Manuscript (TeX or Word only)
medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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.
Click here to access/download;Manuscript (TeX or Word only);VoiceQuality_CochlearImplant.docx

Running title: Talkers' voice quality through cochlear implant

Simulated cochlear-implant processing reveals major loss of acoustic information relevant to talkers' voice quality differences [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

Date of manuscript submission: June 3, 2020

[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_khalilarjmandi@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NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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 July 5, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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: Talkers' voice quality through cochlear implant

ABSTRACT

The ability to discern variations in talkers' voice quality 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 cochlear implant (CI) speech processing. This study analyzed unprocessed and CI-simulated versions of sustained vowel sounds /a/ from two groups of individuals with normal and disordered voice qualities to investigate the effects of CI speech processing on acoustic information relevant to the talkers' voice quality distinction. The CI-simulated stimuli were created by processing the vowel sounds using 4-, 8-, 12-, 16-, 22-, and 32-channel noise-vocoders. The voice quality for each stimulus was characterized by calculating *mel-frequency cepstral coefficients* (MFCCs). Then, the effects of CI speech processing on the acoustic distinctiveness between normal and disordered voices was measured by calculating the *Mahalanobis distance* and classification accuracy of *support vector machines* (SVMs) on their MFCC features. The results showed that CI noise vocoding is substantially detrimental to acoustic information involved in voice quality distinction, suggesting that CI listeners likely experience difficulties in perceiving voice quality variations. The results underscore challenges CI users may face for effective recognition of talkers and processing their speech.

Keywords: Talkers' voice quality, cochlear implant, acoustic distance

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

I. INTRODUCTION

The abundant spectro-temporal cues in speech serve as rich sources of information for listeners to learn and retrieve a variety of linguistic and indexical cues important for robust speech comprehension and language development. Talkers' voice quality is one such aspect that listeners with normal hearing (NH) have access for capturing several indexical and sociolinguistics information and, thus establish successful spoken communication. Variations in talkers' voice quality facilitate human spoken communication in multiple ways. These 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). Furthermore, listeners use acoustic information relevant to talkers' voice quality to encode and discern various indexical information related to talkers' identity such as gender, age, and race (Abberton & Fourcin, 1978; Eidsheim, 2012; Laver, 1968). This encoded information is actively incorporated by listeners to identify and recognize talkers (Gussenhoven & Rietveld, 1998; Latinus & Belin, 2011; Van Dommelen, 1987) and understand their speech (e.g., Creel & Tumlin, 2011). Talkers' voice quality also contributes to speech understanding through constructing stance in communicative interaction, particularly by conveying talkers' affective states during communication (e.g., happiness, anger, etc.; Aubergé & Cathiard, 2003; Guzman, Correa, Muñoz, & Mayerhoff, 2013; Podesva, 2007; Sicoli, 2010; Tsai et al., 2010; Zimman, 2012). In addition, listeners use talkers' voice quality to infer their medical conditions (i.e., physical, psychological, and mental status; e.g., Kreiman, Gerratt, Kempster, Erman, & Berke, 1993; Laver, 1968). While listeners with normal hearing (NH) can often readily infer this information from talkers' voice, little is known about how CI speech processing may impact faithful transmission of acoustic information relevant to talkers' voice quality distinction.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

The present study examined unprocessed and CI-simulated sustained vowel sound /a/ uttered by two groups of individuals with normal and disordered voice qualities to examine the effect of CI speech processing on voice quality distinction.

A. Talkers' voice quality and robust speech comprehension

Acoustic information relevant to talkers' voice quality signals a range of attributes, which together facilitates robust perception of speech. In facts, variations in talkers' voice translate into perceptual attributes of voice quality such as breathiness, roughness, creakiness, and nasality (Childers & Lee, 1991; Eskenazi, Childers, & Hicks, 1990; Klatt & Klatt, 1990), which can directly provide linguistic cues for speech sound contrast (Dicanio, 2009; Garellek & Keating, 2011; Gordon, 2001; Gordon & Ladefoged, 2001). For instance, speakers of English and Finish modify their voicing behaviors and use creaky voice to 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). Spectral measures such as *H1-H2*, *H1-A2*, and cepstral peak prominence (CPP) are acoustic correlates of talkers' voice quality that were shown to be able to reflect these variations and distinguish between phonation types (Dicanio, 2009; Garellek & Keating, 2011). Furthermore, access to acoustic information related to talkers' voice quality is crucial for effective speech comprehension because learning talker-specific information is essential for dealing with talker variability and, thus robust perception of speech (Johnson, 2005; Kleinschmidt & Jaeger, 2015; Pisoni, 1992). These findings highlight the importance of having access to voice quality-related acoustic information in talkers' voice for robust speech comprehension.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

B. Talkers' voice quality and talker identification

Robust speech comprehension is also indirectly facilitated through talker identification and recognition. Listeners actively attend to variations in the talkers' voice quality to index variability across talkers in their identity. Several studies have demonstrated the connection between talkers' voice quality and their gender (e.g., Gussenhoven, 2004; Ohala, 1983; Puts, Hodges, Cárdenas, & Gaulin, 2007), race (Alim, 2004; Irwin, 1977; Moisik, 2013; Thomas & Reaser, 2004), ethnicity (Purnell, Idsardi, & Baugh, 1999), social group (Esling, 1978; Sicoli, 2007; Stuart-Smith, 1999), and cultural status (Rilliard, Shochi, Martin, Erickson, & Aubergé, 2009; Stross, 2013), which are all indices of talkers' identity (Eidsheim, 2012; Podesva & Callier, 2015). Perceptual and acoustic correlates of voice quality such as pitch, phonation (e.g., breathiness, harshness, and creakiness), formant frequencies, and difference between amplitude of the first and second harmonics (*H1-H2*) have been shown to reflect talkers anatomical differences in the size of their larynx and vocal tracts, and thus are incorporated by listeners to identify talkers' voice (Henton & Bladon, 1985; Ohala, 1983; Simpson, 2009). Therefore, preserving the integrity of spectro-temporal information in talkers' voice is important for robust understanding of their speech by listeners.

Listeners are able to determine talkers' gender from listening to isolated vowels (Coleman, 1971), as well as to continuous speech (Gelfer & Bennett, 2013). In addition, studies on African American English demonstrated a connection between talkers' voice quality and their race (Irwin, 1977; Thomas & Reaser, 2004). For example, 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). Purnell et al., (1999) showed that landlords use acoustic information from a short segment of speech in the word "hello" to identify tenants with African American racial background and inappropriately discriminate against them. Acoustic analysis of African American

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

vowels showed that this racial judgment is likely facilitated through acoustic features relevant to talkers' phonatory and articulatory behaviors (Arjmandi, Dilley, & Wagner, 2018; Thomas, 2007). Recognition of talkers' social class is another dimension whereby voice quality contributes to the recognition of talkers' identity (Esling, 1978; Henton, 1986; Stuart-Smith, 1999). For example, the level of harshness and pharyngealization in talkers' voice were found to be associated with talkers' social class. It is therefore important for listeners to have access to acoustic information relevant to talkers' voice quality variations for talker identification and recognition, and thus effective speech understanding.

C. The role of talkers' voice quality in language development

Learning talker's voice quality also plays an important role in spoken language development. Children and adults incorporate acoustic information relevant to talkers' voice quality to develop and construct talker-specific identity. The ability to map talker-specific acoustic information to talker identity is fundamental to children's successful language development. Infants use talker-specific information for word recognition and language learning (Houston & Jusczyk, 2000, 2003; Ramírez-Esparza, García-Sierra, & Kuhl, 2017), suggesting that they likely develop mental representations specific to talkers' voice information. Time-course of attending to talker-specific cues in speech begins very early before birth where infants recognize their mother's voice (Kisilevsky et al., 2003) and was also shown in newborn infants who preferred listening to their mothers voice over that of a stranger (DeCasper & Fifer, 1980). Models of long-term memory for voice quality perception suggest that listeners encode and use patterns of voice quality variations in talkers' speech for perception of voice quality (Papcun, Kreiman, & Davis, 2005). Talkers' voice quality contributes to learning several other information about talkers such as their physical status (e.g., body size; Podesva & Callier, 2015), behavioral traits (e.g., smoking behaviors), and

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

their physical, psychological and mental health (e.g., vocal folds disorders, anxiety level, mood; e.g., Kreiman, Vanlancker-Sidtis, & Gerratt, 2005). One example of relying on voice quality for medical diagnosis is the auditory perceptual assessment of individuals' voice by speech-language pathologists (SLPs) for differential diagnosis of disordered voices (Kreiman & Gerratt, 1996; Kreiman, Gerratt, & Ito, 2007). Although some of this information is partially retrievable in the absence of voice quality cues (e.g., Fellowes, Remez, & Rubin, 1997), the significant contribution of these cues in multiple aspects from establishing robust and effective spoken communications to differential diagnosis of individuals with disordered voice quality is indisputable.

D. Talkers' voice quality through CIs

While listeners with NH have access to acoustic information related to talkers' voice quality, it is yet unknown how users of cochlear implants may be at disadvantage in receiving this information due to limited spectral resolution in CI speech processing. Cochlear implants device has enabled many individuals with severe-to-profound sensorineural hearing loss to experience a partially restored hearing and develop language. However, due to the limited spectral bands in the vocoding process imposed by CI speech processing, listeners with CIs have partial access to fine-grained acoustic information in speech (Cleary, Pisoni, & Kirk, 2005; Houston, Pisoni, Kirk, Ying, & Miyamoto, 2003; Houston et al., 2012; Svirsky, 2017). Lack of access to these fine-grained acoustic cues such as pitch and low-frequency voice harmonics (Chatterjee & Peng, 2008; Deroche, Kulkarni, Christensen, Limb, & Chatterjee, 2016; Fu, Chinchilla, & Galvin, 2004; Gaudrain & Baskent, 2018; McDermott, 2004), as well as temporal fine structure cues from individual harmonics (Carlyon, Deeks, & McKay, 2010) may disrupt the ability of listeners with CIs to receive information contributing to signaling variations in talkers' voice quality.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

Listeners utilize acoustic information in snippets of talker's voice during vowel phonation to draw inferences about talkers' voice quality (e.g., Gerratt, J. Kreiman, 2016) and identity (Adams, 2009; Ladefoged & Broadbent, 1957; Purnell et al., 1999; Thomas, 2007). Acoustic information derived from sustained phonation of a vowel sound may reflect structural and functional properties of vocal folds vibrations (static and/or dynamic) and carries important information about identity and quality of talkers' voice (Ladefoged & Broadbent, 1957; Morrison & Assmann, 2012; Winn, Chatterjee, & Idsardi, 2012). For example, fundamental frequency (F_0) and low-frequency harmonic components (e.g., *H1-H2*, *H1-A1*) are acoustic cues that reflect changes in the quality of talkers' voice (e.g., breathy, strained, rough, etc.), and thus contribute to recognition of talkers' identity (Chatterjee & Peng, 2008; Deroche et al., 2016; Fu et al., 2004; Gaudrain & Baskent, 2018; Gelfer & Bennett, 2013; McDermott, 2004; Van Dommelen, 1987). A similar role in talker recognition was shown for pitch dynamics (i.e., intonational patterns; Abberton & Fourcin, 1978). Chang et al., (2006) showed that the voice pitch cue in the low-frequency spectrum and the temporal envelop cues are important for segregation of the target voice from the competing voice and performing a robust recognition of speech. These acoustic cues, however, are poorly perceived by CI users due to the vocoding process in CI speech processing that involves a limited number of frequency channels (Carlyon et al., 2010; Laneau, Wouters, & Moonen, 2004). This was observed in prelingually deafened children with CIs who showed poorer performance compared to their NH peers in distinguishing between talkers with similar voices (Cleary & Pisoni, 2002). Low-numbered resolved harmonics are not faithfully transmitted via CI device, which negatively affects the performance of CI users in speaker identification and recognition (Gaudrain & Baskent, 2018), perception of prosodic information (Davidson, Geers, Uchanski, & Firszt, 2019; Deroche et al., 2016), and speech understanding in multi-talker

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

conditions (Rosen, Souza, Ekelund, & Majeed, 2013; Stickney, Zeng, Litovsky, & Assmann, 2004). Taken together, these prior findings indicate that some aspects of voice quality cues may be degraded through CI processing. Such findings may represent a combination of altered input at peripheral and central auditory pathways for sound and language processing. We currently have limited evidence about how CI speech processing may be detrimental to acoustic information relevant to talkers' voice quality variations. Degradation of this information potentially disrupts the bridge between recognizing talkers' voice and effective processing of their speech.

E. Present study

The present study analyzed voices from two groups of talkers with normal and disordered voice qualities to investigate the effect of simulated cochlear-implant speech processing and the number of spectral channels on the acoustic information involved in distinguishing between these two classes of voice qualities. Noise-vocoder was used to simulate the limited resolution of CIs in representation of acoustic information relevant to voice quality distinction. Voices spoken by individuals with voice disorders (e.g., vocal fold polyps, nodules, etc.) represent an instance where voice quality is abnormally altered compared to the voice of individuals with healthy vocal folds function (Arjmandi & Pooyan, 2012; Arjmandi, Pooyan, Mikaili, Vali, & Moqarehzadeh, 2011; Ghasemzadeh & Arjmandi, 2019; Umapathy, Krishnan, Parsa, & Jamieson, 2005). Abnormal physiological, neurological, and/or functional changes in vocal folds affect talkers' voice quality, which in turn would be reflected as certain variations in the properties of talkers' voice spectrum (Arjmandi & Pooyan, 2012; Behroozmand & Almasganj, 2007; Eskenazi et al., 1990; Fukazawa, el-Assuooty, & Honjo, 1988; Hammarberg, Fritzell, Gaufin, Sundberg & Wedin, 1980; Hanson, 1997; Klatt & Klatt, 1990; Kreiman, Gerratt, & Berke, 1994; Kreiman, Gerratt, & Precoda, 1990; Krom, 1995; O'Leidhin & Murphy, 2005; Sasaki & Okamura, 1984; Yumoto, Gould, & Baer,

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

1982). These abnormalities, that signal the distinction between disordered and normal voice qualities, generally appear as time-varying noise-like patterns in different ranges of low-, mid- and high-frequency bands of voiced signals (e.g., Arjmandi & Pooyan, 2012; Behroozmand & Almasganj, 2007; Fukazawa, el-Assuooty, & Honjo, 1988; Hammarberg, Fritzell, Gaufin, Sundberg, & Wedin, 1980; Hanson, 1997; Klatt & Klatt, 1990; O'Leidhin & Murphy, 2005; Sasaki & Okamura, 1984; Yumoto, Gould, & Baer, 1982). Spectral changes in the vowel spectrum have shown to be important in monitoring variations in talkers' voice quality such as breathiness, laryngealization, harshness, and nasality (Gerratt, Kreiman, 2016; Coleman, 1971; Klatt & Klatt, 1990; Park et al., 2016). In the present study, we investigated the variations in the magnitude spectrum of normal and disordered voices to understand how CI speech processing may impact the distinctive patterns of the voice quality. We further quantified the acoustic distance between normal and disordered voices as a function of the number of spectral channels to study the effect of the number of spectral channels on the degradation of voice quality information.

II. MATERIALS

A. Voice samples

The voice samples in this study were sustained vowel sounds /a/ 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/ while their 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 phonations. The vowel sounds from 293

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

individuals were analyzed where 53 talkers had normal voice quality (21 Males) and the remaining 240 talkers (96 Males) were diagnosed with one or multiple voice disorders. The voice disorders were the result of abnormal physiological, neurological, and/or functional changes that affected the regular function of talkers' vocal folds.

B. Creation of noise-vocoded voice samples

The CI-simulated versions of the unprocessed voice stimuli were created using a noise-excited envelope vocoder in AngelSim^{TM} Cochlear Implant and Hearing Loss Simulator (Fu, 2019; Emily Shannon Fu Foundation, www.tigerspeech.com), which follows the CI-simulated vocoding process in prior studies (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). The process involved dividing each voice spectrum into a variable number of logarithmically-spaced frequency bands between absolute lower and higher-frequency of 200 Hz and 7000 Hz (24 dB/Octave analysis filter slopes), corresponding to the frequency-place map simulated by Greenwood function (Greenwood, 1990). These frequency limits approximate the corner frequencies in the 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 to simulate the performance of the average CI listeners 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 filtered using a similar 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 course-grained temporal structures. The quality of CI-simulated voice is a function of the number of spectral

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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 in the vocoder (e.g., Friesen, Shannon, Baskent, & Wang, 2001; Xu, Thompson, & Pfingst, 2005). The noise-excited envelope vocoder was used in *AngelSim* software to process unprocessed/natural voices and create their noise-vocoded versions with 4-, 8-, 12-, 16-, 22-, and 32-channel. Therefore, the simulated cochlear implant voices were created with seven levels of spectral degradation (unprocessed, 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.

III. METHODS

A. Analysis of voice spectra

We first explored the average spectra of two groups of normal and disordered voice quality to understand how noise-vocoder and the number of spectral channels affect the distinctive features of two classes of voice qualities within different 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 linear predictive coding (LPC) with 12th order (Rabiner & Schafer, 1978). These average spectra were comparatively used to investigate how the distribution of energy of the voice signal is changed over various frequency regions due to CI noise vocoding process. In fact, the variations in the characteristics of the average voice spectra were investigated

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

under seven levels of spectral degradations (unprocessed, 32, 22, 16, 12, 8, and 4-channels noise vocoder) to identify how CI filtering process affects the acoustic information signaling distinction between two classes of normal and disordered voice quality.

B. Using MFCC features to characterize acoustic information relevant to voice quality variation

We used mel-frequency cepstral coefficients (MFCCs) to characterize variations in voice quality for normal and disordered voices and their corresponding CI-simulated versions. The CI-simulated voices were created at 6 levels of spectral degradations (i.e., number of spectral channels). These features were inspired by human auditory processing and were originally developed to 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). These acoustic features have been effective in representation of the variations in voice quality, particularly in discrimination between 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 even absent in the CI-simulated voices. Using MFCC features, we were able to deal with such methodological challenges.

Fig. 1A shows the schematic of the approach for calculating MFCCs for samples of normal and disordered voices. To calculate MFCCs, each vowel stimulus /a/ (i.e., S_{n_i} or S_{d_j}) was first segmented into frames of 30 ms with a frame shift of 15 ms. Here the *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,

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

…,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, the mel-filterbank were generated and applied to the voice power spectrum. MFCCs were eventually derived by calculating the DCT of the logarithm of all filterbank energies (Rabiner & Schafer, 1978). Eventually, the first twelve components were preserved as MFCC features for each frame of a voice stimulus (i.e., $MFCCs_{n_i}$ or $MFCCs_{d_j}$). Each voice stimulus was eventually represented by an MFCC matrix with the size of $F\times 12$, where F indicates the number of frames in each voice stimulus. The same procedure was performed to calculate MFCC matrices for CI-simulated versions of natural/unprocessed normal and disordered voices, as shown by dashed lines in Fig. 1A.

>>>>>>>>**Fig. 1 about here** <<<<<<<<

C. Acoustic distance quantification using Mahalanobis distance measure

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) 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). This multivariate statistical approach uses two feature matrices (or vectors) from two separate classes to evaluate the extent that the two classes are distinguished, after sphering the distance matrix between 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 relatively larger distance

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

between 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 voice). 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 (i.e., S_{n_i} or S_{d_j}) were characterized by a single, time-averaged MFCC vector (i.e., $\mu_{MFCCs_{n_i}}$ 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 sound (Mckinney & Breebaart, 2003; Davis and Mermelstein, 1980; Terasawa, Slaney, & Berger, 2005). Therefore, 53 normal voices were presented by 53 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 eventually measured by calculating 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 July 5, 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.

D. Acoustic distance quantification based on the classification accuracy of support vector machines (SVMs)

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). An application of SVM classifier is the evaluation of features in distinguishing between two classes, where the classification accuracy of SVMs classifier is used as a criterion for feature evaluation (Heijden, Ferdinand, Ridder, & Tax, 2005). We took advantage of this property of SVM to examine the effect of CI speech processing and the number of channels on the acoustic distinctiveness between two classes of normal and disordered voices as 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 to be used in 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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

imposed 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*. Compared to MD, the training phase in SVM allowed us to simulate the effect of exposure to voice samples from two classes of voice qualities in their acoustic distinction in a computational fashion. The training phase in SVM is expected to benefit from seen data in a supervised fashion to associate degraded patterns in CI-simulated voices to their corresponding class of voice quality (i.e., normal or disordered). This effect is basically very important in listeners with CIs as studies have shown that the central auditory system plays a significant role in decoding the degraded acoustic patterns at the outputs of CI electrodes in CI users (Basura, Hu, Juan, Tessier, & Kovelman, 2018; Fallon, Irvine, & Shepherd, 2008; Houston et al., 2012).

IV. RESULTS

A. Effects of CI noise vocoding on spectral information relevant to voice quality distinction

We first examined the vowel /a/ spectra from normal and disordered voices under 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 (orange 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/natural (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 noise vocoding process on spectral information that could signal differences in talkers' voice quality. The patterns of variation in average spectral energy of disordered voices compared to

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

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 orange or light gray line) in Fig. 1A reveals distinctive patterns of spectral energy within low-, mid-, and high-frequency regions. For unprocessed/natural 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 degraded low-frequency patterns in the disordered voice spectrum may be signs of partial closure of vocal folds. These differences in the spectral level in frequency bands covering the first formant (*e.g.*, breakdown in formant structure) can 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 disordered voice qualities compared to that of 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 be attributed to a rough voice quality in disordered voices (Krom, 1995). Finally, the higher degree of variability in voice spectrum across the average

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

spectrum in disordered class (the wider orange (light gray) shaded area) compared to the one for normal voice quality (blue (dark gray) shaded area) can be attributed to variable levels of abnormality in vocal folds vibrations in disordered voices, as well as the variety of the analyzed voice disorders.

More importantly, 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 22-channel in panel (C)) demonstrates that CI noise vocoding substantially degrades acoustic information involved in voice quality distinction. In general, the noise vocoding process in CI speech processing caused major loss of acoustic information distinctive of talkers' voice quality at low-, mid-, and high-frequency ranges of their voice spectra. The detrimental effect of CI noise-vocoder increases as the number of channels decreases to the extent that spectra of two classes of voice qualities become almost visually indistinguishable at 4-channel CI-simulated voices, highlighting the substantially 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 frequency regions is discarded due to the filtering process in CI. 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 low-numbered resolved harmonics, which are important for robust pitch perception (Bernstein & Oxenham, 2003; Smurzynski, 1990). The noise level elevated in the mid-frequency band as the number of spectral channels decreases in the CI-simulated voices, likely leading to lack of access to acoustic cues relevant to variations in talkers' voice quality in listeners with CIs.

>>>>>>>>**Fig. 2 about here** <<<<<<<<

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

B. Effects of CI noise vocoding on acoustic information distinctive of talkers' voice quality

As discussed in the Method section, MFCCs for each voice stimulus were calculated to characterize talkers' voice quality and measure the effect of CI noise-vocoding on acoustic information relevant to voice quality distinction. 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 energy variations at different frequency sub-bands.

>>>>>>>>**Fig. 3 about here** <<<<<<<<

Fig. 4 shows the calculated *Mahalanobis distance* 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 due to the CI noise vocoding process, suggesting that CI speech processing is detrimental to voice quality-related acoustic information. On average, there was an approximately 33% decline in MD due to CI noise-vocoding process when comparing the MD at 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 noise-vocoder potentially discards an important portion of acoustic information responsible for signaling variation in talkers' voice qualities. An unexpected 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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

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 CI 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.

>>>>>>>>**Fig. 4 about here** <<<<<<<<

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 cochlear implant 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. This decline is much smaller than the ~33% decline measured by MD, which potentially highlights the key role of classifier exposure to seen data during training phase. Classification accuracy in Fig. 5 shows three categories of performance between 80-85%, 85-90%, and 90-95%. These simulated results suggest that the current CI technology falls within the second category in terms of the number of spectral channels (12 channels in MED-EL devices, 16 channels in Advanced Bionics devices, and 22 channels in Cochlear), where the classification accuracy is still at least 5% below the one for the unprocessed condition. It is notable that classification accuracies for CI-simulated conditions fall within a good performance range (80% to 90%), even for a highly degraded CI-simulated voice with 4 spectral channels. We speculate that this

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

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 about here** <<<<<<<<

V. DISCUSSION

This study investigated how CI speech processing affects acoustic information involved in signaling variations in talkers' voice quality. We analyzed vowel sounds /a/ spoken by 53 talkers with normal voice and 240 talkers with disordered voice qualities at seven levels of spectral degradation, simulated by CI noise-excited envelope vocoder, to examine the effect of CI speech processing on acoustic information that distinguishes talkers' voice quality. To our knowledge, this is the first study that examines the effect of CI speech processing on acoustic information involved in conveying variations in talkers' voice qualities. Overall, the results of the current study highlighted the detrimental effect of CI noise vocoding process on acoustic information that signals voice quality contrast. This unfaithful transmission of acoustic information relevant to voice quality distinction through CIs may negatively impact the performance of CI listeners in identification and recognition of talkers' voice and, thus comprehension of their speech.

Our investigation of vowel /a/ spectra within different frequency sub-bands across two groups of normal and disordered voice qualities 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 noise-vocoder processing substantially degraded spectral information in low- (<2 kHz), mid- (~4-12 kHz), and high-frequency ranges (>12 kHz) that could contribute to voice

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

quality distinction. 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). Our results suggest that CI speech processing substantially degrades this distinctive spectral information, likely leading to a degraded perception of talkers' voice quality variations. These patterns of loss in voice quality-related information due to CI voice processing suggest that listeners with CIs potentially do not receive a large portion of acoustic information signaling changes in talkers' voice quality due to the partial transformation of fine-grained spectral structures.

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 calculating the distance between MFCC features using MD. The MDs 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 through CI. 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

The results of SVM accuracy in classification between normal and disordered voice qualities corroborated the detrimental effect of CI noise vocoding on acoustic information involved in voice quality distinction. However, the CI noise-vocoder processing and the number of channels in the CI noise-vocoder resulted in lower degrees of drop based on SVM classification compared to the amount of decline quantified by MD. These results also 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 highlights the detrimental effect of CI speech processing to acoustic information relevant to recognition of talkers' voice.

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, creakiness, and nasality), suggests that CI listeners may perform poorer than 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 noise vocoding substantially degrades spectral properties signaling voice quality

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

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 et al., 2016), and speech recognition in complex listening conditions such as multi-talker situations (Rosen et al., 2013; Stickney, Assmann, Chang, & Zeng, 2007; Stickney et al., 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 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 channels in the vocoder, which restricts spectral and temporal resolution of 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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

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.

Despite the limitations of CI 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 CI 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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

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 and language development (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 advantageous in general, these findings should be viewed as the general trend rather than the actual performance of CI listeners in perception of talkers' voice quality. 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 negatively impacts acoustic cues involved in voice quality transmission, leading to subsequent poor perception of talkers' voice quality in listeners with CIs. This degraded transmission of acoustic cues relevant to talkers' voice quality is particularly detrimental for speech processing and auditory scene analysis in prelingually deaf children who heavily rely on the bottom-up sensory information in speech to develop mental representations specific to talkers'

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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. 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, especially in pediatric CI users.

REFERENCES

Abberton, E., & Fourcin, A. J. (1978). Intonation and speaker identification. *Language and Speech*, *21*(4), 305–318. https://doi.org/10.1177/002383097802100405

Adams, C. A. (2009). *An Acoustic Phonetic Analysis of African American English : A Comparative Study of Two Dialects. (master's thesis)*. Eastern Michigan University, Ypsilanti, Michigan.

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. https://doi.org/10.1016/j.bspc.2015.02.008

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. https://doi.org/10.1109/IEEEGCC.2013.6705792

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*,

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

17(1), 149–151. https://doi.org/10.1525/jlin.2007.17.1.149

Anand, S., Kopf, L. M., Shrivastav, R., & Eddins, D. A. (2019). Objective Indices of Perceived Vocal Strain. *Journal of Voice*, *33*(6), 838–845. https://doi.org/10.1016/j.jvoice.2018.06.005

Arjmandi, M. K., 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 African American Dialect. *11th International Conference on Voice Physiology and Biomechanics*, (August), 36–37.

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. https://doi.org/10.1016/J.BSPC.2011.03.010

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. https://doi.org/10.1016/J.JVOICE.2010.08.003

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. https://doi.org/10.1109/IRANIANCEE.2010.5507106

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.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

Aubergé, V., & Cathiard, M. (2003). Can we hear the prosody of smile? *Speech Communication*, *40*(1–2), 87–97. https://doi.org/10.1016/S0167-6393(02)00077-8

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. https://doi.org/10.1121/1.5034171

Basura, G. J., Hu, X. S., Juan, J. S., Tessier, A. M., & Kovelman, I. (2018). Human central auditory plasticity: A review of functional near-infrared spectroscopy (fNIRS) to measure cochlear implant performance and tinnitus perception. *Laryngoscope Investigative Otolaryngology*. https://doi.org/10.1002/lio2.185

Behroozmand, R., & Almasganj, F. (2007). Optimal selection of wavelet-packet-based features using genetic algorithm in pathological assessment of patients' speech signal with unilateral vocal fold paralysis. *Computers in Biology and Medicine*, *37*(4), 474–485. https://doi.org/10.1016/j.compbiomed.2006.08.016

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. https://doi.org/10.1121/1.1572146

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. https://doi.org/10.1016/j.heares.2008.04.012

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

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. https://doi.org/10.1017/S0047404511000042

Brown, J., Geers, A., Herrmann, B., Kirk, I., Tomblin, J. B., & Waltzman, S. (2004). Cochlear Implants. *Asha Supplement*, 1–39. https://doi.org/10.1044/policy.TR2004-00041

Cameron, D. (2001). Designer voices. *Critical Quarterly*, *43*(4), 81–85. https://doi.org/10.1111/1467-8705.00392

Carlyon, R. P., Deeks, J. M., & McKay, C. M. (2010). The upper limit of temporal pitch for cochlear-implant listeners: Stimulus duration, conditioner pulses, and the number of electrodes stimulated. *The Journal of the Acoustical Society of America*, *127*(3), 1469–1478.

Chang, J. E., Bai, J. Y., & Zeng, F. G. (2006). Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise. *IEEE Transactions on Biomedical Engineering*, *53*(12), 2598–2601. https://doi.org/10.1109/TBME.2006.883793

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.

Cleary, M., & Pisoni, D. (2002). Talker discrimination by prelingually deaf children with cochlear implants: preliminary results. *Annals of Otology, Rhinology & Laryngology*, *111*(5_suppl),

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

113–118.

Cleary, M., Pisoni, D. B., & Kirk, K. I. (2005). Influence of Voice Similarity on Talker Discrimination in Children With Normal Hearing and Children With Cochlear Implants. *Journal of Speech Language and Hearing Research*, *48*(1), 204. https://doi.org/10.1044/1092-4388(2005/015)

Coleman, R. O. (1971). Male and Female Voice Quality and Its Relationship to Vowel Formant Frequencies. *Journal of Speech and Hearing Research*, *14*(3), 565–577.

Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language*, *65*(3), 264–285. https://doi.org/10.1016/j.jml.2011.06.005

Crew, J. D., & Galvin, J. J. (2012). Channel interaction limits melodic pitch perception in simulated cochlear implants. *The Journal of the Acoustical Society of America*, *132*(5), EL429–EL435.

Davidson, L. S., Geers, A. E., Uchanski, R. M., & Firszt, J. B. (2019). Effects of early acoustic hearing on speech perception and language for pediatric cochlear implant recipients. *Journal of Speech, Language, and Hearing Research*, *62*(9), 3620–3637. https://doi.org/10.1044/2019_JSLHR-H-18-0255

DeCasper, A. J., & Fifer, W. P. (1980). Of Human Bonding: Newborns Prefer their Mothers' Voices. *American Association for the Advancement of Science*, *208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928

Deroche, M. L. D., Kulkarni, A. M., Christensen, J. A., Limb, C. J., & Chatterjee, M. (2016).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

Deficits in the sensitivity to pitch sweeps by school-aged children wearing cochlear implants. *Frontiers in Neuroscience*, *10*(MAR), 1–15. https://doi.org/10.3389/fnins.2016.00073

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. https://doi.org/10.1017/S0025100309003879

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. https://doi.org/10.1121/1.4969743

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. https://doi.org/10.1006/jpho.1996.0023

Dolar, M. (2006). *A voice and nothing more*. MIT press. https://doi.org/10.1207/s15327868ms1001_3

Eddins, D. A., Anand, S., Lang, A., & Shrivastav, R. (2020). Developing Clinically Relevant Scales of Breathy and Rough Voice Quality. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2019.12.021

Eidsheim, N. (2012). Voice as Action: Toward a Model for Analyzing the Dynamic Construction of Racialized Voice. *Current Musicology*, *6*(93), 9–34.

Eskenazi, L., Childers, D. G., & Hicks, D. M. (1990). Acoustic correlates of vocal quality. *Journal*

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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 Speech and Hearing Research, *33*(2), 298–306. https://doi.org/10.1044/jshr.3302.298

Esling, J. (1978). The identification of features of voice quality in social groups. *Journal of the International Phonetic Association*, *8*(1–2), 18–23. https://doi.org/10.1017/S0025100300001699

Fallon, J. B., Irvine, D. R. F., & Shepherd, R. K. (2008). Cochlear implants and brain plasticity. *Hearing Research*, *238*(1–2), 110–117. https://doi.org/10.1016/j.heares.2007.08.004

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. https://doi.org/10.1016/S0892-1997(05)80308-X

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.

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America*, *110*(2), 1150–1163.

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. https://doi.org/10.1044/jslhr.4001.183

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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. (2019). *AngelSim: Cochlear implant and hearing loss simulator*. [Computer program], available at 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.

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. https://doi.org/10.1121/1.396012

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. https://doi.org/10.1017/S0025100311000193

Gaudrain, E., & Baskent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear and Hearing*, *39*(2), 226–237. https://doi.org/10.1097/AUD.0000000000000480

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. https://doi.org/10.1044/0161-1461(2002/015)

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. https://doi.org/10.1016/j.jvoice.2012.11.008

Gerratt, B. R., Kreiman, J., & Garellek, M. (2016). Comparing Measures of Voice Quality From

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

Sustained Phonation and Continuous Speech. *American Journal of Speech-Language Pathology*, *59*(5), 994–1001. https://doi.org/10.1044/2016

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. https://doi.org/10.1016/J.BSPC.2015.07.002

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. https://doi.org/10.1109/TASLP.2019.2959222

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. https://doi.org/10.1109/TBME.2006.871883

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. https://doi.org/10.1006/jpho.2001.0147

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. https://doi.org/10.1017/CBO9780511616983

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

Gussenhoven, C., & Rietveld, T. (1998). On the speaker-dependence of the perceived prominence of F0 peaks. *Journal of Phonetics*, *26*(4), 371–380. https://doi.org/10.1006/jpho.1998.0080

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. https://doi.org/10.1016/j.jvoice.2012.08.008

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. https://doi.org/10.1121/1.417991

Hanson, Helen M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America*, *101*(1), 466–481. https://doi.org/10.1121/1.417991

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., & Bladon, R. A. W. (1985). Breathiness in normal female speech: Inefficiency versus desirability. *Language and Communication*, *5*(3), 221–227. https://doi.org/10.1016/0271-5309(85)90012-6

Henton, Caroline G. (1986). Creak as a sociophonetic marker. *The Journal of the Acoustical Society of America*, *80*(S1), S50–S50. https://doi.org/10.1121/1.2023837

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic Correlates of Breathy Vocal

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

Quality. *Journal of Speech Language and Hearing Research*, *37*(4), 769. https://doi.org/10.1044/jshr.3704.769

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570–1582. https://doi.org/10.1037/0096-1523.26.5.1570

Houston, D. M., & Jusczyk, P. W. (2003). Infants' Long-Term Memory for the Sound Patterns of Words and Voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(6), 1143–1154. https://doi.org/10.1037/0096-1523.29.6.1143

Houston, D. M., Pisoni, D. B., Kirk, K. I., Ying, E. A., & Miyamoto, R. T. (2003). Speech perception skills of deaf infants following cochlear implantation: A first report. *International Journal of Pediatric Otorhinolaryngology*, *67*(5), 479–495.

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. https://doi.org/10.3766/jaaa.23.6.7.The

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. https://doi.org/10.1177/002383097702000307

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

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).

Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., … Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*(3), 220–224. https://doi.org/10.1111/1467-9280.02435

Kitzing, P., & Åkerlund, L. (1993). Long-time average spectrograms of dysphonic voices before and after therapy. *Folia Phoniatrica et Logopaedica*, *45*(2), 53–61. https://doi.org/10.1159/000266213

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). https://doi.org/10.1121/1.398894

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. https://doi.org/10.1067/mod.2000.109031

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

Kreiman, Jody, & Gerratt, B. R. (1996). The perceptual structure of pathologic voice quality. *The*

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

Journal of the Acoustical Society of America, *100*(3), 1787–1795. https://doi.org/10.1121/1.416074

Kreiman, Jody, Gerratt, B. R., & Berke, G. S. (1994). The multidimensional nature of pathologic vocal quality. *The Journal of the Acoustical Society of America*, *96*(3), 1291–1302. https://doi.org/10.1121/1.410277

Kreiman, Jody, Gerratt, B. R., & Ito, M. (2007). When and why listeners disagree in voice quality assessment tasks. *The Journal of the Acoustical Society of America*, *122*(4), 2354–2364. https://doi.org/10.1121/1.2770547

Kreiman, Jody, Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *Journal of Speech, Language, and Hearing Research*, *36*(1), 21–40.

Kreiman, Jody, Gerratt, B. R., & Precoda, K. (1990). Listener experience and perception of voice quality. *Journal of Speech and Hearing Research*, *33*(1), 103–115. https://doi.org/10.1121/1.2025887

Krom, G. 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. https://doi.org/10.1044/jshr.3804.794

Ladefoged, P., & Broadbent, D. E. (1957). Information Conveyed by Vowels. *The Journal of the Acoustical Society of America*, *29*(1), 98–104. https://doi.org/10.1121/1.1908694

Laneau, J., Wouters, J., & Moonen, M. (2004). Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. *The Journal of the*

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

Acoustical Society of America, *116*(6), 3606–3619. https://doi.org/10.1121/1.1823311

Latinus, M., & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, *2*(JUL), 1–12. https://doi.org/10.3389/fpsyg.2011.00175

Laver, J. D. M. (1968). Voice Quality and Indexical Information. *International Journal of Language & Communication Disorders*, *3*(1), 43–54. https://doi.org/10.3109/13682826809011440

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

McDermott, H. J. (2004). Music Perception with Cochlear Implants: A Review. *Trends in Amplification*, *8*(2), 49–82. https://doi.org/10.1177/108471380400800203

Mckinney, M. F., & Breebaart, J. (2003). Features for Audio and Music Classification. *Proc ISMIR*, *4*, 151–158. https://doi.org/10.1016/j.asoc.2010.08.017

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*

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

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. https://doi.org/10.1177/0003489400109S1220

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. https://doi.org/10.1159/000351059

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

Moore, D. R., & Shannon, R. V. (2009). Beyond cochlear implants: Awakening the deafened brain. *Nature Neuroscience*, *12*(6), 686–691. https://doi.org/10.1038/nn.2326

Morrison, G. S. & Assmann, P. F. (2013). *Vowel Inherent Spectral Change*. Springer Science & Business Media.

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. https://doi.org/10.1109/ICASSP.2005.1415252

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

https://doi.org/10.1017/S0025100301001116

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. https://doi.org/10.1177/2331216514553783

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. https://doi.org/10.1515/amcs-2015-0046

Papcun, G., Kreiman, J., & Davis, A. (2005). Long- term memory for unfamiliar voices. *The Journal of the Acoustical Society of America*, *85*(2), 913–925. https://doi.org/10.1121/1.397564

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. https://doi.org/10.21437/Interspeech.2016-523

Pisoni, D. B. (1992). Talker normalization in speech perception. *Speech Perception, Production 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. https://doi.org/10.1111/j.1467-9841.2007.00334.x

Podesva, R. J., & Callier, P. (2015). Voice quality and identity. *Annual Review of Applied*

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

Linguistics, *35*, 173–194. https://doi.org/10.1017/S0267190514000270

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. https://doi.org/10.1177/0261927X99018001002

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. https://doi.org/10.1016/j.evolhumbehav.2007.05.002

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

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). Look who's talking NOW! Parentese speech, social context, and language development across time. *Frontiers in Psychology*, *8*(JUN), 1–12. https://doi.org/10.3389/fpsyg.2017.01008

Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, *29*(4), 407–429. https://doi.org/10.1006/jpho.2001.0145

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. https://doi.org/10.1177/0023830909103171

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

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. https://doi.org/10.1177/000348947508400518

Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, *133*(4), 2431–2443. https://doi.org/10.1121/1.4794379

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. https://doi.org/10.1016/S0892-1997(05)80333-9

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. https://doi.org/10.1109/TASSP.1980.1163420

Sasaki, Y., & Okamura, H. (1984). Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness. *Journal of Speech and Hearing Research*, *27*(March), 2–6.

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-*

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

Laryngologica, Supplement, *124*(552), 50–54. https://doi.org/10.1080/03655230410017562

Sicoli, M. A. (2007). *Tono: A linguistic ethnography of tone and voice in a Zapotec region (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. https://doi.org/10.1017/S0047404510000436

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Linguistics and Language Compass*, *3*(2), 621–640. https://doi.org/10.1111/j.1749-818X.2009.00125.x

Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *Journal of the Acoustical Society of America*, *87*(1), 304–310. https://doi.org/10.1121/1.399297

Souza, P., Gehani, N., Wright, R., & McCloy, D. (2013). The advantage of knowing the talker. *Journal of the American Academy of Audiology*, *24*(8), 689–700.

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. https://doi.org/10.1121/1.2750159

Stickney, G. S., Zeng, F.-G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

recognition with speech maskers. *The Journal of the Acoustical Society of America*, *116*(2), 1081–1091. https://doi.org/10.1121/1.1772399

Stross, B. (2013). Falsetto voice and observational logic: Motivated meanings. *Language in Society*, *42*(2), 139–162. https://doi.org/10.1017/S004740451300002X

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

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. https://doi.org/10.1121/1.4989293

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. (2007). Phonological and Phonetic Characteristics in African American Vernacular English. *Language and Linguistics Compass*, *5*, 450–475.

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. https://doi.org/10.1111/j.1467-9841.2004.00251.x

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.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

(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. https://doi.org/10.1109/TBME.2004.842962

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. https://doi.org/10.1007/s10772-017-9471-8

Van Dommelen, W. A. (1987). The contribution of speech rhythm and pitch to speaker recognition. *Language and Speech*, *30*(4), 325–338. https://doi.org/10.1177/002383098703000403

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., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *The Journal of the Acoustical Society of America*, *131*(2), 1465–1479. https://doi.org/10.1121/1.3672705

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.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

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.

Xu, L., Thompson, C. S., & Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *The Journal of the Acoustical Society of America*, *117*(5), 3255–3267. https://doi.org/10.1121/1.1886405

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.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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 the approach 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 at seven levels of spectral degradation from no degradation (unprocessed vowel sound) to CI-simulated voices with 4 channels based on calculating *Mahalanobis distance* between MFCCs matrices of the two classes, (C) evaluate the acoustic distance between two classes of normal and disordered voice qualities as a function of levels of spectral degradation based on the classification accuracy derived from applying 5-fold validation to SVM classifier. The SVM classifiers were trained and tested on MFCCs features obtained from voice stimuli from two classes of normal and disordered voice qualities. The dashed lines refer to the process for creating and analyzing the CI-simulated versions of the unprocessed 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 orange (light gray) color show the paths for processing voices with disordered quality. S_{n_i} or S_{d_j} in panel (A) represent normal and disordered voice stimuli for subject number *i* (*i*={1,2,…, 53}) and subject number *j* (*j*={1,2,…, 240}), respectively. MFCCs_{n_i} and MFCCs_{d_j} are MFCC matrices extracted from frames of S_{n_i} and S_{d_j} signals. $\mu_{MFCCs_{n_i}}$ and $\mu_{MFCCs_{d_j}}$ in panel (B) and (C) are average 12-dimensional MFCCs obtained over frames of $MFCCs_{n_i}$ and $MFCCs_{d_j}$ matrices. *MD* in panel (B) is the *Mahalanobis distance* calculated on average MFCCs matrices derived from voice stimuli, belonging to two classes of normal and disordered voice qualities. Panel (C) shows the procedure of 5-fold cross-validation in 5 iterations applied to SVM to evaluate the effect of spectral degradation in CI-simulated voices on the acoustic separation between two classes of normal and disordered voices. F1 to F5 stand for Fold 1 to Fold 5 where the entire feature

medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

data (i.e., MFCCs features) was split into 5 folds; 4 folds for training the SVM classifier and the remaining fold for testing the SVM. Dashed rectangles indicate the folds for testing the classifier and solid rectangles refer to those for training.

Fig. 2. (Color online) Average magnitude spectra of voice stimuli with normal (blue or dark gray lines) and disordered (orange 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 orange (light gray) shaded area across the average orange line (light gray) for the group of disordered voices.

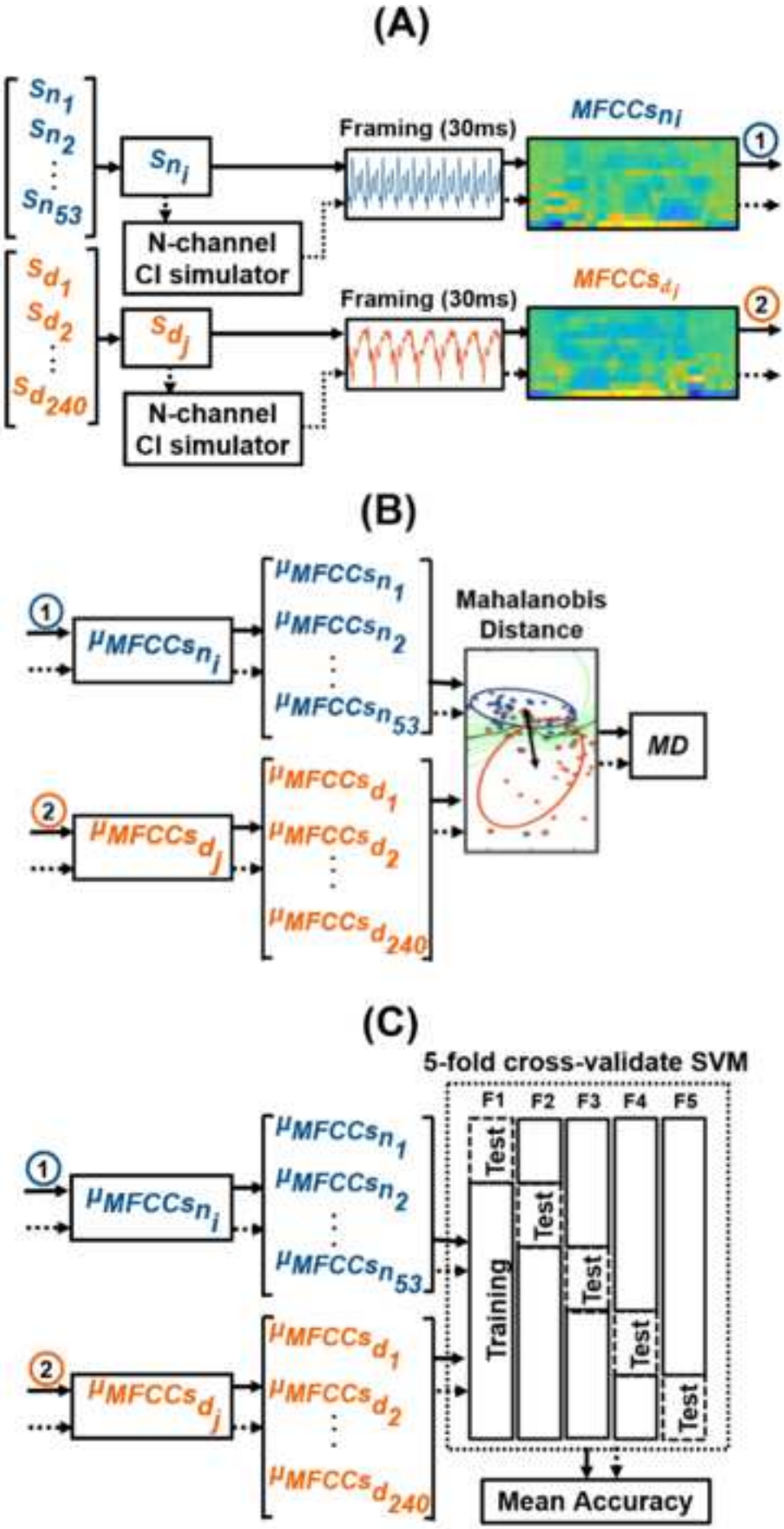
Fig. 3. (Color online) The filterbank of mel-spaced triangular filters (green or dark gray dotted lines) superimposed on average magnitude spectra of voices with normal (blue or dark gray line) and disordered (orange or light gray line) qualities. In this example, mel-filterbak contains 12 filters, which 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. (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). 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 "NT", which stands for natural/unprocessed stimuli.

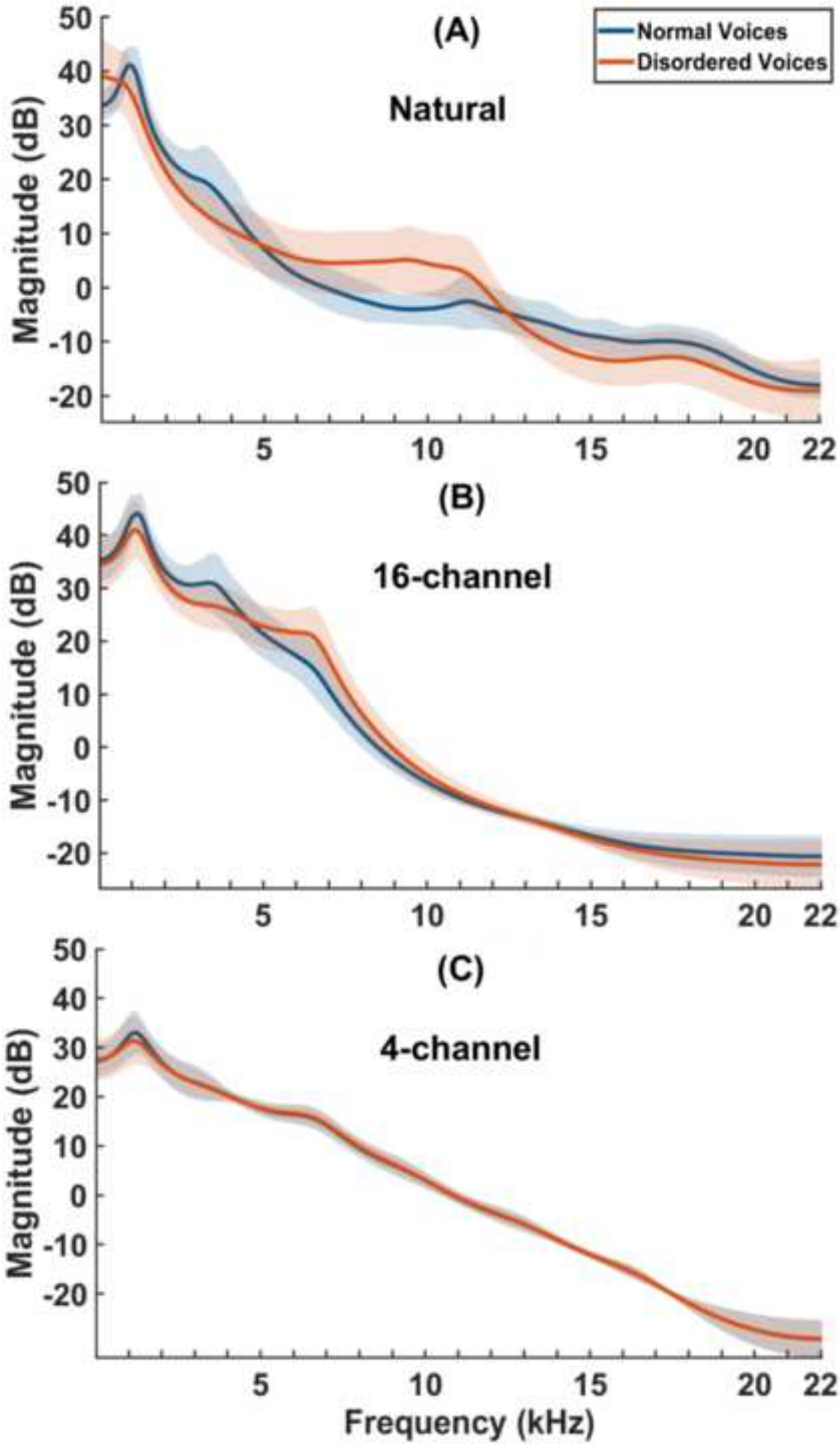
medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fig. 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, 32-, 22-, 16-, 12-, 8-, and 4-channels noise-vocoder). The unprocessed condition is labeled as "NT", which stands for natural/unprocessed stimuli.

Figures (PDF, TIFF, EPS, PS, or JPEG only) medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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. Click here to access/download Figures (PDF, TIFF, EPS, PS, or JPEG only) Figure 1.tif

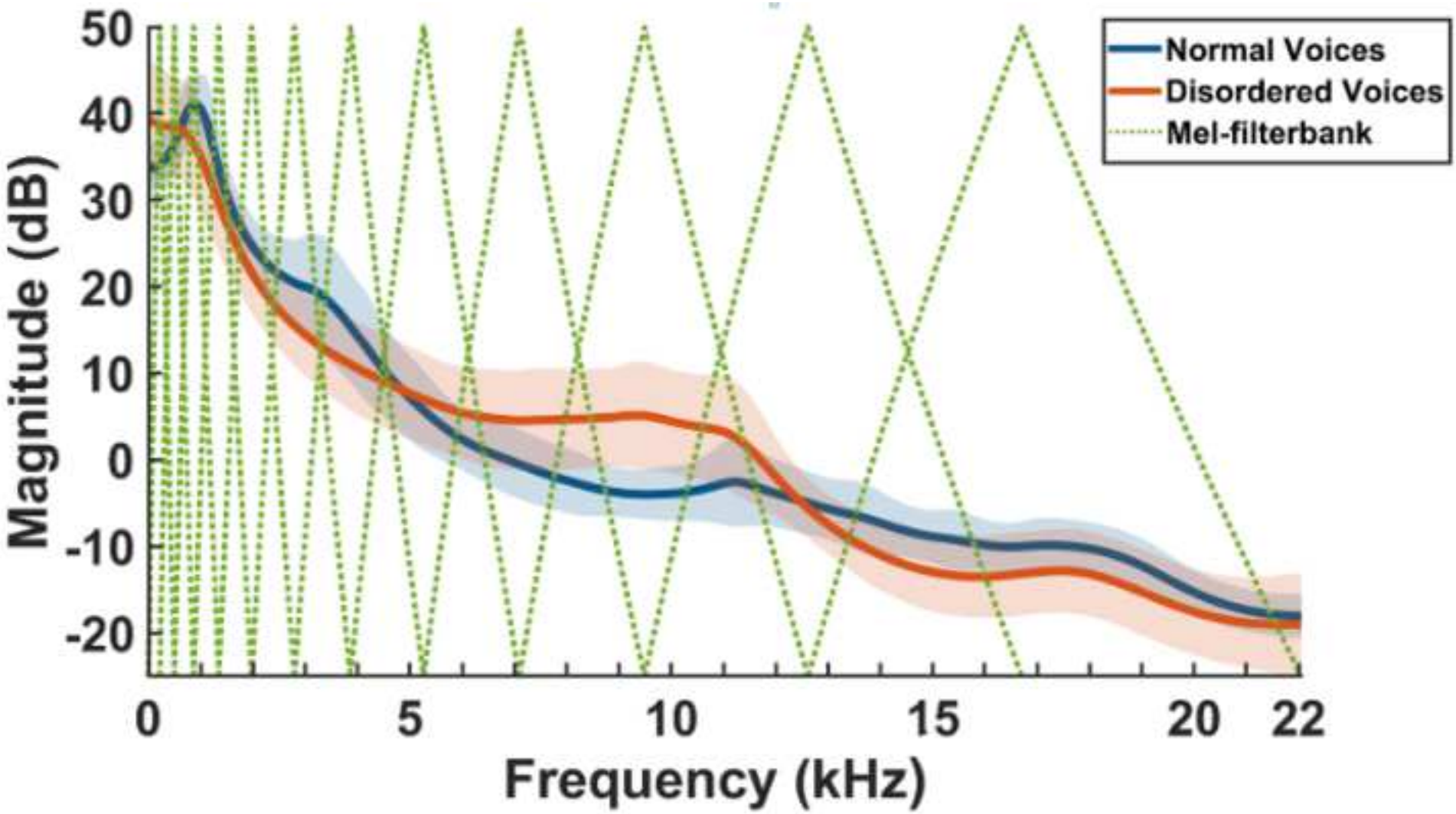


Figures (PDF, TIFF, EPS, PS, or JPEG only)
medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.
Click here to access/download Figures (PDF, TIFF, EPS, PS, or JPEG only) Figure2.tif



Figures (PDF, TIFF, EPS, PS, or JPEG only)

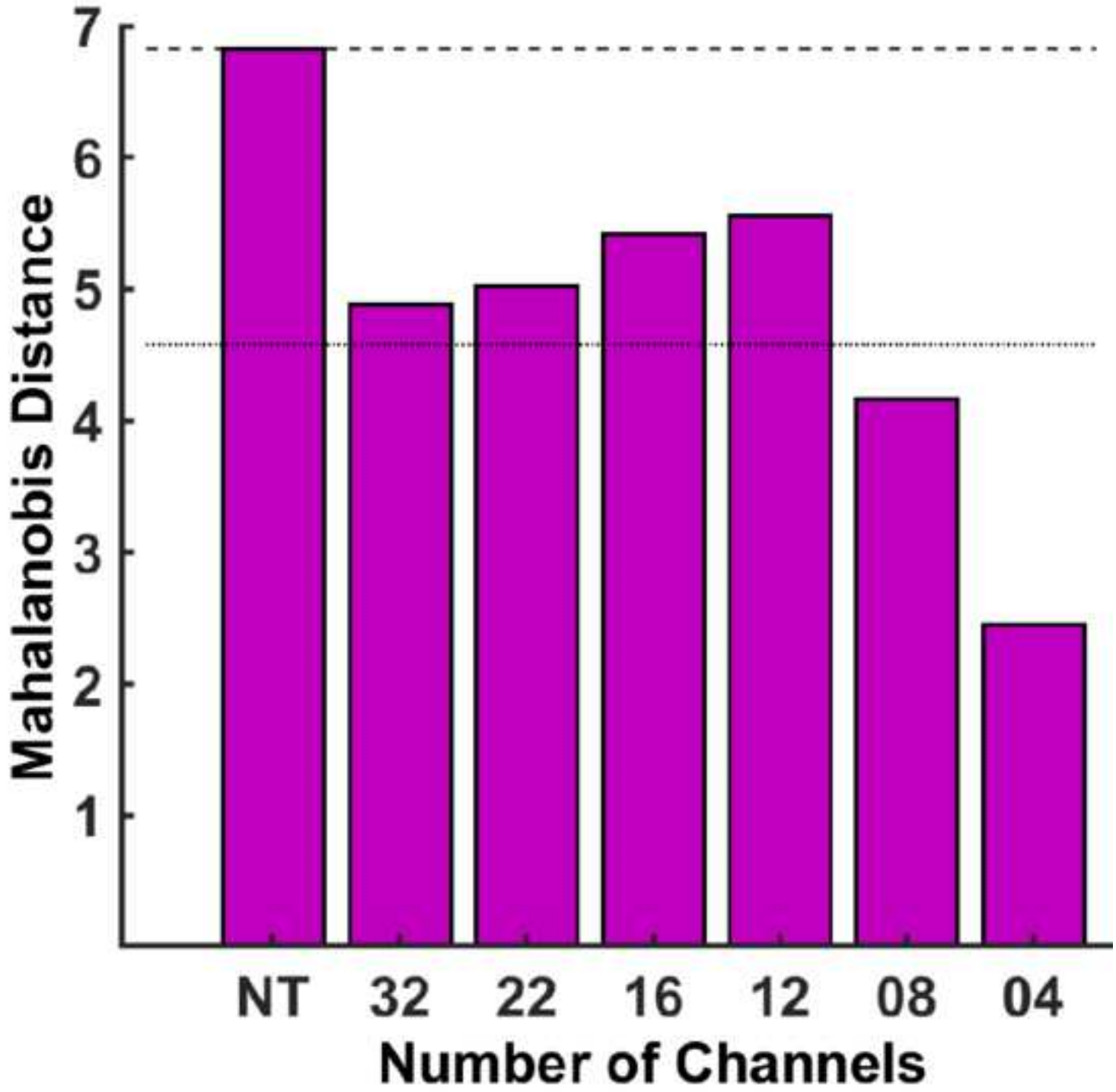
Click here to access/download;Figures (PDF, TIFF, EPS, PS, or JPEG only);Figure3.tif



medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

Figures (PDF, TIFF, EPS, PS, or JPEG only)

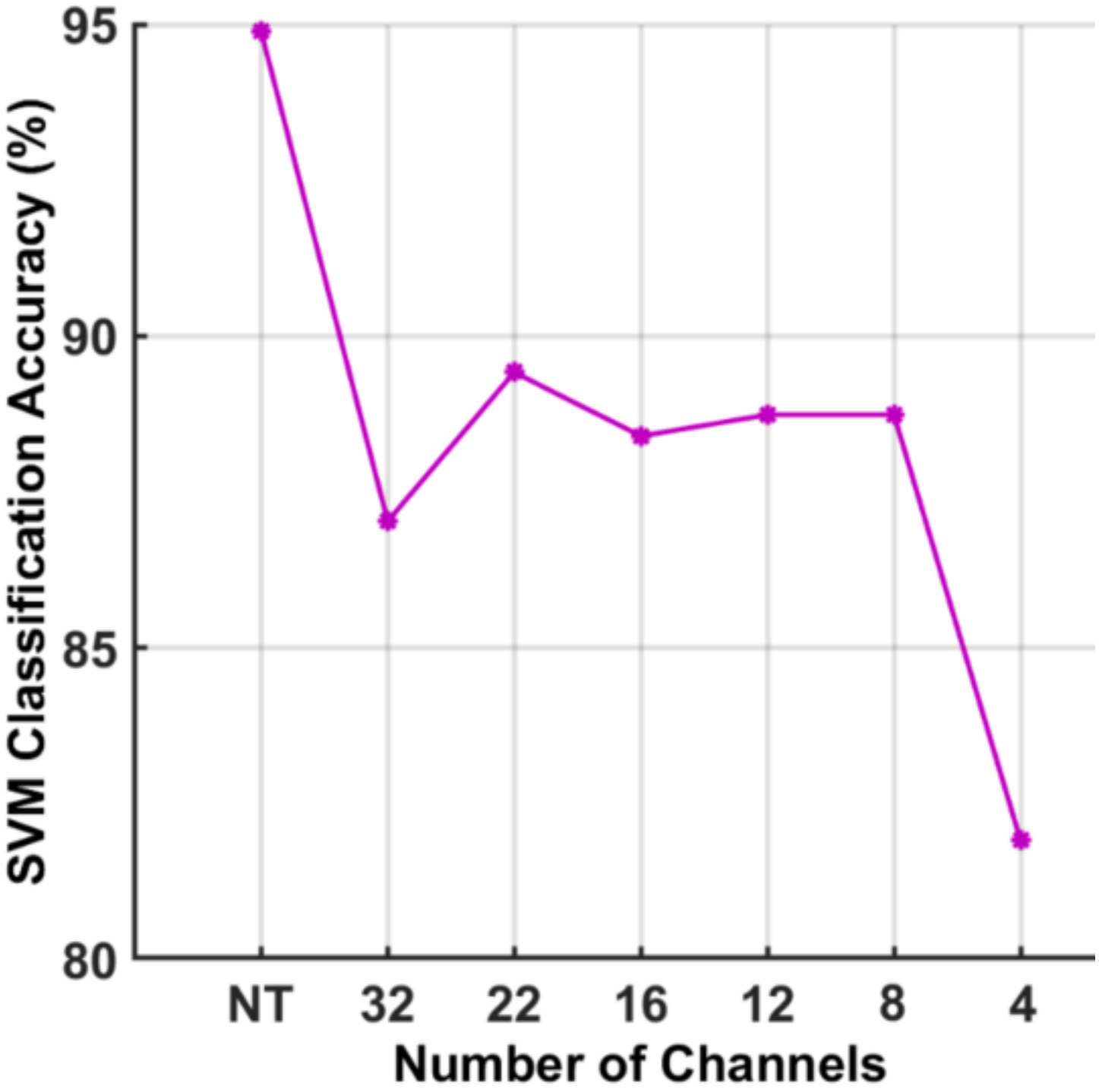
Click here to access/download;Figures (PDF, TIFF, EPS, PS, or JPEG only);Figure4.tif



medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.

Figures (PDF, TIFF, EPS, PS, or JPEG only)

Click here to access/download;Figures (PDF, TIFF, EPS, PS, or JPEG only);Figure5.tif



medRxiv preprint doi: https://doi.org/10.1101/2020.06.29.20142885; this version posted July 5, 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.