5), 2017–2020.
- 865 Terasawa, H., Slaney, M., & Berger, J. (2005). Perceptual Distance in Timbre Space. *Proceedings*
866 *of ICAD 05-Eleventh Meeting Of the International Conference on Auditory Display*, 6–9.
- 867 Thomas, E. R., & Reaser, J. (2004). Delimiting perceptual cues used for the ethnic labeling of
868 African American and European American voices. *Journal of Sociolinguistics*, 8(1), 54–87.
- 869 Tsai, C. G., Wang, L. C., Wang, S. F., Shau, Y. W., Hsiao, T. Y., & Auhagen, W. (2010).
870 Aggressiveness of the growl-like timbre: Acoustic characteristics, musical implications, and
871 biomechanical mechanisms. *Music Perception*, 27(3), 209–222.
- 872 Tyler, R. S., Teagle, H. F., Kelsay, D. M., Gantz, B. J., Woodworth, G. G., & Parkinson, A. J.
873 (2000). Speech perception by prelingually deaf children after six years of cochlear implant
874 use: effects of age at implantation. *Annals of Otology, Rhinology & Laryngology*,
875 109(12_suppl), 82–84.
- 876 Umopathy, K., Krishnan, S., Parsa, V., & Jamieson, D. G. (2005). Discrimination of pathological
877 voices using a time-frequency approach. *IEEE Transactions on Biomedical Engineering*,
878 52(3), 421–430.
- 879 Umopathy, S., Rachel, S., & Thulasi, R. (2018). Automated speech signal analysis based on feature
880 extraction and classification of spasmodic dysphonia: a performance comparison of different
881 classifiers. *International Journal of Speech Technology*, 21(1), 9–18.

Running title: Voice quality through cochlear implant

- 882 Vongpaisal, T., Trehub, S. E., Schellenberg, E. G., Lieshout, P. Van, & Papsin, B. C. (2010).
883 Children With Cochlear Implants Recognize Their Mother's Voice. *Ear and Hearing, 31*(4),
884 555–566.
- 885 Winn, M. B., & Litovsky, R. Y. (2015). Using speech sounds to test functional spectral resolution
886 in listeners with cochlear implants. *The Journal of the Acoustical Society of America, 137*(3),
887 1430–1442.
- 888 Wolfe, V., Cornell, R., & Palmer, C. (1991). Acoustic Correlates of Pathologic Voice Types.
889 *Journal of Speech Language and Hearing Research, 34*(3), 509.
- 890 Wolfe, V. I., & Bacon, M. (1971). Spectrographic comparison of two types of spastic dysphonia.
891 *Journal of Speech and Hearing Disorders, 41*(3), 325–332.
- 892 Xiang, S., Nie, F., & Zhang, C. (2008). Learning a Mahalanobis distance metric for data clustering
893 and classification. *Pattern Recognition, 41*(12), 3600–3612.
- 894 Yanagihara, N. (1967). Significance of harmonic changes and noise components in hoarseness.
895 *Journal of Speech, Language, and Hearing Research, 10*(3), 531–542.
- 896 Yumoto, E., Gould, W. J., & Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree
897 of hoarseness. *The Journal of the Acoustical Society of America, 71*(6), 1544-1550.
- 898 Zimman, L. (2012). Voices in transition: Testosterone, transmasculinity, and the gendered voice
899 among female-to-male transgender people. *Linguistics Graduate Theses & Dissertations,*
900 (24), 1–253.

901

902