

University of Groningen

## The discrimination of voice cues in simulations of bimodal electro-acoustic cochlear-implant hearing

Başkent, Deniz; Luckmann, Annika; Ceha, Jessy; Gaudrain, Etienne; Tamati, Terrin N

*Published in:*  
Journal of the Acoustical Society of America

*DOI:*  
[10.1121/1.5034171](https://doi.org/10.1121/1.5034171)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2018

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Başkent, D., Luckmann, A., Ceha, J., Gaudrain, E., & Tamati, T. N. (2018). The discrimination of voice cues in simulations of bimodal electro-acoustic cochlear-implant hearing. *Journal of the Acoustical Society of America*, 143(4), [EL292]. <https://doi.org/10.1121/1.5034171>

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# The discrimination of voice cues in simulations of bimodal electro-acoustic cochlear-implant hearing

Deniz Başkent, Annika Luckmann, Jessy Ceha, Etienne Gaudrain, and Terrin N. Tamati

Citation: *The Journal of the Acoustical Society of America* **143**, EL292 (2018); doi: 10.1121/1.5034171

View online: <https://doi.org/10.1121/1.5034171>

View Table of Contents: <http://asa.scitation.org/toc/jas/143/4>

Published by the [Acoustical Society of America](#)

---

## Articles you may be interested in

[Arrays of rectangular subcritical speech bands: Intelligibility improved by noise-vocoding and expanding to critical bandwidths](#)

*The Journal of the Acoustical Society of America* **143**, EL305 (2018); 10.1121/1.5034170

[Speaking rhythmically improves speech recognition under “cocktail-party” conditions](#)

*The Journal of the Acoustical Society of America* **143**, EL255 (2018); 10.1121/1.5030518

[Musician effect on perception of spectro-temporally degraded speech, vocal emotion, and music in young adolescents](#)

*The Journal of the Acoustical Society of America* **143**, EL311 (2018); 10.1121/1.5034489

[Song recorded near a super-group of humpback whales on a mid-latitude feeding ground off South Africa](#)

*The Journal of the Acoustical Society of America* **143**, EL298 (2018); 10.1121/1.5032126

[Unknown beaked whale echolocation signals recorded off eastern New Zealand](#)

*The Journal of the Acoustical Society of America* **143**, EL285 (2018); 10.1121/1.5032127

[Comparison of passive acoustic soniferous fish monitoring with supervised and unsupervised approaches](#)

*The Journal of the Acoustical Society of America* **143**, EL278 (2018); 10.1121/1.5034169

---

# The discrimination of voice cues in simulations of bimodal electro-acoustic cochlear-implant hearing

Deniz Başkent,<sup>a),b)</sup> Annika Luckmann,<sup>b)</sup> Jessy Ceha,<sup>c)</sup>  
Etienne Gaudrain,<sup>d)</sup> and Terrin N. Tamati<sup>b)</sup>

Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen,  
University Medical Center Groningen, Groningen, The Netherlands  
*d.baskent@umcg.nl, annikaluckmann@gmail.com, jessy.ceha@gmail.com,*  
*etienne.gaudrain@cns.fr, t.n.tamati@umcg.nl*

**Abstract:** In discriminating speakers' voices, normal-hearing individuals effectively use two vocal characteristics, vocal pitch (related to fundamental frequency,  $F_0$ ) and vocal-tract length (VTL, related to speaker size). Typical cochlear-implant users show poor perception of these cues. However, in implant users with low-frequency residual acoustic hearing, this bimodal electro-acoustic stimulation may provide additional voice-related cues, such as low-numbered harmonics and formants, which could improve  $F_0$ /VTL perception. In acoustic noise-vocoder simulations, where added low-pass filtered speech simulated residual hearing, a strong bimodal benefit was observed for  $F_0$  perception. No bimodal benefit was observed for VTL, which seems to mainly rely on vocoder spectral resolution.

© 2018 Acoustical Society of America

[QJF]

**Date Received:** February 2, 2018    **Date Accepted:** April 5, 2018

## 1. Introduction

Speaker-specific acoustic characteristics (Abercrombie, 1967) are important for identifying a speaker (e.g., their gender), and can greatly contribute to speech communication (e.g., in cocktail-party listening). In normal hearing, two orthogonal voice cues seem to provide crucial information for speaker discrimination; the vocal pitch (related to fundamental frequency,  $F_0$ ), and vocal-tract length (VTL, related to the size of the speaker; Smith and Patterson, 2005).

In users of cochlear implants (CIs), a deficit in  $F_0$  perception is routinely observed (for a review, see Moore and Carlyon, 2005), as a result of the spectro-temporal degradations in the transmitted signal (for a review, see Başkent *et al.*, 2016). Recently, a direct manipulation of  $F_0$  and VTL that maintained other speaker-related acoustic characteristics intact showed a more severe deficiency for VTL perception than  $F_0$  perception, both in CI users (Fuller *et al.*, 2014; Meister *et al.*, 2016; Gaudrain and Başkent, 2018; Zaltz *et al.*, 2018) and acoustic CI simulations (Gaudrain and Başkent, 2015). Thus, overall voice perception in CI users seems to be poorer than previously shown. In bimodal CI users, most commonly with a CI in one ear and acoustic hearing in the contralateral ear (Blamey *et al.*, 2015), the combined electro-acoustic hearing may improve  $F_0$  perception (Marx *et al.*, 2015). Yet no study has systematically investigated perception of  $F_0$  and VTL in bimodal hearing, limiting our knowledge on potential benefits from bimodal stimulation on voice perception in CI users.

In this study, we have systematically investigated perception of  $F_0$  and VTL in acoustic noise-vocoder simulations of electric hearing (vocoder-only; simulating the CI electric hearing only), also with added low-frequency acoustic information (bimodal; simulating electro-acoustic hearing). Previous studies have shown that adding even very limited low-frequency acoustic information to vocoded speech provides usable  $F_0$  cues (Brown and Bacon, 2009) or information about the first formant (Verschuur *et al.*, 2013), and increasing spectral resolution of the vocoded speech improves VTL

---

<sup>a)</sup> Author to whom correspondence should be addressed.

<sup>b)</sup> Also at: Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, University of Groningen, Groningen, The Netherlands.

<sup>c)</sup> Also at: Department of Artificial Intelligence, University of Groningen, Groningen, The Netherlands.

<sup>d)</sup> Also at: CNRS UMR5292, Inserm UMRS 1028, Lyon Neuroscience Research Center, Université Claude Bernard Lyon 1, Université de Lyon, Lyon, France.

perception (Gaudrain and Başkent, 2015). What remains unknown is if these two manipulations provide complementary voice information, leading to a beneficial, but perhaps differential, bimodal effect for  $F_0$  and VTL perception.

## 2. Methods

$F_0$  and VTL perception was measured in just-noticeable-differences (JNDs), using the same stimuli, voice manipulation, vocoding, and adaptive procedure as Gaudrain and Başkent (2015).

### 2.1 Participants

Sixteen (12 female; 19–28 yrs, average 21.7 yrs, standard deviation 2.5 yrs) native Dutch speakers, with hearing thresholds  $\leq 20$  dB hearing level at frequencies between 125 and 8000 Hz, participated in the study.

### 2.2 Stimuli

Sixty-one consonant-vowel (CV) syllables, uttered by a Dutch female speaker, were spliced from meaningful short Dutch words, taken from the Nederlandse Vereniging voor Audiologie (NVA) corpus (Bosman and Smoorenburg, 1995). Before each JND measurement, a preview was provided for the upcoming condition, using the sentence (“We kunnen weer even vooruit”; English “For now, we can continue”), uttered by a female speaker, and taken from the “Vrij Universiteit” corpus (Versfeld et al., 2000). All stimuli were presented at 62 dB sound pressure level in a sound attenuated booth, via HD600 headphones (Sennheiser, Wedemark, Germany), AudioFire4 sound card (Echo, Santa Barbara, CA), and DA10 D/A converter (Lavry Engineering, Rollingbay, WA) through S/PDIF output.

*Voice manipulation.*  $F_0$  and VTL were directly manipulated to re-synthesize a range of voices from a single speaker (e.g., Fuller et al., 2014). Using STRAIGHT (Kawahara and Irino, 2005), implemented in MATLAB, different voices along a continuum, going from a reference voice similar to that of the original female speaker to that of a typical male speaker, were created by artificially lowering  $F_0$  and simulating a longer VTL. For the reference voice, the average  $F_0$  of each CV-syllable was set to 242 Hz, the overall average  $F_0$  for the original speaker in the NVA corpus.  $F_0$  was manipulated by directly modifying the  $F_0$  contours by a number of semitones (st). VTL was manipulated by applying a “spectral envelope ratio” (also expressed in st) inversely proportional to the ratio between new and original VTL values.

*Low-frequency acoustic hearing (LPF-only).* The low-pass filtered (LPF) speech (Fig. 1) was produced using a sixth order, zero-phase, low-pass filter, at two different cutoff frequencies. One cutoff frequency was set at 150 Hz, complementing the vocoder frequency range (150–7000 Hz). The other was set at the slightly larger value of 300 Hz, covering a range of frequencies that most bimodal CI users would have access to (Gantz et al., 2016, and also within our own patient population, Clarke, 2017). The

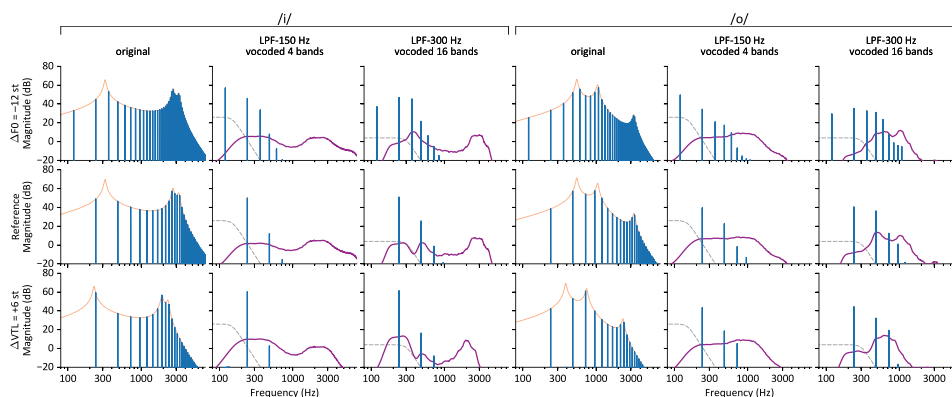


Fig. 1. (Color online) Schematic spectra of artificial vowels /i/ (left panel) and /o/ (right panel) with various manipulations. The vowels were generated for illustration purposes using typical formant values in the NVA corpus. For each vowel, the spectral envelope (first column, continuous orange line) was first generated and used to modulate harmonic frequencies (vertical blue lines). The middle row corresponds to the original voice, while the top row corresponds to a decrease in  $F_0$  of 12 st, and the bottom row corresponds to an increase in VTL of 6 st (resulting in a shift of 6 st of all formants toward the low frequencies). For each vowel, the second column shows the 4-band noise-vocoded signal (continuous purple line) and the LPF signal with a cutoff of 150 Hz (vertical blue lines). The frequency response of the filter used for the LPF signal, including the chosen gain, is displayed as a dashed gray line. The third column is the same for the 16-band noise-vocoded signal and the 300 Hz LPF signal.

level of the LPF signal was calibrated using the reference female voice and adjusted for each cutoff frequency so that the average intensity was identical to that of the vocoded signal. However, the level was not further adjusted to compensate the variations in intensity in the LPF signal that resulted from  $F_0$  or VTL manipulations. That led the acoustic signal to be amplified by 26 dB in the 150 Hz LPF condition, and 4 dB in the 300 Hz LPF condition. With these amplifications and with the relatively shallow filter slope used, LPF cues remained audible well beyond the cutoff frequency in a way that might be similar to actual bimodal CI users. With such spectral profile it was expected that not only  $F_0$ , but also the lower end of the spectral envelope containing the first formant for at least some of the vowels, would be audible (Fig. 1).

*Vocoder (vocoder-only)*. A noise-vocoder was implemented with 4, 8, or 16 spectral bands, within the frequency range of 150 to 7000 Hz (Greenwood, 1990), and with parameters based on previous studies (Bingabr et al., 2008; Gaudrain and Başkent, 2015). Filtering stimuli and white noise with eighth order, zero-phase, bandpass filters produced analysis and synthesis (carrier) bands, respectively. The temporal envelope in each analysis band was extracted using half-wave rectification and a fourth order, zero-phase, low-pass filter. A relatively low cutoff frequency of 50 Hz was used to minimize temporal pitch cues in the envelope. The synthesis carrier bands were first modulated with envelopes, then summed to produce the vocoded stimulus, and finally adjusted to the same level as the unprocessed stimulus. See Fig. 1 for spectra of two vocoded sample vowels.

*Bimodal (vocoder-only + LPF-only)*. To represent the most common configuration of bimodal listeners, the vocoded stimuli and the LPF speech were always presented contralaterally to the right and left ears, respectively.

### 2.3 Experimental procedure

$F_0$  JNDs and VTL JNDs were each measured in 12 conditions (unprocessed,  $2 \times$  LPF-only,  $3 \times$  vocoder-only,  $6 \times$  bimodal), with two repetitions, resulting in 48 measurements in total. The order of the test conditions was randomized per participant, and testing was completed in 2–3 sessions, with a maximum duration of 2.5 h per session.

In each condition, first the preview sentence was presented, once unprocessed, and once modified with the upcoming condition's parameters. Following, the JND was measured using processed stimuli in a 3-interval 3-alternative forced choice (3AFC) 2 down-1 up adaptive procedure (based on Gaudrain and Başkent, 2015, 2018). For each trial, three syllables, randomly chosen from 61 spliced CV syllables, were concatenated with 50 ms of silence in-between, to form a triplet. The participants were presented with this triplet three times, separated by 200 ms, with one triplet modified in voice cue and randomly assigned to one of the three presentation intervals. Each threshold measurement started with the deviant triplet differing from the standard triplets by 12 st in  $F_0$  or VTL. The initial step size was 2 st. The step size was divided by  $\sqrt{2}$ , if 15 trials elapsed with the same step size, or when the difference between the stimuli became smaller than 2 times the step size. The JND measurement ended after eight reversals and the JND was calculated as the mean of the last six reversals.

## 3. Results

Figures 2 and 3 show the  $F_0$  and VTL JNDs, respectively, for unprocessed, LPF-only, vocoder-only, and combined bimodal conditions. The unprocessed condition produced the baseline JNDs,  $F_0$  JND =  $0.99 \pm 0.60$  st and VTL JND =  $1.45 \pm 0.95$  st. The effect of each manipulation of LPF and vocoding alone was investigated by comparing JNDs in LPF-only and vocoder-only conditions, respectively, to an unprocessed condition, using paired comparison  $t$ -tests (with Bonferroni correction  $\alpha = 0.05/5 = 0.01$ ). In LPF-only, LPF did not significantly affect  $F_0$  JNDs (Fig. 2, left side; 150 Hz:  $1.38 \pm 1.16$  st,  $t(15) = 1.68$ ,  $p = 0.114$ ; 300 Hz:  $1.34 \pm 1.05$  st;  $t(15) = 1.71$ ,  $p = 0.109$ ), but significantly increased VTL JNDs (Fig. 3, left side; 150 Hz:  $14.33 \pm 3.51$  st,  $t(15) = 15.32$ ,  $p < 0.001$ ; 300 Hz:  $11.72 \pm 4.11$  st;  $t(15) = 9.80$ ,  $p < 0.001$ ). In vocoder-only, vocoding significantly increased both  $F_0$  JNDs (Fig. 2, right side; 4 bands:  $21.18 \pm 2.20$  st;  $t(15) = 37.72$ ,  $p < 0.001$ ; 8 bands:  $22.00 \pm 2.58$  st;  $t(15) = 3.03$ ,  $p < 0.001$ ; 16 bands:  $14.33 \pm 6.23$  st;  $t(15) = 9.09$ ,  $p < 0.001$ ) and VTL JNDs (Fig. 3, right side; 4 bands:  $8.53 \pm 2.96$  st,  $t(15) = 11.81$ ,  $p < 0.001$ ; 8 bands:  $4.25 \pm 1.47$  st;  $t(15) = 8.00$ ,  $p < 0.001$ ; 16 bands:  $3.13 \pm 1.95$  st;  $t(15) = 4.07$ ,  $p = 0.001$ ).

Next, we focus on the JNDs with bimodal conditions, the main interest of the study. They were analyzed with a repeated-measures analysis of variance (ANOVA) with the main within-subject factors of number of bands in the vocoder (3 levels; 4, 8, 16) and the cutoff frequency of the added LPF (3 levels; none, 150 Hz, 300 Hz), along with the

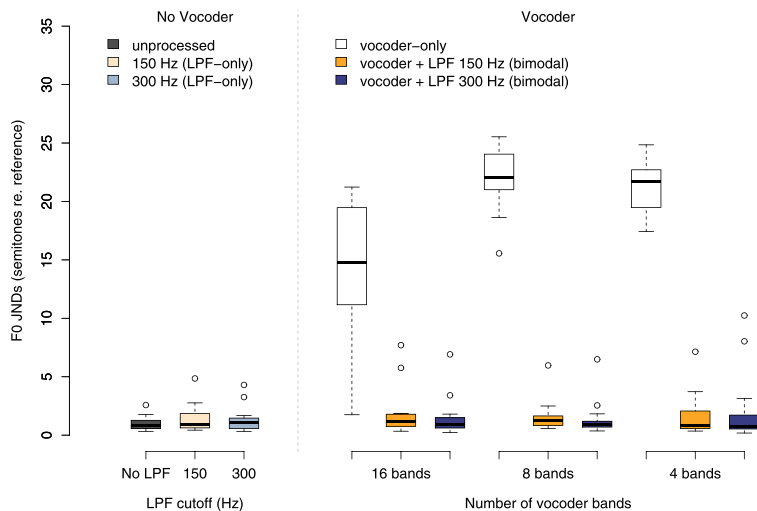


Fig. 2. (Color online) Average  $F_0$  JNDs shown for each condition. On the left: No Vocoder conditions (unprocessed, 150 Hz LPF-only, and 300 Hz LPF-only). On the right: Vocoder conditions (16, 8, and 4 bands), as vocoder-only, or combined with LPF (bimodal). The boxes extend from the lower to the upper quartile (the interquartile range, IQ), and the midline indicates the median. The whiskers indicate the highest and lowest values no greater than 1.5 times the IQ, and the dots indicate the outliers, i.e., data points larger than 1.5 times the IQ.

generalized eta-squared ( $\eta_G^2$ ) measure of effect size. The Greenhouse-Geisser correction was applied when the sphericity assumption was violated. The bimodal benefit *per se*, namely the change in JNDs as a result of adding LPF speech to vocoder-only speech, was investigated with planned paired comparison *t*-tests (using the False Discovery Rate, FDR, correction), by comparing each bimodal condition (150 and 300 Hz) to the corresponding vocoder-only condition for each number of bands (4, 8, and 16 bands).

For  $F_0$  JNDs (Fig. 2, right side), the ANOVA showed significant main effects of number of bands in the vocoder [ $F(2,30) = 27.49$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.15$ ] and added LPF cutoff [ $F(2,30) = 644.98$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.90$ ], with a significant interaction [ $F(4,60) = 24.84$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.26$ ]. The *t*-tests revealed significant differences between bimodal and vocoder-only conditions, for both LPF cutoffs and for all three vocoder bands [all  $p_{\text{FDR}}$ 's  $< 0.001$ ], but not between the two bimodal conditions of 150 and 300 Hz [all  $p_{\text{FDR}}$ 's  $> 0.630$ ]. Hence, there was a bimodal benefit for  $F_0$  perception, and  $F_0$  JNDs were mostly determined by LPF.

For VTL JNDs (Fig. 3, right side), the ANOVA showed a significant main effect of number of bands in the vocoder [ $F(2,30) = 111.77$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.52$ ], but no significant effect of added LPF cutoff [ $F(2,30) = 2.48$ ,  $p = 0.101$ ,  $\eta_G^2 = 0.17$ ], and no significant interaction [ $F(4,60) = 1.27$ ,  $p = 0.294$ ,  $\eta_G^2 = 0.12$ ]. The *t*-tests indicated that there was no effect of adding LPF on VTL JNDs in any bimodal condition [all  $p_{\text{FDR}}$ 's  $> 0.129$ ]. Hence, there was no bimodal benefit for VTL perception, and VTL JNDs were mostly determined by the spectral resolution in the vocoder.

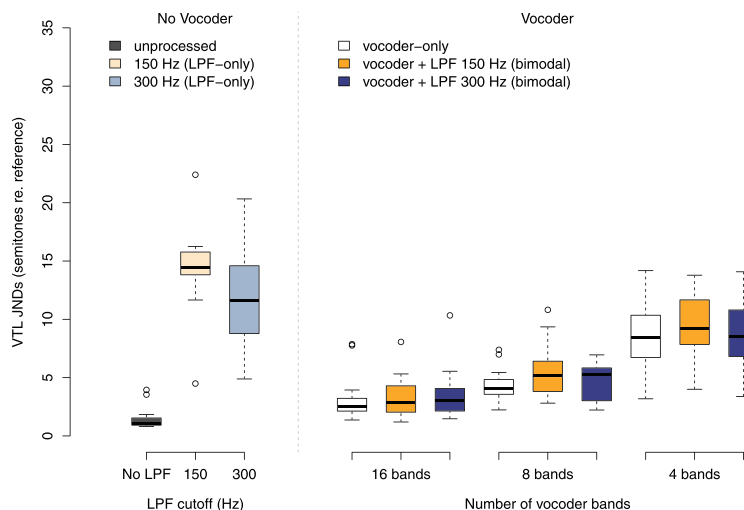


Fig. 3. (Color online) Same as Fig. 2, except average VTL JNDs shown for each condition.

#### 4. Discussion

The overall aim of this study was to systematically investigate, using acoustic simulations of CIs, how the LPF and vocoder manipulations would affect  $F0$  and VTL perception, and whether combining the two manipulations would lead to a bimodal benefit for  $F0$  and VTL perception.

The smallest  $F0$  and VTL JNDs were observed in the unprocessed condition (similar values to [Gaudrain and Başkent, 2015](#)). The first manipulation, LPF (LPF-only), had differing effects on  $F0$  and VTL JNDs, likely as a result of acoustic cues needed for the perception of each voice cue. LPF with either cutoff frequency did not change  $F0$  JND compared to the unprocessed condition, suggesting low frequencies provided sufficient  $F0$  information to do the task (in line with [Brown and Bacon, 2009](#)). The lack of a differential effect between the two LPF-only conditions perhaps came from the adjusted intensity levels. For the 300 Hz condition, with the reference voice average  $F0$ , 242 Hz, the first harmonic would lay well within the LPF range, but not necessarily for the 150 Hz condition. It is likely, then, that the extra 22 dB amplification applied in the 150 Hz condition compared to the 300 Hz condition in order to equate intensity somewhat compensated for the effect of the varying LPF cutoff. In contrast to  $F0$  JNDs, LPF made VTL JNDs significantly larger compared to the unprocessed condition, implying that VTL cues were more broadly distributed across a wider frequency range. The second manipulation, vocoding (vocoder-only) made both  $F0$  and VTL JNDs larger, indicating this manipulation made relevant acoustic voice cues poorer (in line with [Gaudrain and Başkent, 2015](#)).

The potential bimodal benefit as a result of the combination of the two manipulations, namely adding LPF-only to vocoder-only speech, was the main interest of the study. There was a bimodal advantage for  $F0$  JNDs, consistent with previous findings with bimodal CI users ([Marx \*et al.\*, 2015](#)).  $F0$  perception was affected by the spectral resolution of the vocoder when there was no LPF added (also in line with [Gaudrain and Başkent, 2015](#)). However, when LPF was added, the added acoustic cues seemed to be so strong that good  $F0$  JNDs were achieved, regardless of degraded spectral resolution. In contrast, there was no bimodal advantage for VTL JNDs, indicating that LPF speech did not provide sufficiently salient VTL cues. In the previously mentioned study by [Fuller \*et al.\* \(2014\)](#), the difference between a typical male and female voice in VTL was 3.6 st. The results in [Fig. 3](#) indicate that this differentiation is possible by using at least 16 bands in the vocoder (with or without added LPF), but not for lower spectral resolution. Thus, the residual low-frequency acoustic hearing or a hearing aid, if limited to very low frequencies, may not improve perception of VTL or its use for other tasks, such as gender categorization, if the spectral resolution in the speech delivered from the CI is also limited, but may improve perception of  $F0$  regardless of spectral resolution.

Interestingly, the bimodal benefit in  $F0$  perception seemed to be mainly due to the LPF speech, rather than an interactive bimodal effect. Similarly, VTL perception seemed to be mainly explained by the vocoder spectral resolution, without an effect of added LPF speech. This observation perhaps implies that  $F0$  and VTL perception rely on different cues and perceptual mechanisms. Alternatively, the lack of interaction may also be specific to the present study, affected by the cutoff values chosen for LPF and vocoder envelope filters. These values were on purpose selected to be relatively limited, to both simulate low-frequency hearing available to most bimodal implant users, as well as to truly reduce the spectral cues in LPF and temporal cues in vocoder envelopes, to be able to better observe any potential complementary bimodal effect. In addition, the use of a female voice—with a relatively high  $F0$ —as reference in the 3AFC task may also have limited the possibility to observe a bimodal benefit for VTL. A male voice, with both a lower  $F0$  and with lower frequency formants, might have a better chance at retaining some of its defining characteristics through the LPF. That the study already showed a bimodal improvement in  $F0$  perception despite such limited parameters may be a promising outcome for actual bimodal CI users. Especially for those users with more residual hearing or envelope information transmission capabilities than simulated here, the temporal and spectral cues may overlap more, leading to stronger bimodal benefits in voice perception.

#### Acknowledgments

The research was funded in part by Research School of Behavioral and Cognitive Neuroscience (BCN-Brain), University of Groningen, and a VICI Grant (Grant No. 918-17-603), and a VENI Grant (Grant No. 275-89-035) 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 the Otorhinolaryngology Department of the University Medical Center Groningen: Healthy Aging and Communication, as well as 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.

### References and links

- Abercrombie, D. (1967). *Elements of General Phonetics* (Edinburgh University Press, Edinburgh).
- Başkent, D., Gaudrain, E., Tamati, T. N., and Wagner, A. (2016). “Perception and psychoacoustics of speech in cochlear implant users,” in *Scientific Foundations of Audiology Perspectives from Physics, Biology, Modeling, and Medicine*, edited by A. T. Cacace, E. de Kleine, A. Genene Holt, and P. van Dijk (Plural Publishing, San Diego, CA), Chap. 12.
- 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.
- Blamey, P. J., Maat, B., Başkent, D., Mawman, D., Burke, E., Dillier, N., Beynon, A., Kleine-Punte, A., Govaerts, P. J., Skarzynski, P. H., Huber, A. M., Sterkers-Artières, F., Van de Heyning, P., O’Leary, S., Fraysse, B., Green, K., Sterkers, O., Venail, F., Skarzynski, H., Vincent, C., Truy, E., Dowell, R., Bergeron, F., and Lazard, D. S. (2015). “A retrospective multicenter study comparing speech perception outcomes for bilateral implantation and bimodal rehabilitation,” *Ear Hear.* **36**, 408–416.
- Bosman, A., and Smoorenburg, G. (1995). “Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment,” *Int. J. Audiol.* **34**, 260–284.
- Brown, C., and Bacon, S. (2009). “Low-frequency speech cues and simulated electric-acoustic hearing,” *J. Acoust. Soc. Am.* **125**, 1658–1665.
- Clarke, J. (2017). “The Pitch Hunt: The role of vocal characteristics in top-down repair of interrupted speech,” Ph.D. thesis, University of Groningen, Groningen, The Netherlands.
- Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J. I., Fu, Q.-J., Free, R., and Başkent, D. (2014). “Gender categorization is abnormal in cochlear implant users,” *J. Assoc. Res. Otolaryngol.* **15**, 1037–1048.
- Gantz, B. J., Dunn, C., Oleson, J., Hansen, M., Parkinson, A., and Turner, C. (2016). “Multicenter clinical trial of the Nucleus Hybrid S8 cochlear implant: Final outcomes,” *Laryngoscope* **126**, 962–973.
- Gaudrain, E., and Başkent, D. (2015). “Factors limiting vocal-tract length discrimination in cochlear implant simulations,” *J. Acoust. Soc. Am.* **137**, 1298–1308.
- Gaudrain, E., and Başkent, D. (2018). “Discrimination of voice pitch and vocal-tract length in cochlear implant users,” *Ear Hear.* **39**, 226–237.
- Greenwood, D. D. (1990). “A cochlear frequency-position function for several species—29 years later,” *J. Acoust. Soc. Am.* **87**, 2592–2605.
- Kawahara, H., and Irino, T. (2005). “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. Divenyi (Kluwer Academic, The Netherlands), pp. 167–180.
- Marx, M., James, C., Foxtan, J., Capber, A., Fraysse, B., Barone, P., and Deguine, O. (2015). “Speech prosody perception in cochlear implant users with and without residual hearing,” *Ear Hear.* **36**, 239–248.
- Meister, H., Fürsen, K., Streicher, B., Lang-Roth, R., and Walger, M. (2016). “The use of voice cues for speaker gender recognition in cochlear implant recipients,” *J. Speech Lang. Hear. Res.* **59**, 546–556.
- Moore, B. C. J., and Carlyon, R. P. (2005). “Perception of pitch by people with cochlear hearing loss and by cochlear implant users,” in *Pitch Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, New York), pp. 234–277.
- Smith, D. R. R., and 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**, 3177–3186.
- Verschuur, C., Boland, C., Frost, E., and Constable, J. (2013). “The role of first formant information in simulated electro-acoustic hearing,” *J. Acoust. Soc. Am.* **133**, 4279–4289.
- Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (2000). “Method for the selection of sentence materials for efficient measurement of the speech reception threshold,” *J. Acoust. Soc. Am.* **107**, 1671–1684.
- Zaltz, Y., Goldsworthy, R. L., Kishon-Rabin, L., and Eisenberg, L. (2018). “Voice discrimination by adults with cochlear implants: The benefits of early implantation for vocal-tract length perception.” *J. Assoc. Res. Otolaryngol* **19**, 193–209.