Factors limiting vocal-tract-length perception in cochlear-implants

Etienne Gaudrain, Deniz Başkent

University Medical Center Groningen, Department of Otorhinolaryngology | University of Groningen, School of Behavioural, Cognitive Neuroscience | Groningen, Netherlands

university of groningen

umcg

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: with vowels, the JND for F0 is about 2%, and between 5 and 7% for VTL (Smith and Patterson, 2005).

Figure 2 shows gender categorization by NH listeners and cochlear implant (CI) users, as a function of F0 and VTL from Gaudrain et al. (2013). This study shows that gender categorization is abnormal in CI users because their perception of VTL is very limited.

The purpose of the present study is to test various possible causes for this lack of VTL sensitivity using different types of vocoders.

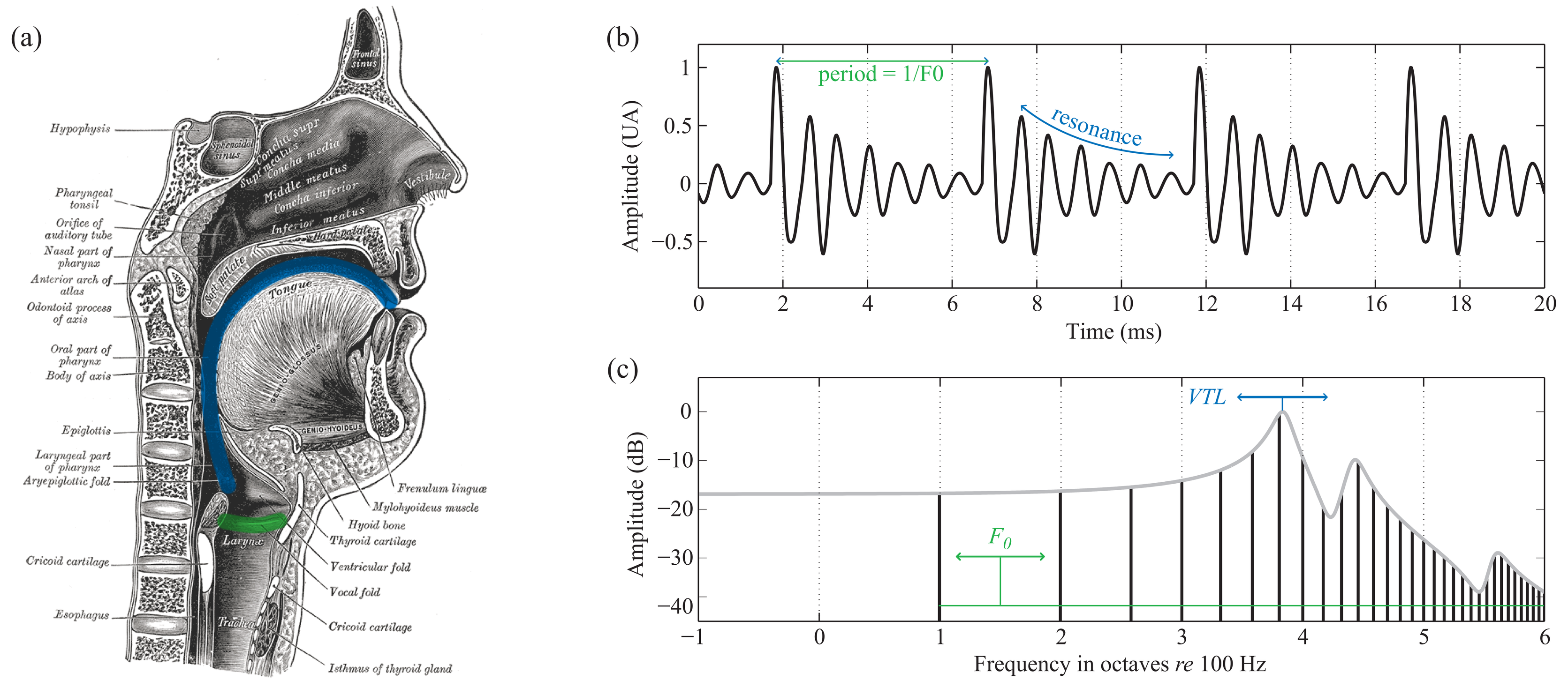


Figure 1 – *(a)* Vocal folds (in green) and vocal tract (in blue) on a sagittal cross-section of the head (adapted from Gray, 1908). *(b)* F0 (or GPR) and resonance (VTL) shown on the waveform of an /a/. *(c)* Same on the spectrum. (Adapted from Patterson et al., 2010)

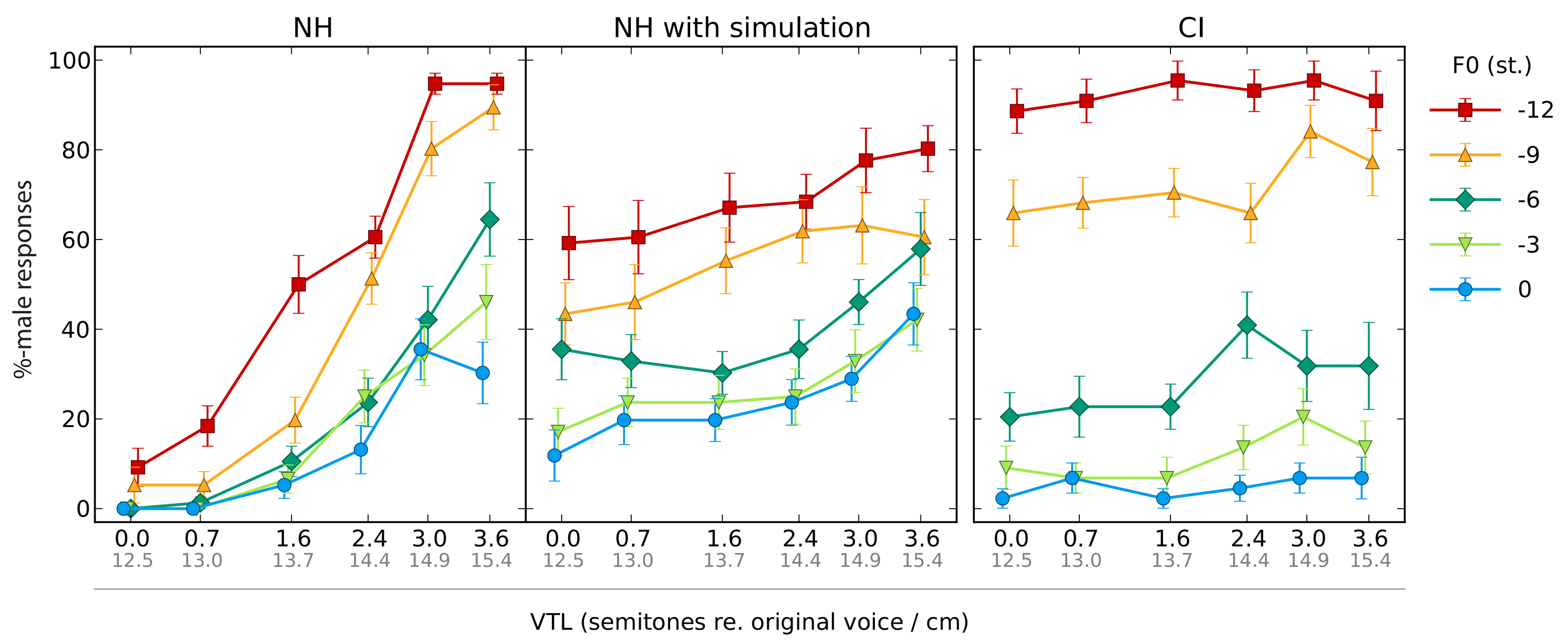


Figure 2 – *Top row:* Proportion of words judged as uttered by a male speaker as a function of VTL (x-axis), and of F0 (color). Both VTL and F0 are shown in semitones relative to the original voice. *Bottom row:* Same data but plotted as a function of F0 (x-axis) and VTL (color). For each subject, each point is the average of 4 responses. The errorbars show the across-subject standard error. From Gaudrain et al. (2013).

Contact details: e.p.c.gaudrain@umcg.nl

Exp. 1: Number of bands

14 NH participants listened to triplets of Dutch CV syllables in an adaptive 3AFC task tracking the 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 stimuli were either presented as is, or vocoded using a sinewave vocoder with 4 or 12 bands equally spaced along the cochlea between 150 and 7000 Hz. The band-pass 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 4th order Butterworth filter). The envelope was then used to modulate the amplitude of a sinewave centered on the frequency band.

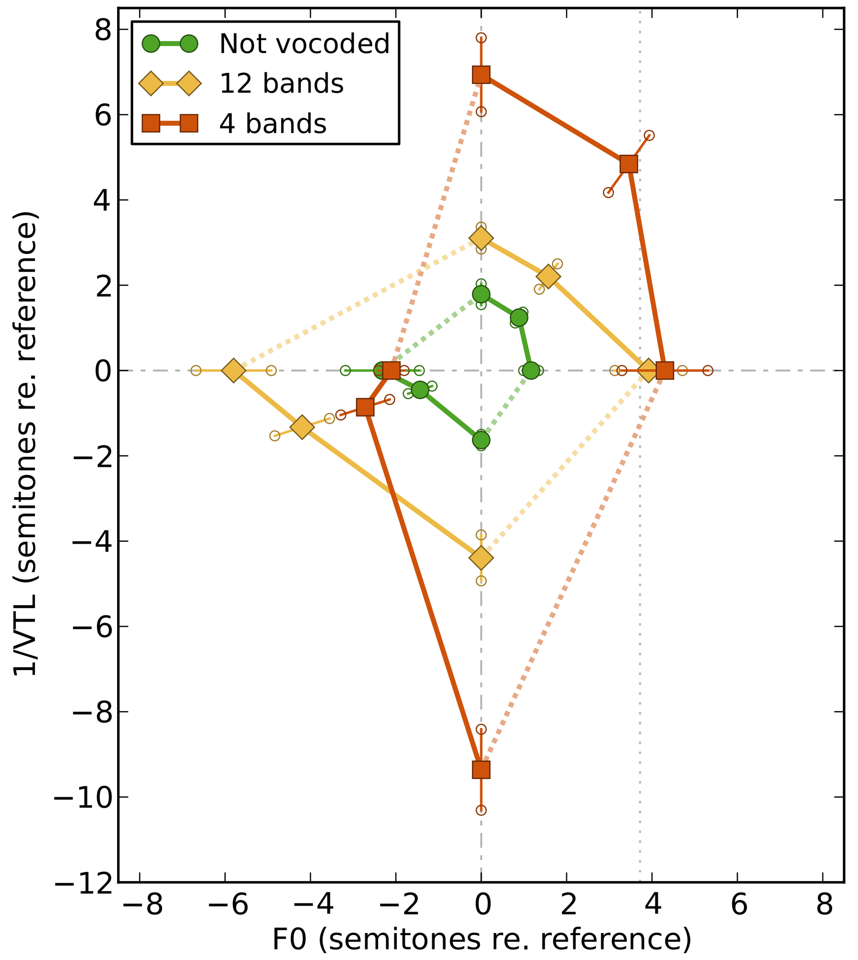
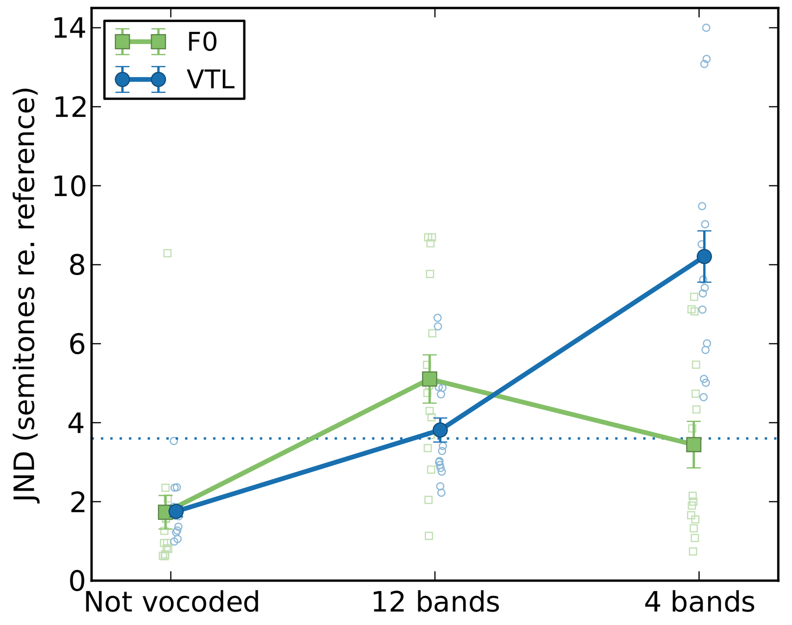


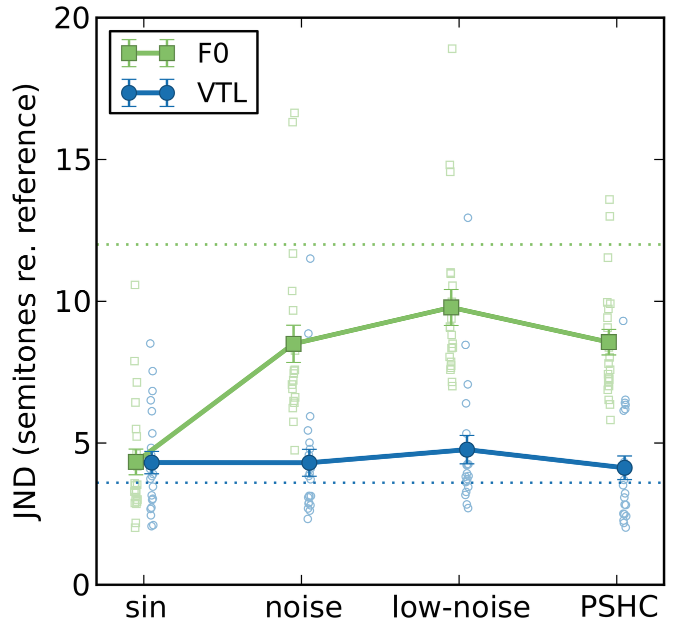
Figure 3 – *Left:* JNDs in the F0-VTL plane for the three conditions non-vocoded, 12 and 4 bands. JNDs in the F0-VTL plane for the three conditions non-vocoded, 12 and 4 bands. *Below:* Average JNDs for F0 and VTL in the three vocoding conditions. The light points show individual data.



Exp. 2: Type of carrier

Sinewave vocoder does not allow simulation of the spread of excitation that happens in actual CIs. It also provides spectral pitch cues which are not present in real implants. Here we test various carriers that do not provide spectral cues, and which have more or less flat envelopes, providing more or less clean F0-modulation information. The main objective is to evaluate the effect on VTL JNDs.

Figure 4 – Average F0 and VTL JNDs for the sinwave, noise, low-noise noise and phase-spreading harmonic complex (PSHC, see Macherey et al., poster P23). Individual results are shown as light symbols. The dotted lines represent the F0 and VTL difference between the male and female talkers from Gaudrain et al. (2013).



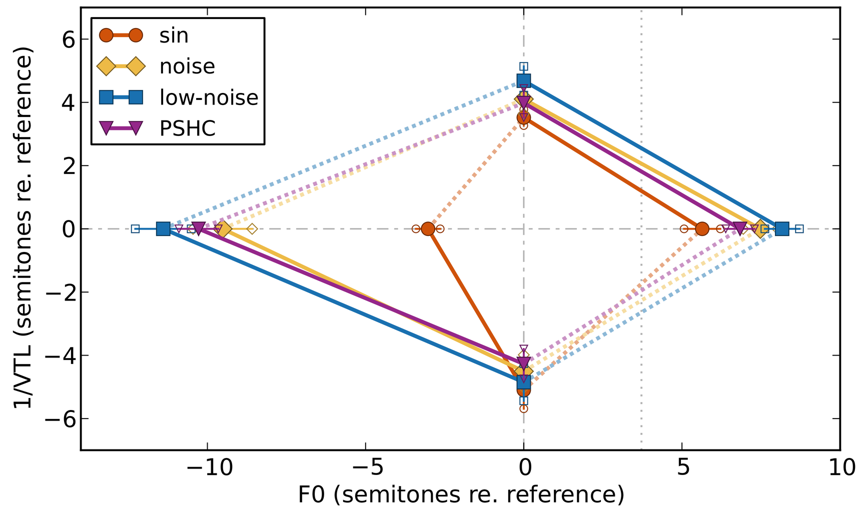
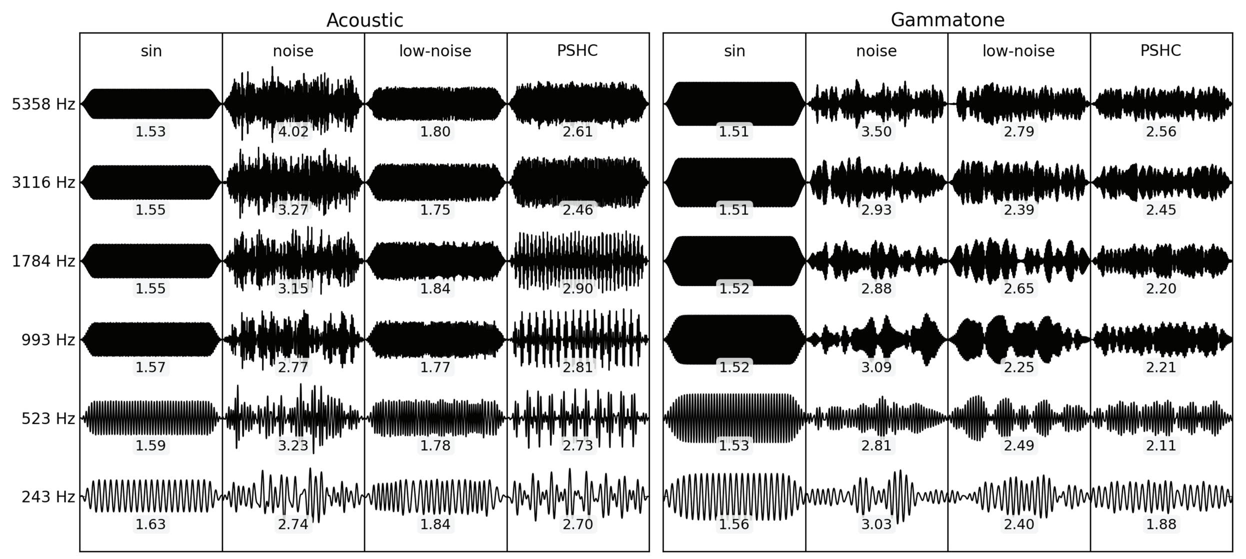


Figure 5 – JNDs in the F0-VTL plane for the different carriers.

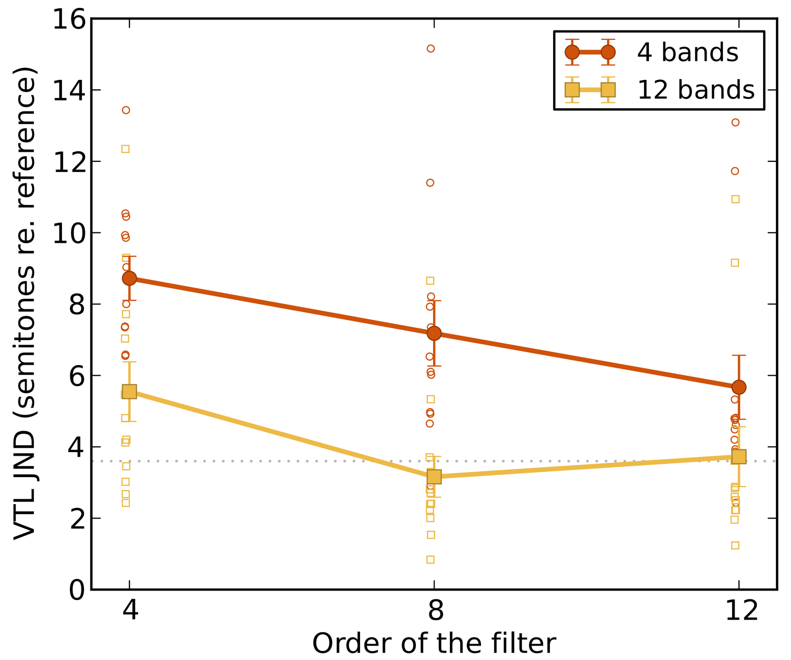
▼ Figure 6 – *Left:* Acoustic waveforms of the different carriers in the 6 bands. *Right:* Waveforms at the output of gammatone filters centered on the frequency band. The crest factor is reported for each waveform.



Exp. 3: Spread of excitation

Using a noise carrier, spread of excitation can be simulated by changing the order of the bandpass filters. 4th order filters (-24 dB/oct.) are very shallow and produce a fair amount of channel interaction. 8th order filters (-48 dB/oct.) are reasonably sharp and 12th order filters (-72 dB/oct.) are very sharp. Only the VTL JNDs were measured in this experiment.

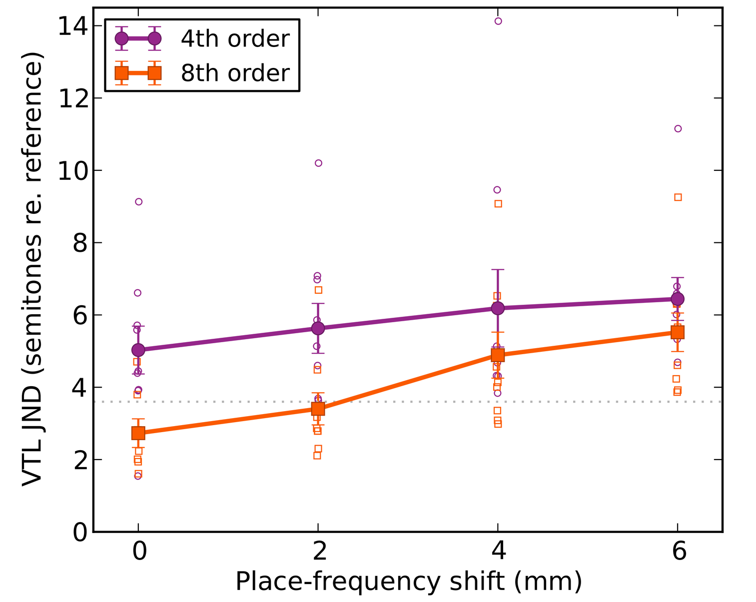
Figure 7 – Average VTL JNDs for 4 and 12 bands vocoders, as a function of filter order. The light symbols represent individual data for the 12 participants.



Exp. 4: Place-frequency shift

Another characteristic of CIs is the fact that the frequency band allocated to a specific electrode may not correspond to frequency to which the place is tuned. This place-frequency mismatch can be quantified in millimeters along the basilar membrane and can be simulated by shifting the synthesis filter in the vocoder by a fixed distance. Twelve bands were used, with two filter orders and noise carriers.

Figure 8 – Average VTL JNDs for 4th and 8th order filters, as a function of place-frequency shift. The light symbols represent individual data for the 9 participants.



Gaudrain E., Fuller C.D., Clarke J., Galvin J.J., Fu Q.-J., Free R.H., Başkent D. (2013). "Little Red Riding Hood was a Cochlear Implant User!!" Conference on Implantable Auditory Prostheses, Lake Tahoe, California, USA. Poster. doi: 10.6084/m9.figshare.871408.

Gray H. (1908). Anatomy of the human body (W. H. Lewis, Ed.) (Lea & Febiger, Philadelphia).

Patterson R.D., Gaudrain E., Walters T.C. (2010). "The Perception of Family and Register in Musical Notes." In M. R. Jones, R. R. Fay, and A. N. Popper (Eds.), Music Perception, Springer Handbook of Auditory Research (Springer), 1st Edition., Vol. 36, pp. 13–50.

Smith D.R.R., Patterson R.D. (2005). "The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age." J. Acoust. Soc. Am. 118, 3177–86.