Manuscript File
medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10
preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a licens
All rights reserved. No reuse allowed without permission.

Running title: Voice quality through cochlear implant

Effects of simulated cochlear-implant processing on voice quality distinction: Evidence from analysis of disordered voices [a)]

Meisam K. Arjmandi [1,b)], Hamzeh Ghasemzadeh [1,c)], Laura C. Dilley [1]

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

[a)] Portions of this work were presented at the 177th Meeting of the Acoustical Society of America, Louisville, Kentucky, USA, May 2019.

[b)] Current affiliation: Department of Otolaryngology – Head and Neck Surgery, Massachusetts Eye and Ear and Harvard Medical School, Boston, Massachusetts, USA. Email: meisam_arjmandi@meei.harvard.edu

[c)] Also affiliated with: The Department of Computational Mathematics Science and Engineering, Michigan State University, East Lansing, Michigan, USA.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Running title: Voice quality through cochlear implant

ABSTRACT

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

Keywords: Voice quality, cochlear implant, acoustic distinctiveness

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

1. INTRODUCTION

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

2. MATERIALS AND METHODS

2.1. Voice samples

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

2.2. Creation of noise-vocoded voice samples

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

2.3. Analysis of voice spectra

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

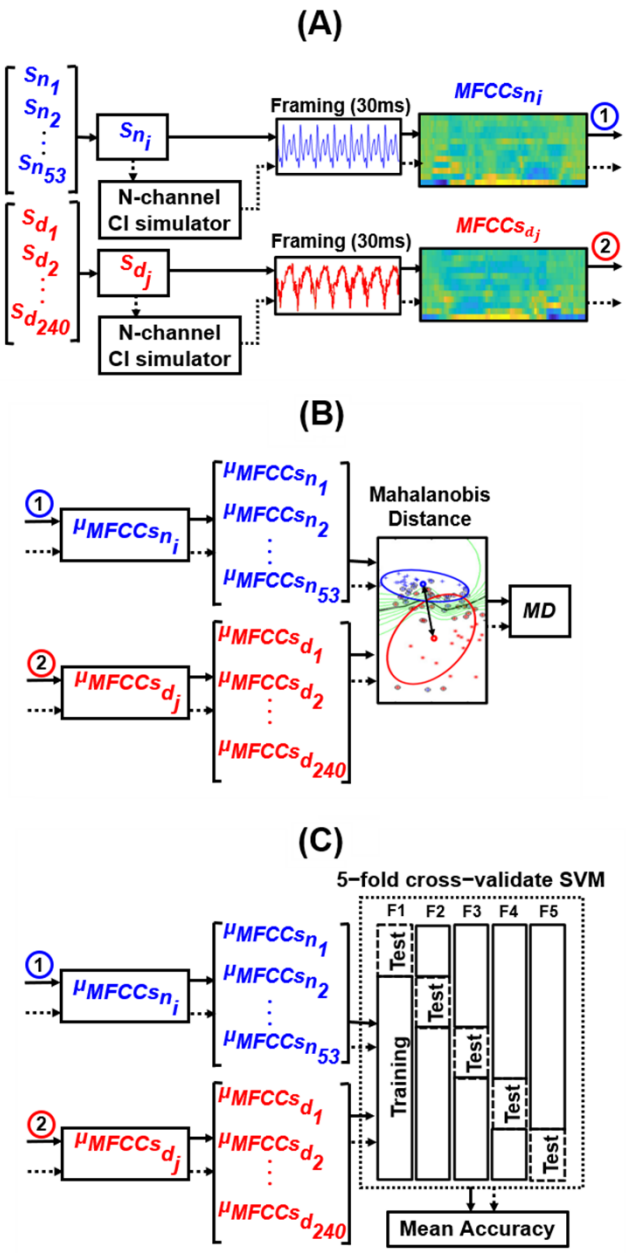
2.4. Using MFCC features to characterize acoustic information

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

Fig. 1A shows the schematic of the approach used for calculating MFCCs for samples of voices with normal and disordered qualities. To calculate MFCCs, each /a/ vowel sound (S_{n_i} or 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 indicate the index of voice stimuli for normal and disordered voices, respectively (i = {1,2,3, …,53} and j = {1,2,3, …,240}). A Hamming window was then applied to each frame to decrease the effect of sidelobes for better frequency-selective analysis (Rabiner & Schafer, 1978). The power spectrum of each frame was calculated based on Fast Fourier transform (FFT) analysis. Then, 32 mel-filterbanks were generated and applied to voice power spectra. MFCCs were derived by calculating the DCT of the logarithm of all filterbank energies (Rabiner & Schafer, 1978).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Eventually, the first twelve components were preserved as MFCC features for each frame of a voice signal ($MFCCs_{n_i}$ or $MFCCs_{d_j}$ in Fig. 1). Each voice signal was represented by an MFCC matrix of size $F\times12$, where F indicates the number of frames in a voice signal. The same procedure was performed to calculate MFCC matrices for CI-simulated versions of the unprocessed normal and disordered voices, as indicated by dashed lines in Fig. 1A.



medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

2.5. Acoustic distance quantification using Mahalanobis distance

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

3. RESULTS

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

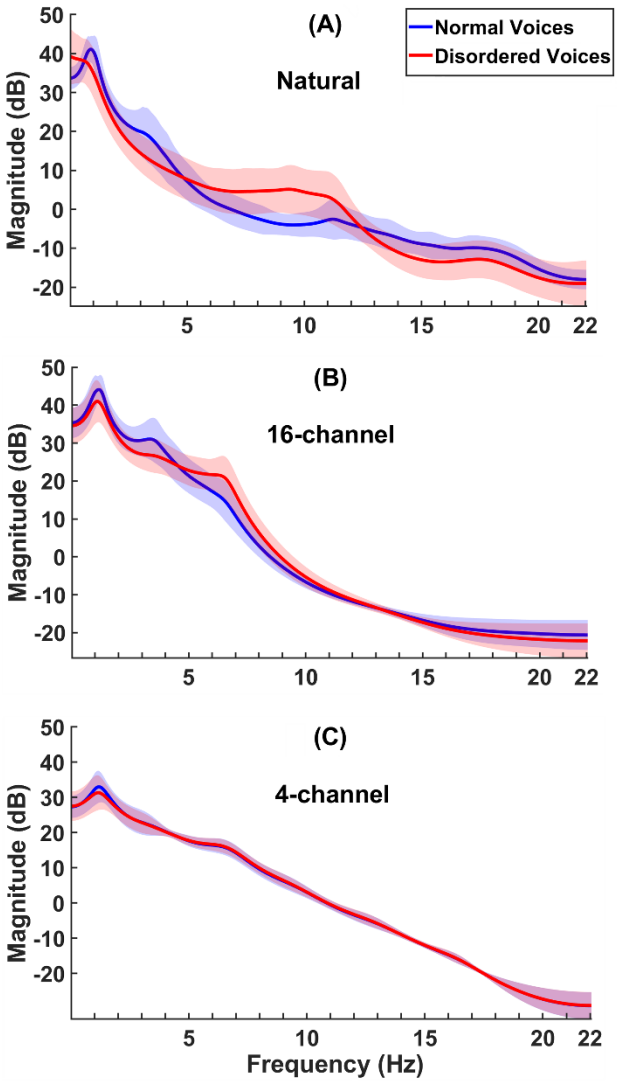


Fig. 2. (Color online) Average magnitude spectra of voice stimuli with normal (blue or dark gray lines) and disordered (red or light gray lines) qualities for (A) unprocessed voices, (B) simulated cochlear-implant voices with 16-channel, and (C) 4-channel in the CI noise-vocoder. The standard deviations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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

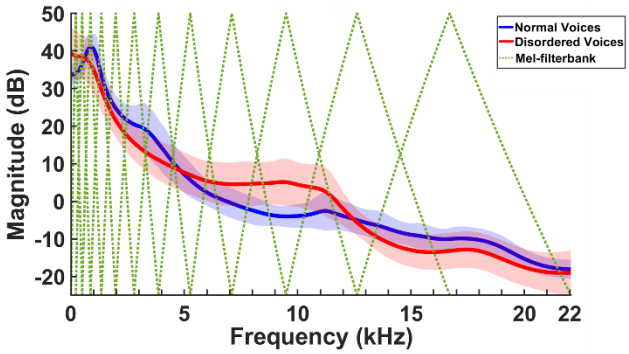


Fig. 3. (Color online) The filterbank of mel-spaced triangular filters (green or dark gray dotted lines) superimposed on average magnitude spectra of voices with normal (blue or dark gray line) and disordered (red or light gray line) qualities. In this example, mel-filterbak contains 12 filters, which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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

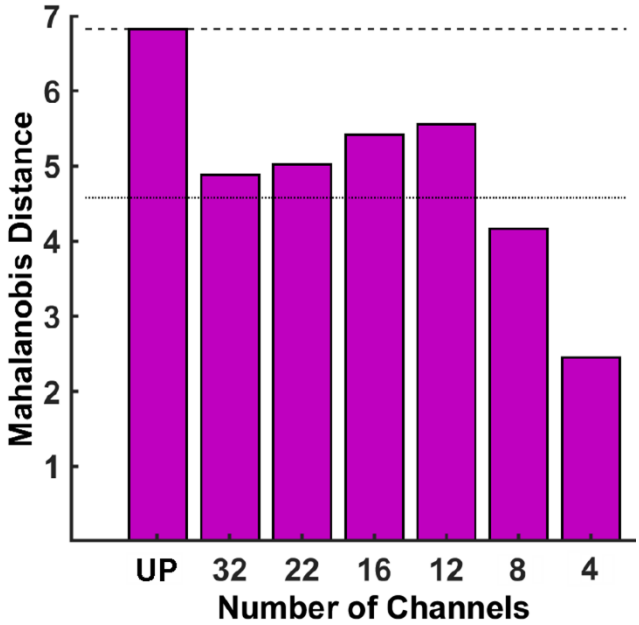


Fig. 4. (Color online) *Mahalanobis distance* between two groups of voices with normal and disordered qualities as a function of spectral degradation in CI-simulated voices (i.e., number of spectral channels in noise-vocoder). The top, horizontal dashed line shows the MD derived from unprocessed voices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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

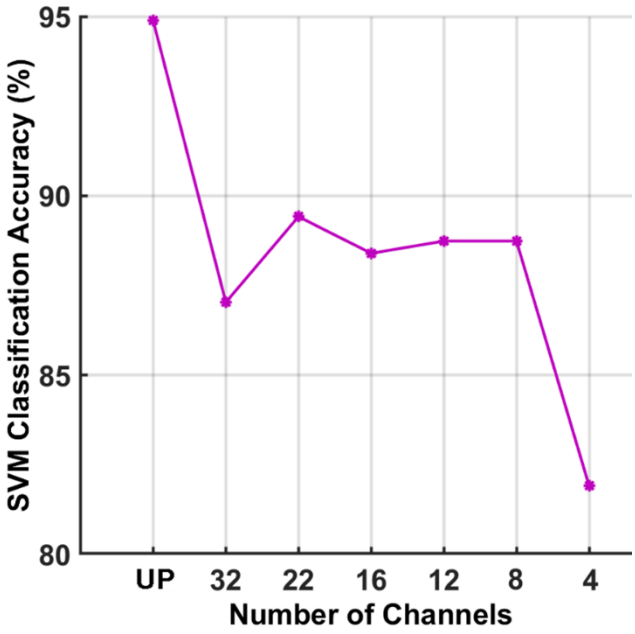


Fig. 5. (Color online) The accuracy of SVM in classification between two groups of normal and disordered voices at seven levels of spectral degradation, corresponding to change in the number of noise-vocoder frequency channels (i.e., unprocessed (UP), 32-, 22-, 16-, 12-, 8-, and 4-channels noise-vocoder).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

4. DISCUSSION AND CONCLUSIONS

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Acknowledgments

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

REFERENCES

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

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

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

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

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

using genetic algorithm in pathological assessment of patients' speech signal with unilateral vocal fold paralysis. *Computers in Biology and Medicine*, *37*(4), 474–485.

Bernstein, J. G., & Oxenham, A. J. (2003). Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *The Journal of the Acoustical Society of America*, *113*(6), 3323.

Bingabr, M., Espinoza-Varas, B., & Loizou, P. C. (2008). Simulating the effect of spread of excitation in cochlear implants. *Hearing Research*, *241*(1–2), 73–79.

Blood, G. W., Mahan, B. W., & Hyman, M. (1979). Judging personality and appearance from voice disorders. *Journal of Communication Disorders*, *12*(1), 63–67.

Britt, E. (2011). Can the church say amen: Strategic uses of black preaching style at the State of the Black Union. *Language in Society*, *40*(2), 211–233.

Brown, J., Geers, A., Herrmann, B., Kirk, I., Tomblin, J. B., & Waltzman, S. (2004). Cochlear Implants. *Asha Supplement*, 1–39.

Cameron, D. (2001). Designer voices. *Critical Quarterly*, *43*(4), 81–85.

Chatterjee, M., & Oberzut, C. (2011). Detection and rate discrimination of amplitude modulation in electrical hearing. *The Journal of the Acoustical Society of America*, *130*(3), 1567–1580.

Chatterjee, M., & Peng, S. C. (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research*, *235*(1–2), 143–156.

Childers, D., & Lee, C. K. (1991). Vocal quality factors: analysis, synthesis, and perception. *Journal of the Acoustical Society of America*, *90*(5), 2394–2410.

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

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

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Eddins, D. A., Anand, S., Lang, A., & Shrivastav, R. (2020). Developing Clinically Relevant Scales of Breathy and Rough Voice Quality. *Journal of Voice*.

Eskenazi, L., Childers, D. G., & Hicks, D. M. (1990). Acoustic correlates of vocal quality. *Journal of Speech and Hearing Research*, *33*(2), 298–306.

Esling, J. (1978). The identification of features of voice quality in social groups. *Journal of the International Phonetic Association*, *8*(1–2), 18–23.

Fant, G. (1973). Acoustic description and classification of phonetic units. *Speech Sounds and Features*, 32–83.

Fex, B., Fex, S., Shiromoto, O., & Hirano, M. (1994). Acoustic analysis of functional dysphonia: Before and after voice therapy (accent method). *Journal of Voice*, *8*(2), 163–167.

Firdos, S., & Umarani, K. (2016). Disordered Voice Classification using SVM and Feature Selection using GA. *2016 Second International Conference on Cognitive Computing and Information Processing (CCIP)*, 1–6.

Fryauf-Bertschy, H., Tyler, R. S., Kelsay, D. M., Gantz, B. J., & Woodworth, G. G. (1997). Cochlear implant use by prelingually deafened children: the influences of age at implant and length of device use. *Journal of Speech, Language, and Hearing Research*, *40*(1), 183–199.

Fu, Q.-J. (2019). *AngelSim: Cochlear implant and hearing loss simulator*. Retrieved from http://www.tigerspeech.com/angelsim/angelsim_about.html

Fu, Q.-J., Chinchilla, S., & Galvin, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology*, *5*(3), 253–260.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fu, Q.-J., Chinchilla, S., Nogaki, G., & Galvin, J. J. (2005). Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *The Journal of the Acoustical Society of America*, *118*(3), 1711–1718.

Fukazawa, T., el-Assuooty, a, & Honjo, I. (1988). A new index for evaluation of the turbulent noise in pathological voice. *The Journal of the Acoustical Society of America*, *83*(3), 1189–1193.

Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association*, *41*(2), 185–205.

Gaudrain, E., & Baskent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear and Hearing*, *39*(2), 226–237.

Geers, A. E. (2002). Factors Affecting the Development of Speech, Language, and Literacy in Children With Early Cochlear Implantation. *Language Speech and Hearing Services in Schools*, *33*(3), 172.

Gelfer, M. P., & Bennett, Q. E. (2013). Speaking fundamental frequency and vowel formant frequencies: Effects on perception of gender. *Journal of Voice*, *27*(5), 556–566.

Ghasemzadeh, H., Khass, M. T., Arjmandi, M. K., & Pooyan, M. (2015). Detection of vocal disorders based on phase space parameters and Lyapunov spectrum. *Biomedical Signal Processing and Control*, *22*, 135–145.

Ghasemzadeh, H., & Arjmandi, M. K. (2019). Toward Optimum Quantification of Pathology-induced Noises : An Investigation of Information Missed by Human Auditory System. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *28*, 519–528.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Godino-Llorente, J. I., Gomez-Vilda, P., & Blanco-Velasco, M. (2006). Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. *IEEE Transactions on Biomedical Engineering*, *53*(10), 1943–1953.

Gordon, M. (2001). Linguistic aspects of voice quality with special reference to Athabaskan. *Proceedings of the Athabaskan Languages Conference*, 163–178.

Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics*, *29*(4), 383–406.

Goupell Matthew J. (2015). Pushing the Envelope of Auditory Research with Cochlear Implants. *Acoustics Today*, *11*(2), 26–33.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *Journal of the Acoustical Society of America*, *87*(6), 2592–2605.

Gussenhoven, C. (2004). The phonology of tone and intonation. *The Phonology of Tone and Intonation*, 1–355.

Guzman, M., Correa, S., Muñoz, D., & Mayerhoff, R. (2013). Influence on spectral energy distribution of emotional expression. *Journal of Voice*, *27*(1), 129.e1-129.e10.

Hammarberg, B., Fritzell, B., Gaufin, J., Sundberg, J., & Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Oto-Laryngologica*, *90*(1–6), 441–451.

Hanson, H. M. (1997). Glottal characteristics of female speakers: acoustic correlates. *The Journal of the Acoustical Society of America*, *101*(1), 466–481.

Hartl, D. M., Hans, S., Vaissière, J., Riquet, M., & Brasnu, D. F. (2001). Objective voice quality

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

analysis before and after onset of unilateral vocal fold paralysis. *Journal of Voice*, *15*(3), 351–361.

Heijden, V. Der, Ferdinand, R. P. D., Ridder, D. De, & Tax, D. M. (2005). *Classification, Parameter Estimation and State Estimation An Engineering Approach Using MATLAB*. John Wiley & Sons.

Henton, C. G. (1986). Creak as a sociophonetic marker. *The Journal of the Acoustical Society of America*, *80*(S1), S50–S50.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, *37*(4), 769–778.

Houston, M., Beer, Bergeson, R., Chin, B., Pisoni, B., & Miyamoto, T. (2012). The ear is connected to the brain: some new directions in the study of children with cochlear implants at Indiana University. *Journal of American Academy of Audiology*, *23*(6), 446–463.

Hunt, N. J., Lennig, N., & Mermeletein, P. (1980). Experiments in syllable-based recognition of continuous speech. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, (5), 880–883.

Irwin, R. B. (1977). Judgments of vocal quality, speech fluency, and confidence of southern black and white speakers. *Language and Speech*, *20*(3), 261–266.

John J. Ohala. (1983). Cross-Language Use of Pitch: An Ethological View. *Phonetica*, Vol. 40, pp. 1–18.

Johnson, K. (2005). Speaker Normalization in Speech Perception. In *The Handbook of Speech Perception* (pp. 363–389).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Kitzing, P., & Åkerlund, L. (1993). Long-time average spectrograms of dysphonic voices before and after therapy. *Folia Phoniatrica et Logopaedica*, *45*(2), 53–61.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. In *The Journal of the Acoustical Society of America* (Vol. 87).

Kleinschmidt, D., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203.

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Appears in the International Joint Conference on Articial Intelligence (IJCAI)*, 1–7.

Kreiman, J., Vanlancker-Sidtis, D., & Gerratt, B. R. (2005). Perception of voice quality. In *Handbook of speech perception* (pp. 338–362). Malden, MA: Blackwell.

Krom, Guus de. (1995). Some Spectral Correlates of Pathological Breathy and Rough Voice Quality for Different Types of Vowel Fragments. *Journal of Speech Language and Hearing Research*, *38*(4), 794.

Laver, J. D. M. (1968). Voice Quality and Indexical Information. *International Journal of Language & Communication Disorders*, *3*(1), 43–54.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Maesschalck, R. D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, *50*(1), 1–18.

Masnan, M. J., Mahat, N. I., Shakaff, A. Y. M., Abdullah, A. H., Zakaria, N. Z. I., Yusuf, N., …

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Aziz, A. H. A. (2015). Understanding Mahalanobis distance criterion for feature selection. *AIP Conference Proceedings*, *1660*, 050075. AIP Publishing LLC.

Mckinney, M. F., & Breebaart, J. (2003). Features for Audio and Music Classification. *Proc ISMIR*, *4*, 151–158.

Mehta, A. H., Lu, H., & Oxenham, A. J. (2020). The Perception of Multiple Simultaneous Pitches as a Function of Number of Spectral Channels and Spectral Spread in a Noise-Excited Envelope Vocoder. *JARO - Journal of the Association for Research in Otolaryngology*, *21*(1), 61–72.

Mehta, A. H., & Oxenham, A. J. (2017). Vocoder Simulations Explain Complex Pitch Perception Limitations Experienced by Cochlear Implant Users. *JARO - Journal of the Association for Research in Otolaryngology*, *18*(6), 789–802.

Meyer, T. A., & Svirsky, M. A. (2000). Speech perception by children with the CLARION (CIS) or nucleus 22 (speak) cochlear implant or hearing aids. *Annals of Otology, Rhinology and Laryngology*, *109*(12 II SUPPL.), 49–51.

Miyamoto, R. T., Osberger, M. J., & Kessler, K. (1996). Cochlear implant in aural re (habilitation) of adults and children. *J Otolaryngology Head and Neck Surgery*, *116*, 1142–1152.

Moisik, S. R. (2013). Harsh voice quality and its association with blackness in popular American media. *Phonetica*, *69*(4), 193–215.

Moore, B. C. J., & Carlyon, R. P. (2005). Perception of Pitch by People with Cochlear Hearing Loss and by Cochlear Implant Users. In *Pitch* (pp. 234–277). https://doi.org/10.1007/0-387-28958-5_7

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Moore, D. R., & Shannon, R. V. (2009). Beyond cochlear implants: Awakening the deafened brain. *Nature Neuroscience*, *12*(6), 686–691.

Naranjo, N. V., Lara, E. M., Rodríguez, I. M., & García, G. C. (1994). High-frequency components of normal and dysphonic voices. *Journal of Voice*, *8*(2), 157–162.

O'Leidhin, E., & Murphy, P. (2005). Analysis of spectral measures for voiced speech with varying noise and pertubation levels. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *I*, 869–872.

Oates, J. (2009). Auditory-perceptual evaluation of disordered voice quality. *Folia Phoniatrica et Logopaedica*, *61*(1), 49–56.

Ogden, R. (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association*, *31*(1), 139–152.

Oxenham, A. J., & Kreft, H. A. (2014). Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing. *Trends in Hearing*, *18*, 1–14.

Panek, D., Skalski, A., Gajda, J., & Tadeusiewicz, R. (2015). Acoustic Analysis Assessment in Speech Pathology Detection. *International Journal of Applied Mathematics and Computer Science*, *25*(3), 631–643.

Park, S. J., Sigouin, C., Kreiman, J., Keating, P., Guo, J., Yeung, G., … Alwan, A. (2016). Speaker identity and voice quality: Modeling human responses and automatic speaker recognition. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *08-12-Sept*, 1044–1048.

Pisoni, D. B. (1992). Talker normalization in speech perception. *Speech Perception, Production*

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

and Linguistic Structure, (1974), 143–151.

Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*, *11*(4), 478–504.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and Phonetic Experiments on American English Dialect Identification. *Journal of Language and Social Psychology*, *18*(10), 10–30.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. C. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, *28*(5), 340–344.

Rabiner, L. R., & Schafer, R. W. (1978). *Digital Processing of Speech Signals*. New Jersey: Prentice-Hal.

Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, *29*(4), 407–429.

Reilly, R. B., Moran, R., & Lacy, P. (2004). Voice pathology assessment based on a dialogue system and speech analysis. *Proceedings of the AAAI Fall Symposium on Dialogue Systems for Health Communication, Washington DC*, 104–109.

Rilliard, A., Shochi, T., Martin, J. C., Erickson, D., & Aubergé, V. (2009). Multimodal Indices to Japanese and French Prosodically Expressed Social Affects. *Language and Speech*, *52*(2–3), 223–243.

Rontal, E., Rontal, M., & Rolnick, M. I. (1975). Objective evaluation of vocal pathology using voice spectrography. *Annals of Otology, Rhinology & Laryngology*, *84*(5), 662–671.

Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, *133*(4), 2431–2443.

Roy, N., & Leeper, H. A. (1993). Effects of the manual laryngeal musculoskeletal tension reduction technique as a treatment for functional voice disorders: Perceptual and acoustic measures. *Journal of Voice*, *7*(3), 242–249.

S. Davis and P. Mermelstein. (1980). Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *28*(4), 357–366.

Santos, J. F., Cosentino, S., Hazrati, O., Loizou, P. C., & Falk, T. H. (2013). Objective speech intelligibility measurement for cochlear implant users in complex listening environments. *Speech Communication*, *55*(7–8), 815–824.

Scott, S., & McGettigan, C. (2015). The voice: From identity to interactions. In *APA Handbook of Nonverbal Communication*.

Shaneh, M., & Taheri, A. (2009). Voice command recognition system based on MFCC and VQ algorithms. *World Academy of Science, Engineering and Technology*, *57*, 534–538.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303–304.

Shannon, Robert V., Fu, Q. J., & Galvin, J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Oto-Laryngologica, Supplement*, *124*(552), 50–54.

Sicoli, M. A. (2007). *Tono: A linguistic ethnography of tone and voice in a Zapotec region*

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(Doctoral dissertation).

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

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

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

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

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

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Svirsky, M. A. (2017). Cochlear implants and electronic hearing. *Physics Today*, *70*(8), 53–58.

Tamati, T. N., Janse, E., Pisoni, D. B., & Baskent, D. (2017). Talker variability in real-life speech recognition by cochlear implant users. *The Journal of the Acoustical Society of America*, *141*(5), 2017–2020.

Terasawa, H., Slaney, M., & Berger, J. (2005). Perceptual Distance in Timbre Space. *Proceedings of ICAD 05-Eleventh Meeting Ofthe International Conference on Auditory Display*, 6–9.

Thomas, E. R., & Reaser, J. (2004). Delimiting perceptual cues used for the ethnic labeling of African American and European American voices. *Journal of Sociolinguistics*, *8*(1), 54–87.

Tsai, C. G., Wang, L. C., Wang, S. F., Shau, Y. W., Hsiao, T. Y., & Auhagen, W. (2010). Aggressiveness of the growl-like timbre: Acoustic characteristics, musical implications, and biomechanical mechanisms. *Music Perception*, *27*(3), 209–222.

Tyler, R. S., Teagle, H. F., Kelsay, D. M., Gantz, B. J., Woodworth, G. G., & Parkinson, A. J. (2000). Speech perception by prelingually deaf children after six years of cochlear implant use: effects of age at implantation. *Annals of Otology, Rhinology & Laryngology*, *109*(12_suppl), 82–84.

Umapathy, K., Krishnan, S., Parsa, V., & Jamieson, D. G. (2005). Discrimination of pathological voices using a time-frequency approach. *IEEE Transactions on Biomedical Engineering*, *52*(3), 421–430.

Umapathy, S., Rachel, S., & Thulasi, R. (2018). Automated speech signal analysis based on feature extraction and classification of spasmodic dysphonia: a performance comparison of different classifiers. *International Journal of Speech Technology*, *21*(1), 9–18.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted September 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Vongpaisal, T., Trehub, S. E., Schellenberg, E. G., Lieshout, P. Van, & Papsin, B. C. (2010). Children With Cochlear Implants Recognize Their Mother's Voice. *Ear and Hearing*, *31*(4), 555–566.

Winn, M. B., & Litovsky, R. Y. (2015). Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *The Journal of the Acoustical Society of America*, *137*(3), 1430–1442.

Wolfe, V., Cornell, R., & Palmer, C. (1991). Acoustic Correlates of Pathologic Voice Types. *Journal of Speech Language and Hearing Research*, *34*(3), 509.

Wolfe, V. I., & Bacon, M. (1971). Spectrographic comparison of two types of spastic dysphonia. *Journal of Speech and Hearing Disorders*, *41*(3), 325–332.

Xiang, S., Nie, F., & Zhang, C. (2008). Learning a Mahalanobis distance metric for data clustering and classification. *Pattern Recognition*, *41*(12), 3600–3612.

Yanagihara, N. (1967). Significance of harmonic changes and noise components in hoarseness. *Journal of Speech, Language, and Hearing Research*, *10*(3), 531–542.

Yumoto, E., Gould, W. J., & Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *The Journal of the Acoustical Society of America*, *71*(6), 1544-1550.

Zimman, L. (2012). Voices in transition: Testosterone, transmasculinity, and the gendered voice among female-to-male transgender people. *Linguistics Graduate Theses & Dissertations*, (24), 1–253.