Perception of vocal characteristics in cochlear implants

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

Contact: etienne.gaudrain@cnrs.fr

1. University Medical Center Groningen, Department of Otorhinolaryngology | Groningen, Netherlands
2. University of Groningen, Research School of Behavioural and Cognitive Neuroscience | Groningen, Netherlands
3. Lyon Neuroscience Research Center, CNRS UMR 5292, INSERM U1028, Université Lyon 1 | Lyon, France

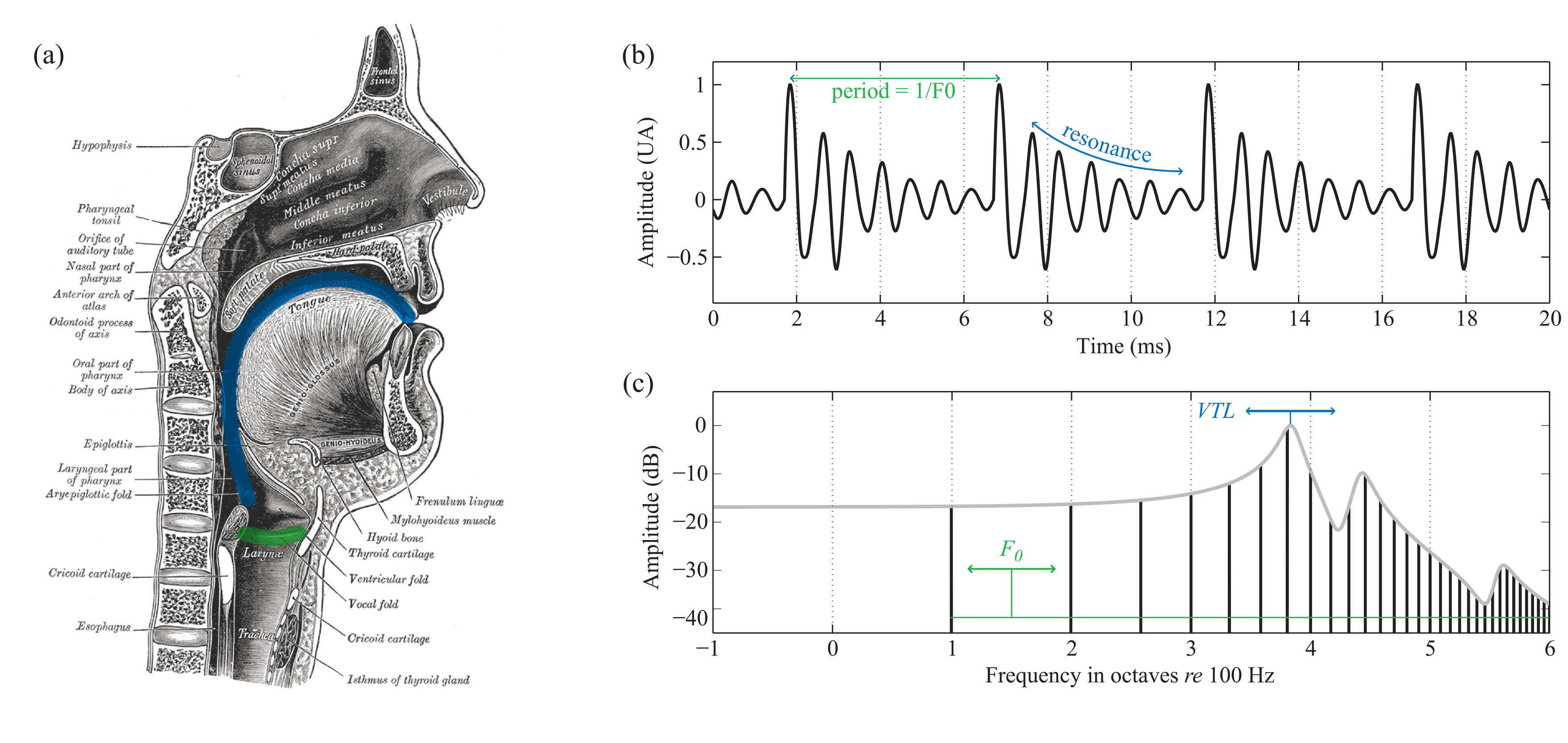
Introduction

To categorize speakers as 'male' or 'female', or to track a voice in a crowded environment, normal-hearing (NH) listeners rely on **two** vocal characteristics: **glottal-pulse rate** (GPR), which defines the F0, and **vocal-tract length** (VTL), related to the size of the speaker.

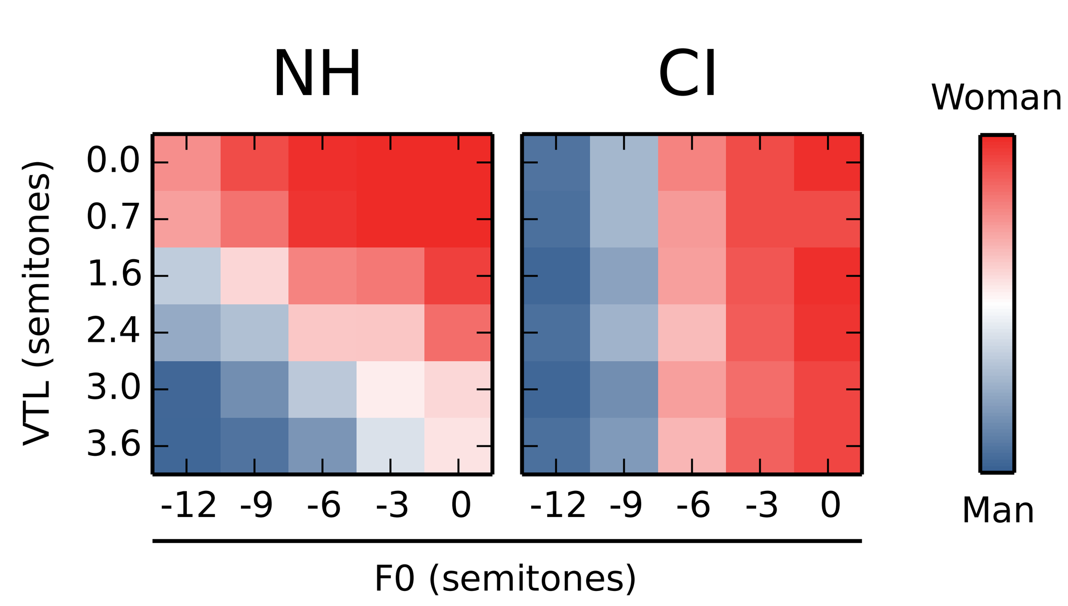
Figure 1 locates the origin of these dimensions on a cross-section of the head and shows their effect on the waveform and the spectrum of a vowel. Previous studies showed that normal-hearing (NH) listeners are extremely sensitive to these two dimensions.

Figure 2 shows gender categorization by NH listeners and cochlear implant (CI) users, as a function of F0 and VTL from Fuller et al. (2014). This study shows that gender categorization is abnormal in CI users because they do not exploit VTL differences. However it is unclear whether CI listeners fail to use the VTL dimension because they cannot detect changes along this dimension, or because the VTL information is so distorted that it cannot be used as such in a gender categorization task.

The purpose of the present study is to directly measure VTL (and F0) sensitivity in CI listeners.



▲ **Figure 1** – *(a)* Vocal folds (in green) and vocal tract (in blue) on a sagittal cross-section of the head (adapted from Gray, 1908). *(b)* F0 (or GPR) and resonance (VTL) shown on the waveform of an /a/. *(c)* Same on the spectrum. (Adapted from Patterson et al., 2010)



▶ **Figure 2** – Proportion of words judged as uttered by a male/female speaker as a function of F0 (x-axis), and of VTL (y-axis). Both VTL and F0 are shown in semitones relative to the original voice. The left panel shows average results for NH listeners. The right panel shows average results for CI listeners. Adapted from Fuller et al. (2014).

Methods

14 NH and 11 CI participants listened to triplets of Dutch CV syllables in an adaptive 3AFC task, tracking the just-noticeable difference (JND) in various directions of the F0-VTL plane. The original utterances were recorded from a female speaker and manipulated with STRAIGHT to effect changes in F0 and/or VTL. All JNDs were measured relative to the original voice. The JNDs were measured three times for each direction.

The CI participants were aged 46 to 73 years old and were all post-lingually deaf. Four of the participants had an Advance Bionics device, while the others all had Cochlear devices. The NH participants were aged 19 to 63 years old. They all had audiometric thresholds at 20 dB HL or below at octave frequencies between 500 and 4000 Hz.

For the CI participants, the JNDs were measured in free field. The stimuli were presented at 60 dB SPL on a Tannoy loudspeaker located about 1 m in front of the subject. For the NH participants, the JNDs were measured with headphones in a sound-attenuated booth. The stimuli were also presented at 60 dB SPL.

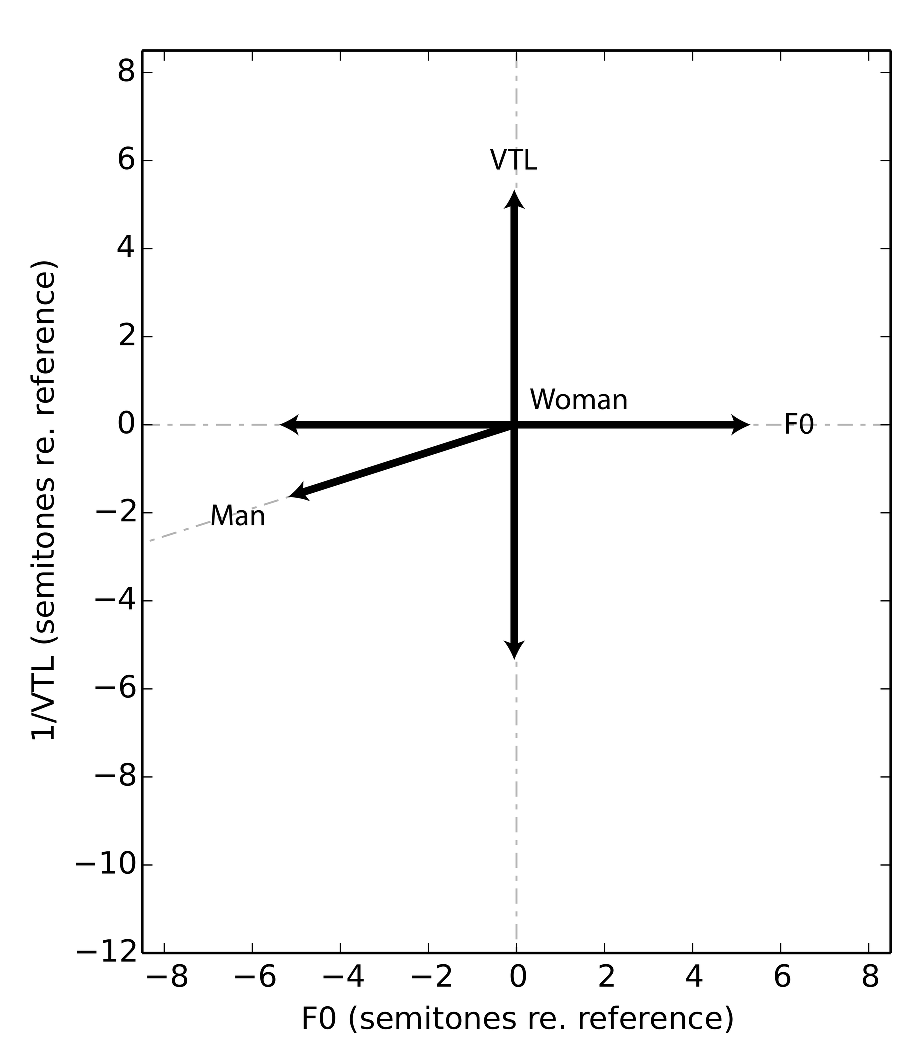
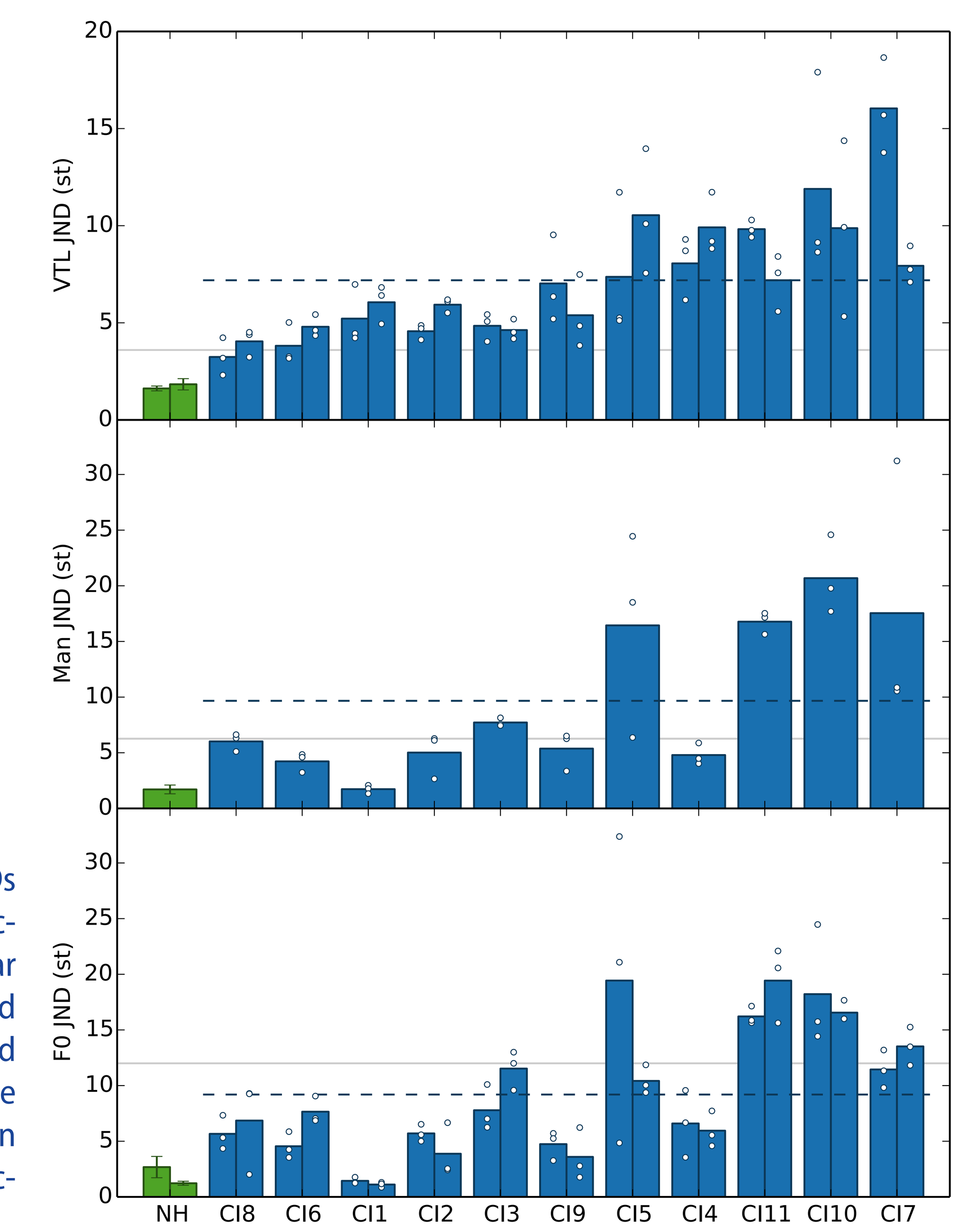


Figure 3 – Directions along which the JND was measured in the F0-VTL plane. The reference voice was that of a woman.

Results

Along the VTL dimension, the JNDs for the CI listeners are all but one above the typical VTL difference found between men and women. Along the F0 dimension, this is the case only for 4 or the 11 CI participants. The CI JNDs are also on average significantly higher than those of the NH listeners [VTL: $t(10.8)=5.94$, $p<0.001$; F0: $t(11.7)=3.95$, $p=0.002$].

The JNDs towards the Man voice seem to be more related to the F0 JNDs [$R^2=0.90$] than to the VTL JNDs [$R^2=0.61$]. The F0 and VTL JNDs are weakly by significantly correlated [$R^2=0.44$, $p=0.025$].



▶ **Figure 4** – Average (NH) and invidiual (CI) JNDs for VTL (top), F0 (bottom) and Man (middle) directions in the F0-VTL plane. For F0 and VTL the first bar corresponds to negative differences, and the second bar corresponds to positive differences. The dashed line shows the CI average. The gray line shows the difference between the man and woman voices in Fuller et al. (2014). The CI participants are ordered according to their VTL JND.

Discussion

In all CI users but one, the VTL JNDs were larger than the typical difference between male and female voices. Unlike the subjective gender categorezation results of Fuller et al. (2014), these JNDs were measured with an objective, unbiased method. The 3AFC task does not require listeners to interpret the VTL cue as such, as they can use any perceived difference between the test and standard stimuli. Our results thus suggest that abnormal gender categorization in CIs results from lack of VTL sensitivity rather than distortion of the VTL cue.

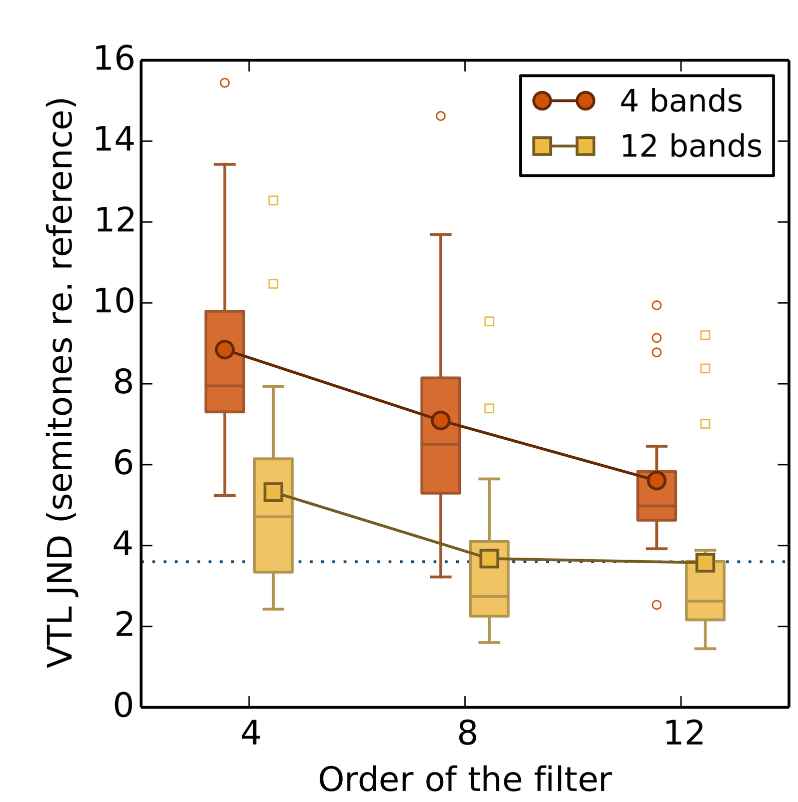


Figure 5 – VTL JNDs in 12 NH listening through 4 and 12 bands vocoders, as a function of filter order. From Gaudrain and Başkent (2014).

The exact cause of this lack of sensitivity, however, remains unclear. Analysis of electrical patterns in Fuller et al. (2014) indicates that VTL differences are available at the output of the implant. The average VTL JND reported here corresponds to a shift of about 2 electrodes in the implant's stimulation pattern. Vocoders designed to investigate the role of spectral resolution in VTL perception showed that both number of bands and amount of channel interaction affect VTL perception (**Figure 5**; Gaudrain and Başkent, 2014). The JNDs measured in CIs correspond to either 12 bands, 4th order filters, or 4 bands, 8th order filter. Similar vocoders have been shown to also yield speech-in-noise intelligibility comparable to that of actual CI users. These values suggest that considerable current spread in CIs impairs the effective spectral resolution and inflates VTL JNDs compared to NH listeners.

VTL perception is not only useful to discriminate male from female talkers. Previous work showed that VTL is a powerful cue for speaker separation in speech-on-speech situations (Darwin et al., 2003). The lack of VTL sensitivity thus deprives CI listeners from one the segregation cues available to NH listeners in cocktail party situations. It is therefore necessary to develop and investigate new methods to improve VTL representation in CIs.

Moreover, the measurement method was found to be rather reliable and repeatable within subjects. Unlike other spectral resolution measures (like spectral ripple discrimination), it has potential to provide a functional measure directly relevant for speech perception. Such method could be easily implemented in the clinic and could provide indications to clinician during the fitting procedures.

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. doi: 10.1121/1.1616924

Fuller C, Gaudrain E, Clarke J, et al (2014) Gender categorization is abnormal in cochlear-implant users. J Assoc Res Otolaryngol. doi: 10.1007/s10162-014-0483-7

Gaudrain E, Başkent D (2014) Factors limiting perception of vocal characteristics in cochlear-implants. 37th Annual Mid-winter Meeting of the Association for Research in Otolaryngology, San Diego, California. doi: 10.6084/m9.figshare.964805

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