## Title
Gender Categorization Is Abnormal in Cochlear Implant Users



## Journal
Journal of the Association for Research in Otolaryngology, 15(6)

## ISSN
1525-3961

## Authors
Fuller, Christina D
Gaudrain, Etienne
Clarke, Jeanne N
et al.

## Publication Date
2014-12-01

## DOI
10.1007/s10162-014-0483-7

Peer reviewed

JARO 15: 1037–1048 (2014)
DOI: 10.1007/s10162-014-0483-7


JARO
Journal of the Association for Research in Otolaryngology

*Research Article*

# Gender Categorization Is Abnormal in Cochlear Implant Users

Christina D. Fuller,[1,3] Etienne Gaudrain,[1,3] Jeanne N. Clarke,[1,3] John J. Galvin,[2] Qian-Jie Fu,[2] Rolien H. Free,[1,3] and Deniz Başkent[1,3]

[1]*Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, P.O. Box 30.001, BB21, 9700 RB, Groningen, The Netherlands*

[2]*David Geffen School of Medicine, Department of Head and Neck Surgery, University of California, Los Angeles, Los Angeles, CA, USA*

[3]*University of Groningen, Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, Groningen, The Netherlands*

Received: 11 September 2013; Accepted: 29 July 2014; Online publication: 30 August 2014

## ABSTRACT

In normal hearing (NH), the perception of the gender of a speaker is strongly affected by two anatomically related vocal characteristics: the fundamental frequency (F0), related to vocal pitch, and the vocal tract length (VTL), related to the height of the speaker. Previous studies on gender categorization in cochlear implant (CI) users found that performance was variable, with few CI users performing at the level of NH listeners. Data collected with recorded speech produced by multiple talkers suggests that CI users might rely more on F0 and less on VTL than NH listeners. However, because VTL cannot be accurately estimated from recordings, it is difficult to know how VTL contributes to gender categorization. In the present study, speech was synthesized to systematically vary F0, VTL, or both. Gender categorization was measured in CI users, as well as in NH participants listening to unprocessed (only synthesized) and vocoded (and synthesized) speech. Perceptual weights for F0 and VTL were derived from the performance data. With unprocessed speech, NH listeners used both cues (normalized perceptual weight: F0=3.76, VTL=5.56). With vocoded speech, NH listeners still made use of both cues but less efficiently (normalized perceptual weight: F0=1.68, VTL=0.63). CI users relied almost exclusively on F0 while VTL perception was profoundly impaired (normalized perceptual weight: F0=6.88, VTL=0.59). As a result, CI users' gender categorization was abnormal compared to NH listeners. Future CI signal processing should aim to improve the transmission of both F0 cues and VTL cues, as a normal gender categorization may benefit speech understanding in competing talker situations.

**Keywords:** cochlear implants, gender categorization, fundamental frequency, vocal tract length, vocal characteristics

## INTRODUCTION

In "cocktail party" listening conditions, normal hearing (NH) listeners use the voice characteristics of different talkers to track and listen to a target talker. The ability to identify the gender of a voice may help to sort out various talkers in a multi-talker environment, especially when two talkers are speaking at the same time. Voice differences across speakers of the same gender can improve intelligibility of the target speech by more than 20 percentage points (Brungart 2001). Voice differences across gender can increase intelligibility by 50 percentage points (Brungart 2001; Festen and Plomp 1990).

NH listeners use two anatomically related vocal characteristics to identify the gender of a talker: (i) the fundamental frequency (F0) of the voice, related to perceived vocal pitch and determined by the glottal

*Correspondence to*: Christina Fuller · Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen · University of Groningen · P.O. Box 30.001, BB21, 9700 RB, Groningen, The Netherlands. email: c.d.fuller@umcg.nl

pulse rate, and (ii) vocal tract length (VTL),[1] mainly related to the height of the speaker (Fitch and Giedd 1999). F0 and VTL have been shown to similarly influence NH listeners' voice gender identification (Skuk and Schweinberger 2013) and concurrent speech perception (Darwin et al. 2003).

Unlike NH listeners, cochlear implant (CI) users do not benefit from differences in speaker's gender in competing talker situations (Luo et al. 2009; Stickney et al. 2004). This may be partly due to poor representation and/or perception of voice characteristics. Previous studies have shown that CI users' gender categorization performance is highly variable and generally poorer than that of NH listeners (Fu et al. 2004, 2005; Kovačić and Balaban 2009, 2010; Massida et al. 2013; Wilkinson et al. 2013). It was argued in these studies that CI users might rely more on F0 than NH listeners. In Fu et al. (2005), when the F0s of the talkers were overlapping, CI users' gender categorization performance was poorer than that of NH participants listening to sinewave-vocoded stimuli (68 vs. 92 % correct). Subsequently, Kovačić and Balaban (2009) also observed that gender categorization was particularly difficult for CI listeners when the F0 was within the overlap region between the male and female ranges. Recently, Massida et al. (2013) created a continuum between a typical female voice and a typical male voice using a morphing technique. They observed that CI users had shallower psychometric functions than NH listeners and concluded that categorization of ambiguous voices, around the middle point of the continuum, was more difficult for CI users than for NH listeners.

However, the origins of these difficulties are, as yet, unknown. The studies cited above essentially focus on the role of F0, but VTL could also play a crucial role in the categorization of voices, especially when the F0 cue is ambiguous. For instance, although F0 values were estimated and reported in Fu et al. (2005), there was no attempt to estimate talker VTL values. This is probably explained by the fact that, unlike F0, it is difficult to estimate VTL from recordings. To date, the best estimators only achieve between 10 and 30 % root-mean-square-error accuracy (Lammert et al. 2013), which is similar to differences between males and females when measured anatomically (15 %, according to Fant 1970). Thus, it is unclear in Fu et al. (2005) and Massida et al. (2013) to what degree VTL cues might have contributed to CI and NH performance. Moreover, although F0 and VTL seem to be the most important cues for gender categorization in NH listeners (Skuk and Schweinberger 2013), other cues also contribute to gender categorization in recordings of real speech, such as breathiness (Holmberg et al. 1988; Van Borsel et al. 2009) or intonation (Fitzsimons et al. 2001). These cues may be used differently by CI users, further complicating the interpretation of past studies based on natural utterances by male and female speakers. One indication that VTL cues might be particularly degraded comes from a study by Mackersie et al. (2011) who observed that listeners with mild to severe hearing loss above 1 kHz could not benefit from VTL differences in a concurrent sentence experiment. By extension, it seems likely that CI listeners might also have difficulties with this cue, but this remains to be shown.

In the present study, we focused on the role of F0 and VTL for gender categorization in NH and CI listeners, by artificially manipulating these two dimensions in stimuli resynthesized from one single female voice. Although the reduced spectral resolution inherent to CI sound transmission notoriously degrades the F0 representation, pitch percept remains possible on the basis of temporal cues (see Moore and Carlyon 2005 for a review). In particular, it can be expected that F0 differences of about one octave that separate typical male from typical female voices would be accessible. However, when the F0 difference is smaller, this cue might become more ambiguous and less useful. VTL, on the other hand, affects the location of the formants (see Fig. 1). In other words, accurate perceptual estimates of VTL rely on accurate perception of the formant peak locations. The limited spectral resolution of the implant, therefore, would be expected to severely hinder the perception of this cue, although such an effect has not been documented. The electrodograms in Fig. 1 suggest that the typical VTL difference between a male and a female voice results in a shift of the electrical stimulation pattern by one electrode. Different spectral resolution measures yield slightly different predictions regarding the detectability of such a shift (see "DISCUSSION" for more details). It could thus also be the case that impaired VTL perception prevents voices with ambiguous F0s from being properly categorized.

The purpose of the present study was to directly measure and characterize the contribution of F0 and VTL cues to gender categorization by CI users as compared to NH listeners. Because VTL cannot be easily estimated from recordings of real speech,

[1] VTL affects the center frequency of the formants and is sometimes referred to as 'formant dispersion': lengthening the vocal tract by a given factor results in dividing all formant frequencies by that same factor, equivalent to an homothetic translation of the spectral envelope on a log-frequency axis (a detailed explanation can be found in Patterson et al. 2010). One of the main differences between VTL and F0, unlike for glottal pulse rate, F0, and pitch, there are no commonly defined terms to denote the acoustic and perceptual analogs of VTL. In the present study, we therefore used the term VTL to refer to the physical dimension, the apparent acoustic dimension, as well as the perceived quantity related to this anatomical property.

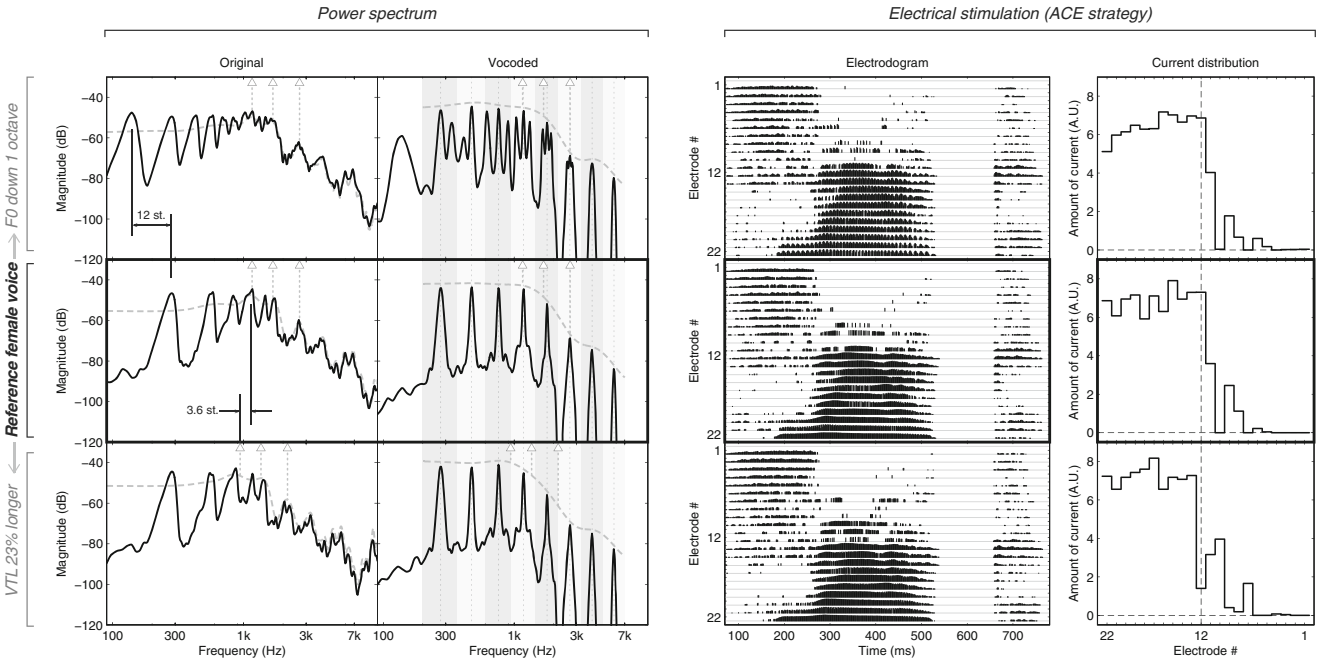


**FIG. 1.** Power spectrum, waveform, and electrodogram of the vowel/aa/in "Vaak." A different voice is represented per row. The stimulus resynthesized with the original parameters of the female voice is shown in the *middle row*. The *top row* shows the F0 changes only, by an octave down. The *bottom row* shows the VTL changed to be made 23 % longer, which results in shifting all the formants down by 3.6 semitones (st). The *left panel* shows, over the duration of the vowel, the spectra, for the non-vocoded (*left column*, noted "Original") and vocoded (*right column*) versions of the stimulus. The spectrum itself is shown by the *solid black line*, visualizing the harmonics and/or the sinusoidal carriers of the vocoder. The spectral envelope is represented by the *dashed gray line* as extracted by STRAIGHT for the non-vocoded sounds on the *left* and as an interpolation between the carriers for the vocoded sounds on the *right*. The locations of the first three formants, based on a visual inspection of the envelope, are pointed out by the *triangles and stems*, for both the *left and the right columns*. The analysis filter bands of the vocoder are shown in the *gray areas in the right column*, whereas the sinewave carrier's frequency is shown with a *dotted line*. The *right panel* shows the electrical stimulation as obtained with the Nucleus Matlab Toolbox (v4.31, Cochlear Limited, Australia) using an ACE strategy with a default frequency map. The *left column* shows the electrodogram for the whole word, while the *right column* shows the total amount of current per channel accumulated over the duration of the vowel. The *vertical dashed line* in this column locates the middle electrode.

speech stimuli were resynthesized to effect systematic manipulation of F0 and apparent VTL cues. Gender categorization with resynthesized speech was measured as a function of VTL and F0 in CI users and in NH subjects listening to non-vocoded and vocoded versions of the synthesized stimuli. Perceptual weights for F0 and VTL were derived from the CI, NH, and NH-vocoded gender categorization data. We predicted that the poor spectral resolution of the implant would affect the relative weights attributed to VTL and F0. A similar prediction was also made for NH listeners tested with degraded spectral cues in the vocoded condition.

## METHODS

### Participants

Nineteen postlingually deafened CI users (11 male and 8 female, mean age=64.6 years, range=28–78 years) with more than 1 year of CI experience (mean experience=4.6 years, range=1–12 years) were recruited. One CI user was bilaterally implanted. The details of all CI participants are shown in Table 1.

This study was conducted in parallel with Fuller et al. (2014), where a musician effect was explored on gender categorization, and the same non-musician NH listeners comprised the control group in both studies. The criterion for non-musician was to have not received musical training within the 7 years preceding the study. The motivation for excluding musicians was that it was suspected that musicians might make different use of voice cues than non-musicians, especially in degraded conditions (which was confirmed by Fuller et al. 2014). As such, non-musician NH listeners were thought to be a better control group for CI listeners, who also tend to be not musically involved post-implantation (e.g., Fuller et al. 2012), than NH listeners with extensive musical expertise.

**TABLE 1**
Details of the CI participants

| *Subject number* | *Gender* | *Years of CI use* | *Cochlear implant* | *Speech processor* | *Rate of stimulation* |
|---|---|---|---|---|---|
| 1 | Male | 9 | CI24R CS | CP810 | 900 |
| 2 | Male | 5 | HiRes 90K Helix | Harmony | 3,712 |
| 3 | Male | 4 | HiRes 90K Helix | Harmony | 849 |
| 4 | Male | 1 | CI24RE CA | CP810 | 900 |
| 5 | Female | 4 | HiRes 90K Helix | Harmony | 2,184 |
| 6 | Female | 12 | CI24R k | CP810 | 900 |
| 7 | Male | 2 | CI24RE CA | CP810 | 900 |
| 8 | Male | 5 | CI24RE CA | Freedom | 900 |
| 9 | Female | 2 | CI24RE CA | CP810 | 900 |
| 10 | Female | 3 | CI512 | CP810 | 900 |
| 11 | Male | 6 | HiRes 90K Helix | Harmony | 2,900 |
| 12 | Male | 4 | HiRes 90K Helix | Harmony | 1,740 |
| 13 | Female | 3 | CI24RE CA | CP810 | 900 |
| 14 | Male | 8 | CI24R CA | CP810 | 900 |
| 15 | Male | 5 | CI 11+11+2M | Freedom | 900 |
| 16 | Female | 2 | CI24RE H | CP810 | 900 |
| 17 | Male | 2 | CI24RE CA | CP810 | 900 |
| 18 | Female | 1 | CI24RE CA | CP810 | 900 |
| 19 | Female | 9 | CI24R CA | Freedom | 900 |

The NH control group of the present study comprised 19 NH participants (3 male and 16 female; mean age=22.1 years, range=19–28 years), who were a subset of the 25 NH non-musician listeners reported in Fuller et al. (2014). NH participants were audiometrically selected to have pure tone thresholds better than 20 dB HL at frequencies between 250 and 4,000 Hz. All participants were native Dutch speakers, with no neurological disorders.

The study protocol was approved by the Medical Ethical Committee of the University Medical Center Groningen. Detailed information about the study was provided to the participants before data collection, and written informed consent was obtained. All subjects received financial reimbursement for their participation.

## Stimuli

**Speech Synthesis.** The sources for subsequent speech synthesis were four meaningful Dutch words in CVC format ("bus," "vaak," "leeg" and "pen," meaning "bus," "often," "empty," and "pencil," respectively), taken from the NVA corpus (Bosman and Smoorenburg 1995). The source speech tokens were spoken by a single Dutch female talker. The average word duration was 0.83 s and the average F0 was 201 Hz. The VTL was estimated to be 13.5 cm, based on an average height of 169 cm for Dutch women and the regression between VTL and height reported by Fitch and Giedd (1999).

The source speech tokens were manipulated using the STRAIGHT software (v40.006b; Kawahara et al. 1999), implemented in MATLAB. Both the F0 and the VTL of the source female voice were manipulated to obtain a male voice at the extreme parameter values, where the F0 was decreased by an octave and the VTL was increased by 23 % (resulting in a downward spectral shift of 3.6 semitones). To achieve this in STRAIGHT, the speech signal was first decomposed into the F0 contour and the spectral envelope. All values of the F0 contour were then multiplied by a specific factor, resulting in a change in the average F0 while preserving the relative fluctuations. The VTL lengthening was effected by compressing the extracted spectral envelope toward the low frequencies. The modified components were then recombined via a pitch synchronous overlap-add resynthesis method. In previous studies with similar manipulations, Clarke et al. (2014) confirmed that the chosen F0 and VTL values, applied together, indeed made the listeners perceive a talker of a different gender than the original one, and Fuller et al. (2014) confirmed these values provided a full characterization of gender categorization from the female's voice to that of a man's.

In the present study, similar to the studies by Clarke et al. and Fuller et al., intermediate steps were created between the source female voice and the target male talker. The F0 was varied to be 0, 3, 6, 9, or 12 semitones below the F0 of the original female source, which corresponds to changes of 0, 19, 41, 68, and 100 % or average F0 values of 201, 169, 142, 119, and 100 Hz. The VTL was varied to be 0.0, 0.7, 1.6, 2.4, 3.0, or 3.6 semitones, i.e., 0, 4, 7, 14, 19, and 23 % longer than the VTL of the female source, corre-

sponding to lengths of 13.5, 14.1, 14.8, 15.5, 16.1, and 16.6 cm. These combinations produced 30 different voices and resulted in a total of 120 stimuli (5 F0 values×6 VTL values×4 words). All stimuli were resynthesized, even when the original values of F0 and VTL were used. Smith et al. (2007) estimated distributions of natural voices in the F0–VTL plane based on Peterson and Barney (March 1952) and Fitch and Giedd (1999). Using these estimates, we calculated that all the synthesized voices were within 99.7 % of the adult population, and 22 of the 30 voices were within 95 %.

**Vocoder Processing.** Similar to the studies by Fu et al. (2004, 2005), a simple acoustic CI simulation was used in the form of an eight-channel, sinewave vocoder. The vocoder was based on the continuous interleaved sampling strategy (Wilson et al. 1991) and was implemented using the Angelsound™ software (Emily Shannon Fu Foundation, http://www.angelsound.tigerspeech.com/). An eight-channel vocoder was used because it has been shown to yield both gender categorization and speech intelligibility performance similar to that of the best performing CI users (Fu et al. 2004, 2005; Friesen et al. 2001). Both of these are an indication that the eight-band vocoder likely delivers spectral resolution functionally similar to that of better-performing CI users. Despite this functional similarity, it should be noted that this type of vocoder does not accurately reflect the processes happening in actual implants and is here merely used to provide an indication of how degraded spectral cues can affect the task in normal hearing.

The input frequency range was 200–7,000 Hz. The acoustic input was bandpass-filtered into eight frequency analysis bands using fourth order Butterworth filters. The band cutoff frequencies were distributed according to Greenwood (1990) frequency-place formula. For each band, a sinusoidal carrier was generated; the frequency of the sinewave carrier was equal to the center frequency of the analysis filter (i.e., the geometric mean of the band cutoff frequencies). The temporal envelope was extracted from each band using half-wave rectification and lowpass filtering with a Butterworth filter (cutoff frequency=160 Hz, fourth order). These envelopes modulated the corresponding sinusoidal carriers. Finally, the modulated carriers were summed and the overall level was adjusted to be the same level as the original speech token. Figure 1 shows from the left to the right panel the spectra of the generated sounds, the electrodograms, and the total amount of current per channel accumulated over the duration of the vowel, respectively. The *middle row* shows the stimulus resynthesized in STRAIGHT, with the F0 and VTL of the original female voice. The *top row* shows the stimulus resynthesized with only the F0 shifted by an octave down. The *bottom row* shows the stimulus with only the VTL made 23 % longer, which resulted in all formants being shifted down by 3.6 semitones.

## Procedure

All synthesized stimuli, with or without vocoding, were presented using AngelSound™ software (Emily Shannon Fu Foundation, http://www.angelsound.tigerspeech.com/). The stimuli were routed via a PC with an Asus Virtuoso Audio Device soundcard (ASUSTeK Computer Inc, Fremont, USA), converted to an analog signal via a DA10 digital-to-analog converter of Lavry Engineering Inc. (Washington, USA), and then played at 65 dB SPL in free field in an anechoic chamber. The participants were seated at a distance of 1 m from the speaker (Tannoy Precision 8D; Tannoy Ltd., North Lanarkshire, UK). During testing, the participant heard a randomly selected stimulus, and their task was to select one of two response buttons shown on screen labeled "man" or "vrouw" (i.e., "man" or "woman," in Dutch), to indicate the gender of the talker. The participants replied on an A1 AOD 1908 touch screen (GPEG International, Woolwich, UK). CI users were tested with their own clinical processor. The CI participants were instructed to use their everyday clinical volume and sensitivity settings and to use these settings throughout testing. CI listeners were tested with non-vocoded stimuli. NH listeners were tested first with non-vocoded stimuli and then with vocoded stimuli.

Participant responses were directly scored by the program. NH listeners were not naïve to the vocoding processing as they had participated in similar experiments before. No training was provided to either participant group for the gender recognition task. The gender categorization task lasted for 10 min. This resulted in a total testing time of approximately 20 min for NH participants and 10 min for CI users.

## Statistical Analysis

All statistical analyses were done in R (version 3.01, R Foundation for Statistical Computing, Vienna, Austria) using the lme4 package (version 1.0-5, Bates et al. 2013). A generalized linear mixed effects model with a logit link function was used following the method described by Jaeger (2008). The model selection started from the full factorial model in lme4 syntax:

$$\text{score} \sim f0 * \text{vtl} * \text{moh} + (1 + f0 * \text{vtl} \mid \text{subject})$$

The variable *score* is the proportion of "man" responses. The *f0* and *vtl* factors are normalized dimensions defined as $f0 = -\Delta F0/12 - 1/2$ and $vtl = \Delta VTL/3.6 - 1/2$ where $\Delta F0$ and $\Delta VTL$ represent

the F0 and VTL difference in semitones relative to the original voice. With these normalized dimensions, the point (*f0*=–0.5, *vtl*=–0.5) represents the original female voice, while the point (*f0*=0.5, *vtl*=0.5) represents the artificially created male voice. The factor *moh* codes the mode of hearing (NH, NH-vocoded, or CI). The notation "(…|…)" denotes the random effect, here per subject, with "1" thus representing a random intercept per subject. The full factorial model had an Akaike information criterion (AIC)=6,342, a Bayesian information criterion (BIC)=6,492, and a log-likelihood=–3,149. The full factorial model was not significantly different from the simpler model below [$\chi^2(7)=13.45$, $p=0.062$], which was then retained as reference:

$$\text{score} \sim (f0 + \text{vtl}) * \text{moh} + (1 + (f0 + \text{vtl}) \mid \text{subject})$$

This model had an AIC=6,341, a BIC=6,443, and a log-likelihood=–3,155. This model has random intercept per subject, as well as random slopes for *f0* and *vtl*, also per subject. Effects for each factor were then tested using the $\chi^2$ statistic and *p* values obtained from the likelihood ratio test comparing the model without the factor of interest against the reference model. In order to compare modes of hearing, the model above was applied to subsets of the data, excluding one mode of hearing at a time and testing the *moh* effect and its interactions within the remaining dataset. Because there were only three comparisons, no correction for multiple comparisons was applied but note that none of the obtained statistics would have changed significance even with a correction as stringent as the Bonferroni correction.

To quantify the contribution of the F0 and VTL, a simpler logistic regression model was used (as described, for instance, by Peng et al. 2009). The "perceptual weights" for each cue were estimated as the coefficients for the *f0* and *vtl* factors in the logistic regression model. In other words, the cue weights are expressed as *a* and *b* in the equation $\text{logit}(score) = a\ f0 + b\ vtl + \varepsilon$, where $\varepsilon$ is the subject-dependent random intercept. Given the coding of the *f0* and *vtl* variables, the cue weights represent variations in log odd ratios over the entire course of change along each of the cues. Cue weights for groups of subjects are accompanied with their associated Wald statistic *z*. Individual cue weights were also obtained using the model used for the statistical analyses, i.e., with random *f0* and *vtl* effects. These are reported in Table 2.

**TABLE 2**

Individual logistic regression coefficients for each subject in each mode of hearing. The "Intercept," "F0," and "VTL" columns correspond, respectively, to $\varepsilon$, *a*, and *b* coefficients of the regression equation given in the "METHODS" section. Summary statistics are given at the bottom of the table. See the section on statistical analyses for details about the calculation of these coefficients. Note that the average of the individual coefficients do not exactly match the coefficients reported in text which result from fitting the logistic regression model to the group data (i.e., without F0 and VTL as random effects)

| | *NH* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Non-vocoded* | | | *Vocoded* | | | *CI* | | |
| | *Intercept* | *F0* | *VTL* | *Intercept* | *F0* | *VTL* | *Intercept* | *F0* | *VTL* |
| 1 | −0.79 | 1.52 | 6.17 | 0.12 | 3.55 | −0.34 | −0.75 | 7.41 | 0.44 |
| 2 | −0.43 | 4.44 | 5.68 | 0.09 | 3.77 | 1.31 | −0.17 | 5.41 | 1.19 |
| 3 | −0.48 | 5.29 | 5.29 | 0.96 | 0.30 | 2.03 | 0.01 | 3.96 | 1.33 |
| 4 | −1.01 | 3.70 | 6.13 | 0.14 | 1.23 | 1.44 | −1.19 | 10.33 | 0.05 |
| 5 | −1.36 | 2.50 | 6.21 | 0.34 | 0.44 | 0.93 | −1.08 | 8.62 | 0.07 |
| 6 | −1.04 | 2.39 | 5.77 | 0.25 | 3.07 | −0.08 | −0.88 | 10.03 | 0.39 |
| 7 | −1.72 | 4.45 | 5.87 | −0.19 | 4.83 | 0.68 | −0.78 | 7.73 | 0.42 |
| 8 | −2.17 | 4.30 | 5.75 | 0.48 | 0.71 | −0.01 | −0.66 | 9.16 | 0.59 |
| 9 | −1.29 | 6.31 | 5.42 | 0.14 | 1.48 | 0.20 | −0.28 | 6.80 | 0.93 |
| 10 | −0.51 | 3.42 | 6.07 | 0.05 | 0.35 | 0.78 | −0.16 | 8.28 | 1.24 |
| 11 | −3.16 | 5.36 | 5.52 | −0.33 | 1.31 | 1.48 | −1.52 | 9.15 | −0.32 |
| 12 | −1.39 | 3.92 | 6.05 | 0.46 | 2.68 | 0.59 | −0.08 | 5.90 | 1.23 |
| 13 | 0.21 | 5.55 | 5.23 | 0.04 | 4.38 | 0.45 | −1.04 | 7.75 | 0.23 |
| 14 | −0.50 | 2.40 | 6.10 | 0.17 | 0.96 | 0.41 | −0.34 | 5.99 | 0.88 |
| 15 | −2.35 | 6.04 | 5.36 | 1.63 | 0.50 | −0.24 | −1.01 | 2.42 | 0.05 |
| 16 | −0.30 | 3.10 | 5.86 | 0.02 | −0.23 | 0.23 | −1.19 | 10.33 | 0.05 |
| 17 | −0.42 | 4.33 | 5.62 | 0.03 | 2.48 | 0.52 | −0.49 | 8.18 | 0.74 |
| 18 | −0.98 | 3.92 | 5.41 | 0.20 | 0.83 | 1.39 | −0.45 | 7.86 | 0.83 |
| 19 | −0.74 | 2.83 | 6.07 | 0.48 | 1.84 | 0.54 | −0.76 | 9.09 | 0.61 |
| Min | −3.16 | 1.52 | 5.23 | −0.33 | −0.23 | −0.34 | −1.52 | 2.42 | −0.32 |
| Max | 0.21 | 6.31 | 6.21 | 1.63 | 4.84 | 2.03 | 0.01 | 10.33 | 1.33 |
| Mean | −1.08 | 3.99 | 5.77 | 0.27 | 1.82 | 0.65 | −0.68 | 7.60 | 0.58 |
| Std. dev. | 0.82 | 1.34 | 0.33 | 0.43 | 1.51 | 0.65 | 0.44 | 2.12 | 0.48 |

## RESULTS

In this study, there was no "correct" answer for gender categorization, as all stimuli were resynthesized to be between a woman's voice and a man's voice. Therefore, the categorization judgment of NH group was considered to be the "normal" gender categorization, and CI and NH-vocoded performance were evaluated with respect to this normal performance. Figure 2 shows the results for the three modes of hearing in relation to the normal performance in this test, as is defined by the performance of NH listeners. The normal data are the NH results that are ordered from most strongly judged female voice conditions in the left to most strongly judged male voice conditions in the right. The figure clearly shows a more variable and abnormal pattern for the gender categorization in CI users compared to both the NH and the NH-vocoded modes of hearing. The NH-vocoded mode of hearing also differs from the normal categorization, but there was less variation in their judgment than the real CI users.

Figure 3 shows the average and individual results in more detail, for all conditions tested, and separately for the NH (top), the NH-vocoded (middle), and CI (bottom) modes of hearing. The comparison between the top and bottom panels again shows the discrepancy between NH and CI listeners. With non-vocoded speech (top panel), NH responses gradually shift from female to male as the VTL or F0 are increased. With the vocoded speech (middle panel) or with real CI users (bottom panel), VTL had little effect on gender categorization. Compared to VTL, F0 had a stronger

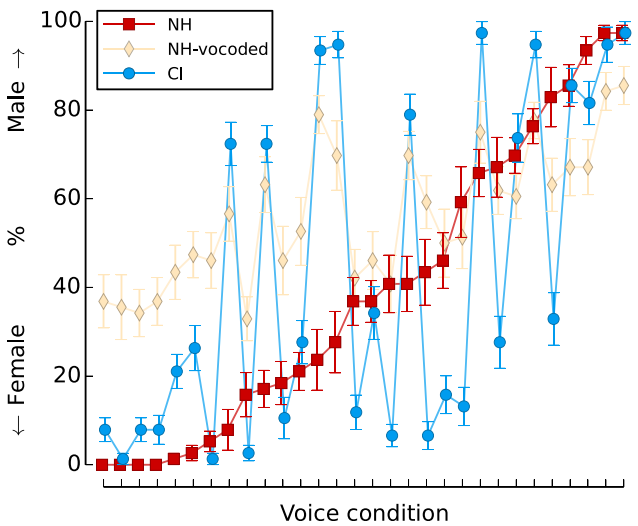


**FIG. 2.** Gender categorization results of NH listeners (*red squares*), NH listeners tested with vocoded stimuli (CIsim, *yellow diamonds*), and CI users (*blue circles*). The *x*-axis represents the 30 voice conditions ordered according to the NH listeners' average gender categorization, from female on the *left* to male on the *right*. The *circles and diamonds* show the data for the actual and simulated CI listeners for the same voice conditions. The *error bars* represent the standard error.

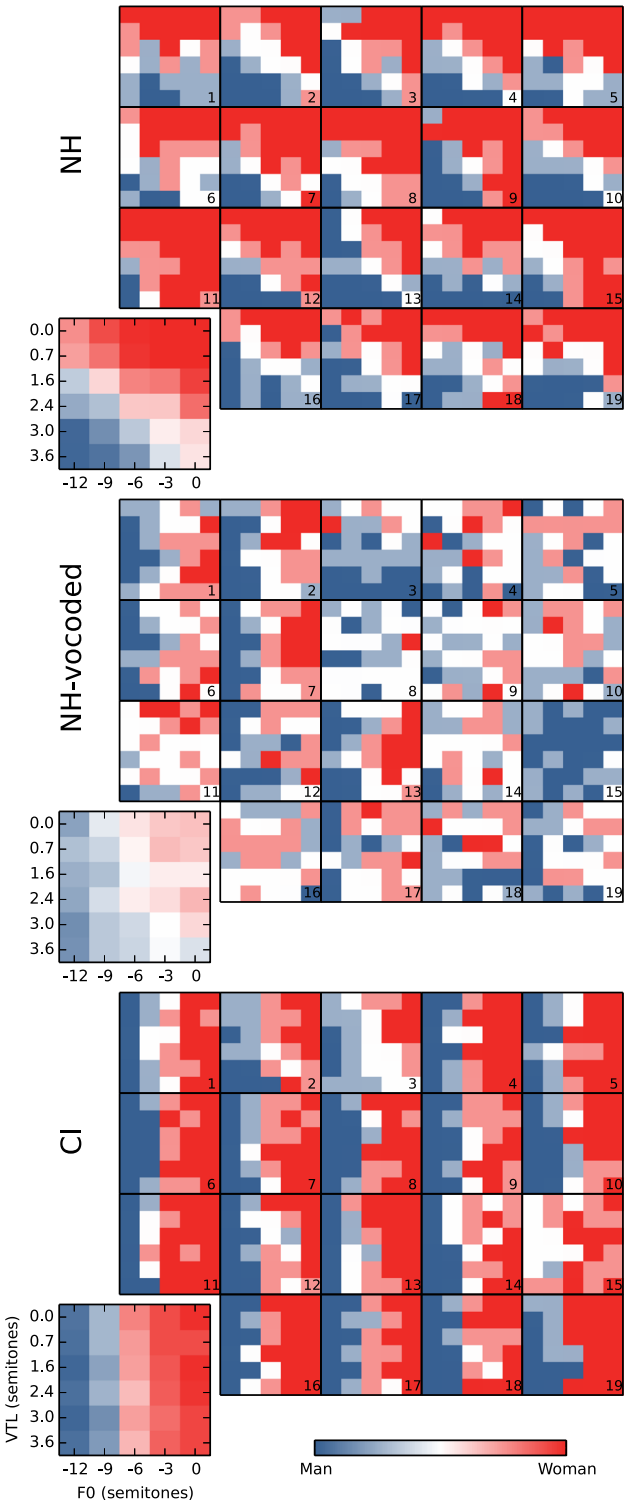


**FIG. 3.** Individual and average gender categorization judgments, presented as maps in the F0–VTL plane. For each mode of hearing, the *smaller panels* numbered 1 to 19 show the individual maps where each pixel corresponds to a combination of F0 and VTL, while *blue* corresponds to 100 % "man" responses and *red* corresponds to 100 % "woman" responses.

effect on performance both for NH-vocoded group (middle panel) and for real CI users (bottom panel).

On average, F0 [$\chi^2_{(6)}$=2,184, $p$<0.0001] and VTL [$\chi^2_{(6)}$=958.4, $p$<0.0001] both had a significant effect on gender categorization and both interacted with the mode of hearing [F0: $\chi^2_{(2)}$=105.3, $p$<0.0001; VTL: $\chi^2_{(2)}$=420.1, $p$<0.0001]. Mode of hearing itself also had a main effect on the results [$\chi^2_{(2)}$=271.2, $p$<0.0001]. These effects are detailed in the following sections, and perceptual weights are reported for each of these cues and modes of hearing.

Individual logistic regression coefficients are reported in Table 2.

## Comparisons of Modes of Hearing

NH listeners (top panel of Fig. 3) gave high weights both to F0 (3.76, $z$=18.1) and VTL (5.56, $z$=22.6), indicating that they used both dimensions to estimate the gender of the voices. For NH subjects to completely perceive the female voice as male, both F0 and VTL needed to be changed; changing F0 alone or VTL alone produced less reliable categorization in most cases. In particular, a change of −12 semitones in F0 with no change of VTL produced a male judgment only in 10 % of the trials, illustrating the importance of VTL for gender categorization. Individual weights for VTL (see Table 2) were also remarkably similar across participants (ranging from 5.23 to 6.21, s.d. 0.33) while those for F0 showed larger variability (1.52 to 6.31, s.d. 1.34).

In contrast, CI listeners (bottom panel of Fig. 3) relied more on F0 (6.88, $z$=25.4) than the NH listeners [$\chi^2_{(1)}$=94.51, $p$<0.0001] and less on VTL (0.59, $z$=3.27) than the NH listeners [$\chi^2_{(1)}$=301.2, $p$<0.0001]. The CI listeners showed a somewhat larger variability across listeners in their sensitivity to both F0 (weights ranging from 2.42 to 10.33, s.d. 2.12) and VTL (weights ranging from −0.32 to 1.33, s.d. 0.48). There was no main effect of mode of hearing between these two groups [$\chi^2_{(1)}$=2.87, $p$=0.0888] indicating that mode of hearing did not bias gender categorization toward one sex or the other.

In the NH-vocoded condition (middle panel of Fig. 3), the weights were reduced both for F0 [weight 1.68, $z$=12.8; vs. NH $\chi^2_{(1)}$=66.70, $p$<0.0001] and VTL [weight 0.63, $z$=4.87; vs. NH $\chi^2_{(1)}$=382.2, $p$<0.0001]. These perceptual weights obtained for F0 were also different from the one obtained for actual CI listeners [$\chi^2_{(1)}$=404.8, $p$<0.0001], but those obtained for VTL were not significantly different [$\chi^2_{(1)}$=0.034, $p$=0.85]. Finally, in the NH-vocoded condition, listeners showed large inter-individual variability: weights for F0 ranged from −0.23 to 4.84 (s.d. 1.51), and weights for VTL ranged from −0.34 to 2.03 (s.d. 0.65).

## Within Group Factors for the CI Listeners

Although the variability across CI listeners was relatively small, a number of factors were tested for significance by adding them to the reference model. We found that the type of *speech processor* of the implant had a significant main effect on gender categorization [$\chi^2_{(2)}$=12.929, $p$=0.0016], but this effect did not interact with either F0 or VTL. The Freedom and CP810 processors from Cochlear Limited (Australia) were not different from each other [$p$=0.84], but the users of the Harmony processor from Advanced Bionics AG (Switzerland) were significantly more likely to answer "female" than the other participants [$p$<0.0001]. This could be a confound with the effect of *rate of stimulation* [$\chi^2_{(1)}$= 6.893, $p$=0.0087], which also did not interact with F0 and VTL: overall, participants with higher stimulation rates (i.e., using the Harmony processor) had a higher tendency to answer "female" than those with lower rates. This effect was not significant anymore when the effect of processor was partialled out.

Another factor that could potentially influence gender categorization is the *type of electrode array* of the implant. Some arrays are designed to place electrodes closer to the modiolus and limit cochlear damage during insertion. In our group of subject, this might be the case for users of "CI24R CS" and "CI24RE H." However, only two of the 19 CI participants had electrode arrays that differed from the others, and inspection of the individual regression coefficients for these participants did not reveal a particular pattern.

Further examining individual results, it appears that four participants had perceptual weights greater than 1.0 for VTL (subject numbers 2, 3, 10, and 12). Looking at the history, device, duration of implantation, age, or gender of these participants, however, we could not find a common trait. Similarly, the four listeners who had the highest perceptual weights for F0 had nothing in common: they used different devices, had different ages, and were of different sex.

Finally, two of the participants used the Fidelity 120 strategy of Advanced Bionics. This strategy involves current steering and thus offers the possibility to deliver peaks of the spectrum at their exact location, which could provide a significant advantage for VTL perception. However, these two listeners showed among the smallest perceptual weights for VTL.

## Measures of Sensitivity

To perform the gender categorization task, the listeners integrate the manipulated cues F0 and VTL (in addition to other non-manipulated cues) into a single judgment. This process yields data that can be

represented in a three-dimensional space with F0, VTL, and gender categorization as the three dimensions (as displayed in Fig. 3). For each participant, the two perceptual weights, resulting from the cue weighting analysis, define a plane in the logit F0–VTL space. The slope of this surface represents the sensitivity in perceiving the gender difference in stimuli. The maximal slope, or the score gradient, represents the absolute sensitivity independent of the cue that is used and can be calculated as $s_{\max} = \sqrt{a^2 + b^2}$ where $a$ and $b$ are the coefficients for *f0* and *vtl* as defined in the logistic regression. Another slope can be calculated along the straight line between the male and the female voice. This diagonal is similar to the line followed by the continuum of voices used in Massida et al. (2013). The slope along this line, calculated as $s_{\text{diag}} = (a + b)/\sqrt{2}$ , thus reflects the sensitivity in a way that is comparable to that of Massida et al. (2013). Note that none of these slopes give any indication about the normal behavior by themselves, and they only bear information about how sensitive participants are to any of the cues used in a specific task.

The values for $s_{\max}$ and $s_{\text{diag}}$ were calculated for each participant and compared across groups. We found that maximal slopes $s_{\max}$ were similar for NH (7.12, s.d. 0.56) and CI (7.65, s.d. 2.09) listeners [$t_{(20.6)}$=1.06, $p$=0.29]. However, when comparing slopes along the diagonal, CI users (5.78, s.d. 1.39) did show lower slopes than NH listeners [6.90, s.d. 0.77; $t_{(28.07)}$=−3.07, $p$=0.0048].

## DISCUSSION

In this study, gender categorization by CI users was shown to be abnormal relative to NH performance with unprocessed speech. By systematically varying F0 and VTL cues with synthesized stimuli, we found that CI users' gender categorization mainly depends on F0 cues, with nearly no contribution of VTL cues. This is an important finding, as F0 alone or VTL alone is not sufficient for the normal categorization of gender.

### Normal Gender Categorization

In this study, "normal" gender categorization was defined as NH performance with non-vocoded speech. These results are in accordance with data previously reported in literature that also showed NH subjects to rely equally strongly on both F0 and VTL cues for gender categorization (Skuk and Schweinberger 2013; Smith and Patterson 2005; Smith et al. 2007). Only when both VTL and F0 were changed was the source female voice completely perceived as male. When the source female VTL was retained, even the largest F0 change (−12 semitones) only resulted in a "male" judgment in less than 10 % of the trials. Reciprocally, when the source female F0 was retained and only VTL was changed (by 3.6 semitones), the voice was judged as "male" only in about 30 % of the trials. These results are comparable to those obtained in previous gender categorization studies (Smith and Patterson 2005; Smith et al. 2007) and emphasize the importance of both vocal characteristics.

### Gender Categorization by CI Listeners

CI gender categorization was abnormal relative to NH performance with unprocessed speech. Different from NH performance, CI users' weighted F0 cues very strongly and VTL cues almost not at all in the categorization. These results therefore bring strong evidence to what was indirectly suggested in previous studies, namely, that CI users primarily rely on F0 cues for gender categorization (Fu et al. 2004, 2005; Kovačić and Balaban 2009, 2010). However, further, the present results also showed that overreliance on F0 cues may cause CI users to make abnormal judgments of a talker's gender.

Unlike for the NH listeners, the voice presented in the experiment never seemed to be ambiguous to the CI participants. For NH listeners, 7 of the 30 voices produced average male judgments between 35 and 65 %. For the CI listeners, none of the voices produced a judgment in that range. This is in apparent contrast with the results of Massida et al. (2013) who reported that the gender categorization deficit in CI compared to NH listeners was "stronger for ambiguous stimuli" in the continuum between a male and a female voice. This conclusion was supported by the fact that the psychometric functions for their CI participants were 58 % shallower than for their NH participants. In our study, instead of using a unidimensional continuum, we measured gender categorization on a bidimensional space. Sensitivity in such a space is captured by the maximal slope of the two-dimensional psychometric function, i.e., the norm of the gradient of the plane fitted to the logit scores as described in the last part of the "RESULTS" section. With this sensitivity measure, we found that CI listeners showed at least as high sensitivity as NH listeners on average. In other words, the psychometric functions were equally steep for CI and NH listeners, but their orientation in the F0–VTL plane was different. However, when measuring sensitivity along a unidimensional continuum between our female and male voices similar to the one used by Massida et al. (2013), we found results consistent with their findings: that sensitivity along that continuum was smaller for CI

listeners than for NH listeners. Our results now bring further explanation that this weaker sensitivity to voice gender is due to a deficit in VTL perception.

It is perhaps surprising that CI listeners showed such a strong reliance on F0 cues when pitch perception has been repeatedly reported as defective, or at best, weak, with an implant (see Moore and Carlyon 2005 for a review). However, it is worth noting that the F0 difference separating our male and female voices—one octave—is extremely large compared to F0 difference limens in NH listeners (e.g., Rogers et al. 2006, report F0 difference limens in words of about half a semitone) or even in CI listeners (3.4 semitones, reported in that same study). In other words, while F0 perception is indeed degraded in CI listeners, it remains sufficiently robust to discriminate the pitch of a male voice from that of a female voice.

VTL, on the other hand, could be expected to be more clearly perceived in CIs, as changes along this dimension do not affect the spectral fine structure but the spectral envelope, which is better preserved in the implant. The right-most column of Figure 1 shows electrical stimulation patterns for the voice with the unmodified VTL and the elongated VTL of the male voice. Frequency channels in CIs are typically separated by 2.5 to 3.0 semitones. The VTL separation between the male and female voice, 3.6 semitones, thus results in a shift of the stimulation pattern along the electrode array of about one electrode (Fig. 1, right-most column). Using stimulation patterns comprising one to eight adjacent electrodes (the latter is relatively similar to the stimulation pattern of the vowels in our experiment), Laneau and Wouters (2004) found that CI listeners have just-noticeable differences for place shifts of about 0.5 electrodes. Yet, the CI users in our experiment did not use the VTL cue for gender categorization. Another measure of spectral resolution uses broadband spectral ripple discrimination, where listeners have to discriminate between a spectral ripple pattern and its inverse-phase counterpart. With this method, Anderson et al. (2011) showed that, on average, CI listeners could discriminate phase-inverted spectral ripples up to 1.68 ripple/octave. The detection of the 3.6-semitone shift in our experiment would require discrimination of 1.67 ripple/octave, so average CI listeners could perhaps just detect this VTL shift. However, on a larger population of CI users, Won et al. (2007) observed that only about 35 % of their participants had discrimination thresholds above 1.44 ripple/octave. Therefore, it remains unclear whether the VTL shift could be detected at all by the CI listeners.

From these considerations, two hypotheses can thus be formulated. The first one is that although the difference of VTL is visible on the electrodogram, the wide spread of excitation of electrical stimulation prevents this cue from being available in the neural activity pattern. In other words, the effective spectral resolution of electrical stimulation is not sufficient for this cue to be perceived. A direct way to test this hypothesis would be to measure VTL difference limens in CI listeners. The second hypothesis is that this cue remains available to some extent in the neural representation but is either too weak or too distorted to be reliably used for gender categorization. The place–frequency mismatch that results from the fact that electrode arrays cannot be inserted all the way to the apex, for instance, could distort (without removing) the representation of this cue, as previously suggested by Kovačić and Balaban (2009, 2010). In such a context, CI listeners would overly rely on the more robust cue that is available, i.e., pitch. If this hypothesis was verified, i.e., the VTL cue was only distorted but not entirely destroyed, specific training could improve its usability.

## Gender Categorization with Vocoded Stimuli

Compared to NH performance with non-vocoded speech, the NH-vocoded performance was much poorer, hewing close to 50 % "man"/"woman" responses at all F0–VTL combinations. Such a pattern can be interpreted as increased uncertainty in the responses or lack of agreement across participants. Examination of the logistic regression coefficients showed that F0 and VTL were used less efficiently than in the non-vocoded condition. This is expected since the sinewave vocoder weakened both F0 and VTL cues, compared to unprocessed speech.

However, performance in the NH-vocoded condition was markedly different from real CI users' performance, suggesting that sinewave vocoding might be too simple a simulation for gender categorization tasks. A notable difference between actual and simulated CI hearing is that, for conditions where the F0 was below 160 Hz, the sinewave vocoder provided not only temporal but also spectral F0 cues to the NH listeners, which are not available to actual CI users. Nevertheless, NH participants did not seem to make a strong use of these F0 cues as the results below and above F0=160 Hz are not markedly different. More importantly, even when F0 cues were present (below 160 Hz), these cues were weaker than in the non-vocoded condition. Because the same NH subjects did the task first with non-vocoded stimuli and then with the vocoded set, they were aware that the voice cues were weaker in the vocoded case relative to the non-vocoded condition, and this could have, in turn, resulted in them relying less on these cues.

Regarding VTL, as the carrier center frequencies of the vocoder were separated by 7.5 semitones on

average (or 2.7 mm in cochlear distance, according to Greenwood 1990), VTL differences as small as 3.6 semitones were not expected to be detectable in the vocoded stimuli. Yet, the cue weight for VTL was larger in the NH-vocoded condition than for CI users. This suggests that CI users' functional spectral resolution was probably poorer than that achieved by the eight independent frequency channels of the vocoder. The specific role of channel interaction in CIs could be investigated in NH listeners using a more elaborate vocoder (e.g., Churchill et al. 2014).

## CONCLUSION

The main finding of our study is that CI users have an abnormal gender categorization compared to NH listeners. CI users strongly and almost exclusively use the F0 cue, while NH listeners use both vocal characteristics, F0 and VTL, for gender categorization. This can have practical consequences on everyday situations for CI users as, for a given voice, they may judge gender differently than what it should be. Further, this could also mean that CI users may not be able to use VTL differences to segregate competing talkers, thus contributing to difficulties understanding speech in multi-talker environments. Consequently, although the CI users achieve some gender categorization, as was also shown previously, the present study emphasizes that their ability to do so is not complete and must be considered impaired.

At this point, it remains unclear whether the observed deficiency in VTL perception is because VTL differences are not transmitted by the CI to the auditory nerve (e.g., because of spread of exCitation and channel interaction) or, alternatively, whether they are actually transmitted and detected but not reliable enough for accurate gender categorization. Further research is therefore needed to explore whether VTL differences can be detected at all or whether they are simply not interpreted as talker-size differences. Based on such knowledge, appropriate coding schemes or better fitting algorithms for CIs can be developed and abnormal judgment of gender identification can perhaps be corrected.

Another point that will require further investigation is the extent to which other cues may contribute to gender categorization. Although F0 and VTL seem to be the most important factors for gender categorization in NH listeners (Skuk and Schweinberger 2013), other cues such as breathiness (Holmberg et al. 1988; Van Borsel et al. 2009) or intonation (Fitzsimons et al. 2001) could play a more important role in CI listeners.

Finally, the protocol used in the present study was a quick test (10 min only) that characterized how CI users' gender categorization deviates from normal and what specific vocal cues are underutilized. Using such a quick test, new coding strategies or fitting algorithms can be improved to achieve a normal gender categorization, which will likely indicate that vocal characteristics are fully utilized. Because gender categorization and specifically F0 and VTL differences have been shown to facilitate concurrent speech perception, improving their representation in the implant could, in turn, lead to improved speech-in-noise perception by CI users.

## ACKNOWLEDGMENTS

We would like to thank the participants in this study. Furthermore, we would like to thank Joeri Smit and Karin van der Velde for their help with collecting the data, as well as Anita Wagner for her advice regarding statistical methods. The fourth author is supported by a NIH R01-DC004792 grant. The sixth author is supported by an otological/neurotological stipendium from the Heinsius-Houbolt Foundation. The last author is supported by a Rosalind Franklin Fellowship from the University Medical Center Groningen, University of Groningen and the VIDI grant 016.096.397 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw). The study is part of the research program of our department: Healthy Aging and Communication.

### *Conflict of Interest*

There is no conflict of interest regarding this manuscript.

## REFERENCES

Anderson ES, Nelson DA, Kreft H, Nelson PB, Oxenham AJ (2011) Comparing spatial tuning curves, spectral ripple resolution, and speech perception in cochlear implant users. J Acoust Soc Am 130:364–375. doi:10.1121/1.3589255

Bates D, Maechler M, Bolker B, Walker S (2013) lme4: Linear mixed-effects models using Eigen and S4. http://cran.r-project.org/package=lme4. Version 1.1-6

Bosman AJ, Smoorenburg GF (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 DS (2001) Informational and energetic masking effects in the perception of two simultaneous talkers. J Acoust Soc Am 109:1101–1109

Churchill TH, Kan A, Goupell MJ, Ihlefeld A, Litovsky RY (2014) Speech perception in noise with a harmonic complex excited vocoder. J Assoc Res Otolaryngol 15:265–278. doi:10.1007/s10162-013-0435-7

Clarke J., Gaudrain E., Chatterjee M., Başkent D (2014) T'ain't the way you say it, it's what you say—perceptual continuity and top-down restoration of speech. Hear Res, 315:80–387.

Darwin CJ, Brungart DS, Simpson BD (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

Fant G (1970) Acoustic theory of speech production. Walter de Gruyter, The Hague

Festen JM, Plomp R (1990) Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. J Acoust Soc Am 88:1725–1736

Fitch WT, Giedd J (1999) Morphology and development of the human vocal tract: a study using magnetic resonance imaging. J Acoust Soc Am 106:1511–1522

Fitzsimons M, Sheahan N, Staunton H (2001) Gender and the integration of acoustic dimensions of prosody: implications for clinical studies. Brain Lang 78:94–108

Friesen LM, Shannon RV, Başkent D, Wang X (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 QJ, Chinchilla S, Galvin JJ III (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 QJ, Chinchilla S, Nogaki G, Galvin JJ III (2005) Voice gender identification by cochlear implant users: the role of spectral and temporal resolution. J Acoust Soc Am 118:1711–1718

Fuller CD, Free RH, Maat B, Başkent D (2012) Musical background not associated with self-perceived hearing performance or speech perception in postlingual cochlear-implant users. J Acoust Soc Am 132:1009–1016. doi:10.1121/1.4730910

Fuller CD, Galvin JJ III, Free RH, Başkent D (2014) Musician effect in cochlear implant simulated gender categorization. J Acoust Soc Am 135:EL159–EL165

Greenwood DD (1990) A cochlear frequency-position function for several species—29 years later. J Acoust Soc Am 87:2592

Holmberg EB, Hillman RE, Perkell JS (1988) Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. J Acoust Soc Am 84:511

Jaeger TF (2008) Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. J Mem Lang 59:434–446. doi:10.1016/j.jml.2007.11.007

Kawahara H, Masuda-Katsuse I, de Cheveigné A (1999) Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds. Speech Commun 27:187–207

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

Kovačić D, Balaban E (2010) Hearing history influences voice gender perceptual performance in cochlear implant users. Ear Hear 31:806–814

Lammert A, Proctor M, Narayanan S (2013) Morphological variation in the adult hard palate and posterior pharyngeal wall. J Speech Lang Hear Res 56:521–530

Laneau J, Wouters J (2004) Multichannel place pitch sensitivity in cochlear implant recipients. J Assoc Res Otolaryngol 5:285–294. doi:10.1007/s10162-004-4049-y

Luo X, Fu QJ, Wu HP, Hsu CJ (2009) Concurrent-vowel and tone recognition by Mandarin-speaking cochlear implant users. Hear Res 256:75–84

Mackersie CL, Dewey J, Guthrie LA (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. doi:10.1121/1.3605548

Massida Z, Marx M, Belin P et al (2013) Gender categorization in cochlear implant users. J Speech Lang Hear Res 56:1389–1401. doi:10.1044/1092-4388(2013/12-0132)

Moore BC, Carlyon RP (2005) Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In: Pitch. Springer, pp 234-277

Patterson RD, Gaudrain E, Walters TC (2010) The perception of family and register in musical notes. In: Jones MR, Fay RR, Popper AN (eds) Music perception, 1st Edition. Springer, pp 13–50

Peng S-C, Lu N, Chatterjee M (2009) Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. Audiol Neurotol 14:327–337. doi:10.1159/000212112

Peterson GE, Barney HL (1952) Control methods used in a study of the vowels. J Acoust Soc Am 24:175–184

Rogers CF, Healy EW, Montgomery AA (2006) Sensitivity to isolated and concurrent intensity and fundamental frequency increments by cochlear implant users under natural listening conditions. J Acoust Soc Am 119:2276–2287

Skuk VG, Schweinberger SR (2013) Influences of fundamental frequency, formant frequencies, aperiodicity and spectrum level on the perception of voice gender. J Speech Lang Hear Res. doi: 10.1044/1092-4388(2013/12-0314)

Smith DR, Patterson RD (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 DR, Walters TC, Patterson RD (2007) Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled. J Acoust Soc Am 122:3628–3639

Stickney GS, Zeng F, Litovsky R, Assmann P (2004) Cochlear implant speech recognition with speech maskers. J Acoust Soc Am 116:1081

Van Borsel J, Janssens J, De Bodt M (2009) Breathiness as a feminine voice characteristic: a perceptual approach. J Voice 23:291–294

Wilkinson EP, Abdel-Hamid O, Galvin JJ III, Jiang H, Fu QJ (2013) Voice conversion in cochlear implantation. Laryngoscope 123(Suppl 3):S29–S43

Wilson BS, Finley CC, Lawson DT et al (1991) Better speech recognition with cochlear implants. Nature 352:236–238. doi:10.1038/352236a0

Won JH, Drennan WR, Rubinstein JT (2007) Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. J Assoc Res Otolaryngol 8:384–392. doi:10.1007/s10162-007-0085-8