Discrimination of Voice Pitch and Vocal-Tract Length in Cochlear Implant Users

Gaudrain, Etienne; Başkent, Deniz

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Link to publication in University of Groningen/UMCG research database

Discrimination of Voice Pitch and Vocal-Tract Length in Cochlear Implant Users

Etienne Gaudrain[1,2] and Deniz Başkent[1,3]

Objectives: When listening to two competing speakers, normal-hearing (NH) listeners can take advantage of voice differences between the speakers. Users of cochlear implants (CIs) have difficulty in perceiving speech on speech. Previous literature has indicated sensitivity to voice pitch (related to fundamental frequency, F0) to be poor among implant users, while sensitivity to vocal-tract length (VTL; related to the height of the speaker and formant frequencies), the other principal voice characteristic, has not been directly investigated in CIs. A few recent studies evaluated F0 and VTL perception indirectly, through voice gender categorization, which relies on perception of both voice cues. These studies revealed that, contrary to prior literature, CI users seem to rely exclusively on F0 while not utilizing VTL to perform this task. The objective of the present study was to directly and systematically assess raw sensitivity to F0 and VTL differences in CI users to define the extent of the deficit in voice perception.

Design: The just-noticeable differences (JNDs) for F0 and VTL were measured in 11 CI listeners using triplets of consonant–vowel syllables in an adaptive three-alternative forced choice method.

Results: The results showed that while NH listeners had average JNDs of 1.95 and 1.73 semitones (st) for F0 and VTL, respectively, CI listeners showed JNDs of 9.19 and 7.19 st. These JNDs correspond to differences of 70% in F0 and 52% in VTL. For comparison to the natural range of voices in the population, the F0 JND in CIs remains smaller than the typical male–female F0 difference. However, the average VTL JND in CIs is about twice as large as the typical male–female VTL difference.

Conclusions: These findings, thus, directly confirm that CI listeners do not seem to have sufficient access to VTL cues, likely as a result of limited spectral resolution, and, hence, that CI listeners' voice perception deficit goes beyond poor perception of F0. These results provide a potential common explanation not only for a number of deficits observed in CI listeners, such as voice identification and gender categorization, but also for competing speech perception.

Key words: Fundamental frequency, Psychophysics, Voice perception.

INTRODUCTION

When trying to understand two competing talkers, normal-hearing (NH) listeners greatly benefit from voice differences among the talkers. Brungart (2001) showed that this benefit could reach 50 percentage-points for speakers of differing sexes. For cochlear implant (CI) users, in a similar task, Stickney et al. (2004) showed that this benefit was less than 20 percentage-points. Although it is debatable whether these percentage-point differences are indeed comparable because both the reference performance and target-to-masker ratios were not the same (Deroche et al. 2017), this observation nevertheless raises the question whether CI listeners have difficulties in perceiving and effectively utilizing voice cues.

Most of the literature on perception of voice cues by CI users indicates that CI listeners notoriously suffer from poor pitch perception (see Moore & Carlyon 2005 for a review). CI listeners have greater difficulties than NH listeners with distinguishing questions from statements (Green et al. 2005; Chatterjee & Peng 2008), discriminating pitch contours in tonal languages (He et al. 2016), and do not benefit from fundamental frequency (F0) differences between competing speakers (Stickney et al. 2007). F0 gives rise to the voice-pitch percept and may be coded either through temporal cues or through place cues. The perception of temporal pitch is based on the periodicity of the signal or of the temporal envelope of the signal. Temporal pitch perception seems to be relatively preserved in CI listeners, and they demonstrate functional use of this pitch cue similar to that of NH listeners (Hong & Turner 2009; Deroche et al. 2014; Gaudrain et al. 2017), as long as F0 remains below 300 Hz (Shannon 1983; Carlyon et al. 2002, 2010; Zeng 2002). In contrast, place pitch, that is, the pitch that results from exciting different segments of the cochlea, seems to be more difficult to use for CI listeners (Geurts & Wouters 2001; Laneau et al. 2004), in particular in speech-like stimuli where dynamic spectral envelope fluctuations may interfere with spectral changes induced by F0 differences (Green et al. 2002, 2004).

However, F0 is not the only cue that differentiates voices. Among many potential speaker-specific characteristics (Abercrombie 1982), not one but two principal vocal characteristics seem to mostly contribute to voice differences: F0, but also vocal-tract length (VTL). VTL is highly correlated with the height of the speaker (Fitch & Giedd 1999) and is related to the formant frequencies in the spectral envelope of the speech signal. Many voice-related perceptual phenomena observed in NH listeners can only be explained by considering both F0 and VTL together. For instance, the concurrent speech intelligibility increase observed by Brungart (2001) cannot be explained by F0 differences alone but can be replicated when F0 and VTL differences are combined (Darwin et al. 2003). F0 and VTL are also the two principal voice characteristics that are associated with gender and age perception (Smith & Patterson 2005; Smith et al. 2005; Skuk & Schweinberger 2014), or more generally, with voice identity (Gaudrain et al, Reference Note 3).

[1]University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology-Head and Neck Surgery, Groningen, The Netherlands; [2]CNRS UMR 5292, Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics, Université Lyon, Lyon, France; and [3]Research School of Behavioral and Cognitive Neurosciences, Graduate School of Medical Sciences, University of Groningen, Groningen, The Netherlands.

<zdoi; 10.1097/AUD.0000000000000480>

It is, thus, clear that not only F0 but also VTL plays an important role in voice perception in NH listeners. However, the relative importance of these two voice cues may be different in CI listeners. Using different talkers, Stickney et al. (2004) found that CI listeners could, modestly, benefit from speaker differences in a speech-on-speech task. However, when the voices only differed in F0, no such benefit was observed (Stickney et al. 2007). Although it can be debated whether the NH and CI data can be directly compared, as discussed above, this discrepancy led the authors to speculate that VTL differences may play a role for speech-on-speech perception in CI users. In contrast, some studies on voice gender perception suggested that, unlike NH listeners, CI listeners may be relying strongly on F0 cues to identify the sex of a speaker (Fu et al. 2004, 2005; Kovačić & Balaban 2009; Li & Fu 2011).

Studies involving real speakers may give general indications of which voice dimensions may contribute to specific phenomena. However, only with careful manipulation of these voice characteristics is it possible to ascertain, and quantify, the role that F0 and VTL may play. Massida et al. (2013) showed, using morphing between recorded male and female voices, that CI listeners were less sensitive to voice gender differences than NH listeners. To clarify the role of F0 and VTL, Fuller et al. (2014) varied the two variables orthogonally in a voice gender categorization task and found that while NH listeners made use of both F0 and VTL, CI listeners relied exclusively on the 1-octave F0 difference separating the male voice from the female voice. Fuller et al., hence, concluded that the abnormal voice gender categorization in CI listeners is mainly due to a deficit in VTL

perception. This conclusion was recently confirmed by Meister et al. (2016) using a wider range of stimuli than words only, also including word quadruples and sentences. Hence, unlike F0, VTL has been seldom studied in CI users, and very little is known about the factors that may limit its perception.

The acoustic cue for VTL lies in the formant frequencies, which result from resonances in the vocal tract. For a given formant configuration, shortening the VTL by a given ratio r results in shifting all formants up in frequency by that same ratio (Fig. 1). This results into a translation of all formants—as a unit—on a logarithmic frequency axis. If represented onto a linear frequency axis, a VTL change is reflected as an expansion/contraction of the formant distances. To make it a frequency-compatible unit, this ratio can be expressed in semitones (st, the 12th of an octave) using $12 \cdot \log_2(r)$. Semitones are used in music and represent an intuitive frequency increment, thus providing a perceptually relevant unit while not relying on a specific perceptual model like Bark or ERB (equivalent rectangular bandwidth). Note that the VTL is a distance, related to wavelength, and, thus, inversely related to frequency; positive VTL ratios in semitones correspond to negative formant frequency shifts in semitones. In adult speakers, VTL is, on average, longer in men than in women by about 23%, leading male formants to be about 3.6 st lower in frequency than female formants. Examining the output of the implant, Fuller et al. (2014) observed that this 3.6-st frequency shift in formant frequency was roughly equivalent to a one-electrode shift in the electrical pattern of excitation. Because speech has a complex spectral
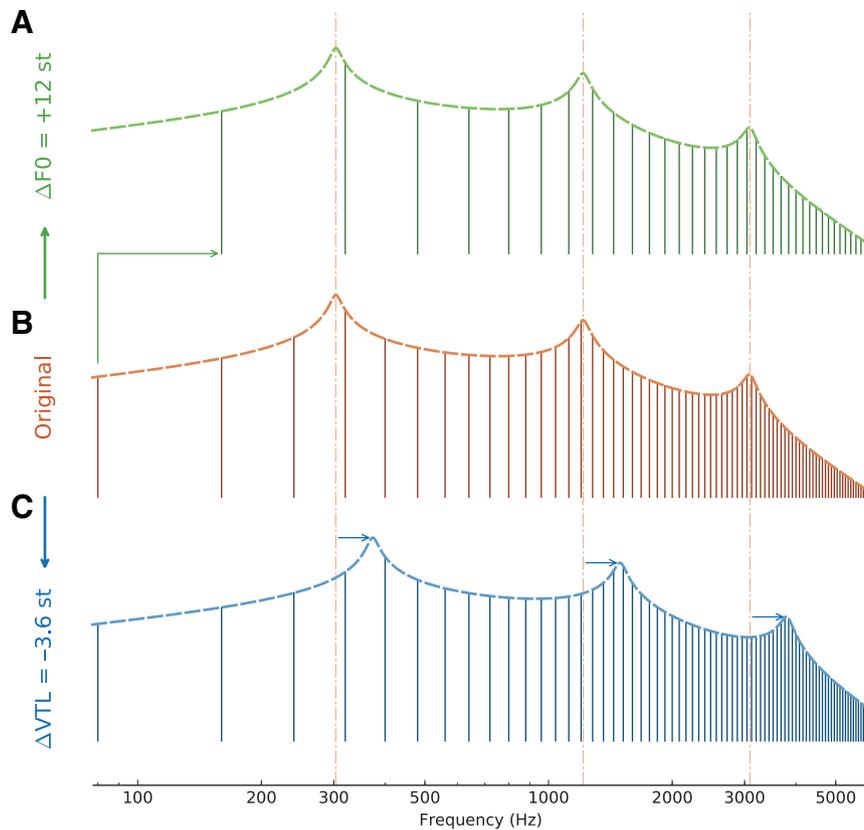
Fig. 1. Harmonic spectra (vertical lines) and vocal-tract resonances (dashed line) of an idealized vowel. The vocal-tract resonance profiles were generated using a three-tube model (mimicking different vocal-tract lengths) and were then used to define the spectral envelope of the harmonics. Comparing (**B**) and (**A**) shows the effect of increasing F0 by 12 st. Comparing (**B**) and (**C**) shows the effect of decreasing vocal-tract length by 3.6 st. The vertical dash-dotted lines mark the formant positions in the original vowel.

structure, it is hard to predict whether such a VTL difference can be detected by CI users.

A first potential limitation to the transmission of VTL cues through the implant is the poor spectral resolution available due to the limitations of the electrode–neuron interface (see Başkent et al. 2016 for a review). This may severely reduce precise access to individual formants and make it impossible for listeners to detect small—but consistent—changes in their position. An additional complication in the implant comes from the fact that, because the spectral envelope is quantized into a number of discrete frequency bands in the implant, not all formants shift by the same amount. Depending on the formant frequency, a 3.6-st frequency shift can remain within a frequency band or jump up to two electrodes. In other words, the VTL cues may not be transmitted due to coarse spectral resolution available, and if transmitted, the VTL cues could still be severely distorted.

Previous studies using acoustic simulations of CIs gave only a partial answer. Gaudrain and Başkent (2015) showed perception of VTL to be more vulnerable than that of F0 to the kind of degradations that the CI stimulation imposes on the speech signal. However, the simulations of Fuller et al. (2014) showed a pattern that differed from that of actual CI users. In their study, the uses of F0 and VTL for gender categorization were both significantly reduced as a result of vocoding; however, NH listeners still seemed to make some use of both cues. CI users, on the other hand, only exclusively utilized F0 cues and no VTL cues. Hence, it remains unclear whether the CI participants in Fuller et al.'s study were unable to use VTL cues for gender categorization because VTL cues are not represented through electrical stimulation, or if VTL cues are represented, but in a form that is too distorted to be used for the subjective labeling of the gender of a voice.

In the present study, we aimed to directly answer the question of to what degree actual CI users can hear F0 and VTL cues. More specifically, in a design similar to Gaudrain and Başkent (2015), we measured F0 and VTL just-noticeable differences (JNDs) in actual CI users. In a gender categorization task, the listener has to correctly hear and interpret the F0 and VTL cues as such to make the correct labeling of the sex. Here, unlike in the gender task, we aimed to directly measure the raw sensitivity to hearing any perceptual difference that results from a F0 and VTL change, in an odd-one-out task. On the basis of previous literature on pitch perception, gender categorization, and the simulation study on F0 and VTL perception, we hypothesized that CI users would show a higher threshold for perception of F0 and VTL than NH individuals; however, this difference would be more pronounced for VTL perception.

METHODS

Participants

Eleven CI listeners were recruited (6 men), aged 47 to 74 years (average: 60.1 years). They all had more than 1 year experience with their implant (details provided in Table 1). The participants had no substantial residual hearing, except CI1 who had postoperative thresholds of 35 and 50 dB HL at 250 and 500 Hz, respectively, in the nonimplanted ear. CI1 was not wearing a hearing aid in the nonimplanted ear during testing, and that ear was also not plugged. All participants were native speakers of Dutch. No other speech perception or language-processing performance was used for inclusion criterion. As a result, the participants had a range of scores from the clinical speech perception assessment (Table 1, rightmost column). This range ensured that the study population represented good, but not only star, users of CIs.

Each participant provided signed informed consent. The experiment was approved by the ethics committee of the University Medical Center Groningen (METc 2012.392). Finally, the participants received an hourly wage for their participation.

Stimuli

The JNDs were measured using a 3I-3AFC adaptive odd-one-out procedure where the participants had to point to which of three consecutive stimuli was different from the two others. The listeners could use any cue available to perform the task. Depending on the specific condition, the odd one out differed in F0 only, in VTL only, or both together.

To ensure that the stimuli used in this experiment remain relevant for speech perception, each interval of the 3I-3AFC procedure was made of a syllable triplet—a sequence of three short syllables, similar to a pseudo-word. The syllables were consonant–vowel (CV) tokens spliced from meaningful consonant–vowel–consonant words taken from the Nederlandse Vereniging voor Audiologie corpus (Bosman & Smoorenburg 1995). The words were uttered by a female speaker, the same speaker as for the words used by Fuller et al. (2014) to study

TABLE 1. Details of the Participants of the Study

Id	Age (yr)	Use of CI (yr)	Implant	Processor	Strategy	Word Recognition (%-correct)
CI1	47	1	CI422	CP920	MP3000, 6 maxima	80
CI2	56	5	HR90K/HiFocus 1J	Harmony	HiRes-S, no ClearVoice	73
CI3	62	13	CI24M	CP810	ACE, 8 maxima	88
CI4	63	12	CI24R CS/CI422	CP900	MP3000, 6 maxima	42
CI5	63	12	CI24RE CA	CP810	ACE, 8 maxima	85
CI6	69	6	CI24RE CA	Freedom	ACE, 8 maxima	95
CI7	50	3	CI24RE CA	CP810	MP3000, 6 maxima	60
CI8	74	7	HR90K/HiFocus Helix	Naída CI Q70	HiRes-S, no ClearVoice	94
CI9	72	7	HR90K/HiFocus Helix	Naída CI Q70	HiRes-S, no ClearVoice	80
CI10	50	12	CI24R CS	CP810	ACE, 8 maxima	58
CI11	55	13	CI24M	CP810	ACE, 8 maxima	50

The percentage correct word recognition (last column) represents the proportion of correctly repeated phonemes, measured in the clinic after implantation using the Nederlandse Vereniging voor Audiologie word lists (see text for details), presented at 65 dB SPL.

gender categorization. The average F0 of the speaker's voice measured on all the words was 242 Hz. From a selection of the available words in the corpus, 61 CV syllables were obtained, with durations ranging from 142 to 200 ms.

In the course of the adaptive procedure, the F0 and VTL of individual syllables were manipulated using STRAIGHT (Kawahara & Irino 2004). During this process, the duration of each CV token was also normalized to 200 ms.

The syllable triplets were created by forming a sequence where the 200-ms syllables were separated by a 50-ms silence. To make the triplet sound more natural and to force the participants to focus on average—rather than instantaneous—F0, the F0s of the three syllables were not kept identical, but a slight F0 contour was imposed over the sequence (Smith et al. 2005). This was done by imposing random steps of 1/3 st between consecutive tokens in the sequence, while maintaining the average F0 (over the sequence) equal to the value set by the adaptive procedure (Gaudrain & Başkent 2015). While in previous studies the CV syllables changed from one interval to the next (Ives et al. 2005; Smith et al. 2005), we took the approach to keep the same syllable triplet over the three intervals. This adaptation was necessary to keep the task relatively easy, because we were testing CI users, and also to reduce the risk that changes in syllables could be interpreted as changes in voice. In our design, the participants only had to judge three voice intervals, two of which were identical, while in previous studies, it could be argued that the participants had to judge—and thus memorize—as many intervals as there were vowels or syllables (8 syllables in Ives et al. 2005; 8 vowels in Smith et al. 2005).

Finally, the intensity level of each triplet was roved by ±2 dB to prevent the use of loudness cues by the participants. This 2-dB value corresponds to the SD of the maximum loudness evaluated across all the syllables with different F0 and VTL values, as evaluated using the method described by Glasberg and Moore (2002).

Procedure and Apparatus

The JNDs were measured with a two-down, one-up adaptive procedure, yielding an estimate of the voice difference corresponding to 70.7% correct discrimination on the psychometric function (Levitt 1971). In each trial, the participants heard three stimuli as they saw three buttons, numbered one to three, on a computer screen light up as each corresponding stimulus was being played. They were then instructed to choose the one stimulus that differed from the other two by clicking on the corresponding button on the computer screen. Visual feedback was provided by making the correct answer blink either green if the answer was correct or red if it was incorrect.

In all measurements, the test voice became progressively more similar, in terms of F0 and VTL, to the reference female voice, but the reference voice was approached along different axes in the F0-VTL plane and from different directions along these axes, as shown in Figure 2. VTL JNDs were measured in two directions: starting from larger (+VTL) and smaller (−VTL) VTL values than the reference voice. Similarly, F0 JNDs were measured starting from lower (−F0) and from higher (+F0) F0 values than the reference voice. In a fifth condition, combining changes on both F0 and VTL, JNDs were measured along a continuum between an artificial man's voice and the reference female voice. The man's voice was defined as having a VTL
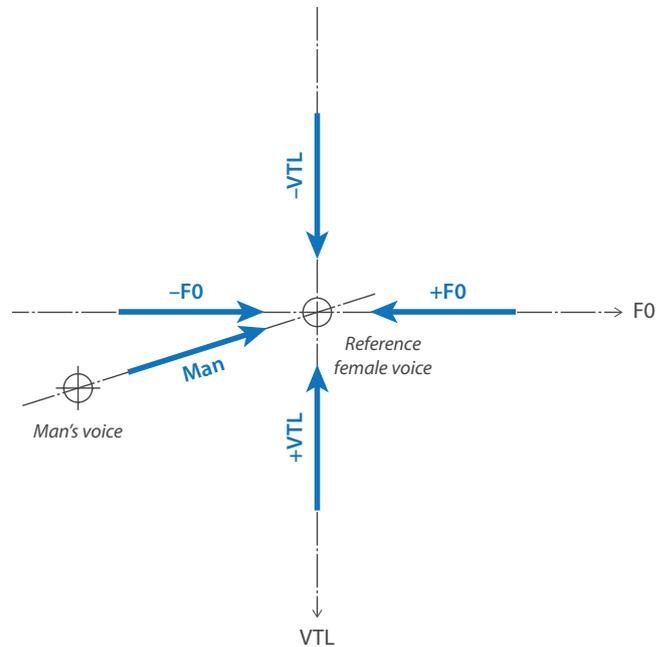

Fig. 2. F0-vocal-tract length (VTL) plane with the reference female voice in the center, and the man's voice in the lower left corner. The axes are shown with dashed lines, while the approach directions and their labels are shown with thick arrows. Note that the VTL axis is represented upside down as it represents a length, which is associated with a wavelength dimension, while F0 represents a frequency. The orthogonality of the two axes illustrates their independence in the physical domain, where they are manipulated.

24.5% (3.8 st) longer than that of the reference female voice, and an F0 half (−12 st) of that of the reference voice (as used by Gaudrain & Başkent 2015). The JNDs were measured in only one direction along this axis: from the man's voice toward the reference female voice.

For all axes and directions, the initial voice difference was 12 st, calculated as the Euclidian distance in the F0-VTL plane represented in semitones relative to the reference voice. After two consecutive correct answers, the voice difference was reduced by a certain step size, while after every incorrect answer, the voice difference was increased by that same step size. The initial step size was 2 st but was also modified during the procedure. After every 15 trials, or when the voice difference became smaller than twice the step size, the step size was reduced by a factor of $\sqrt{2}$. The procedure ended after 8 reversals or after 150 trials. The JND was calculated as the mean of the voice difference from the last 6 reversals. For each axis and direction, 3 repetitions of the JNDs were obtained per participant. The 5 Directions × 3 Repetitions = a total of 15 JND measurements were tested in random order for each participant.

The participants were seated in an anechoic room about 1 m away from a loudspeaker (Precision 80, Tannoy, Coatbridge, United Kingdom). The stimuli were played through an Audio-Fire4 soundcard (Echo Digital Audio Corp, Santa Barbara, CA) connected to a DA10 D/A converter (Lavry Engineering, Poulsbo, WA) through S/PDIF (Sony/Philips Digital Interconnect Format). The sound level was calibrated to 63 dB SPL. At the beginning of each testing session, the participant took part in a short training consisting of the same adaptive procedure but limited to 8 trials. During this training, they were instructed to adjust the gain

setting on their implant so that the sound level would be comfortable and to keep it the same throughout the data collection. The average testing time was 1 hour 35 minutes (SD 15.4 minutes) to which was added about 15 to 20 minutes for reading and completing the informed consent form and for training for the task.

RESULTS

Individual results are shown in Figure 3, ordered according to increasing VTL JNDs. The approach direction along the F0 or VTL axes had no significant effect on the CI listeners' JNDs [(+F0) − (−F0) = −0.12 st, $t(10)$ = −0.12; p = 0.92; (+VTL) − (−VTL) = 0.51 st, $t(10)$ = −0.55; p = 0.59]; hence, in subsequent analyses, and on Figure 3, these data were collapsed by averaging.

For the interpretation of the results, the JNDs from the CI listeners are compared with those of NH listeners (tested without acoustic CI simulations) from Gaudrain and Başkent (2015). In addition, we also compared the JNDs to the typical F0 and VTL differences used to distinguish male from female speakers in the gender categorization study of Fuller et al. (2014).

The mean F0 JND was 9.19 st for CI users, against 1.95 st for NH listeners [$t(12.24)$ = 4.03; p < 0.01]. The mean VTL JND was 7.19 st for CI listeners, against 1.73 st for NH listeners [$t(11.11)$ = 6.28; p < 0.001]. Check Table 2 to see these JND values expressed in other units than semitones. The mean JND along the man voice axis was 9.67 st for CI listeners against 1.71 st for NH listeners [$t(10.82)$ = 3.84; p < 0.01]. This JND, in CI users, can be decomposed as 9.22 st along the F0 axis (almost identical to the measured F0 JND), and 2.92 st along the VTL axis (much smaller than the VTL JND), thus suggesting that CI listeners rely on the F0 cue to perform the task in this condition. In NH users, the 1.71 st JND can be decomposed into 1.63 st along the F0 axis and 0.52 st along the VTL axis, both smaller than the JNDs for F0 alone (2.68 st in the direction of lower F0s) and for VTL alone (1.62 st in the direction of longer VTLs) and thus suggesting an additive effect of the two dimensions (Gaudrain & Başkent 2015).

This conclusion is based on the observation of the average JNDs over CI participants. To assess whether this relationship holds at the individual level, the correlations of the individual JNDs across the different axes were examined. The average F0 JNDs (across directions along this axis) for each participant were correlated with their average VTL JNDs [r^2 = 0.45; $t(9)$ = 2.70; p < 0.05], indicating that participants with larger
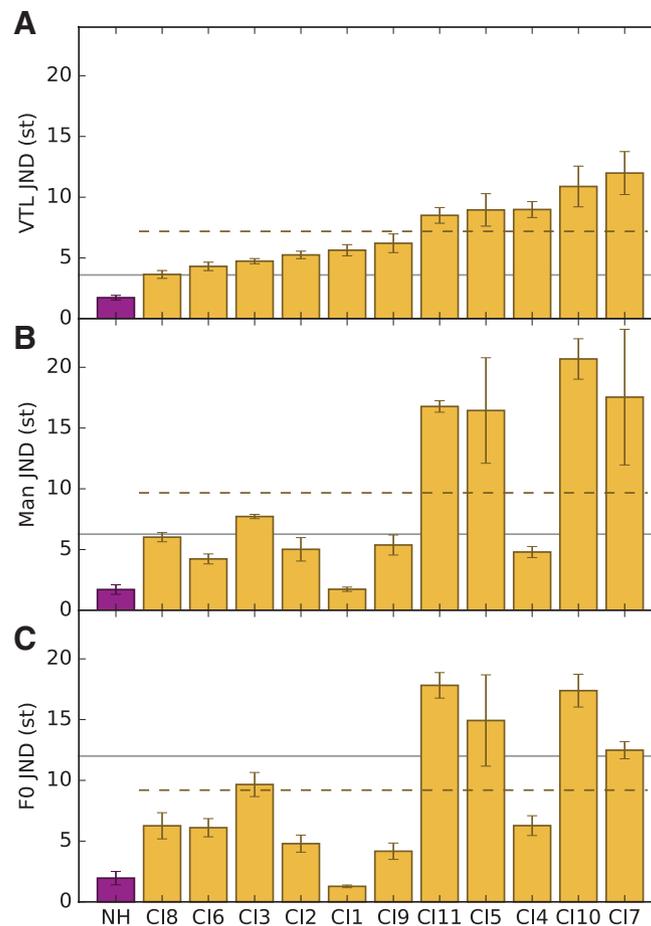

Fig. 3. **A**, Just-noticeable difference (JND) for vocal-tract length (VTL), shown in average for normal-hearing (NH) listeners (dark/purple) and individually for cochlear implant (CI) listeners (light/yellow). The JNDs have been averaged across positive and negative VTL differences (shown in Fig. 2). The data of the CI participants are ordered by increasing VTL JNDs. The error bars for the NH data show the SE across participants. The error bars for the CI individuals show the SE across measurements (directions and repetitions). The dashed lines show the average JNDs for the CI participants. The solid horizontal line represents a 3.6-st VTL difference, corresponding to the average difference between the man and woman voices in Fuller et al. (2014). **B**, Same as the top panel but shown for the "man" voice axis. The solid line represents a 6.3-st difference along the "man" voice axis. **C**, Same as the top panel, but shown for F0. The solid line represents a 12-st F0 difference, corresponding to the average difference between the man and woman voices in Fuller et al. (2014).

TABLE 2. Average F0 and VTL JNDs Expressed in Various Units, in NH and CI Listeners

		st	%	Hertz
NH	ΔF0	1.95	11.9	−26/+29 Hz re 242 Hz
	ΔVTL	1.73	10.5	−29/+32 Hz re 300 Hz
				−285/+315 Hz re 3000 Hz
CI	ΔF0	9.19	70.0	−100/+169 Hz re 242 Hz
	ΔVTL	7.19	51.5	−102/+154 Hz re 300 Hz
				−1020/+1545 Hz re 3000 Hz

The "Hertz" column shows the increment in Hertz corresponding to the JND, relative to the average F0 of the reference voice (242 Hz) or to low- (300 Hz) and high-frequency (3000 Hz) formants.
CI, cochlear implant; JND, just-noticeable difference; NH, normal hearing; VTL, vocal-tract length.

F0 JNDs also tended to have larger VTL JNDs. The man JNDs seem more strongly correlated with F0 [$r^2 = 0.91$; $t(9) = 9.27$; $p < 0.001$] than with VTL [$r^2 = 0.63$; $t(9) = 3.91$; $p < 0.01$], although the comparison of these two correlation coefficients only approached significance [$z = 1.89$; $p = 0.059$; Steiger 1980].

In addition to comparing the JNDs measured in CI users to those collected with NH listeners, they can also be compared with actual VTL and F0 differences found between voices in the population. The solid horizontal lines in Figure 3 represent typical F0 and VTL differences separating men voices from women voices and were used to create the man's voice from the original woman's recording in the gender categorization experiment of Fuller et al. (2014). Only 1 of our 11 CI participants had an average VTL JND smaller than the typical man–woman VTL difference (Fig. 3A). Having a JND larger than this typical difference is an indication that these 10 CI participants would not be able to use VTL differences to perceive the sex of a voice. In contrast, 7 of our 11 CI participants had an average F0 JND smaller than the typical man–woman F0 difference (Fig. 3C).

Although the observation is rather anecdotal, it is worth noting that CI1, who had substantial residual hearing in the nonimplanted ear, had F0 JNDs on par with those of the NH listeners. In contrast, their VTL JNDs were at least twice as large as that of the NH listeners, thus suggesting that residual hearing may not contribute as much to VTL perception as it contributes to F0 perception. However, because these results are limited to a single participant, caution is required in drawing that conclusion until a larger population of CI users with residual hearing has been tested. A first element of response is provided by a recent study using noise-band vocoders to simulate electroacoustic stimulation and confirming the present results (Başkent et al., Reference Note 1).

Similarly, the 3 participants equipped with an Advanced Bionics device using a current steering strategy (CI2, CI8, and CI9) seem to perform well on both F0 and VTL JNDs. However, the collected data are too sparse to be able to control for possible confounds and draw reliable conclusions regarding the effect of brand, processor, or strategy.

DISCUSSION

In this study, we have directly and systematically investigated the perception of two principal voice characteristics, F0 and VTL, in CI listeners. Our results have shown that voice perception in CI users not only differs from that of NH, tending to be poorer in general, but also, this deficiency is more complex and serious than was previously reported in the literature, which had mostly focused on F0 perception.

F0 Perception With Cochlear Implants

Pitch perception in CI users has attracted a lot of attention from researchers and has been shown to be poorer than that of NH listeners. In line with this idea, the present JND data indeed show that the F0 JNDs in CI listeners are more than 4 times larger than those observed in NH listeners. However, most studies have used very artificial stimuli and only a few have used broadband speech-like stimuli spanning across all the electrodes.

Gfeller et al. (2002) used 1-second piano tones on a semitone scale and observed JNDs of 1.13 st in NH and 7.56 st in CI listeners. Geurts and Wouters (2001) observed in CI listeners strikingly small F0 JNDs ranging from 0.22 to 2.4 st in steady state synthetic vowels /a/ and /i/, but loudness could have played a role in these JNDs. Laneau et al. (2004) found F0 discrimination JNDs larger than 17 st when only place cues were available in a single-formant idealized vowel and using a clinical Cochlear Corp. ACE filterbank. After making temporal pitch cues available, the JNDs fell to around 4 st. Vandali et al. (2005) found that a 6-st difference yielded 70% correct discrimination—thus equivalent to the JND—for steady state sung vowels. These performances improved when temporal cues were enhanced with various strategies. Green et al. (2004) measured identification of F0 glides on English diphthongs and found thresholds between 7 and 11 st in CI listeners using a standard continuous interleaved sampling strategy. These authors also developed strategies aiming at enhancing temporal F0 cues and, like Vandali et al., observed significant improvements. However, these improvements only modestly transferred to question/statement discrimination (Green et al. 2005). Chatterjee and Peng (2008) also found JNDs in the order of 6.5 to 7.5 st in CI listeners, for question/statement discrimination. More recently, He et al. (2016) found pitch contour discrimination thresholds to be around 10 to 11 st for the syllable /ma/ and 4 to 8 st for artificial complex tones.

The consensus view in these studies is that temporal pitch cues play a major role in F0 perception in CI listeners. Figure 4A and B show how temporal modulation in a single channel changed when the F0 was increased by 9.19 st. The reference voice in our experiment had an average F0 of 242 Hz, which is rather close to the 300-Hz limit for temporal pitch perception reported in the literature (Shannon 1983; Carlyon et al. 2002, 2010; Zeng 2002). Therefore, in the conditions where the test stimulus had a smaller F0, it could be expected that temporal pitch cues would become more available, yielding smaller JNDs than when the test voice had a larger F0. Indeed, −9.19 st from 242 Hz means the F0 of the test voice was 142 Hz, while +9.19 st from 242 Hz means it was 411 Hz. However, there was no difference between positive and negative F0 JNDs (the 2 directions only differed by 0.12 st, and in 10 of 11 participants, the intrasubject variability was larger than the individual direction effect). This could mean that temporal pitch cues were still sufficiently salient to perform the task at 411 Hz. Single-channel temporal pitch JNDs (based on stimulation rate or on temporal envelope periodicity) reported in the literature range from 3 to
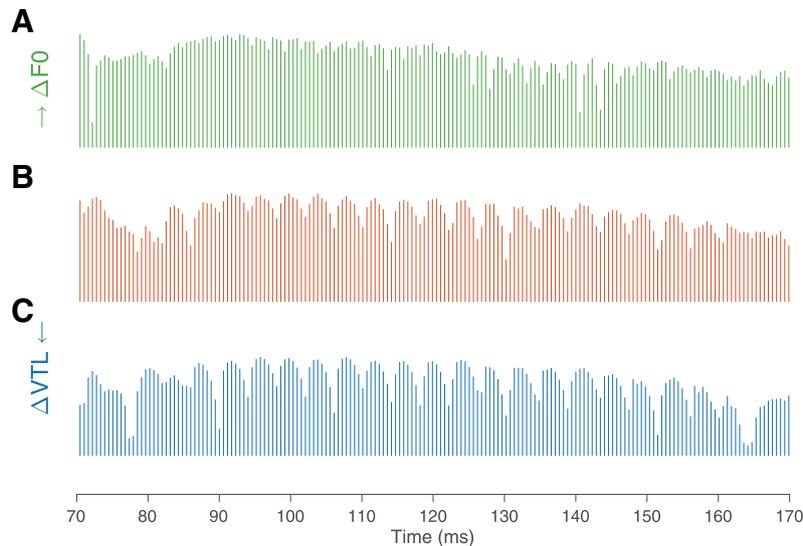
Fig. 4. Excerpt of channel 10 of the electrodogram of the syllable /ki/, for the original voice (**B**), for the F0 just-noticeable difference (JND; **A**) and for the vocal-tract length (VTL) JND (**C**).

6 st for base rates close to 242 Hz (McDermott & McKay 1997; Zeng 2002; Baumann & Nobbe 2004; Chatterjee & Peng 2008; Carlyon et al. 2010; Gaudrain et al. 2017).

Complementing temporal pitch cues, it is also arguable that place pitch cues may have contributed to the observed JNDs. Fielden et al. (2015), using steady state vowels, found that 6-st differences could be reliably discriminated by CI listeners, but not 3 st ones. These authors suggested that participants used a shift in spectral centroid resulting from F0 differences to do the task. The larger centroid shifts they observed were around 0.4%. Because we used different, non–steady state syllables, the spectral centroid in our experiment was constantly shifting. Calculating the distribution of centroids across all the available tokens (while keeping F0 and VTL constant), we obtained a distribution whose SD was 22% of the average (Gaudrain 2016). Spectral centroid is, thus, unlikely to have played a significant role in the F0 JNDs reported here, as has also been previously argued by Green et al. (2002, 2004) in similar conditions. However, place cues, rather than truly pulling the whole center of gravity of the electric stimulation, may still arise in individual channels. Figure 5 shows the summary electrodograms, that is, the average power in each channel of the implant, for three different syllables. The top and middle curves in each parts of Figure 5 differ by 9.19 st. While most of the channels are not strongly affected by the F0 change, a few channels in each panel are affected by the F0 change (e.g., channel 2 and 3 for /ba/, 3 and 4 for /ki/, and 1 to 4 for /po/). While it might be difficult to quantify how much these place cues may have contributed to the observed JNDs, it seems likely that this contribution was not negligible.

Other factors that could have potentially contributed to the difference between NH and CI listeners are task difficulty and age. The difficulty of the 3AFC task depends on the sensory input that the participant receives—and this is the effect we aim to capture—but also depends on the proficiency of cognitive mechanisms. While there is no way to absolutely rule out the possibility that individual differences in cognitive function, for example, due to aging, may have contributed to our pattern of results, it can be argued that this is rather unlikely. Indeed,

3AFC tasks similar to the one used in the present study have been used in other studies where performance of the CI group was found to be equivalent or better than that of the NH group (gap detection in words: Gaudrain et al., Reference Note 2; rate pitch discrimination: Gaudrain et al. 2017). As for age, the NH listeners were on average 22.7 years younger than the CI listeners. In their vocoder study, Gaudrain and Başkent (2015) assessed whether age could play a role in the F0-VTL JND task by including a number of older participants. While their study population did not allow a systematic study of a potential age effect, they did report that age did not significantly contribute to intersubject variability. On the basis of this report, it, thus, seems unlikely that age would have played a major role in the JND differences observed between the two groups, although a more detailed study would be required to definitely answer this question.

VTL Perception With Cochlear Implants

The cues used for VTL perception in NH listeners are most likely spectral in nature, and this is likely true as well in CI listeners. Using vocoders, Gaudrain and Başkent (2015) have shown that while the nature of temporal cues did not affect VTL JNDs, reducing spectral resolution (by reducing the number of channels or by increasing channel overlap) drastically increased VTL JNDs. Examining the output of the implant, it seems likely to also be the case with actual implants. Figure 4B and C show the output of one channel of a standard continuous interleaved sampling strategy for a VTL change of 7.19 st. The amplitude modulation pattern of the channel is largely unaffected by the change in VTL. In contrast, the middle and bottom curves of Figure 5 show average stimulation profiles along the electrode array for the same change in VTL and for three different syllables. Changing VTL seems to result in a shift of the stimulation profile by 2 to 3 channels.

We could not identify, in the literature, any psychophysical measurement that could be directly compared with VTL JNDs of the present study. The best comparison likely lies in broadband estimates of spectral resolution in implants. The observation that, at threshold, the stimulation profile shifts by 2 to 3 channels implies
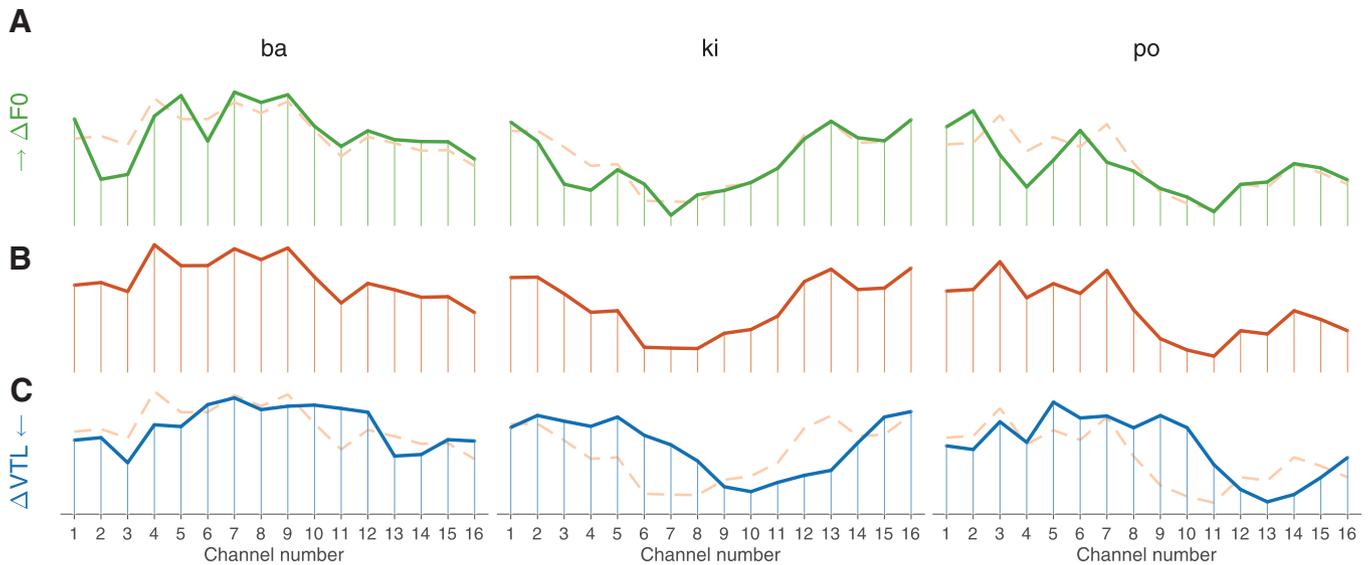
Fig. 5. Summary electrodograms for three syllables. The summary electrodogram shows the average power in each channel over the course of the syllable. The middle curve (B) shows the original version of the stimulus, while the top (A) and bottom (C) curves correspond to an F0 shift of 1 just-noticeable difference (JND) and a vocal-tract length (VTL) shift of 1 JND, respectively. The curves corresponding to the original version of the stimulus are repeated, as dashed lines, behind the F0 and VTL shifted version to facilitate comparison. The average power is normalized and, thus, presented without units here.

that only 6 to 8 independent channels are really used by the CI listeners to perform this task. This is in agreement with an equivalent estimate obtained from spectral-ripple discrimination concluding that only 8 independent channels contribute to spectral resolution (Henry & Turner 2003). With a similar method, Henry et al. (2005) reported an average spectral-ripple discrimination threshold of 0.62 ripple/octave, which would be equivalent to a JND of 9.67 st. However, spectral-ripple discrimination data obtained by applying rectified sinusoidal ripples to the logarithm of the magnitude (in dB) rather than to the magnitude directly yields smaller equivalent JNDs of 3.4 st on average (Won et al. 2007).

Another method that might be worth comparing our results to is formant discrimination. Fitzgerald et al. (2007) measured F1, F2, and F3 discrimination by altering the frequencies of individual formants while holding the other formants constant. Unfortunately, the data are only reported in millimeters that were specific to individual frequency maps and electrode array. Sagi et al. (2010) reported that the JNDs found by Fitzgerald et al. were "about 50 to 100 Hz in the F1 frequency range," which corresponds to 3.2 to 5.8 st given the base F1 of 250 Hz used in their experiment. For F2, they report a 10% JND, which corresponds to 1.7 st. Note that Sagi et al. (2010) also measured formant discrimination, in a more systematic way (varying F1 frequency over a range), but they reported the data in the form of an averaged difference in Hertz (over all reference F1 values), which, thus, cannot be compared with the other data in the literature or with the data presently reported. More recently, Winn et al. (2012) reported psychometric functions for "heat"–"hit" discrimination from which we estimated a JND of 1.24 st for F2 discrimination (at onset), although the cue was mixed with other cues, such as F1/F2 profile frequency and duration.

These estimates of spectral resolution vary greatly depending on whether they are local (single channel or single formant) or global (spectral ripple). Limited spectral resolution in the implant is often pointed at as the culprit for poor speech understanding, especially in noise (Friesen et al. 2001; Qin & Oxenham 2003; Stickney et al. 2004; Clarke et al. 2016). As a result, much effort has been spent

to try to increase the spectral resolution in the implant (Nogueira et al. 2009, 2016; Bhattacharya et al. 2011). To evaluate these new strategies, researchers have used either one of the aforementioned techniques or speech understanding in quiet or in noise. The latter can capture the full benefits of these new strategies only after their chronic use over a few months, because speech understanding requires a rather long adaptation. For this reason, the other measures mentioned above might be more appealing. One advantage of the local, single-formant measures is that they are based on speech stimuli. However, because of their local nature, they incur the risk of being strongly sensitive to small-frequency allocation map variations, which are unlikely to matter for CI users once they have adapted to it. Therefore, these measures might not be reliable predictors of success of a strategy for general speech perception. In contrast, the global spectral-ripple measure is likely less sensitive to local processing modifications but does not focus on spectral contrasts that are specifically relevant for speech.

The measure investigated here, VTL discrimination thresholds, constitutes a global, objective spectral resolution measure while maintaining relevance to speech perception. It could, thus, prove particularly useful for the evaluation of novel speech-processing strategies and for clinical assessment of CI patients, especially if the relationship between VTL JNDs and speech intelligibility (in quiet or in noise) can be demonstrated.

Voice Discrimination[1] in Implants

Our present data show a more complex picture than was previously indicated by studies that mostly reduced voice discrimination to voice-pitch ranking. Instead, our results are more in line with the gender categorization data of Fuller et al. (2014) who showed that while NH listeners use both F0 and VTL to determine the sex of a voice, CI listeners use almost exclusively F0 and rely very little on VTL.

---

1 By "voice discrimination," we mean the discrimination of voices emanating from different speakers, real, or simulated through vocal cue manipulation. We do not refer to the ability to distinguish whether a sound is a human voice or not, as was done by Massida et al. (2011).

To understand the importance of the relative roles of F0 and VTL, it is worth noting that the typical F0 difference between male and female speakers is about 1 octave—or 12 st—while the average F0 JND in CI users is just above 9 st, that is, 4.7 times larger than for the NH listeners. In other words, although pitch perception is greatly reduced in CI listeners, it seems to remain sufficient for sex categorization purposes. Unfortunately, the same cannot be said of VTL: while the VTL JNDs are also just 4.2 times larger in CI than in NH listeners, the typical VTL difference between male and female speakers is much smaller. Estimates in the literature vary from 13% (Fitch & Giedd 1999) or 15% (Fant 1970) to about 18% (Turner et al. 2009, based on the data of Peterson & Barney 1952). To produce a clear distinction between their female and male voices, Fuller et al. (2014) used 23%. Expressed in semitones, these male–female VTL differences represent, respectively, the following: 2.0, 2.4, 2.9, and 3.6 st. In other words, even when using this last, more conservative estimate, the present JND data indicate that CI listeners do not perceive the VTL difference between male and female speakers.

Fuller et al. (2014) discussed the possible origins for the abnormal gender categorization pattern they observed in CI listeners: (1) either VTL cues are not transmitted through the implant and the electrode–neuron interface or (2) the VTL cues are transmitted, but in a distorted form that makes it impossible to correctly interpret them as speaker size information and use for categorizing speaker's gender. The conclusions from this study suggest that the first hypothesis is more likely than the second and that the deficit originates at the sensory level. Because a 3AFC task was used here, to do the task, the participants were free to use any type of cue available in the signal, which requires no or minimal interpretation of the VTL-related cues into speaker size information. The large VTL JNDs observed, therefore, suggest that the VTL cues are too degraded in electric hearing to be picked up even at a rather primitive level of the auditory system.

This study measured voice cue discrimination from the recording of a single speaker. Yet, the large JNDs observed in this experiment will likely have consequences beyond gender categorization, extending to speaker identification in general. A number of studies have shown that CI listeners had greater difficulty discriminating actual speakers than NH listeners (Cleary & Pisoni 2002; Mühler et al. 2009). Gaudrain et al. (Reference Note 3) investigated that F0 and VTL difference would lead participants to judging they were hearing two different speakers. They found that speakers were judged as different when their voice differed by 3.8 st in F0 or 2.2 st in VTL or more. Because the JNDs observed in the present study for CI listeners are larger than these values, one may predict that speaker discrimination based on F0 and VTL differences would be altered in CI users. Indeed, Cleary et al. (2005) reported that pediatric CI users needed larger combined F0 and VTL differences than their NH peers to discriminate voices. Unfortunately, the two cues were manipulated together so their individual contribution cannot be separated in these data.

Intuitively, it is assumed that good voice discrimination should entail large voice difference benefit in speech-on-speech perception. A few studies have shown that the sensitivity to F0 differences does correlate with the F0-difference advantage in a concurrent speech-listening situation (Summers & Leek 1998; Mackersie et al. 2001; Mackersie 2003; Gaudrain et al. 2012),

and systematic voice differences have been shown to improve perception of concurrent speech (Brungart 2001; Darwin et al. 2003; Mackersie et al. 2011; Başkent & Gaudrain 2016). A deficit in voice discrimination is, thus, likely to hinder concurrent speech perception. Again, this was shown for F0 or actual voice differences (Stickney et al. 2004, 2007; Qin & Oxenham, 2005; Gaudrain et al. 2008; Luo et al. 2009), but such consequences for a deficit in VTL sensitivity has not been documented yet.

Relation to Acoustic Simulations of Implants

Gaudrain and Başkent (2015) measured F0 and VTL JNDs using a collection of vocoders to systematically investigate which parameters had an effect on VTL and F0 JNDs. The main conclusion that they drew was that VTL JNDs are more affected by a loss of spectral resolution than F0 JNDs (their experiment 1). They, thus, predicted that VTL JNDs would be more likely than F0 JNDs to be larger than the typical male–female difference along these respective dimensions. In actual CI users, when we compared JNDs directly to each other, we found that F0 and VTL JNDs are equally degraded in CI listening, relatively to NH, that is, contrary to the prediction, VTL JNDs are not more degraded than F0 JNDs. However, the second part of the data analysis verified the prediction that VTL JNDs would be more likely than F0 JNDs to be larger than the typical male–female difference.

As mentioned in Introduction, the vocoder simulations of CIs had produced differing patterns in utilization of F0 and VTL cues for gender categorization than the patterns of actual CI users in the study by Fuller et al. (2014). The results from actual CI users of the present study, in contrast, partially overlapped with the results from simulations used for F0 and VTL JNDs in the study by Gaudrain and Başkent (2015). This difference perhaps is closely related to the different perceptual processes related to these tasks. For JNDs, especially with the design of the odd-one-out experiment, detecting any voice difference is sufficient. For gender categorization, this is not sufficient—the listener has to make a correct interpretation of this difference, such as extracting the height of the speaker, as well. Hence, it is possible that JNDs are more governed by the degradations in the sensory processes, such as degraded temporal–spectral cues, factors that are more easily captured in vocoder simulations, while gender categorization is governed by both these degradations and the higher-level cognitive mechanisms needed for further interpretation, which are not captured by these simulations.

One advantage of this overlap between actual and simulated CI data, however, is that this situation offers a good opportunity for developing a robust simulation tool that can be used for further research. Here, for this purpose, we examine which of the vocoders used by Gaudrain and Başkent (2015) best match the present CI data. Among the tested vocoders, the 4-band sine wave or noise vocoders (with filter slopes of 48 to 72 dB/octave) they used in experiments 1 and 3 seem to provide the closest match to the actual CI VTL JND data, both in terms of average values and across-subject variability. The closest match for the F0 JNDs was the 6-band noise vocoder used in experiment 2. The 4-band sine wave vocoder they used in experiment 1 yielded F0 JNDs that are markedly better than the average F0 JND obtained by the present CI participants. Unfortunately, Gaudrain and Başkent (2015) did not measure F0 JNDs with their 4-band noise vocoder. However, it is likely that performance would have been slightly better with that vocoder than with the 6-band noise

vocoder because wider noise bands allow for deeper amplitude modulation, which means better coding of the F0.

It, thus, seems that noise vocoders with 4 to 6 bands should be able to capture the gist of the behavior of actual CI listeners in the VTL and F0 JND tasks. These numbers of bands in the vocoder are similar to those reported by Stickney et al. (2004) as yielding speech comprehension performance in NH listeners that is comparable to that of actual CI listeners.

CONCLUSIONS

The perception of F0 and VTL, together or separately, is not only related to gender perception (Smith & Patterson 2005; Hillenbrand & Clark 2009; Meister et al. 2009; Fuller et al. 2014; Skuk & Schweinberger 2014) but also related to speaker size perception (Ives et al. 2005; Smith et al. 2005), speaker identification (Gaudrain et al, Reference Note 3), pragmatic prosody (Chatterjee & Peng 2008; Meister et al. 2009), and emotion perception (Chuenwattanapranithi et al. 2008; Sauter et al. 2010). In addition to providing all of these extra information supplementing speech communication, the vocal cues also play an important role for understanding speech in noise—F0 and VTL differences between competing voices represent crucial cues that NH listeners use to segregate one voice from the other, which seem to significantly improve speech-on-speech perception (Darwin et al. 2003; Mackersie et al. 2011; Başkent & Gaudrain 2016).

Cochlear implant users have been shown to have specific difficulties with perceiving speaker gender (Massida et al. 2013; Fuller et al. 2014) and vocal emotions (Luo et al. 2007; Chatterjee et al. 2015; Gilbers et al. 2015) and with taking advantage of voice differences for competing speech perception (Stickney et al. 2004, 2007; Pyschny et al. 2011). The F0 and VTL JNDs we are reporting provide a common explanation to all these deficits: the principal voice cues, although represented through the implant, are not available with sufficient precision to make them useful for many real-life situations.

This is particularly true for VTL whose variations, both within and across speakers, are relatively small compared with those of F0. Yet, within this small range of variation, VTL produces perceptual effects of the same magnitude as F0 does over a much larger range. For instance, in Başkent and Gaudrain (2016), a VTL difference of 1 st produced the same speech-on-speech intelligibility increase as an F0 difference of 2.3 st. In Fuller et al. (2014), a VTL difference of 1 st affected perceived voice sex as much as an F0 difference of 4.8 st. Therefore, while CI listeners present abnormally enlarged JNDs both for VTL and F0, in some situations, the loss in VTL sensitivity could have more dramatic consequences on voice and speech perception than the reduction in F0 sensitivity does.

Finally, because VTL perception relies primarily on spectral resolution (Gaudrain & Başkent 2015), VTL JNDs constitute a measure of spectral resolution, which could be compared with spectral-ripple detection, but more directly related to speech perception. This measure could, thus, be used in clinical setting for fitting purposes or in developing of the new strategies.

REFERENCES

Abercrombie, D. (1982). *Elements of General Phonetics*. Edinburgh: University Press.

Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *J Acoust Soc Am, 139*, EL51–EL56.

Başkent, D., Gaudrain, E., Tamati, T. N., & Wagner, A. (2016). Perception and psychoacoustics of speech in cochlear implant users. In A. T. Cacace, E. de Kleine, A. G. Holt, & P. van Dijk (Eds.), *Scientific Foundations of Audiology: Perspectives From Physics, Biology, Modeling, and Medicine* (pp. 285–319). San Diego, CA: Plural Publishing, Inc.

Baumann, U., & Nobbe, A. (2004). Pulse rate discrimination with deeply inserted electrode arrays. *Hear Res, 196*, 49–57.

Bhattacharya, A., Vandali, A., Zeng, F. G. (2011). Combined spectral and temporal enhancement to improve cochlear-implant speech perception. *J Acoust Soc Am, 130*, 2951–2960.

Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology, 34*, 260–284.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am, 109*, 1101–1109.

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. *J Acoust Soc Am, 127*(3), 1469–1478.

Carlyon, R. P., van Wieringen, A., Long, C. J., et al. (2002). Temporal pitch mechanisms in acoustic and electric hearing. *J Acoust Soc Am, 112*, 621–633.

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

Chatterjee, M., Zion, D. J., Deroche, M. L., et al. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hear Res, 322*, 151–162.

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., et al. (2008). Encoding emotions in speech with the size code. A perceptual investigation. *Phonetica, 65*, 210–230.

Clarke, J., Başkent, D., Gaudrain, E. (2016). Pitch and spectral resolution: A systematic comparison of bottom-up cues for top-down repair of degraded speech. *J Acoust Soc Am, 139*, 395–405.

Cleary, M., & Pisoni, D. B. (2002). Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results. *Ann Otol Rhinol Laryngol Suppl, 189*, 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. *J Speech Lang Hear Res, 48*, 204–223.

Darwin, C. J., Brungart, D. S., Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J Acoust Soc Am, 114*, 2913–2922.

Deroche, M. L., Lu, H. P., Limb, C. J., et al. (2014). Deficits in the pitch sensitivity of cochlear-implanted children speaking English or Mandarin. *Front Neurosci*, 8, 282.

Deroche, M. L., Culling, J. F., Lavandier, M., et al. (2017). Reverberation limits the release from informational masking obtained in the harmonic and binaural domains. *Atten Percept Psychophys*, 79, 363–379.

Fant, G. C. M. (1970). *Acoustic Theory of Speech Production*. The Hague: Mouton.

Fielden, C. A., Kluk, K., Boyle, P. J., et al. (2015). The perception of complex pitch in cochlear implants: A comparison of monopolar and tripolar stimulation. *J Acoust Soc Am*, 138, 2524–2536.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *J Acoust Soc Am*, 106(3 Pt 1), 1511–1522.

Fitzgerald, M. B., Shapiro, W. H., McDonald, P. D., et al. (2007). The effect of perimodiolar placement on speech perception and frequency discrimination by cochlear implant users. *Acta Otolaryngol*, 127, 378–383.

Friesen, L. M., Shannon, R. V., Baskent, D., et al. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am*, 110, 1150–1163.

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. *J Assoc Res Otolaryngol*, 5, 253–260.

Fu, Q. J., Chinchilla, S., Nogaki, G., et al. (2005). Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *J Acoust Soc Am*, 118(3 Pt 1), 1711–1718.

Fuller, C. D., Gaudrain, E., Clarke, J. N., et al. (2014). Gender categorization is abnormal in cochlear implant users. *J Assoc Res Otolaryngol*, 15, 1037–1048.

Gaudrain, E. (2016). Can spectral centroid explain voice pitch and vocal-tract length perception in normal-hearing and cochlear implant listeners? *J Acoust Soc Am*, 140, 3439–3439.

Gaudrain, E., & Başkent, D. (2015). Factors limiting vocal-tract length discrimination in cochlear implant simulations. *J Acoust Soc Am*, 137, 1298–1308.

Gaudrain, E., Deeks, J. M., Carlyon, R. P. (2017). Temporal regularity detection and rate discrimination in cochlear-implant listeners. *J Assoc Res Otolaryngol*, 18, 387–397.

Gaudrain, E., Grimault, N., Healy, E. W., et al. (2008). Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation. *J Acoust Soc Am*, 124, 3076–3087.

Gaudrain, E., Grimault, N., Healy, E. W., Béra, J.-C. (2012). The relationship between concurrent speech segregation, pitch-based streaming of vowel sequences, and frequency selectivity. *Acta Acust United Acust*, 98, 317–327.

Geurts, L., & Wouters, J. (2001). Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *J Acoust Soc Am*, 109, 713–726.

Gfeller, K., Turner, C., Mehr, M., et al. (2002). Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants Int*, 3, 29–53.

Gilbers, S., Fuller, C., Gilbers, D., et al. (2015). Normal-hearing listeners' and cochlear implant users' perception of pitch cues in emotional speech. *i-Perception*, 6, 0301006615599139.

Glasberg, B. R., & Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *J Audio Eng Soc*, 50(5), 331–342.

Green, T., Faulkner, A., Rosen, S. (2002). Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants. *J Acoust Soc Am*, 112(5 Pt 1), 2155–2164.

Green, T., Faulkner, A., Rosen, S. (2004). Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *J Acoust Soc Am*, 116(4 Pt 1), 2298–2310.

Green, T., Faulkner, A., Rosen, S., et al. (2005). Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification. *J Acoust Soc Am*, 118, 375–385.

He, A., Deroche, M. L., Doong, J., et al. (2016). Mandarin tone identification in cochlear implant users using exaggerated pitch contours. *Otol Neurotol*, 37, 324–331.

Henry, B. A., & Turner, C. W. (2003). The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *J Acoust Soc Am*, 113, 2861–2873.

Henry, B. A., Turner, C. W., Behrens, A. (2005). Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners. *J Acoust Soc Am*, 118, 1111–1121.

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f0 and formant frequencies in distinguishing the voices of men and women. *Atten Percept Psychophys*, 71, 1150–1166.

Hong, R. S., & Turner, C. W. (2009). Sequential stream segregation using temporal periodicity cues in cochlear implant recipients. *J Acoust Soc Am*, 126, 291–299.

Ives, D. T., Smith, D. R., Patterson, R. D. (2005). Discrimination of speaker size from syllable phrases. *J Acoust Soc Am*, 118, 3816–3822.

Kawahara, H., & Irino, T. (2004). Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In P. L. Divenyi (Ed.), *Speech Separation by Humans and Machines* (pp. 167–180). Massachusetts: Kluwer Academic.

Kovačić, D., & Balaban, E. (2009). Voice gender perception by cochlear implantees. *J Acoust Soc Am*, 126, 762–775.

Laneau, J., Wouters, J., Moonen, M. (2004). Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. *J Acoust Soc Am*, 116, 3606–3619.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*, 49(2B), 467–477.

Li, T., & Fu, Q.-J. (2011). Voice gender discrimination provides a measure of more than pitch-related perception in cochlear implant users. *Int J Audiol, 50*, 498–502.

Luo, X., Fu, Q. J., Galvin, J. J. 3rd. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends Amplif*, 11, 301–315.

Luo, X., Fu, Q. J., Wu, H. P., et al. (2009). Concurrent-vowel and tone recognition by Mandarin-speaking cochlear implant users. *Hear Res*, 256, 75–84.

Mackersie, C. L. (2003). Talker separation and sequential stream segregation in listeners with hearing loss: Patterns associated with talker gender. *J Speech Lang Hear Res*, 46, 912–918.

Mackersie, C. L., Dewey, J., Guthrie, L. A. (2011). Effects of fundamental frequency and vocal-tract length cues on sentence segregation by listeners with hearing loss. *J Acoust Soc Am*, 130, 1006–1019.

Mackersie, C. L., Prida, T. L., Stiles, D. (2001). The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss. *J Speech Lang Hear Res*, 44, 19–28.

Massida, Z., Belin, P., James, C., et al. (2011). Voice discrimination in cochlear-implanted deaf subjects. *Hear Res*, 275, 120–129.

Massida, Z., Marx, M., Belin, P., et al. (2013). Gender categorization in cochlear implant users. *J Speech Lang Hear Res*, 56, 1389–1401.

McDermott, H. J., & McKay, C. M. (1997). Musical pitch perception with electrical stimulation of the cochlea. *J Acoust Soc Am*, 101, 1622–1631.

Meister, H., Fürsen, K., Streicher, B., et al. (2016). The use of voice cues for speaker gender recognition in cochlear implant recipients. *J Speech Lang Hear Res*, 59, 546–556.

Meister, H., Landwehr, M., Pyschny, V., et al. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *Int J Audiol*, 48, 38–48.

Moore, B. C. J., & Carlyon, R. P. (2005). Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In C. J. Plack, A. J. Oxenham, R. R. Fay, A. N. Popper (Eds.), *Pitch: Neural Coding and Perception* (pp. 234–277). New-York, NY: Springer/Birkhäuser.

Mühler, R., Ziese, M., Rostalski, D. (2009). Development of a speaker discrimination test for cochlear implant users based on the Oldenburg Logatome corpus. *ORL J Otorhinolaryngol Relat Spec*, 71, 14–20.

Nogueira, W., Litvak, L. M., Edler, B., Ostermann, J., Büchner, A.(2009). Signal processing strategies for cochlear implants using current steering. *EURASIP J Adv Signal Process*, 2009, 1–21.

Nogueira, W., Rode, T., Büchner, A. (2016). Spectral contrast enhancement improves speech intelligibility in noise for cochlear implants. *J Acoust Soc Am*, 139, 728–739.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *J Acoust Soc Am*, 24(2), 175–184.

Pyschny, V., Landwehr, M., Hahn, M., et al. (2011). Bimodal hearing and speech perception with a competing talker. *J Speech Lang Hear Res*, 54, 1400–1415.

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *J Acoust Soc Am*, 114, 446–454.

Qin, M. K., & Oxenham, A. J. (2005). Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear Hear*, 26, 451–460.

Sagi, E., Meyer, T. A., Kaiser, A. R., et al. (2010). A mathematical model of vowel identification by users of cochlear implants. *J Acoust Soc Am, 127*, 1069–1083.

Sauter, D. A., Eisner, F., Calder, A. J., et al. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Q J Exp Psychol (Hove), 63*, 2251–2272.

Shannon R. V. (1983). Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. *Hear Res, 11*, 157–189.

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *J Speech Lang Hear Res, 57*, 285–296.

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *J Acoust Soc Am, 118*, 3177–3186.

Smith, D. R., Patterson, R. D., Turner, R., et al. (2005). The processing and perception of size information in speech sounds. *J Acoust Soc Am, 117*, 305–318.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychol Bull, 87*(2), 245–251.

Stickney, G. S., Zeng, F. G., Litovsky, R., et al. (2004). Cochlear implant speech recognition with speech maskers. *J Acoust Soc Am, 116*, 1081–1091.

Stickney, G. S., Assmann, P. F., Chang, J., et al. (2007). Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. *J Acoust Soc Am, 122*, 1069–1078.

Summers, V., & Leek, M. R. (1998). F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *J Speech Lang Hear Res, 41*, 1294–1306.

Turner, R. E., Walters, T. C., Monaghan, J. J., et al. (2009). A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data. *J Acoust Soc Am, 125*, 2374–2386.

Vandali, A. E., Sucher, C., Tsang, D. J., et al. (2005). Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies. *J Acoust Soc Am, 117*, 3126–3138.

Winn, M. B., Chatterjee, M., Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *J Acoust Soc Am, 131*, 1465–1479.

Won, J. H., Drennan, W. R., Rubinstein, J. T. (2007). Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *J Assoc Res Otolaryngol, 8*, 384–392.

Zeng, F. G. (2002). Temporal pitch in electric hearing. *Hear Res, 174*, 101–106.

REFERENCE NOTES

Başkent, D., Luckmann, A., Ceha, J. M., Gaudrain, E., Tamati, T. N. (2017). *The Discrimination of Voice Cues in Simulations of Bimodal Electro-acoustic Cochlear-Implant Hearing.* Poster presented at the 9th Speech in Noise Workshop, Oldenburg, Germany. https://doi.org/10.6084/m9.figshare.5142280.v1.

Gaudrain, E., Bhargava, P., Başkent, D. (2016). Why does gap detection performance in cochlear implant users differ between free-field and direct-stimulation? Presented at the Meeting of the Acoustical Society of America.

Gaudrain, E., Li, S., Ban, V., Patterson, R. (2009). The role of glottal pulse rate and vocal tract length in the perception of speaker identity. Interspeech 2009: 10th Annual Conference of the International Speech Communication Association, *1–5*, 152–155.

To Ear and Hearing Reviewers: Thank You

The *Ear and Hearing* Editorial Board and I would like to give our sincere thanks to all of the peer reviewers who volunteered their time and expertise in 2017. These reviewers provided insightful, rigorous, and constructive critiques for over 700 new and revised manuscripts. It is due in no small part to their efforts that *Ear and Hearing* continues to be ranked as one of the top journals in the categories of Otolaryngology and Audiology & Speech-Language Pathology (2016 *Journal Citation Reports Science Edition [Thomson Reuters, 2016])*.

The names of our 2017 reviewers are listed in Reviewers 2017 in Supplemental Digital Content (http://links.lww.com/EANDH/A406). These reviewers' freely-given time and advice contributes to the overall scientific excellence and clinical relevance of the articles published in *Ear and Hearing* and for that we are truly grateful.

Brenda M. Ryals
Editor-in-Chief
James Madison University
Harrisonburg, VA