

## Letter to the Editor

# Letter to the Editor Concerning Skuk et al., “Parameter-Specific Morphing Reveals Contributions of Timbre and Fundamental Frequency Cues to the Perception of Voice Gender and Age in Cochlear Implant Users”

Hartmut Meister,<sup>a</sup>  Katrin Fuersen,<sup>a,b</sup> Barbara Streicher,<sup>b</sup>  
Ruth Lang-Roth,<sup>b</sup> and Martin Walger<sup>a,b</sup>

**Purpose:** The purpose of this letter is to compare results by Skuk et al. (2020) with Meister et al. (2016) and to point to a potential general influence of stimulus type. **Conclusion:** Our conclusion is that presenting sentences may give cochlear implant recipients the opportunity

to use timbre cues for voice perception. This might not be the case when presenting brief and sparse stimuli such as consonant–vowel–consonant or single words, which were applied in the majority of studies.

With great interest, we read the article, “Parameter-Specific Morphing Reveals Contributions of Timbre and Fundamental Frequency Cues to the Perception of Voice Gender and Age in Cochlear Implant Users,” by Verena G. Skuk et al. (2020). This interesting work focuses on areas vital for cochlear implant (CI) rehabilitation, since the perception of voice gender and the identification of speakers based on voice characteristics is of great importance for everyday communication. Thus, knowledge in this area is essential for the further development of CI technology, such as coding strategies or rehabilitation measures like training programs.

Skuk et al. (2020) describe two experiments dedicated to the recognition of the voice gender and vocal age. In the first experiment, brief vowel–consonant–vowel (VCV) stimuli (i.e., /aba/ and /igi/) were systematically varied with respect to the acoustic characteristics, and the resulting

gender perception was assessed. Main voice characteristics are the fundamental frequency (F0) and timbre, the latter relying strongly on the position of the formant frequencies as well as the ratio of harmonic signal energy to noise. In comparison to a normal-hearing (NH) control group, whose results were based on both F0 and timbre, CI users’ gender perception almost exclusively relied on F0—a finding that confirms the results of earlier studies (e.g., Fuller et al., 2014; Meister et al., 2016). In comparison, the examination of the vocal age represents a novelty. Here, Skuk et al. varied one sentence (i.e., “Keine Antwort ist auch eine Antwort”) with respect to different acoustic characteristics, and the resulting age estimation was determined. Notably, this experiment yielded an outcome that obviously differed from the first one: At least, high-performing CI users were indeed able to use timbre cues in addition to F0. As a possible explanation, the authors suggested that “their ability to utilize acoustic parameters appears to depend on the specific social signal that defines the perceptual task” (Skuk et al., 2020, p. 3170).

Based on our own work on voice recognition, we would additionally like to point to a potential general influence of stimulus type on voice perception: In Meister et al. (2016), the speaker gender perception of NH listeners and CI users

<sup>a</sup>Jean Uhrmacher Institute, University of Cologne, Germany

<sup>b</sup>Clinic of Otorhinolaryngology, Head and Neck Surgery, University of Cologne, Germany

Correspondence to Hartmut Meister: hartmut.meister@uni-koeln.de

Editor-in-Chief: Frederick (Erick) Gallun

Editor: Christian E. Stilp

Received September 23, 2020

Accepted October 9, 2020

[https://doi.org/10.1044/2020\\_JSLHR-20-00563](https://doi.org/10.1044/2020_JSLHR-20-00563)

**Disclosure:** The authors have declared that no competing interests existed at the time of publication.

was investigated considering different types of utterances. In an approach similar to Skuk et al. (2020), F0 and timbre cues (i.e., formant frequencies based on simulated changes in vocal tract length [VTL]) were systematically modified. While the resulting perception of CI users was based almost exclusively on F0 for brief stimuli (i.e., single words), in sentences, the combined modification of F0 and VTL actually showed an improved recognition of the speaker gender compared to the isolated F0 modification (cf., Figure 1 in Meister et al., 2016). Our interpretation was that the perception of the specific voice cues depends significantly on the stimuli presented, with sentences providing a larger variety of phonetic characteristics and more suprasegmental cues. This might facilitate a more detailed analysis of both F0 and timbre, in contrast to brief and sparse stimuli such as VCV or single words.

Studies investigating voice perception in CI users by applying modifications of acoustic cues almost exclusively relied on brief syllable or word stimuli. Exceptions using sentence stimuli, which might provide a greater ecological validity, are the examinations by Skuk et al. (2020) and Meister et al. (2016). Though the acoustic characteristics in these two studies are not identical, both give evidence that timbre cues

may generally be utilized by CI recipients, albeit to a lesser degree than in NH listeners. This opens up options to further improve the perception of important vocal characteristics with CI—for example, by developing novel coding strategies.

## References

- Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R. H., & Başkent, D. (2014). Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research Otolaryngology*, 15(6), 1037–1048. <https://doi.org/10.1007/s10162-014-0483-7>
- Meister, H., Fürsen, K., Streicher, B., Lang-Roth, R., & Walger, M. (2016). The use of voice cues for speaker gender recognition in cochlear implant recipients. *Journal of Speech, Language, and Hearing Research*, 59(3), 546–556. [https://doi.org/10.1044/2015\\_JSLHR-H-15-0128](https://doi.org/10.1044/2015_JSLHR-H-15-0128)
- Skuk, V. G., Kirchen, L., Oberhoffner, T., Guntinas-Lichius, O., Dobel, C., & Schweinberger, S. R. (2020). Parameter-specific morphing reveals contributions of timbre and fundamental frequency cues to the perception of voice gender and age in cochlear implant users. *Journal of Speech, Language, and Hearing Research*, 63(9), 3155–3175. [https://doi.org/10.1044/2020\\_JSLHR-20-00026](https://doi.org/10.1044/2020_JSLHR-20-00026)

## Letter to the Editor

# The Role of Stimulus Type and Social Signal for Voice Perception in Cochlear Implant Users: Response to the Letter by Meister et al.

Stefan R. Schweinberger,<sup>a,b,c</sup>  Celina I. von Eiff,<sup>b</sup> Louisa Kirchen,<sup>b,d</sup> Tobias Oberhoffner,<sup>e,f</sup> Orlando Guntinas-Lichius,<sup>e</sup> Christian Dobel,<sup>e</sup> Christine Nussbaum,<sup>b</sup> Romi Zäske,<sup>a,b,e</sup> and Verena G. Skuk<sup>a,b,e</sup>

**Purpose:** In their letter, Meister et al. (2020) appropriately point to a potential influence of stimulus type, arguing cochlear implant (CI) users may have the ability to use timbre cues only for complex stimuli such as sentences but not for brief stimuli such as vowel–consonant–vowel or single words. While we cannot exclude this possibility on the basis of Skuk et al. (2020) alone, we hold that there is a strong need to consider type of social signal (e.g.,

gender, age, emotion, speaker identity) to assess the profile of preserved and impaired aspects of voice processing in CI users. We discuss directions for further research to systematically consider interactive effects of stimulus type and social signal. In our view, this is crucial to understand and enhance nonverbal vocal perception skills that are relevant to successful communication with a CI.

We are grateful for the opportunity to respond to the letter by Meister et al. (2020), who point out that results from two experiments on the perception of two different social signals (vocal gender and age) in our current article (Skuk et al., 2020) were obtained with different stimulus types. Based on findings reported by Meister et al. (2016), the authors direct attention to a possible influence of stimulus type. Their study had suggested that the availability of timbre/vocal tract length cues

to cochlear implant (CI) users is very limited in the case of single words but more preserved for sentences, presumably because these exhibit larger phonetic variability and more suprasegmental cues. In short, we agree that the role of stimulus type for the use of acoustic cues in voice perception should be considered. Although the focus of our article was different, we see why this might have deserved further discussion by Skuk et al. (2020). We appreciate this opportunity to offer our comments.

Voice perception research should use naturalistic and ecologically valid material where possible, and this is particularly obvious when the aim is to understand and improve human communication in the context of hearing impairment. In tune with Meister et al.'s (2020) letter, our own work on speaker identification with normal-hearing listeners studied the relationship between complexity/duration of an utterance and identification performance (Schweinberger et al., 1997). The resulting function provided strong reasons for using complex sentence stimuli of at least 1.5–2 s duration for speaker identification tasks, because this duration matched the point at which performance reached an asymptote toward ceiling. We recently explicated these recommendations for research on speaker identification in more detail (Schweinberger & Zäske, 2019).

<sup>a</sup>DFG Research Unit Person Perception, Friedrich Schiller University of Jena, Germany

<sup>b</sup>Department for General Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University of Jena, Germany

<sup>c</sup>Swiss Center for Affective Science, Geneva, Switzerland

<sup>d</sup>Social-Pediatric Centre and Centre for Adults with Special Needs, Trier, Germany

<sup>e</sup>Department of Otorhinolaryngology, Institute of Phoniatry and Pedaudiology, Jena University Hospital, Germany

<sup>f</sup>Department of Otorhinolaryngology, Head and Neck Surgery “Otto Körner,” University Medical Center Rostock, Rostock, Germany

Correspondence to Stefan R. Schweinberger:  
stefan.schweinberger@uni-jena.de

Editor-in-Chief: Frederick (Erick) Gallun  
Editor: Christian E. Stip

Received October 8, 2020

Accepted October 9, 2020

[https://doi.org/10.1044/2020\\_JSLHR-20-00595](https://doi.org/10.1044/2020_JSLHR-20-00595)

**Disclosure:** The authors have declared that no competing interests existed at the time of publication.

At the same time, we admit that we used voice gender vowel–consonant–vowel (VCV) stimuli based on our previous research (Skuk & Schweinberger, 2014), whereas we had expected that such brief stimuli would not be appropriate to study a more difficult perceptual task on speaker age. In general, we motivated our choice of stimulus types by previous research showing how (much) duration and/or additional phonetic variation contributes to performance overall for different social signals—but based on performance in normal-hearing listeners. On that basis, gender (but not age) perception is already close to ceiling with VCV stimuli. But, we agree that this is a limitation here, as it holds for original recordings and normal-hearing listeners (Skuk & Schweinberger, 2014), whereas the intriguing findings by Meister et al. (2016) suggest that CI users benefit from sentences for gender perception.

While it seems important to consider the role of stimulus type in the future, we wish to hold that the type of social signal also remains of critical importance when assessing nonverbal voice perception abilities in CI users, particularly when considering the role of these abilities for daily communication. While perception of the age or gender of a communication partner can be important, perception of emotional information in the voice may be even more relevant here: Emotional signals are important during an interaction with the same partner, and vocal emotion perception skills may be tightly related to life quality in CI users (Schorr et al., 2009). In follow-up research, we therefore now study vocal emotion perception and briefly mention one finding that seems directly relevant to the present discussion. One recent study from our lab (von Eiff et al., 2019) used single bisyllabic pseudoword stimuli (e.g., /belam/, /molen/) with emotional intonation (angry or fearful) and with similar parameter-specific morphs as in Skuk et al. (2020). With those stimuli and social signals, we found remarkably preserved use of timbre cues—but strongly impaired use of fundamental frequency (F0)—in CI users who recognized emotional intonation from single pseudowords. This pattern, obtained with brief stimuli, potentially forms a double dissociation compared with the pattern of results from vocal gender perception by Skuk et al. (2020). Even considering the limitations involved in cross-experiment comparisons, this suggests that CI users' ability to use timbre cues is not generally abolished by brief stimuli, but can depend on the social signal perceived.

Overall, we see at least two priorities as this research is further refined. First, more systematic evidence is needed to better understand the relative contributions of stimulus type and social signal to the utilization of F0 and timbre cues by CI users. This should use different stimulus types that differ systematically in phonetic variability, including simple stimuli and sentences with ecological validity. From the perspective of experimental design, it would be ideal to investigate orthogonal combinations of social cues (e.g., gender, age, emotion) and stimulus types (e.g., single word/pseudoword vs. sentence) in the same study and participants. Second, since individual differences in voice perception in CI

users are so prevalent, it would be desirable to study these more systematically in order to potentially identify individual “profiles” of abilities. Ultimately, we envisage that understanding individual differences can be helpful to designing tailor-made perceptual training programs or coding strategies. Although this will require substantial research efforts, remaining obstacles to such a research program may be of practical rather than methodological nature. In any case, we are confident that powerful methodological tools such as voice-morphing technology have improved to a stage that makes this research possible in principle (Kawahara & Skuk, 2019).

## References

- Kawahara, H., & Skuk, V. G. (2019). Voice morphing. In S. Frühholz & P. Belin (Eds.), *The Oxford handbook of voice perception* (pp. 685–706). Oxford University Press.
- Meister, H., Fuersen, K., Streicher, B., Lang-Roth, R., & Walger, M. (2016). The use of voice cues for speaker gender recognition in cochlear implant recipients. *Journal of Speech, Language, and Hearing Research, 59*(3), 546–556. [https://doi.org/10.1044/2015\\_JSLHR-H-15-0128](https://doi.org/10.1044/2015_JSLHR-H-15-0128)
- Meister, H., Fuersen, K., Streicher, B., Lang-Roth, R., & Walger, M. (2020). Letter to the editor concerning Skuk et al., “Parameter-specific morphing reveals contributions of timbre and fundamental frequency cues to the perception of voice gender and age in cochlear implant users.” *Journal of Speech, Language, and Hearing Research*. Advance online publication. [https://doi.org/10.1044/2020\\_JSLHR-20-00563](https://doi.org/10.1044/2020_JSLHR-20-00563)
- Schorr, E. A., Roth, F. P., & Fox, N. A. (2009). Quality of life for children with cochlear implants: Perceived benefits and problems and the perception of single words and emotional sounds. *Journal of Speech, Language, and Hearing Research, 52*(1), 141–152. [https://doi.org/10.1044/1092-4388\(2008\)07-0213](https://doi.org/10.1044/1092-4388(2008)07-0213)
- Schweinberger, S. R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech, Language, and Hearing Research, 40*(2), 453–463. <https://doi.org/10.1044/jslhr.4002.453>
- Schweinberger, S. R., & Zäske, R. (2019). Perceiving speaker identity from the voice. In S. Frühholz & P. Belin (Eds.), *The Oxford handbook of voice perception* (pp. 539–560). Oxford University Press.
- Skuk, V. G., Kirchen, L., Oberhoffner, T., Guntinas-Lichius, O., Döbel, C., & Schweinberger, S. R. (2020). Parameter-specific morphing reveals contributions of timbre and fundamental frequency cues to the perception of voice gender and age in cochlear implant users. *Journal of Speech, Language, and Hearing Research, 63*(9), 3155–3175. [https://doi.org/10.1044/2020\\_JSLHR-20-00026](https://doi.org/10.1044/2020_JSLHR-20-00026)
- Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity and spectrum level on the perception of voice gender. *Journal of Speech, Language, and Hearing Research, 57*(1), 285–296. [https://doi.org/10.1044/1092-4388\(2013\)12-0314](https://doi.org/10.1044/1092-4388(2013)12-0314)
- von Eiff, C. I., Skuk, V. G., Zäske, R., Wulf, C., Frühholz, S., Feuer, U., Guntinas-Lichius, O., & Schweinberger, S. R. (2019). *Perception of emotional expression in voices by cochlear implant users: A parameter-specific voice morphing approach*. Paper presented at the Meeting of the Experimental Psychology Society, London, England. <https://eps.ac.uk/wp-content/uploads/2018/11/EPS-January-2019-meeting-programme.pdf>