Barycentre spectral et perception de la hauteur de la voix et de la longueur de tractus vocal chez les normo-entendants et les utilisateurs d'implants cochléaires

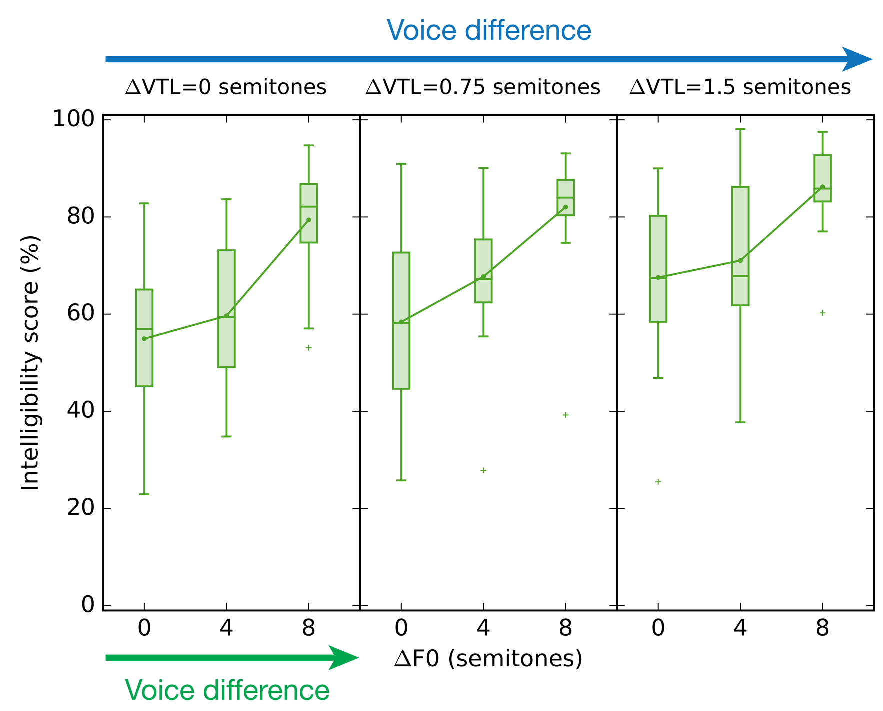
Etienne Gaudrain

Using models to evaluate whether spectral centroid could play a role in F0 and VTL perception in acoustic and electric hearing

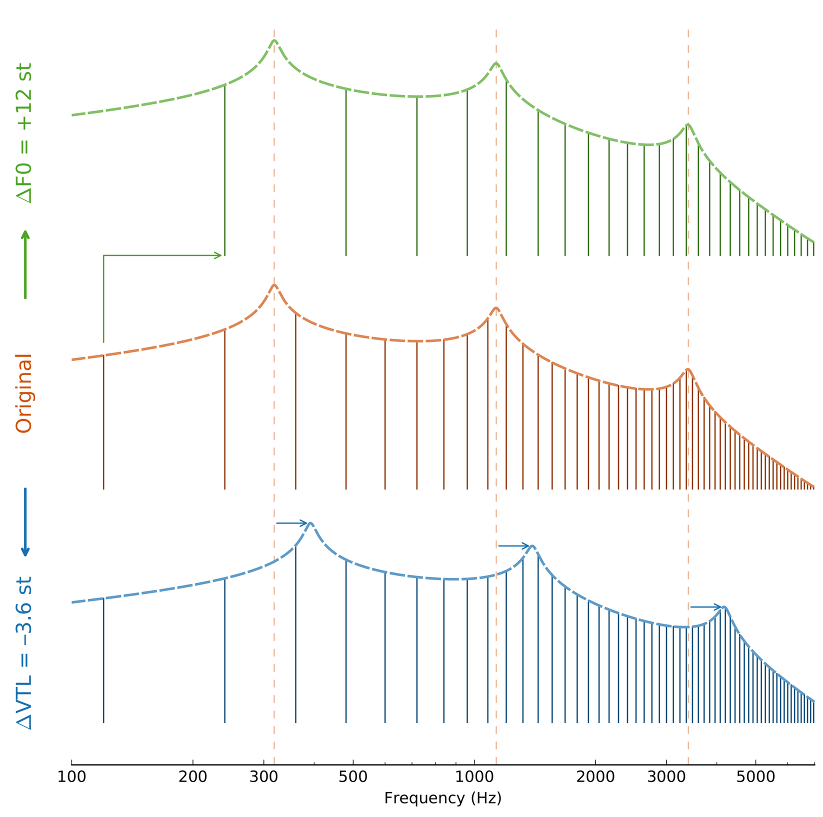
Etienne Gaudrain <etienne.gaudrain@cnrs.fr> Lyon Neuroscience Research Center, CNRS, Université Lyon 1, France | University Medical Center Groningen, University of Groningen, Netherlands

Introduction

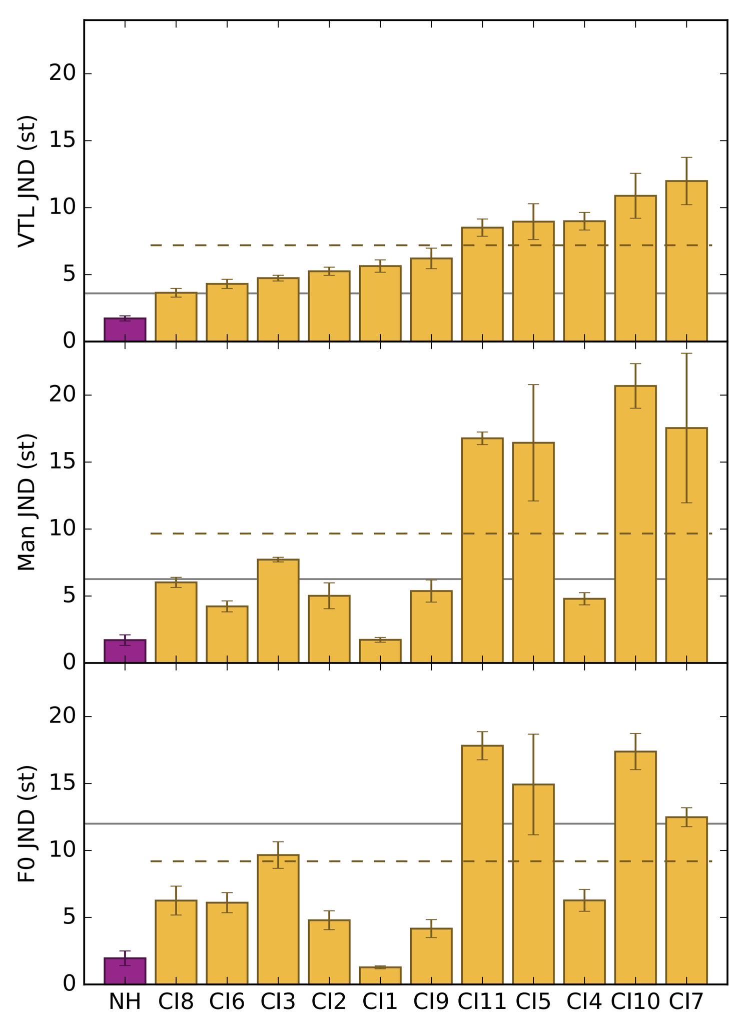
Voice differences are useful for listeners to discriminate competing sentences (Brungart, 2001, Darwin et al., 2003, Başkent and Gaudrain, 2016; see Figure 1). The perception and discrimination of voices relies on two principal cues: the fundamental frequency (F0) and vocal-tract length (VTL). F0 is related to voice pitch. VTL is related to the size of the speaker and affects the spectral envelope such that all the formants are shifted by the same ratio when VTL is changed (Figure 2).



▲ Figure 1 — Effect of voice difference on speech-on-speech intelligibility (adapted from Başkent and Gaudrain, 2016).



▲ Figure 2 — Schematic spectrum of an idealized vowel (vertical bars) and the spectral envelope (dashed line), for different voice manipulations.



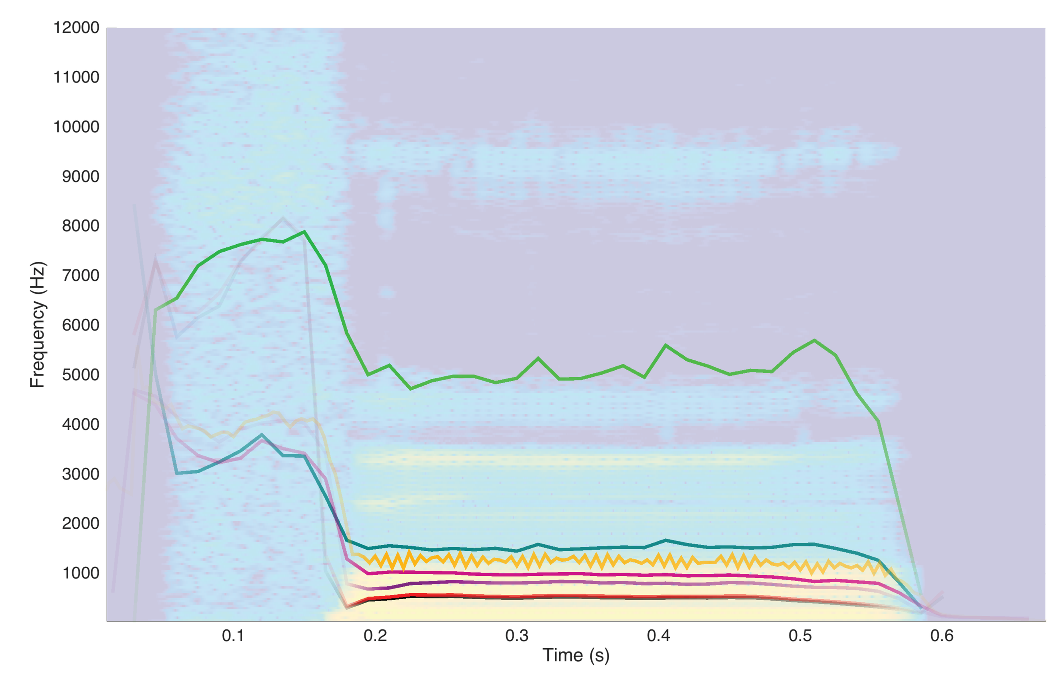
▲ Figure 3 — VTL (top) and F0 (bottom) just-noticeable differences (JNDs) for NH (purple) and CI (yellow) listeners. The dashed line shows the average for CI listeners (from Gaudrain and Başkent, 2015).

While normal-hearing (NH) listeners have good access to these cues, cochlear-implant (CI) users show much enlarged discrimination thresholds along these dimensions (Figure 3). This deficit, especially along the VTL dimension, is likely responsible for difficulties in identifying speaker gender (Massida et al., 2013, Fuller et al., 2014) and in taking advantage of voice differences in competing voices (Stickney et al., 2004, 2007).

Yet, the underlying mechanisms of VTL perception remain largely unknown. A first theory is that VTL requires phonological knowledge as it occurs at a processing stage that follows phonetic categorization. In this theory, VTL perception would be achieved during an interactive speaker normalization process where linguistic and indexical cues interact. Some evidence supporting this theory is provided by studies showing that listeners are better at recognising speakers in their own native language (e.g. Köster and Schiller, 1997), or by studies showing that speaker adaptation is a dynamic process (e.g. Ladefoged and Broadbent, 1957). A second theory is that VTL perception is independent of language (pre-categorical; Sjerps et al., 2013). This theory is particularly appealing to psychoacousticians as, according to the acoustics terminology, VTL should be a component of timbre, independent of language. As such, it has been suggested that VTL perception could rely on the spectral centroid (or spectral center of gravity), which is a common factor in multidimensional scaling of timbre perception (e.g. Marozeau et al., 2003).

The objective of the present study is to determine whether spectral centroid could be a plausible cue for VTL perception, first in NH listeners, and then in CI listeners. While VTL undoubtedly affects the spectral centroid, the very nature of the speech signal also warrants that the spectral centroid varies with phonetic content. Over the course of a natural utterance, the observer is thus exposed to not a unique spectral centroid value, but a whole distribution. Using various definitions of spectral centroid, the width of these distributions was evaluated in order to estimate whether the JNDs reported in Figure 3 could be explained, according to a signal-detection theory model.

Methods



▲ Figure 4 — Spectrogram of the syllable /fa/ with spectral centroid time courses superimposed. See text for different colours.

The following methods were implemented:

- linear frequency, power
- log frequency, power
- linear frequency, dB
- log frequency, dB
- gammatone, power (gtfb)
- neural activity pattern (hcl)
- Marozeau et al., 2003

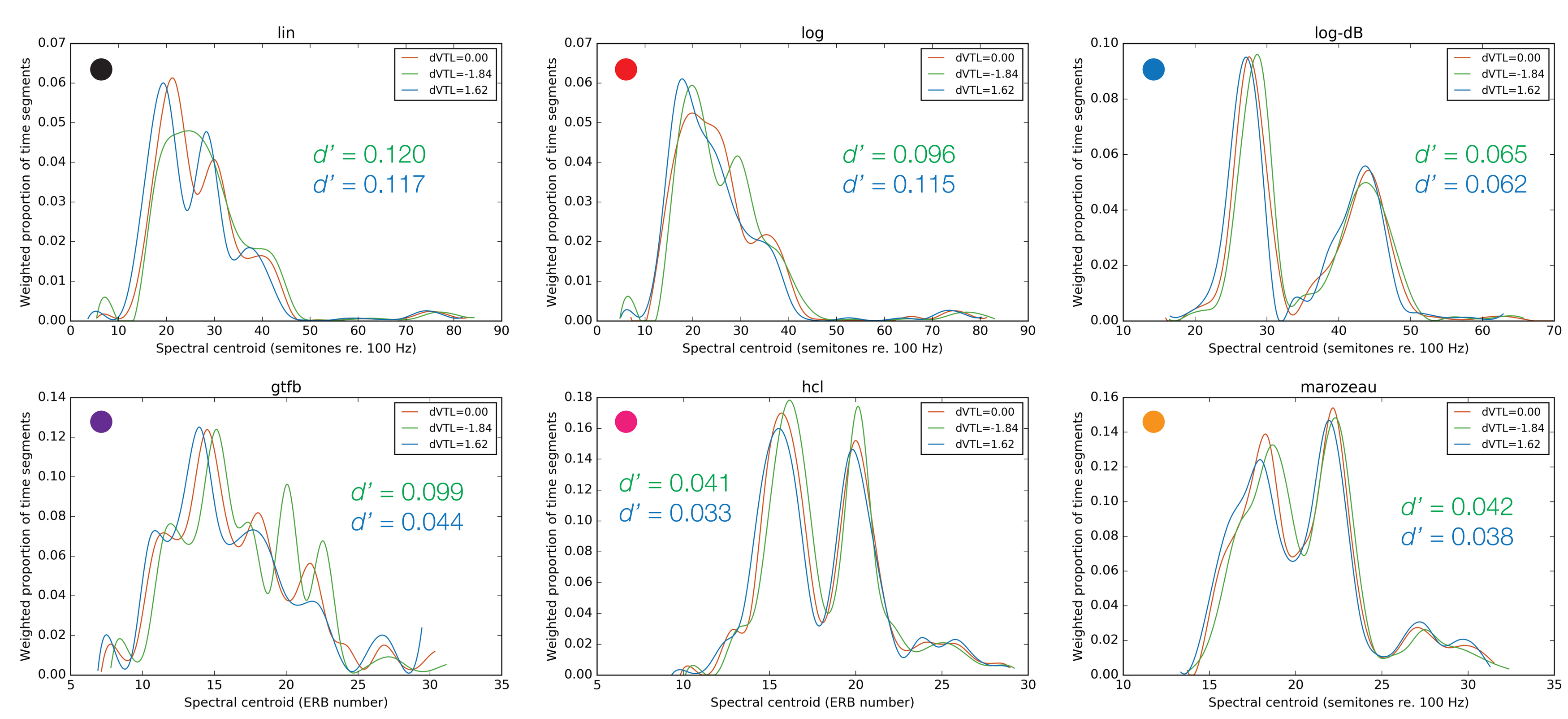
Except for the last one, all centroids were calculated every 30 ms.

The spectral centroid was calculated using 7 different methods. Spectral centroid is defined by the formula on the right.

The weights w and the frequency f can be expressed in various units that lead to different centroid values. For each method, distributions were obtained over the corpus of 61 Dutch CV-syllables used to obtain the JNDs shown in Figure 3.

$$c = \frac{\sum w \cdot f}{\sum w}$$

Results



▲ Figure 5 — Spectral centroid distributions for various calculation methods. In each panel, the red curve shows the distribution of the original corpus, while the green and blue curves show the distribution for negative and positive VTL differences corresponding to the NH JNDs (corresponding to a perceptual d' of 1). The corresponding d' estimates are shown in each panel. F0 differences did not lead to any discernable distribution difference.

▼ Table 1 — d' estimates for a 22-channel vocoder simulation and the CI JNDs. The spectral centroid was calculated using the channel envelopes, either in linear power, or in dBs.

Centroid type	ΔVTL = – 7.19 st	ΔVTL = 7.19 st
Channel number, power	0.313	0.257
Channel number, dB	0.235	0.163

Centroid type	ΔF0 = –9.19 st	ΔF0 = 9.19 st
Channel number, power	0.150	0.464
Channel number, dB	0.118	0.333

The d' estimates indicate the spectral centroid is unlikely to contribute to the observed JNDs in NH listeners. However, the cue could be used by CI listerners for both F0 and VTL discrimination tasks.

References

Başkent, D., and Gaudrain, E. (2016). “Musician advantage for speech-on-speech perception,” J. Acoust. Soc. Am., 139, EL51–EL56. doi:10.1121/1.4942628

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

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. doi:10.1121/1.1616924

Fuller, C. D., 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. doi:10.1007/s10162-014-0483-7

Gaudrain, E., and Başkent, D. (2015). Discrimination of vocal characteristics in cochlear implants . Poster presented at the 38th Annual Mid-winter Meeting of the Association for Research in Otolaryngology.

Köster, O., and Schiller, N. O. (1997). “Different influences of the native language of a listener on speaker recognition,” Int. J. Speech Lang. Law, 4, 18–28. doi:10.1558/ijsll.v4i1.18

Ladefoged, P., and Broadbent, D. E. (1957). “Information Conveyed by Vowels,” J. Acoust. Soc. Am., 29, 98–104. doi:10.1121/1.1908694

Marozeau, J., de Cheveigné, A., McAdams, S., and Winsberg, S. (2003). “The dependency of timbre on fundamental frequency,” J. Acoust. Soc. Am., 114, 2946–2957. doi:10.1121/1.1618239

Massida, Z., Marx, M., Belin, P., James, C., Fraysse, B., Barone, P., and Deguine, O. (2013). “Gender categorization in cochlear implant users,” J. Speech Lang. Hear. Res., 56, 1389–1401. doi:10.1044/1092-4388(2013/12-0132)

Sjerps, M. J., McQueen, J. M., and Mitterer, H. (2013). “Evidence for precategorical extrinsic vowel normalization,” Atten. Percept. Psychophys., 75, 576–587. doi:10.3758/s13414-012-0408-7

Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F.-G. (2007). “Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences,” J. Acoust. Soc. Am., 122, 1069–1078. doi:10.1121/1.2750159

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.

Many thanks to Deniz Başkent for her valuable comments. This work was conducted in the framework of the LabEx CeLyA (“Centre Lyonnais d’Acoustique”, ANR-10-LABX-0060/ANR-11-IDEX-0007) operated by the French National Research Agency, and is also part of the research program of the Otorhinolaryngoly Department of the University Medical Center Groningen: Healthy Aging and Communication.