Factors limiting vocal-tract length discrimination in cochlear implant simulations

Etienne Gaudrain, Deniz Başkent

Factors limiting vocal-tract length discrimination in cochlear implant simulations

Etienne Gaudrain[a] and Deniz Başkent[b]
Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

(Received 17 May 2014; revised 19 January 2015; accepted 28 January 2015)

Perception of voice characteristics allows normal hearing listeners to identify the gender of a speaker, and to better segregate speakers from each other in cocktail party situations. This benefit is largely driven by the perception of two vocal characteristics of the speaker: The fundamental frequency (F0) and the vocal-tract length (VTL). Previous studies have suggested that cochlear implant (CI) users have difficulties in perceiving these cues. The aim of the present study was to investigate possible causes for limited sensitivity to VTL differences in CI users. Different acoustic simulations of CI stimulation were implemented to characterize the role of spectral resolution on VTL, both in terms of number of channels and amount of channel interaction. The results indicate that with 12 channels, channel interaction caused by current spread is likely to prevent CI users from perceiving VTL differences typically found between male and female speakers.
 [http://dx.doi.org/10.1121/1.4908235]

[MAS] Pages: 1298–1308

I. INTRODUCTION

When multiple talkers speak at the same time, normal hearing (NH) listeners can use the characteristics of the voices they hear to segregate and track a particular target. This ability is clearly evidenced with target voices of different gender than the competing ones (Brungart, 2001; Festen and Plomp, 1990). These studies showed that when two competing sentences are presented, within-gender voice differences can provide a substantial increase in intelligibility compared to the case with no voice difference (20 percentage points in Brungart, 2001). However, when the voices also differ in gender, the benefit in intelligibility can be as high as 50 percentage points.

Although many studies focus on the role of differences in fundamental frequency (*F*0) for the separation of concurrent voices (Bird and Darwin, 1998; Brokx and Nooteboom, 1982; Summers and Leek, 1998), perceived gender differences in voices actually depends equally on *F*0 and spectral shape parameters, such as formant frequencies (Skuk and Schweinberger, 2013). These two characteristics can be related to major differences in the anatomy of the male and female speech production systems and are quantified as the glottal-pulse rate (GPR) and the vocal-tract length (VTL), respectively. VTL is directly related to the size of the speaker (Fitch and Giedd, 1999; Roers *et al.*, 2009) and is perceived as such by listeners (Smith and Patterson, 2005). More generally, VTL constrains the scale of the acoustic resonators responsible for the formant peaks in the spectral envelope (for samples and illustrations, see Patterson *et al.*, 2010). For a given set of formant frequencies (i.e., for a given vowel), a longer VTL results in shifting the spectral envelope, i.e., all the formants, toward the low frequencies, on a logarithmic frequency axis. Conversely, a shorter VTL results in a shift of the spectral envelope toward the high frequencies. Shifting the spectral envelope does alter the amplitude of the harmonics but does not alter their frequency. Therefore VTL manipulations do not affect the *F*0. On the other hand, GPR determines the fundamental frequency (*F*0) of the voice, and is thus directly related to the perceived pitch of the voice, which is represented in the auditory system through a combination of temporal and place codes (Carlyon and Shackleton, 1994). Both VTL and *F*0 have been shown to contribute to the perceptual separation of concurrent syllables (Vestergaard *et al.*, 2009, 2011) and sentences (Darwin *et al.*, 2003), and seem to explain most of the advantage induced by voice gender differences in competing speech.

The type of multi-talker situation described above is extremely difficult for cochlear-implant (CI) listeners. Unlike NH listeners, CI users do not benefit from gender differences among competing talkers (Luo *et al.*, 2009; Stickney *et al.*, 2004). This could be tied to the fact that CI listeners generally demonstrate poorer voice-gender categorization performance than NH listeners (Fuller *et al.*, 2014; Fu *et al.*, 2004, 2005; Kovačić and Balaban, 2009, 2010). In particular, Fuller *et al.* (2014) showed that the abnormal gender categorization performance was entirely due to poor VTL perception in CI users. In that study, although the 19 CI listeners reported hearing male and female voices, they all based their gender judgment solely on the basis of the *F*0, and, unlike the NH listeners, were unable to use the VTL cue, leading to many erroneous categorizations.

[a]Author to whom correspondence should be addressed. Also at: Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, University of Groningen, Groningen, Netherlands. Electronic mail: etienne.gaudrain@cnrs.fr
[b]Also at: Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics, CNRS UMR 5292, Inserm U1028, Université Lyon 1, Lyon, France.

It remains unknown why CI listeners did not make use of the VTL cue, especially because very little is known about VTL perception in CI users. A first hypothesis is that VTL cannot be perceived through the implant. This could be the case if spectral resolution in the implant is so poor that the spectral changes produced by VTL differences cannot be detected. An alternative hypothesis would be that VTL differences are detected but not used by the CI users for gender categorization. Such phenomenon, where a voice cue is detected but not exploited for the task at hand, has been previously reported in cases where the cue was deemed unreliable by the listeners for that particular task (e.g., Gaudrain *et al*., 2009). The stimulation through the implant could be such that the VTL cue remains salient enough for detection, but is unreliable for gender categorization.

In the present study, we examined VTL perception with a number of acoustic simulations of CIs—vocoders (Dudley, 1939; Shannon *et al*., 1995)—to bring clarity to the potential explanations above. Specifically, we investigated whether reducing spectral resolution causes an increase in just-noticeable-difference (JND) for VTL but not for *F*0, and thereby evaluate whether the mode of stimulation of the implant could explain the pattern of results observed by Fuller *et al*. (2014). The JNDs for VTL and *F*0 were obtained from NH participants in two experiments varying the number of frequency bands (Exp. 1) and the type of carrier (Exp. 2) in vocoders. The number of bands allows coarse, albeit direct, manipulations of spectral resolution. In a third experiment, where only VTL JNDs were measured, both the number of channels and the sharpness of the bandpass filters were manipulated to simulate the two aspects of electrical stimulation that limit spectral resolution: Number of electrodes and current spread in the cochlea, respectively.

II. EXPERIMENT 1: EFFECT OF NUMBER OF BANDS WITH SINEWAVE VOCODING

A. Rationale

In addition to NH and CI listeners, Fuller *et al*. (2014) also tested gender categorization as a function of *F*0 and VTL difference in the same NH participants using an 8-band vocoder with sinewave carriers. The results obtained with this CI simulation showed some sensitivity to *F*0 (likely due to the perception of side-tones, as was discussed by Fuller *et al*., 2014), but little sensitivity to VTL, similar to the results obtained by the actual CI participants. However, the NH participants with CI simulation also showed very little confidence in their judgments: The categorizations were all around 50% and the psychometric functions were much shallower than those obtained from the CI participants. To avoid these issues, in this first experiment we used similar vocoders but a more objective task to evaluate how the number of channels affects VTL and *F*0 JNDs. The JNDs were obtained with adaptive tracking with a three-interval three-alternative forced choice (3I-3AFC) task.

B. Material and methods

1. Participants

Sixteen participants were recruited to take part in the experiment. One participant was excluded because they could not perform the task. Another participant was excluded because their auditory thresholds were between 15 and 40 dB hearing level (HL). All 14 remaining participants, aged 19 to 63 [mean 37.4, standard deviation (s.d.) 17.6], had auditory thresholds ≤ 20 dB HL at octave frequencies between 500 and 4000 Hz. All the participants were either native Dutch speakers, or had Dutch as one of the languages used in their daily childhood environment. The participants provided signed informed consent prior to data collection. The experiment was approved by the ethics committee of the University Medical Center Groningen (METc 2012.392). Finally, the subjects received an hourly wage for their participation.

2. Procedure

Discrimination thresholds were obtained using a 3I-3AFC adaptive procedure. Each threshold measurement is called a *run* in the following description, each run being composed of a number of trials. In each trial, the subjects were presented with three triplets of syllables (see the description of the stimuli in Sec. II B 3). These three triplets were composed of the same syllables in the same order. The two standard triplets were produced with the original (recorded) voice parameters, while the odd triplet (randomly assigned to one of the three presentation intervals) was produced with VTL and *F*0 that differed from the original voice by some amount. In each run, the way the VTL and *F*0 differences were calculated was by following a spoke radiating from the reference female voice, in the *F*0-VTL plane (see the dotted-dashed lines in Fig. 1). The reference voice was set at an *F*0 of 242 Hz (the average *F*0 of the original recordings), and all VTL differences were expressed relative to the actual VTL of the original speaker. The angle of each spoke in the plane thus determined a fixed ratio between *F*0 and VTL difference expressed in semitones (i.e., a 12th of an octave, denoted "st" in the rest of the article). A VTL increase expressed in semitones results in a decrease of the same number of semitones of all formant frequencies. Six different spokes were used: Two horizontal spokes where *F*0 was increased or decreased, while VTL was kept the same as the original voice; two vertical spokes where VTL was increased or decreased, while *F*0 was held constant; a spoke pointing toward a child's voice, i.e., with decreasing VTL and increasing *F*0; a spoke pointing toward a man's voice, i.e., with increasing VTL and decreasing *F*0. The man's voice was defined as having a VTL difference of 3.8 st and an *F*0 difference of −12 st relative to the original female voice. The child's voice was defined as −7 st in VTL and 5 st in *F*0 away from the original voice.

Each run started with a difference of 12 st calculated along the spoke (i.e., using the Euclidian distance in the *F*0-VTL plane). The voice difference was then modified by a given step size according to the subject's response. The

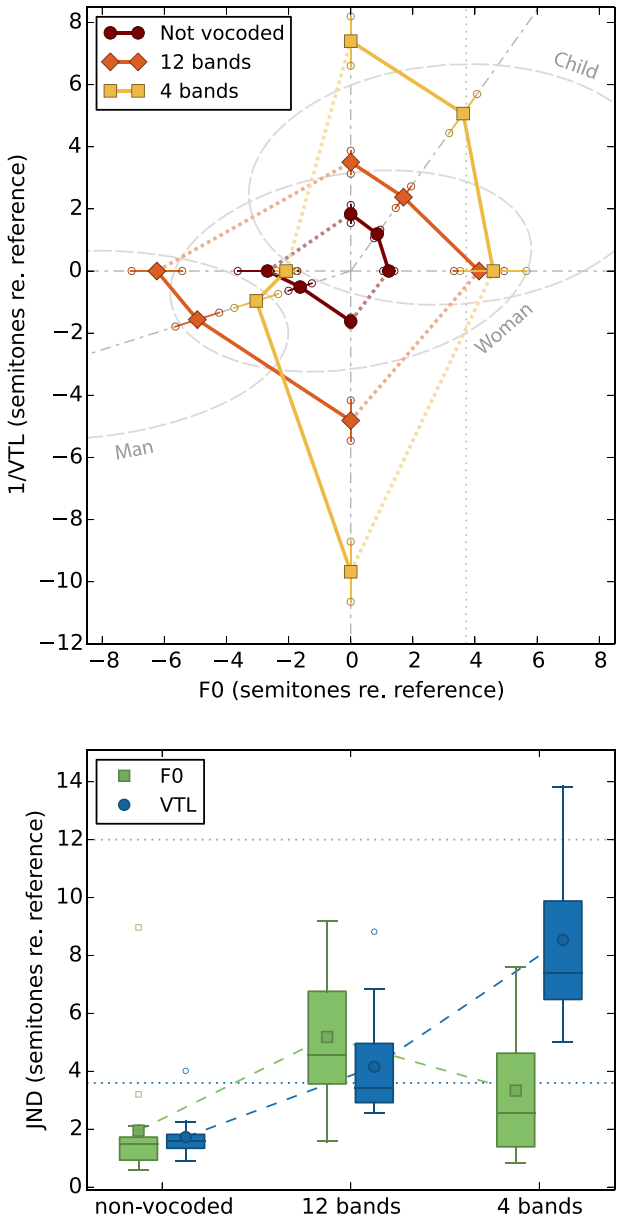

FIG. 1. (Color online) JNDs in $F0$ and VTL as a function of vocoding condition. Top panel: The thresholds expressed in the $F0$-VTL plane. The center of the plane corresponds to the reference female voice. The symbols represent the JND when following a spoke radiating from the reference voice in the direction of the symbol, as shown by the dotted-dashed lines. The correspondence of the symbols is shown in the caption. The (radial) error bars represent the standard error. The vertical dotted line shows the cutoff frequency of the vocoder (300 Hz). The light gray ellipses show estimates of the $F0$/VTL distributions capturing 99% of each group of speakers from Peterson and Barney (1952). Lower panel: The same JNDs plotted separately for $F0$ and VTL as a function of the vocoding condition. The data from the two diagonal spokes were not included, and for each dimension, the data from the negative and positive pointing spokes were averaged. The boxes extend from the lower to the upper quartile, and the middle line shows the median. The filled symbols (square and circle) show the mean (for $F0$ and VTL, respectively). The whiskers show the range of the data within 1.5 times the inner quartile. The empty symbols show the individual data outside of 1.5 times the inner quartile range. The upper and lower dotted lines represent the difference in $F0$ and VTL, respectively, that were used between the male and female voices in Fuller *et al.* (2014).

difference was increased by the step size after each wrong answer, but only reduced (by the step size) after two correct responses (2-down, 1-up) thus converging toward 70.7% of the psychometric function (as would be measured in the same task with the constant stimuli method; Levitt, 1971). The initial step size was 2 st, but was also modified during the run: After every block of 15 trials with the same step size, or when the difference became smaller than 2 times the step size, the step size was divided by $\sqrt{2}$. The procedure ended after eight turn-points (i.e., when the voice difference increased and then decreased again, or vice versa) and the threshold was then calculated as the mean of the last six turn-points.

In each trial the subject was presented with three buttons, numbered 1 to 3, on a computer screen. The buttons lit up successively as the three intervals were presented over headphones. The participants then had to click on the button corresponding to the sound that was different from the other two. Visual feedback was then presented in the form of the correct button blinking green or orange depending on whether the provided answer was correct or not.

Using this method, we obtained voice discrimination thresholds in the $F0$-VTL plane as a function of direction (i.e., spoke) and for different processing conditions (non-vocoded, and vocoded with various numbers of frequency bands). We measured one threshold for each of the following conditions: Non-vocoded, and vocoded with 12 and 4 bands. The 18 thresholds were collected in a single session of 2 h.

3. Stimuli and apparatus

This section describes how the triplets of syllables presented in each trial were produced. The syllables were consonant-vowel (CV) tokens spliced from meaningful Dutch consonant-vowel-consonant (CVC) words taken from the Nederlandse Vereniging voor Audiologie (NVA) corpus (Bosman and Smoorenburg, 1995). This was the same corpus from which the words used by Fuller *et al.* (2014) were extracted. The words, uttered by a female speaker, were selected and spliced to produce 61 CV syllables of duration 142 to 200 ms.

The syllables were first equalized in root-mean-square (rms) and then analyzed with STRAIGHT (Kawahara and Irino, 2004) to obtain the $F0$ contour, the aperiodicity map, and the spectral envelope, which were stored for later use. In the adaptive procedure described above, the $F0$ and VTL of the syllables were varied in each trial. Three randomly selected syllables were resynthesized with STRAIGHT using the new $F0$ and VTL parameters, and normalized to a duration of 200 ms. Note that the syllables were resynthesized even when the $F0$ and VTL were unchanged compared to the original voice.

The three random syllables, separated by 50 ms of silence, were concatenated to form a triplet. To make the triplets sound more natural, an $F0$ contour was applied by altering the $F0$ of each syllable relative to the mean $F0$ of the triplet by random steps of a third of a semitone. The average $F0$ and the VTL were then modified to form the standard and test triplets as described in Sec. II B 2. The three triplets had the same syllables in the same order, but all differed in $F0$ contour, and the test and standard triplets differed in $F0$ and/or VTL. Note that, while previous studies

used different syllables across the intervals to force the participants to make their judgment on VTL rather than on particular spectral differences (Ives *et al.*, 2005; Smith and Patterson, 2005), the same syllables were used in the different intervals in the present experiment because the purpose was to assess whether the spectral information associated with VTL was accessible at all.

Vocoding, when used, was performed by filtering the original signal in n bands between 150 and 7000 Hz. The band boundaries were equally spaced using Greenwood's function, i.e., estimating location on the basilar membrane of a 35-mm long cochlea (Greenwood, 1990). The bandpass filters were implemented as 12th order, zero-phase Butterworth filters. In each frequency band, the temporal envelope was extracted by half-wave rectification and low-pass filtering below 300 Hz (zero-phase fourth order Butterworth filter). This cutoff frequency was used to mimic the average upper limit of temporal pitch perception in CIs (e.g., Zeng, 2002). The envelope was then used to modulate the amplitude of a sinewave centered (in terms of estimated place along the cochlea) on the frequency band. All the sinewaves were then added and the rms level of the composite sound was adjusted to that of the unprocessed stimulus filtered between 150 and 7000 Hz. The number of bands n was 4 or 12 in Experiment 1. Four bands is often described as yielding identification performance similar to that of the relatively less proficient CI users, while 12 bands represents effective spectral resolution better than what can be achieved by the best CI users (e.g., Friesen *et al.*, 2001) but is the smallest number of bands allowing optimal identification of vowels (Xu *et al.*, 2005, with English vowels).

The three triplets were separated by a silence of 200 ms and presented diotically in HD600 headphones (Sennheiser GmbH & Co., Wedemark, Germany), via an AudioFire4 soundcard (Echo Digital Audio Corp, Santa Barbara, CA) connected to a DA10 D/A converter (Lavry Engineering, Poulsbo, WA) through S/PDIF. All the signal processing and stimulus presentations were performed in MATLAB using a sampling frequency of 44.1 kHz. The participants were seated in a sound-attenuated booth. The level was initially adjusted to be most comfortable for the first participant, measured to be 60 dB sound pressure level (SPL), and was fixed for the subsequent participants.

C. Results

The top panel of Fig. 1 shows the JNDs in the $F0$-VTL plane as a radial contour around the original voice. The values are reported in Table I. As expected, the smallest JNDs were obtained in the non-vocoded condition (circles), with an average JND of 1.8 st. When the stimuli were vocoded with 12 bands (diamonds), the JNDs were on average 2.5 times larger. When the 4-band vocoder was used (squares), the JNDs were on average 3.1 times larger than that for non-vocoded. However, it seems that the increase in JND with degraded spectral resolution affects VTL more strongly than for $F0$.

To clarify this point, the lower panel of Fig. 1 shows the average JNDs for $F0$ (square) and VTL (circle) as a function of the vocoding condition. These values were obtained from the horizontal ($F0$) and vertical (VTL) spokes only, by averaging the negative and positive pointing spokes in order to obtain a single JND for $F0$ and VTL each, per vocoding condition. A repeated-measures analysis of variance (ANOVA) on these averaged JNDs with Vocoder (non-vocoded, 12- and 4-bands) and Dimension ($F0$ and VTL) as repeated factors was performed. The reported p-values were corrected with the Greenhouse-Geisser correction when the sphericity assumption was violated, and effect size is also reported in the form of generalized eta-squared η_G^2 (Bakeman, 2005). The analysis confirmed that the vocoding condition altered the discrimination thresholds [Vocoder: $F(2,26) = 34.93$, $p < 0.0001$, $\eta_G^2 = 0.40$]: The JNDs became larger when the spectral resolution decreased. The $F0$ JNDs were also smaller, on average, than the VTL JNDs [Dimension: $F(1,13) = 12.00$, $p = 0.004$, $\eta_G^2 = 0.09$]. More importantly, the difference between the $F0$ and VTL JNDs depended on the vocoding condition [Vocoder × Dimension: $F(2,26) = 21.24$, $p < 0.0001$, $\eta_G^2 = 0.30$]. *Post hoc* comparisons (using the False Discovery Rate, FDR, correction method) confirmed that while $F0$ and VTL JNDs did not significantly differ in the non-vocoded and 12-bands conditions [$t(13) = -0.37$, $p_{FDR} = 0.71$; $t(13) = -1.55$, $p_{FDR} = 0.22$, respectively], with 4 bands, the JND for VTL was significantly larger than that for $F0$ [$t(13) = 6.13$, $p_{FDR} = 0.0001$].

The ages of the participants covered a wide span and the results showed a substantial inter-subject variability as depicted in the lower panel of Fig. 1. However, when adding age as a covariate to the above analysis, none of the effects including age were significant [all $p > 0.19$, $\eta_G^2 < 0.04$].

TABLE I. Average JNDs, in semitones, for each direction and vocoding condition. The value between brackets is the standard deviation. For the diagonal directions "Child" and "Male," the average $F0$ and VTL components of the JND are given.

Voice	Non-vocoded	12 bands	4 bands
Child $F0$	1.23 (0.68)	4.13 (3.03)	4.59 (3.99)
Male $F0$	2.68 (3.58)	6.24 (3.04)	2.08 (1.18)
Child VTL	1.84 (1.09)	3.50 (1.39)	7.40 (2.97)
Male VTL	1.62 (0.47)	4.81 (2.43)	9.68 (3.62)
Child	1.47 (0.63)	2.92 (1.59)	6.23 (2.89)
	0.85 $F0$ + 1.19 VTL	1.70 $F0$ + 2.38 VTL	3.62 $F0$ + 5.07 VTL
Male	1.71 (1.48)	5.18 (2.77)	3.19 (2.77)
	1.63 $F0$ + 0.52 VTL	4.94 $F0$ + 1.56 VTL	3.04 $F0$ + 0.96 VTL

D. Discussion

The most noticeable result, illustrated in the lower panel of Fig. 1, is that the average VTL JND was directly affected by the number of bands in the sinewave vocoder. The average $F0$ JND, on the other hand, while suffering from vocoding, was, on average, less affected by the number of bands. These results are hence compatible with those observed in CI users by Fuller *et al.* (2014): When spectral resolution is reduced, as it also happens in actual CIs, VTL perception is severely degraded while some pitch percept is maintained.

However, it is worth noting that the sinewave vocoder used in the present experiment, as well as by Fuller *et al.* (2014), produces spectral cues related to the $F0$. The

multiplication of the sinusoidal carrier by the temporal envelope produces strong side tones at $F_c - kF_0$ and $F_c + kF_0$ (where F_c is the frequency of the carrier and k is a positive integer). When the carrier spacing is wide enough, such as in the 4-band condition, these side-tones can be used to perceive $F0$ differences in the 3AFC task (see Fuller *et al.*, 2014, for an illustration with 8 bands). When the carrier spacing is smaller, such as in the 12-band condition, the side-tones from successive bands may partially interfere, somewhat reducing the availability of this cue. This could explain the better JNDs observed at 4-bands than at 12-bands in the Male-$F0$ direction. In addition, because the presence or absence of side-tones can also be a strong cue in the 3AFC task, the position of the voice $F0$ relative to the envelope cutoff frequency (300 Hz) can also be critical. In particular, all trials where the test voice had an $F0$ greater than 300 Hz had negligible side-tones, and thus sounded qualitatively different from the reference voice. It is then unsurprising that the $F0$ JND was found to be around this limit (represented as a vertical dotted line in Fig. 1, top panel).

The spectral cues described above, present in sinewave vocoding, are not available to real CI users. Therefore, while the results obtained for VTL may partially explain those of Fuller *et al.* (2014), the comparison between $F0$ and VTL perception may give an unfair advantage to $F0$ when a sinewave vocoder is used. A sign that this is perhaps indeed the case is that, if $F0$ JNDs were estimated from the results of Fuller *et al.* for the CI group, they would be larger than 6 st while we found $F0$ JNDs around 4 st. Although pure discrimination of the voice cues and their use for gender categorization do not necessarily equate, this is nonetheless a potential indication that the JNDs for $F0$ obtained in the present experiment may be unrealistically small. To further assess this hypothesis, different types of vocoder carriers were used in Experiment 2.

It can be noted that the results obtained in the non-vocoded condition are somewhat different from those previously reported. Ives *et al.* (2005) describe the JNDs for VTL as between 4% and 7%, i.e., between 0.8 and 1.2 st, and report the value of 2% (0.3 st) for $F0$. The values obtained in the present experiment are all larger than 1 st. The discrepancy between the two experiments could be due to the type of syllables and sequences used: While the Cambridge group used sequences of 4 long syllables of nearly 700 ms, we used sequences of three 200-ms syllables. In total, our sequences contained 600 ms of speech, while the sequences of Ives *et al.* contained 2.7 s of speech material. Both providing longer signal time and a larger number of different syllables could explain the smaller JNDs reported in their experiment.

Finally, it is worth noting that no age effect was found. Although only 4 of the 14 subjects were aged above 50, visual inspection of their results showed rather better thresholds in all conditions than the younger participants. This was the case even for the $F0$-JND, despite some studies reporting age-related impairment on pitch perception (e.g., Russo *et al.*, 2012). This was also the case when the stimuli were vocoded, despite the fact that vocoded speech has been reported to be more problematic for older than for younger listeners in gender categorization tasks (Schvartz and Chatterjee, 2012) as well as in word identification tasks (Sheldon *et al.*, 2008). We thus conclude from these results that age is unlikely to affect performance negatively in the current experiment. In order to simplify the recruitment of participants in the following experiments, older NH subjects were not involved.

III. EXPERIMENT 2: EFFECT OF CARRIER SIGNAL IN THE VOCODER

A. Rationale

The objective of this second experiment was two-fold. First, it is to confirm that the undesirable spectral cues related to $F0$ played a role in the $F0$-JNDs measured in Experiment 1. To test this, we used a number of different carriers that do not produce such spectral $F0$ cues. The most commonly used carrier is the simple white noise (e.g., Shannon *et al.*, 1995). However, once filtered in a band, this type of noise tends to have an envelope of its own, disturbing the transmission of temporal pitch cues in the signal's envelope (Moore and Glasberg, 2001; Viemeister, 1979). To this purpose, other types of carriers have been produced. In the present study we investigated the potential benefits of "low-noise noise" (Pumplin, 1985) and "pulse-spreading harmonic complex" (PSHC, Hilkhuysen and Macherey, 2014). Previous studies have shown that narrow band low-noise noises (i.e., narrower than a critical band) have less power fluctuations than Gaussian noise, both in the acoustic signal and at the output of auditory filters (Hartmann and Pumplin, 1988; Hilkhuysen and Macherey, 2014). Low-noise noises spectrally broader than a critical band also demonstrate slightly reduced fluctuations compared to Gaussian noise, but this reduction does not seem to translate into sizeable psychophysical measurements when a single channel is used (Hartmann and Pumplin, 1988; Hilkhuysen and Macherey, 2014). The PSHC however displays low internal envelope fluctuations even in the broadband case, although again, this has only been tested with a single band (Hilkhuysen and Macherey, 2014). Based on these observations, it might be expected that low-noise noise would not yield smaller $F0$ JNDs than standard noise while PSHC is expected to. However, using these carriers in a multi-channel setting like a vocoder may yield different results than predicted by the previous mono-channel experiments, as the information across channels may be combined. If any of these carriers can yield $F0$ discrimination performance similar to that observed with speech stimuli in actual CI listeners (e.g., Chatterjee and Peng, 2008), then it will be possible to use a single vocoder to explore the $F0$-VTL space.

The second objective of this study was to assess the effect of these carriers on the VTL-JND. While the defects and benefits of various carriers have been studied for perception of voicing (e.g., Faulkner *et al.*, 2000), the effect of these carriers on VTL perception is not known. It could be the case, for instance, that sinewave carriers, by providing very sharp spectral peaks located at the center of each

frequency band, have disrupted the VTL estimation process, which presumably relies on identifying (formant) peaks. Moreover, in order to study the role of filter sharpness—a proxy for spread of excitation—as done in Experiment 3, a broadband carrier had to be used. Evaluating the effect of the carrier on the VTL-JND also allowed selecting the most appropriate carrier for the third experiment.

B. Material and methods

1. Participants

Twenty-three participants were recruited to take part in the experiment. One was excluded because their audiometric thresholds were too high in both ears. One was excluded because they could not perform the task, saturating the staircase procedure on all conditions. The 21 remaining participants all had NH (audiometric threshold ≤ 20 dB HL at 500, 1000, 2000, and 4000 Hz) and were aged between 20 and 50 (mean: 26.8 yrs, s.d.: 7.3). One subject had also participated in Experiment 1. As for the previous experiment, the participants received monetary compensation for taking part in the experiment.

2. Stimuli and apparatus

The vocoding method used in this experiment was the same as in Experiment 1 except that: (1) Six bands were used (over the same frequency range), (2) the cutoff frequency for envelope extraction was set to half the width of the band, but capped to 300 Hz, and (3) different carrier signals were used (sin, noise, low-noise noise, PSHC). An intermediate number of bands were used compared to Experiment 1 in order to reduce the number of conditions. For each carrier we evaluated the crest factor, i.e., the ratio of the maximum absolute value of the waveform to the rms. The crest factor gives an indication of how flat the temporal envelope of the carrier is: The lower, the flatter. The various carriers and the basilar membrane motion they induce (estimated using linear fourth-order gammatone filters) are shown in Fig. 2. The **sinewave** carriers resulted in crest factors around $\sqrt{2}$. The **noise** was a binary noise with values equal to -1 or $+1$ (in MATLAB), multiplied with the envelope extracted from each frequency band and then filtered again in that band. The carrier itself had crest factors between 2.7 and 4.0 in the acoustic waveforms, and between 2.8 and 3.5 at the output of auditory filters. The **low-noise** noise was created following Method 1 from Kohlrausch *et al.* (1997) by adding pure tones of equal amplitude, separated by 1 Hz, in random phase, between the lower cutoff and the upper cutoff of the frequency band. The signal was then divided by its Hilbert envelope and re-filtered in the frequency band by turning to zero all the frequency bins of the fast Fourier transform (FFT) falling outside the band boundaries. This process was repeated iteratively ten times. This resulted in smaller crest factors than for the Gaussian noise carrier, in the acoustic waveforms (1.7 to 1.9) and, to a lesser degree, at the output of auditory filters (2.2 to 2.8).

The **PSHC** was also generated by adding pure tones by 1-Hz steps within the boundaries of the spectral band, but a more complex phase relationship was applied. For each frequency band an integer factor k was determined to ensure optimal flatness of the envelope following the values reported by Hilkhuysen and Macherey (2014). In practice, the factor was calculated using a third order polynomial fit to the values of Hilkhuysen and Macherey: $k = \lfloor 0.459\, f_c^3 - 0.796\, f_c^2 + 6.91\, f_c + 7.22 \rceil$ where f_c is the center frequency of the frequency band (i.e., the geometric mean of the boundary frequencies) expressed in kilohertz. The phase of each component pure tone was then calculated as

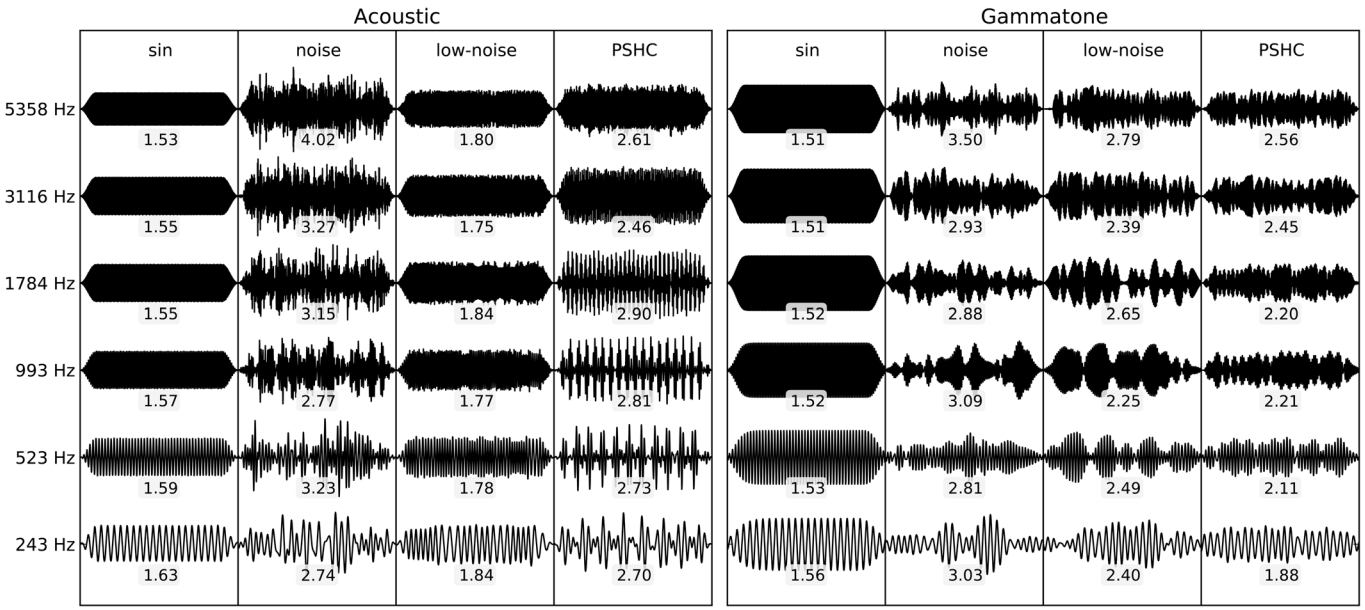

FIG. 2. Left panel: Acoustic waveform of the carrier signals used in Experiment 2 for each of the six frequency bands specified by their center frequency. Each segment shows 100 ms of signal. The number below each waveform is the crest factor. Right panel: Same but showing the output of auditory filters (linear fourth-order gammatone filter) centered on the center frequency of the band.

$$\varphi_i = 2\pi\left(r_i \cdot \frac{i}{k^2} + u_j\right), \text{ for } i \text{ such that :}$$
$$\text{for } j \in \{0\ldots k-1\},\ \exists n \in \mathbb{N},\ i + j = n \cdot \lambda.$$

In this equation, r_i is a random value drawn from $\{0,\ldots,k-1\}$ and u_j is a random value between 0 and 2π, both with uniform distribution. Crest factors obtained for this carrier are between 2.4 and 2.9 in the acoustic signal, but between 1.8 and 2.6 at the output of auditory filters. The PSHC was thus the carrier with the lowest crest factor, except for the sinewave.

The apparatus was the same as in Experiment 1.

3. Procedure

The task and procedure were similar to Experiment 1. The JNDs were obtained with the same method. However this time, in order to keep the length of the experiment reasonable, only two measurements were made per condition and the "diagonal voices" (in the direction of the male voice and of the child's voice) were not included leaving only JND for VTL alone and *F*0 alone. A total of 4 directions and 4 vocoding methods, i.e., 16 conditions, each repeated 2 times leading to a total number of 32 measurements, were performed in a random order. The experiment was divided into two sessions of 2 h that took place on different days within a period of 2 weeks (except one participant who had the second session 20 days after the first one).

C. Results

The JNDs obtained in this experiment are shown in Fig. 3: The left panel shows the JNDs for each carrier in the *F*0-VTL plane, the right panel shows the average JNDs for *F*0 and VTL as a function of carrier. The data presented in this latter panel was analyzed with a repeated-measures ANOVA on average JNDs with carrier and direction (*F*0 or VTL) as repeated factors. The type of carrier had a significant effect on the measured thresholds [$F(3,60) = 32.98$, $p < 0.0001$, $\eta_G^2 = 0.18$], and the JNDs for *F*0 and VTL were significantly different on average [$F(1,20) = 170.2$, $p < 0.0001$, $\eta_G^2 = 0.34$]. Importantly, the difference between *F*0 and VTL JNDs depended on the carrier signal that was used [$F(3,60) = 63.75$, $p < 0.0001$, $\eta_G^2 = 0.16$]. The JND for *F*0 was 4 to 5 semitones larger than that for VTL when noise, low-noise, or PSHC were used as carriers [$t(20) > 10.40$, $p_{FDR} < 0.0001$ for each comparison], but there was no significant difference when the sinewave carrier was used [$t(20) = 0.299$, $p_{FDR} = 0.77$].

One purpose of this experiment was to evaluate whether the type of carrier had an influence on the VTL JND. An ANOVA ran only on the VTL data revealed no effect of carrier type [$F(3,60) = 1.36$, $p = 0.26$, $\eta_G^2 = 0.011$].

D. Discussion

As expected, the noise carrier, the more common form of vocoding in acoustic simulations of CIs, and the sinewave carrier produced very different *F*0 JNDs. The sinewave carrier provides not only spectral cues related to *F*0, but also more salient temporal pitch cues than the noise carrier (as discussed, for instance, by Stone *et al.*, 2008). However, an unexpected outcome is that the low-noise and the PSHC carriers yielded *F*0 JNDs that were not different from those obtained with the noise carrier despite the fact that their temporal envelopes were flatter (as attested by the crest factors reported in Fig. 2).

These results can be explained in a number of ways. The low-noise is iteratively optimized to have a flatter acoustic envelope than regular noise. However, in the 6-band vocoder used here, the bands were 4 to 5 times wider than normal auditory filters (Glasberg and Moore, 1990). Therefore, as shown in Fig. 2 and as reported in other studies (e.g., Hartmann and Pumplin, 1988; Hilkhuysen and Macherey, 2014), although the low-noise noise algorithm is very efficient to flatten the envelope in the acoustic domain, the signal actually received by individual auditory filters is only slightly flatter than regular noise. This minor improvement may not have been sufficient to better transmit periodicity cues than the standard noise.

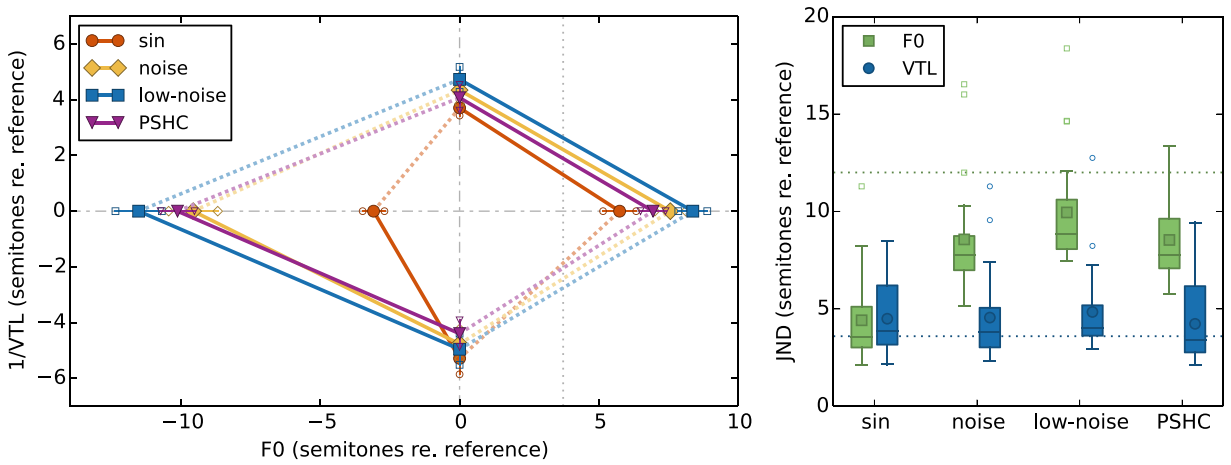

FIG. 3. (Color online) Left panel: JNDs for the four different carriers used in Experiment 2 in the *F*0-VTL plane. Right panel: *F*0 and VTL JNDs for different carrier signals. See Fig. 1, lower panel, for a description of the elements of the boxplot.

As was previously reported by Hilkhuysen and Macherey (2014), the PSHC did not suffer from this defect: Both the acoustic and the auditory crest factors were consistently smaller than those observed for the noise. However, the PSHC is pulsatile, with a pulse rate differing for each channel. The low frequency channels have the slowest pulsation rates, but, in the signal to be vocoded, these are also the ones that contain the strongest envelope fluctuations related to *F*0. With the parameters used in the present experiment, the PSHC had a pulse rate of about 80 Hz in the lowest band, while the *F*0 of the reference voice was 242 Hz on average. In the second band, the pulse rate of the PSHC was 110 Hz, and still only 180 Hz in the third band. Although the pulsatile nature of this carrier, and hence its pulse rate, are barely visible in the output of the auditory filters, it is possible that these low pulse-rates prevented proper coding of the signal's *F*0.

More importantly, the experiment showed no influence of the type of carrier on the VTL JND. This allowed us to use a noise carrier in the following experiment in order to study the effect of spread of excitation on VTL JND.

IV. EXPERIMENT 3: EFFECT OF SPREAD OF EXCITATION

The objective of this experiment was to assess the potential role of the electrical spread of excitation that occurs in the cochlea when current is injected through an intra-cochlear electrode of the implant, and returned through an extra-cochlear electrode, i.e., in "monopolar mode" (Black and Clark, 1980). While the peak of neural activity generally remains located close to the electrode, the amount of spread of excitation defines how much interaction happens between individual stimulation channels. This effect can be simulated in vocoders by modifying the order, and thus the sharpness, of the synthesis filters.

Others have constructed more elaborate vocoders where the current spread was simulated using a simple model of electrical flow in the cochlear fluid (Bingabr *et al.*, 2008; Churchill *et al.*, 2014; Laneau *et al.*, 2006). However, these approaches are limited by the fact that the relationship between the level of current in the cochlea and its specific contribution to loudness is unknown. It is therefore difficult to know how the shape of current spread along the cochlea can be translated into the acoustic domain in the vocoder. For this reason we used the simpler approach of manipulating the order of the filters of the vocoder (e.g., Fu and Nogaki, 2005; Litvak *et al.*, 2007). Although this method may not accurately represent the shape of the spread of excitation, it does qualitatively reproduce the effect of channel interaction.

A. Material and methods

1. Participants

Sixteen volunteers participated in this experiment. They were aged 19 to 51 (mean: 26.6, s.d.: 9.0) and all had audiometric thresholds ≤ 20 dB HL at octave frequencies between 500 and 4000 Hz. One of the volunteers had participated in Experiment 1, and five others had participated in Experiment 2. As for the previous experiment, the participants received a compensation for taking part in the experiment.

2. Stimuli and apparatus

Since the previous experiment showed that the type of carrier does not seem to matter for VTL JNDs, here we used a standard noise-band vocoder, as described in Experiment 2. The effective spectral resolution was manipulated both by the number of frequency bands (4 and 12 bands) and by the order of the filters. The filters were Butterworth filters of 4th, 8th, and 12th order, which correspond to slopes of −24, −48, and −72 dB/octave, respectively. For comparison, Bingabr *et al.* (2008) reviewed the literature for current spread estimates and reported an average value of 2.8 dB/mm, which corresponds to slopes of about −40 dB/octave according to their calculation. Similarly, Churchill *et al.* (2014) used filters with slopes ranging from −40 to −30 dB/octave.

3. Procedure

The VTL JNDs were measured using the same adaptive 3AFC procedure as in Experiments 1 and 2. For each filter-order and number-of-channels combination, two JND measures were obtained for positive VTL differences relative to the female voice (toward male VTLs), and two other measures were obtained for negative VTL differences (toward child VTLs). Therefore in total 12 JND measures were performed with each participant, in a random order. The experiment was divided into two sessions of 2 h taking place on different days.

B. Results

The VTL JNDs for the different vocoders are shown in Fig. 4. The JNDs were analyzed using a repeated measures ANOVA with the number of bands, filter order, and VTL

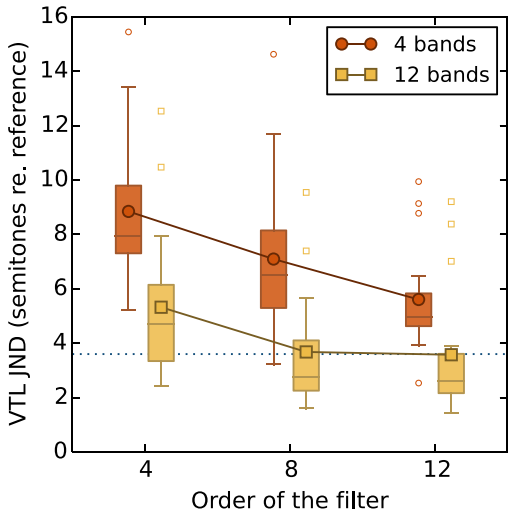

FIG. 4. (Color online) VTL JNDs for 4- (circle) and 12-band (square) vocoders as a function of filter order from shallow (4th order) to sharp (12th order). See Fig. 1, lower panel, for a description of the elements of the boxplot.

change direction (toward a male VTL or toward a child VTL) as repeated factors. Like in Experiment 1, fewer bands in the vocoder was associated with larger VTL JNDs [$F(1,15) = 85.4$, $p < 0.0001$, $\eta_G^2 = 0.233$]. Shallower filters also yielded larger JNDs [$F(2,30) = 55.8$, $p < 0.0001$, $\eta_G^2 = 0.128$]. However, if a sufficient number of bands were used, using sharper filters did not improve the JNDs, resulting in an interaction between the two factors [$F(2,30) = 5.52$, $p = 0.018$, $\eta_G^2 = 0.015$]. Finally, elongating the VTL (toward a male's voice) yielded JNDs on average 0.96 st larger than when shrinking the VTL [toward a child's voice; $F(1,15) = 6.31$, $p = 0.024$, $\eta_G^2 = 0.030$]. Because this effect was very small and did not interact with any of the other factors, the positive and negative VTL differences were averaged in Fig. 4. None of the other interactions were significant.

In Fig. 4, like in previous figures, the dotted line shows the 3.6 st VTL difference that was used between the typical male and female voices by Fuller *et al.* (2014). With 4 channels, the average JNDs were above this limit for all filter orders [$t(15) > 3.98$, $p_{FDR} < 0.0024$]. With 12 channels, the average JND was significantly different from 3.6 for 4th order filters [$t(15) = 2.40$, $p_{FDR} = 0.045$] but not for sharper filters [8th order: $t(15) = 0.143$, $p_{FDR} = 0.97$; 12th order: $t(15) = -0.040$, $p_{FDR} = 0.97$]. With 12 channels and 4th order filters, only 5 of the 16 participants would be able to detect the 3.6 st VTL difference between a male and a female speaker. Still with 12 channels, 10 and 12 of the 16 participants would detect that difference for 8th and 12th order filters, respectively.

Finally using 4th order filters with 12 bands produced, on average, JNDs that were indiscernible from those obtained with 12th order filters, 4 bands [$t(15) = -0.72$, $p = 0.48$, for a difference of -0.28 st].

C. Discussion

In a vocoder, spectral resolution can be manipulated both by the number of channels and by the sharpness of the filters. The former affects how spectral information is quantized along the frequency axis, while the latter determines how much overlap and blending occurs between the channels. Although the two manipulations are physically different, our results show that they had perceptually similar effects on VTL difference detection. The two manipulations are also not completely independent: When the number of channels was high, increasing the order of the filter beyond eight did not further improve the JND.

In actual implants, the number of channels is closer to 12 than to 4, and it has been argued that typical current spread in CIs was equivalent to filter slopes comprised between -40 and -30 dB/octave, i.e., between 4th and 8th order with our filters. In these conditions, it is expected that most listeners would be unable to detect the 3.6 st difference separating male from female voices, which is consistent with the lack of VTL effect on gender categorization observed by Fuller *et al.* (2014) in CI listeners. However by increasing the order from 4 to 8, i.e., when the filter slopes were brought to -48 dB/octave, most of the participants' JNDs dropped below this critical difference. In other words, if current spread could be reduced in CIs, it might be possible to improve VTL perception. According to Bingabr *et al.* (2008), bipolar stimulation produces current spread of about 7.4 dB/mm, which is equivalent to about -100 dB/octave in acoustic terms, and tripolar stimulation produces even a smaller current spread. Our results therefore suggest that using current focusing techniques could potentially improve VTL perception in CI users.

Another conclusion from this experiment is that, as far as VTL difference detection is concerned, using fewer bands with sharp filters can be equivalent to more bands with shallow filters. However, one would have to keep in mind that sharpening the filters does not always improve VTL JNDs, as it happened in our experiment between 8th and 12th order filters with 12 bands.

V. CONCLUSION

The data reported in the present study indicates that VTL perception, as measured by the JND, strongly depends on the spectral resolution available to the listener. In comparison, *F*0 perception seems to be more resilient to a reduction in spectral resolution. This is consistent with previous reports showing that listeners with moderate hearing loss were able, like NH listeners, to take advantage of *F*0 differences to selectively listen to one of two competing sentences even though they could not benefit from VTL differences (Mackersie *et al.*, 2011). This is also consistent with previous gender categorization studies arguing that CI users may rely more on *F*0 differences than NH listeners (Fu *et al.*, 2004, 2005; Kovačić and Balaban, 2009). Finally this is in line with the results of Fuller *et al.* (2014) who showed that, unlike NH listeners who give equal weight to *F*0 and VTL differences for gender categorization, CI users almost exclusively give weight to the *F*0 difference.

The likely explanation to why perception of *F*0 is more robust than that of VTL to spectral degradations is that *F*0 is encoded through both temporal and spectral cues (Carlyon and Shackleton, 1994). When spectral resolution is degraded, temporal *F*0 cues can remain relatively unaffected. In Experiment 2, however, various noise-like carriers were used, providing temporal pitch cues that should have been more or less salient based on the differences in crest factors. Yet, the *F*0 JNDs remained the same across all these carriers, suggesting that this explanation might not be as straightforward as it seems. Importantly, whatever *F*0 cue remains available appears to be sufficient to discriminate the *F*0 of a male voice from that of a female voice as the *F*0 JNDs were all below an octave.

VTL perception was not affected by the type of carrier, suggesting that temporal cues (or fine spectral cues) are not important for this voice dimension. Coding strategies aiming at enhancing the temporal fine structure in the implant are thus unlikely to provide a benefit for VTL perception. Instead, spectral resolution was manipulated both by changing the number of channels in a vocoder and by modifying the amount of interaction between the channels simulating electrical spread of excitation like in CI stimulation. The

results show that these two factors affect VTL perception in a similar fashion. When using values realistic for implants, the results obtained from the vocoder suggest that VTL JNDs for actual CI users would be larger than the typical difference between adult male and female speakers, making it impossible for CI recipients to use this cue to recognize the gender of a speaker or to segregate competing speakers. Furthermore, because the use of sharper filters improved the JNDs, our results suggest that current focusing techniques (Bonham and Litvak, 2008; Srinivasan *et al.*, 2010) could improve VTL perception in CIs.

ACKNOWLEDGMENTS

The authors are thankful to Esmée van der Veen and Maraike Coenen for their help in collecting data, and to Gaston Hilkhuysen, Quentin Mesnildrey, and Olivier Macherey for providing code, support, and comments for the PSHC. The authors are supported by a Rosalind Franklin Fellowship from the University Medical Center Groningen, University of Groningen, and the VIDI Grant No. 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.

Bakeman, R. (**2005**). "Recommended effect size statistics for repeated measures designs," Behav. Res. Methods **37**, 379–384.

Bingabr, M., Espinoza-Varas, B., and Loizou, P. C. (**2008**). "Simulating the effect of spread of excitation in cochlear implants," Hear. Res. **241**, 73–79.

Bird, J., and Darwin, C. J. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 263–269.

Black, R. C., and Clark, G. M. (**1980**). "Differential electrical excitation of the auditory nerve," J. Acoust. Soc. Am. **67**, 868–874.

Bonham, B. H., and Litvak, L. M. (**2008**). "Current focusing and steering: Modeling, physiology, and psychophysics," Hear. Res. **242**, 141–153.

Bosman, A. J., and 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.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

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., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?," J. Acoust. Soc. Am. **95**, 3541–3554.

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

Churchill, T. H., Kan, A., Goupell, M. J., Ihlefeld, A., and Litovsky, R. Y. (**2014**). "Speech perception in noise with a harmonic complex excited vocoder," J. Assoc. Res. Otolaryngol. **15**, 265–278.

Darwin, C. J., Brungart, D. S., and 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.

Dudley, H. (**1939**). "The vocoder," Bell Labs Rec. **18**, 122–126.

Faulkner, A., Rosen, S., and Smith, C. (**2000**). "Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants," J. Acoust. Soc. Am. **108**, 1877–1887.

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

Friesen, L. M., Shannon, R. V., Başkent, D., and 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, Q.-J., Chinchilla, S., and 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., and Galvin, J. J., III (**2005**). "Voice gender identification by cochlear implant users: The role of spectral and temporal resolution," J. Acoust. Soc. Am. **118**, 1711–1718.

Fu, Q.-J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Fuller, C., Gaudrain, E., Clarke, J., Galvin, J. J., Fu, Q.-J., Free, R., and Başkent, D. (**2014**). "Gender categorization is abnormal in cochlear-implant users," J. Assoc. Res. Otolaryngol. **15**, 1037–1048.

Gaudrain, E., Li, S., Ban, V., and Patterson, R. (**2009**). "The role of glottal pulse rate and vocal tract length in the perception of speaker identity," Interspeech 2009 **1**(5), 152–155.

Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Hartmann, W. M., and Pumplin, J. (**1988**). "Noise power fluctuations and the masking of sine signals," J. Acoust. Soc. Am. **83**, 2277–2289.

Hilkhuysen, G., and Macherey, O. (**2014**). "Optimizing pulse-spreading harmonic complexes to minimize intrinsic modulations after auditory filtering," J. Acoust. Soc. Am. **136**, 1281–1294.

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

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

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (**1997**). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," Acta Acust. Acust. **83**, 659–669.

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

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

Laneau, J., Wouters, J., and Moonen, M. (**2006**). "Improved music perception with explicit pitch coding in cochlear implants," Audiol. Neurootol. **11**, 38–52.

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

Litvak, L. M., Spahr, A. J., Saoji, A. A., and Fridman, G. Y. (**2007**). "Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners," J. Acoust. Soc. Am. **122**, 982–991.

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

Mackersie, C. L., Dewey, J., and 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.

Moore, B. C. J., and Glasberg, B. R. (**2001**). "Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 1067–1073.

Patterson, R. D., Gaudrain, E., and Walters, T. C. (**2010**). "The perception of family and register in musical notes," in *Music Perception, Springer Handbook of Auditory Research*, 1st ed., edited by M. R. Jones, R. R. Fay, and A. N. Popper (Springer, New York), Vol. 36, pp. 13–50.

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

Pumplin, J. (**1985**). "Low-noise noise," J. Acoust. Soc. Am. **78**, 100–104.

Roers, F., Mürbe, D., and Sundberg, J. (**2009**). "Voice classification and vocal tract of singers: A study of x-ray images and morphology," J. Acoust. Soc. Am. **125**, 503–512.

Russo, F. A., Ives, D. T., Goy, H., Pichora-Fuller, M. K., and Patterson, R. D. (**2012**). "Age-related difference in melodic pitch perception is probably mediated by temporal processing: Empirical and computational evidence," Ear Hear. **33**, 177–186.

Schvartz, K. C., and Chatterjee, M. (**2012**). "Gender identification in younger and older adults: Use of spectral and temporal cues in noise-vocoded speech," Ear Hear. **33**, 411–420.

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

Sheldon, S., Pichora-Fuller, M. K., and Schneider, B. A. (**2008**). "Effect of age, presentation method, and learning on identification of noise-vocoded words," J. Acoust. Soc. Am. **123**, 476–488.

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

Srinivasan, A. G., Landsberger, D. M., and Shannon, R. V. (**2010**). "Current focusing sharpens local peaks of excitation in cochlear implant stimulation," Hear. Res. **270**, 89–100.

Stickney, G. S., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Stone, M. A., Füllgrabe, C., and Moore, B. C. J. (**2008**). "Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region," J. Acoust. Soc. Am. **124**, 2272–2282.

Summers, V., and Leek, M. R. (**1998**). "FO 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.

Vestergaard, M. D., Fyson, N. R. C., and Patterson, R. D. (**2009**). "The interaction of vocal characteristics and audibility in the recognition of concurrent syllables," J. Acoust. Soc. Am. **125**, 1114–1124.

Vestergaard, M. D., Fyson, N. R. C., and Patterson, R. D. (**2011**). "The mutual roles of temporal glimpsing and vocal characteristics in cocktail-party listening," J. Acoust. Soc. Am. **130**, 429–439.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.

Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.

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