DOI: 10.61782/fa.2023.0837

forum acusticum 2023

Speech-on-speech perception in adult cochlear implant users: Benefits from voice differences

Etienne Gaudrain[1,2] **Eleanor E. Harding**[2,3,4*] **Robert L. Harris**[2,4]
Barbara Tillmann[1] **Bert Maat**[2,5] **Rolien H. Free**[2,5] **Deniz Başkent**[2,3]

[1] Lyon Neuroscience Research Center, CNRS UMR5292, Inserm U1028, Université Lyon 1, Lyon, France

[2] Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[3] Graduate School of Medical Sciences, Research School of Behavioural and Cognitive Neurosciences, University of Groningen, Groningen, The Netherlands

[4] Prins Claus Conservatoire, Hanze University of Applied Sciences, Groningen, The Netherlands

[5] Cochlear Implant Center Northern Netherlands, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

ABSTRACT[*]

Speech-on-speech perception remains a challenge for many cochlear implant (CI) users. In normal-hearing listeners, differences in voice characteristics such as F0 and vocal-tract length (VTL), which indicate age, sex and size of the speaker, are known to support the segregation of competing voices. However, previous research indicated that postlingual adult CI users tend to derive little or no benefit from voice differences, whereas pediatric CI users do. The present study aims to shed some light on this discrepancy by comparing individual voice-difference benefits in speech-on-speech perception to the performance in other voice perception tasks in adult CI listeners. Speech-on-speech perception was evaluated using a coordinate response measure (CRM) paradigm, where participants identify a number and color in a target speech stream competing with a gibberish speech masker. The target voice was female and the masker voice was either identical, or the F0 and VTL were altered parametrically to create voices that sounded progressively more male. Individual performance in this task was compared to F0 and VTL discrimination thresholds as well as voice-gender categorization performance in the same participants. Preliminary results indicate only a weak association suggesting that speech-on-speech perception depends on more than voice segregation alone.

Keywords: *cochlear implants, speech-on-speech perception, vocal characteristics, coordinate response measure.*

1. INTRODUCTION

A cochlear implant (CI) restores partial hearing to deaf individuals (e.g. [1]). Speech-on-speech perception — i.e. target speech presented against a single talker masker — remains difficult for CI users [2]. Differences in vocal characteristics can help discriminate the target from the masker. Fundamental frequency (F0), which is related to voice pitch, and vocal-tract length (VTL), which is correlated with the height of the speaker, have been shown to be efficient cues for stream segregation in normal-hearing (NH) listeners [3], [4].

Results on whether CI users can utilize these voice cues have been mixed, notably because of large variability in the data [5]–[7]. Using a closed-set test such as the coordinate response measure (CRM; [8], [9]) may help reduce this variability [10].

Therefore, the current ongoing exploratory study is assessing voice-difference benefits for speech-on-speech perception in CI listeners at various target-to-masker ratios (TMRs). We further correlated CRM performance with F0 and VTL just-noticeable differences (JNDs), and voice

[*]***Corresponding author***: *e.e.harding@rug.nl*

gender identification performance that varied voice stimuli along F0 and VTL dimensions.

2. METHODS

Twenty-seven Dutch-speaking CI adult users performed a CRM test where the participant had to identify the correct color and number (Figure 1) indicated by the target sentence in competing gibberish speech masker. The participants' age ranged from 21 to 82 (mean: 62.5 years), and they were users of all three CI brands: MED-EL, Cochlear and Advanced Bionics (AB). Two participants were bilaterally implanted, 10 were bimodal users (one ear CI, other ear hearing aid), and 16 were unilateral CI users. Participants were tested with loudspeakers and asked to use their CI daily listening settings together with their usual non-implanted ear/hearing aids, if applicable (i.e., the non-implanted ear was not plugged).

The CRM corpus is made of sentences of the form "Show the dog where the [color] [number] is." (Fig. 1). The target speaker was always the same female voice. The gibberish masker was created by concatenating random segments of sentences from the same corpus [11]. The masker voice was constructed from the target female voice, by shifting the F0 and VTL by a number of semitones [(ΔF0,ΔVTL): (0,0); (-6,+1.8); and (-12,+3.6)], which correspond to a total voice change of 0, 6.3 and 12.5 st respectively.

The TMR was set to 0, +6, or +12 dB. A condition without masker (in quiet) was also presented. The test was conducted in person using an online interface. The voice manipulations and gibberish masker generation were performed on the fly using the WORLD vocoder [12] via a VTServer [13].

The same participants also performed a JND task, which was an adaptive three-interval three-alternative forced choice (3I3AFC) where voice cue differences started at 12 st and progressively decreased or increased depending on the participant's responses using a 2-down, 2-up rule [14]. The stimuli were triplets of consonant-vowel syllables. In each trial, while the syllables remained the same across the presentation intervals, the voice cue of one of the triplets chosen randomly was manipulated to differ from the two others. The procedure ended after 8 reversals, and the JND was calculated as the geometric mean over the last 6 reversals.

Finally, the participants also performed a voice gender categorization task [15]. Four words uttered by a female speaker were used. In each trial the participant heard one of these words where the F0 and VTL cues had been altered to various degrees in the direction of a male voice, and had to judge whether the voice was more male or female.

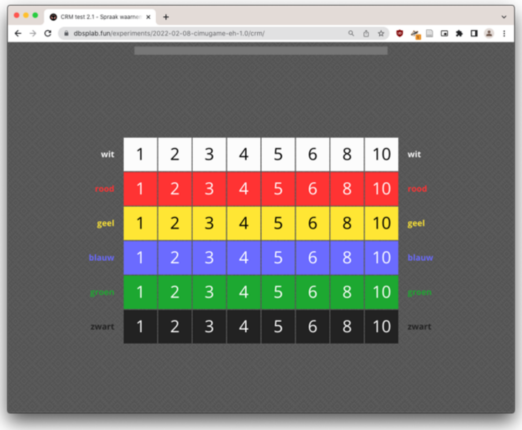

Figure 1. Online interface of the CRM test, in Dutch. The digits 7 ("zeven") and 9 ("negen") have been excluded because they contain two syllables. The test thus has 8 digits and 6 colors. After hearing the sentence "Laat de hond zien waar de [color] [number] is", the participants had to click on the corresponding cell in the matrix.

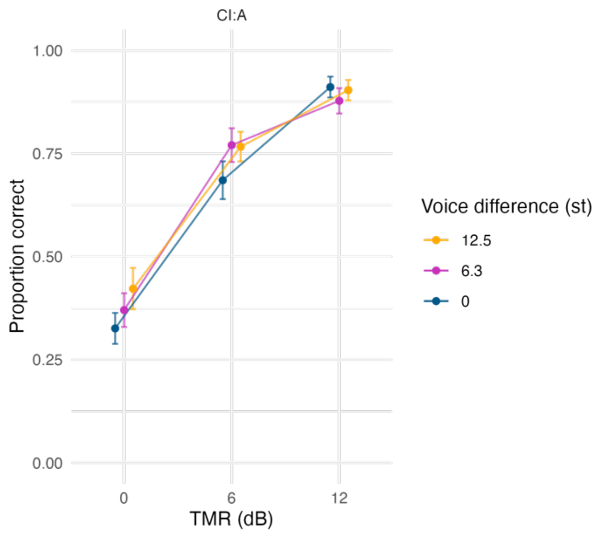

Figure 2. Average proportion correct (both color and number) across participants as a function of TMR (x-axis) and voice difference (color).

3. RESULTS

One participant could not perform the task, even in the absence of masker, and was excluded from all analyses. The results were analyzed using a generalized mixed model (gLMM; binomial distribution, logit link function):

```
score ~ tmr * voice + ( 1 | participant )
```

The model showed a significant effect of TMR (Figure 2, [$\chi 2(1) = 441$, $p<0.001$]), as well as a significant effect of voice difference (Figure 3, [$\chi 2(1)=8.53$, $p<0.01$]), but no significant interaction [$\chi 2(1)=1.72$, $p=0.19$]; see Fig. 3.

Despite the interaction not reaching significance, taking TMRs individually, the voice difference had no effect for +12 dB TMR, but did provide significant benefit for lower TMRs.

CRM performance was associated with VTL JNDs (TMR 0, $R = -.55$, TMR 6, $R = -.61$; TMR 12, $R = -.48$; $p_{FDR}<0.05$ for all TMRs), while F0 JNDs showed no significant correlation with CRM performance ($p_{FDR}>0.63$ for all TMRs; see Figure 4). For voice gender categorization, however, no significant correlation was observed at any of the TMRs and for any of the cues ($p_{FDR}>0.33$; see Figure 5).

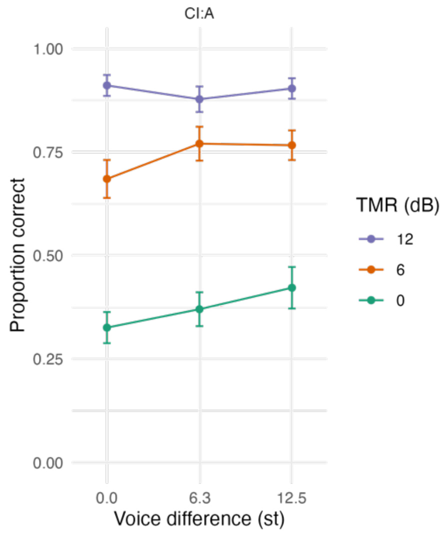

Figure 3. Average proportion correct across participants as a function of voice difference (x-axis) and TMR (color).

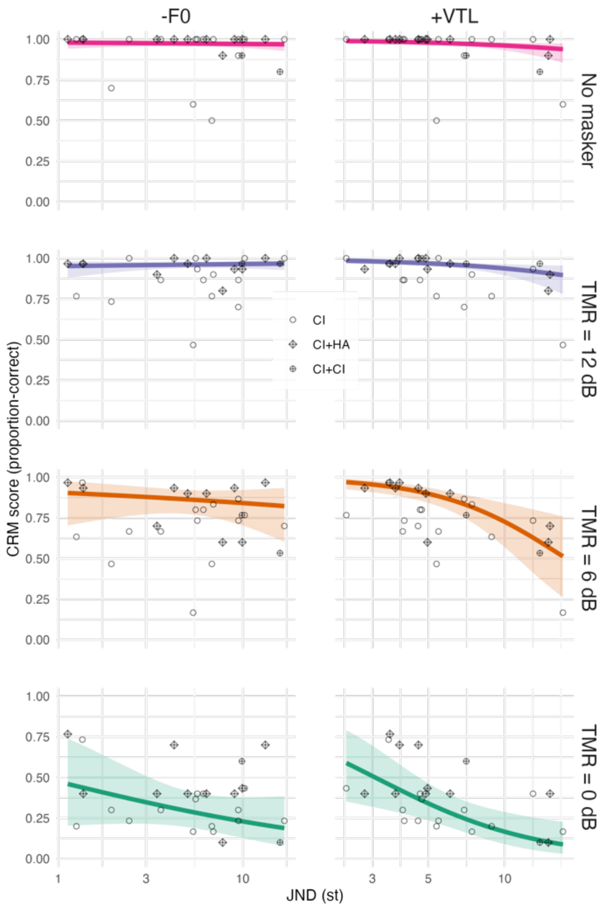

Figure 4. Relation between the CRM scores (averaged across voice conditions) and the voice JNDs.

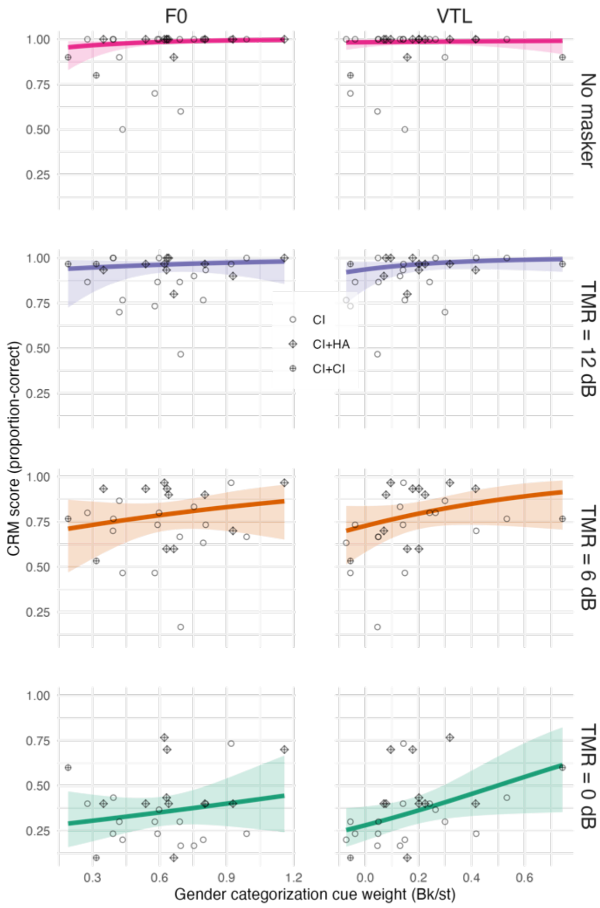

Figure 5. Relation between the CRM scores (averaged across voice conditions) and the weight given to each voice cue in the voice gender categorization task.

4. DISCUSSION

While the voice-difference benefit observed in the present study with adult CI listeners is small, it is clearly significant even at TMRs as low as 0 dB, where performance could be discouraging for the participants.

Inter-individual variability remains large despite the close-set nature of the test. This is consistent with previous reports of large heterogeneity among CI users [16] and may also be due in part to the wide age range (spanning 61 years) of participants. However, for most participants there is a combination of TMR and voice difference that seems beneficial. This brings the question of how to assess individual performance variations, e.g. after training. Moreover, the current dataset contained 2 bilateral and 10 bimodal CI users, possibly further contributing to the spread in the data.

A statistical relationship was found between VTL JNDs and speech-on-speech performance. This could be an indication that speech intelligibility and VTL discrimination rely on partially overlapping mechanisms (e.g. spectral profile discrimination). On the other hand, this could also suggest that improving VTL JND thresholds, for example via a perceptual training, might improve multiple aspects of speech hearing including speech-on-speech perception.

5. ACKNOWLEDGMENTS

We would like to thank Lisa Hoogeveen, Mirthe Ronde and Isa Poortman for assistance in data collection. We further thank Esther Steenbergen-Bethlehem for administrative support.

This study was funded by the Dorhout Mees Foundation, the Gratama Foundation, Heinsus Houbolt Funds, and the VICI Grant No. 016.VICI.170.111 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw). The team Auditory Cognition and Psychoacoustics of the Lyon Neuroscience Research Center is part of the LabEx CeLyA (Centre Lyonnais d'Acoustique, ANR-10-LABX-60).

6. REFERENCES

[1] O. Macherey and R. P. Carlyon, "Cochlear implants," *Current Biology*, vol. 24, no. 18, pp. R878–R884, 2014, doi: 10.1016/j.cub.2014.06.053.

[2] D. Başkent, E. Gaudrain, T. Tamati, and A. E. Wagner, "Perception and psychoacoustics of speech in cochlear implant users," in *Scientific Foundations of Audiology: Perspectives from Physics, Biology, Modeling, and Medicine*, A. T. Cacace, E. de Kleine, A. G. Holt, and P. van Dijk, Eds., Plural Publishing, 2016.

[3] D. Başkent and E. Gaudrain, "Perception and Psychoacoustics of Speech in Cochlear Implant Users," *Scientific Foundations of Audiology. Perspectives from Physics, Biology, Modelling, and Medicine*, pp. 185–320, 2016.

[4] C. J. Darwin, D. S. Brungart, and B. D. Simpson, "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," *J. Acoust. Soc. Am.*, vol. 114, no. 5, p. 2913, 2003, doi: 10.1121/1.1616924.

[5] N. El Boghdady, E. Gaudrain, and D. Başkent, "Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users?," *The Journal of the Acoustical Society of America*, vol. 145, no. 1, pp. 417–439, Jan. 2019, doi: 10.1121/1.5087693.

[6] N. El Boghdady, F. Langner, E. Gaudrain, D. Başkent, and W. Nogueira, "Effect of spectral contrast enhancement on speech-on-speech intelligibility and voice cue sensitivity in cochlear implant users," *Ear Hear.*, vol. 42, no. 2, pp. 271–289, Mar. 2021, doi: 10.1097/AUD.0000000000000936.

[7] W. Nogueira, N. El Boghdady, F. Langner, E. Gaudrain, and D. Başkent, "Effect of channel interaction on vocal cue perception in cochlear implant users," *Trends Hear.*, vol. 25, p. 233121652110301, Jan. 2021, doi: 10.1177/23312165211030166.

[8] R. S. Bolia, W. T. Nelson, M. A. Ericson, and B. D. Simpson, "A speech corpus for multitalker communications research," *The Journal of the Acoustical Society of America*, vol. 107, no. 2, pp. 1065–1066, Feb. 2000, doi: 10.1121/1.428288.

[9] V. Hazan, S. Messaoud-Galusi, S. Rosen, S. Nouwens, and B. Shakespeare, "Speech perception abilities of adults with dyslexia: is there any evidence for a true deficit?," *J Speech Lang Hear Res*, vol. 52, no. 6, pp. 1510–1529, Dec. 2009, doi: 10.1044/1092-4388(2009/08-0220).

[10] L. G. Potts, A. M. Olivo, R. M. Reeder, and J. B. Firszt, "Evaluation of the American English Matrix Test with cochlear implant recipients," *International Journal of Audiology*, pp. 1–7, Mar. 2023, doi: 10.1080/14992027.2023.2185757.

[11] L. Nagels, E. Gaudrain, D. Vickers, P. Hendriks, and D. Başkent, "School-age children benefit from voice gender cue differences for the perception of speech in competing speech," *The Journal of the Acoustical Society of America*, vol. 149, no. 5, pp. 3328–3344, May 2021, doi: 10.1121/10.0004791.

[12] M. Morise, F. Yokomori, and K. Ozawa, "WORLD: A vocoder-based high-quality speech synthesis system for real-time applications," *IEICE Trans. Inf. & Syst.*, vol. E99.D, no. 7, pp. 1877–1884, 2016, doi: 10.1587/transinf.2015EDP7457.

[13] E. Gaudrain, "VTServer." Zenodo, Mar. 05, 2022. doi: 10.5281/ZENODO.5801905.

[14] E. Gaudrain and D. Başkent, "Factors limiting vocal-tract length discrimination in cochlear implant simulations," *J. Acoust. Soc. Am.*, vol. 137, no. 3, pp. 1298–1308, Mar. 2015, doi: 10.1121/1.4908235.

[15] C. D. Fuller *et al.*, "Gender categorization is abnormal in cochlear-implant users," *J. Assoc. Res. Otolaryngol.*, vol. 15, no. 6, pp. 1037–1048, Aug. 2014, doi: 10.1007/s10162-014-0483-7.

[16] P. Blamey *et al.*, "Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients," *Audiology and Neurotology*, vol. 18, no. 1, pp. 36–47, 2012, doi: 10.1159/000343189.