Parameter-Specific Morphing Reveals Contributions of Timbre to the Perception of Vocal Emotions in Cochlear Implant Users

Celina I. von Eiff,[1,2] Verena G. Skuk,[1–3] Romi Zäske,[1–3] Christine Nussbaum,[1,2] Sascha Frühholz,[4,5] Ute Feuer,[6] Orlando Guntinas-Lichius,[3] and Stefan R. Schweinberger[1,2]

Objectives: Research on cochlear implants (CIs) has focused on speech comprehension, with little research on perception of vocal emotions. We compared emotion perception in CI users and normal-hearing (NH) individuals, using parameter-specific voice morphing.

Design: Twenty-five CI users and 25 NH individuals (matched for age and gender) performed fearful-angry discriminations on bisyllabic pseudoword stimuli from morph continua across all acoustic parameters (Full), or across selected parameters (F0, Timbre, or Time information), with other parameters set to a noninformative intermediate level.

Results: Unsurprisingly, CI users as a group showed lower performance in vocal emotion perception overall. Importantly, while NH individuals used timbre and fundamental frequency (F0) information to equivalent degrees, CI users were far more efficient in using timbre (compared to F0) information for this task. Thus, under the conditions of this task, CIs were inefficient in conveying emotion based on F0 alone. There was enormous variability between CI users, with low performers responding close to guessing level. Echoing previous research, we found that better vocal emotion perception was associated with better quality of life ratings.

Conclusions: Some CI users can utilize timbre cues remarkably well when perceiving vocal emotions.

Key words: Cochlear implants, Emotion perception, Parameter-specific morphing, Quality of life, Timbre.

(Ear & Hearing 2022;43;1178–1188)

This article has received an OSF badge for Open Data.

[1]Department for General Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University Jena, Jena, Germany; [2]Voice Research Unit, Institute of Psychology, Friedrich Schiller University Jena, Jena, Germany; [3]Department of Otorhinolaryngology, Jena University Hospital, Jena, Germany; [4]Department of Psychology (Cognitive and Affective Neuroscience), Faculty of Arts and Social Sciences, University of Zurich, Zurich, Switzerland; [5]Department of Psychology, University of Oslo, Oslo, Norway; and [6]Cochlear Implant Rehabilitation Centre Thuringia Erfurt, Herbert Feuchte Stiftungsverbund, Germany.

INTRODUCTION

Hearing loss can be a disabling condition. Severe hearing impairment increases the risk of depression (Kim et al. 2017), is linked with cognitive decline (Ray et al. 2018), and is associated with an annual global cost of 980 billion dollars (World Health Organization 2021). Cochlear implants (CIs)—hearing prostheses designed to functionally replace damaged parts of the inner ear—are a highly successful way to treat severe hearing loss. This technology has advanced to a great extent in the last decades. Refinements in the processing strategies of the devices enabled striking improvements in transferring speech (Wilson et al. 1991) so that today, CIs can promote recovery of remarkable speech understanding abilities (Peterson et al. 2010; Jiam et al. 2017). However, CIs still show crucial limitations in transmitting paralinguistic sounds, such as music or social aspects in voices (for music perception, see Limb & Roy 2014; Thomas & Tripathi 2014). This is thought to be partially due to the limited number of stimulation electrodes. Even though the devices make use of tonotopic representation in the cochlea, only between 6–12 and 22 electrodes are used, depending on the specific CI device. In fact, most devices only cover the frequency range between 200 Hz and 7 kHz within the normal human hearing range between 20 Hz and 20 kHz (Kirtane et al. 2010; Peterson et al. 2010). Thus, CIs have a reduced spectral resolution (Wilson & Dorman 2008), impoverishing sound representation (Moore & Shannon 2009). Other variations in technology, such as the specific processing strategies of the devices (specifically, ACE versus MP3000 strategies; cf. Agrawal et al. 2012, 2013), also seem to affect CI users' ability to perceive nonverbal signals, and emotional prosodies in particular. Accordingly, CIs degrade prosodic cues because of their constraints in extraction, processing, and transmission of pitch and timbre cues (Kong et al. 2004; Galvin et al. 2007; Xu et al. 2009; Kang et al. 2010).

Underlining the limitations of CIs in transmitting paralinguistic social-communicative vocal signals, previous research suggested general deficits in CI users when perceiving *emotions* (e.g., Luo et al. 2007; Schorr et al. 2009; Agrawal et al. 2013; See et al. 2013; Volkova et al. 2013; Wiefferink et al. 2013; Jiam et al. 2017; Kim & Yoon 2018; Paquette et al. 2018; Tinnemore et al. 2018; Waaramaa et al. 2018), *age* (Skuk et al. 2020), or *gender* (e.g., Fu et al. 2004, 2005; Kovacić & Balaban 2009, 2010; Meister et al. 2009, 2016; Li & Fu 2011; Massida et al. 2013; Fuller et al. 2014a; Hazrati et al. 2015; Gaudrain & Baskent 2018; Skuk et al. 2020) in other people's voices. Actually, not only do CI users perform less well than normal-hearing (NH) individuals in perceiving nonverbal social cues, they also seem to employ different perceptual strategies: for example, while

<zdoi; 10.1097/AUD.0000000000001181>

NH individuals rely on both timbre and F0 when discriminating a speaker's sex (Skuk & Schweinberger 2014), CI users rely more on F0 alone (Massida et al. 2013; Fuller et al. 2014a; Skuk et al. 2020). In addition, a recent study indicates that CI users can efficiently use F0 cues for speaker segregation to support speech perception in multitalker situations with noise-vocoded speech maskers (Meister et al. 2020; but see also Stickney et al. 2004).

Perceiving vocal emotions is essential to the accurate understanding of other human beings' messages (Frick 1985; Scherer 1986; Banse & Scherer 1996). A large body of research suggests that the neurocognitive mechanisms for perceiving and producing vocal emotions are tightly interwoven (Frühholz & Schweinberger 2021), and combined research on perception and production is considered increasingly important in CI research (Jiam et al. 2017). Importantly, both abilities are highly relevant for daily communication, and impairments in the perception and production of vocal emotion often cause extensive ramifications on both social interactions and development (Trainor et al. 2000). Considering the relevance of vocal emotion perception, its tight connection to quality of life does not seem surprising: in fact, whereas there is only a weak relationship between life quality and speech understanding abilities in CI users (Huber 2005), perceived quality of life and vocal emotion perception skills are distinctively and positively correlated (Schorr et al. 2009; Luo et al. 2018).

Yet, despite its importance, vocal emotion perception in CI users remains relatively understudied, especially when compared to speech comprehension. Some studies suggest large interindividual differences between CI users in their ability to perceive vocal emotions—with some CI users' performance approximating the level of NH individuals (Chatterjee et al. 2015; Jiam et al. 2017). On average, NH individuals perform better in vocal emotion recognition than CI users even if CI-simulated voice stimuli are presented (Chatterjee et al. 2015; Gilbers et al. 2015). Various factors may influence interindividual differences in CI users (Jiam et al. 2017). For example, early auditory access to the variability of speech seems to be crucial to prevent deprivation in children and to promote speech intelligibility (Artières et al. 2009; Schorr et al. 2009). It is interesting that the performance of children who were congenitally deaf and early implanted was similar to that in late-implanted CI users who had experienced normal hearing early in life (Chatterjee et al. 2015). Identifying specific acoustic parameters relevant for CI users' vocal emotion recognition, Gilbers et al. (2015) reported a bias toward pitch range cues in CI users, whereas NH individuals seem to rely more on mean pitch than pitch range. Other researchers suggested that CI users may rely on tempo-information and intensity (Luo et al. 2007; Kalathottukaren et al. 2015).

In the present study, we planned to gain information on the relative impact of specific acoustic parameters on the perception of vocal emotions in CI users. We also aimed at a detailed quantification of individual differences in CI users' performance. In particular, we planned to compare CI users and NH listeners not only regarding their overall performance but also regarding their reliance on specific acoustic cues to recognize vocal emotion. To quantify the specific acoustic parameters utilized for task performance, we applied a parameter-specific voice morphing approach based on the TANDEM-STRAIGHT algorithm (Kawahara et al. 2013), extending our previous research on the perception of speaker gender and speaker age in CI users (Skuk et al. 2020). Considering the increased scientific attention to the relevance of social-communicative abilities for daily functioning, we additionally assessed the relationship between the CI users' perceived quality of life (Guyatt et al. 1993; Hinderink et al. 2000) and their ability to perceive emotional expression in voices.

Based on previous published findings regarding related topics (i.e., use of F0 and timbre cues by CI users in the context of vocal age and gender perception), we hypothesized that CI users rely more on F0 information in emotion perception, while NH individuals can efficiently use both F0 and timbre information. We furthermore hypothesized that perceived quality of life would be possibly related to vocal emotion perception in CI users.

MATERIALS AND METHODS

Participants

Twenty-five (14 female) CI users between 20 and 83 years old ($M = 61.0$, $SD = 17.0$) and 25 (14 female) individuals with NH abilities aged between 19 and 81 years old ($M = 63.6$, $SD = 16.4$), matched to CI users by age and gender, participated in this study. All CI users and 10 NH individuals were recruited locally and tested in the Cochlear Implant Rehabilitation Centre Thuringia in Erfurt, Germany. Fifteen NH individuals were tested at the Friedrich Schiller University Jena, Germany, and these participants received a small financial reimbursement to compensate them for local travel expenses. All participants were native German speakers; none reported a neurological or psychiatric diagnosis. CI users (for details see Table 1) reported no other otologic disorders and had either bilateral implants or unilateral implants together with severe to profound hearing loss in the nonimplanted ear. The NH individuals were recruited based on self-report of normal hearing, did not report any hearing disorders, and none was using a hearing aid.

Voice Stimuli

We selected all original audio recordings (sampling rate = 44.1 kHz) from a database that was similar to the one described in Frühholz et al. (2015). The database consists of recordings of eight different bisyllabic, five-letter, and phonetically balanced pseudowords, spoken by eight vocal actor portrayals (four female) in 10 different emotional expressions (neutral, anger, fear, happiness, disgust, sadness, achievement, pain, pleasure, and surprise). For the present study, we used four different pseudowords (/belam/, /namil/, /molen/, /loman/), spoken by four speakers (two female) with two different emotional expressions (fearful and angry). This subset of emotions and stimuli was chosen based on both high classification rates and low confusability between selected emotions in a pilot study, in which 10 NH raters performed a 10-alternative-forced choice task (mean correct classification 86.6%). The criterion for selection of an individual stimulus was a minimum performance of 60% correct for a given stimulus. Please note that anger and fear are both high-arousal negative emotions, while they are still characterized by systematic differences in acoustical parameters (cf. Table 2).

To create the experimental stimuli for this study, we applied a parameter-specific voice morphing approach to the selected original recordings, using the speech analysis, modification, and resynthesis framework TANDEM-STRAIGHT (Kawahara et al. 2013). TANDEM-STRAIGHT dissects a speech signal in source and filter information; STRAIGHT-based morphing generates highly naturally sounding synthesized voices (for further information, cf. Skuk & Schweinberger 2014; Kawahara & Skuk 2019). We systematically manipulated

TABLE 1. Demographic characteristics of all CI users

CI user	Performance rank	Sex	Age	Civil status	Pre-/post-deaf	Age at deafness	Age at first CI (yr)	Mode of hearing	Left CI Wear time (hr)	Left CI Manufacturer*	Left CI Processor	Right CI Wear time (hr)	Right CI Manufacturer*	Right CI Processor
11	4 (HP)	Female	61	Widowed	Post	40	60	CI-left	12–16	Cochlear	CP910	/	/	/
12	21 (LP)	Female	47	Single	Pre	0	45	CI-bi	>16	Advanced Bionics	Naida Q90	>16	Advanced Bionics	Naida Q90
13	20 (LP)	Female	76	Widowed	Post	19	70	CI-bi	12–16	MED-EL	OPUS2	12–16	MED-EL	Sonnet
14	12 (HP)	Male	68	Married	Post	40	67	CI-right	/	/	/	>16	Cochlear	CP910
15	19 (LP)	Female	56	Married	Pre	0	54	CI-bi	12–16	Cochlear	CP910	12–16	Cochlear	CP910
16	7 (HP)	Female	50	Divorced	Post	3	46	CI-bi	>16	MED-EL	OPUS2	>16	MED-EL	Sonnet
17	17 (LP)	Male	68	Married	Post	66	68	CI-bi	12–16	Cochlear	Kanso	12–16	Cochlear	Kanso
18	11 (HP)	Male	67	Married	Post	40	66	CI-left	12–16	Cochlear	CP910	/	/	/
19	6 (HP)	Female	57	Divorced	Post	12	55	CI-bi	12–16	Advanced Bionics	Naida Q90	12–16	Advanced Bionics	Naida Q90
20	10 (HP)	Female	81	Widowed	Post	65	81	CI-right	/	/	/	>16	MED-EL	Sonnet
21	1 (HP)	Male	64	Married	Post	59	63	CI-right	/	/	/	12–16	MED-EL	Sonnet
22	24 (LP)	Male	41	Single	Post	3	41	CI-right	/	/	/	8–12	Advanced Bionics	Naida Q90
23	18 (LP)	Female	69	Married	Post	40	69	CI-right	/	/	/	8–12	Cochlear	Kanso
24	15 (LP)	Male	67	Single	Post	50	53	CI-bi	12–16	MED-EL	OPUS2	0–4	MED-EL	Sonnet
25	9 (HP)	Female	76	Widowed	Post	62	74	CI-bi	12–16	MED-EL	Sonnet	12–16	MED-EL	Sonnet
26	23 (LP)	Female	51	Divorced	Post	25	50	CI-bi	12–16	Cochlear	CP910	12–16	Cochlear	CP910
27	16 (LP)	Female	80	Married	Post	40	79	CI-right	/	/	/	8–12	Cochlear	CP910
28	14 (LP)	Male	27	Single	Pre	0	25	CI-bi	12–16	MED-EL	OPUS2	12–16	MED-EL	Sonnet
29	25 (LP)	Female	36	Single	Pre	0	29	CI-bi	8–12	Cochlear	Sonnet	8–12	Cochlear	Kanso
30	13 (HP)	Male	76	Married	Post	44	75	CI-right	/	/	/	12–16	Cochlear	CP910
31	2 (HP)	Female	72	Married	Post	50	72	CI-left	12–16	Cochlear	Kanso	/	/	/
32	22 (LP)	Male	83	Married	Post	20	83	CI-left	8–12	Cochlear	CP910	/	/	/
33	8 (HP)	Male	20	Widowed	Pre	0	2	CI-bi	12–16	Advanced Bionics	Harmony	12–16	Advanced Bionics	Harmony
34	3 (HP)	Male	77	Single	Post	74	76	CI-left	4–8	MED-EL	Sonnet	/	/	/
35	5 (HP)	Female	54	Married	Post	3	53	CI-left	12–16	Advanced Bionics	Naida Q90	/	/	/
M			61.0			30.20	58.24							
SD			17.0			25.06	19.37							
Minimum			20			0	2							
Maximum			83			74	83							
N		25	25	25	25	25	25	25	25	25		25	25	25

TABLE 2. Acoustic characteristics of stimuli used as continuum endpoints

	Female speakers		Male speakers		Paired t test; anger vs. fear*	
	Anger	Fear	Anger	Fear	t(15)	p
F0 mean (in Hz)	366	288	264	205	3.69	0.002
F0 SD (in Hz)	67.1	19.0	45.6	26.8	6.50	0.001
F0 intonation (in Hz)	259	79	171	95	6.99	0.001
F0 glide (in Hz)	−60	3	−41	12	−3.08	0.007
Formant dispersion (in Hz)	910	1043	819	1074	−5.61	0.001
Alpha ratio	1.0	2.3	1.1	2.28	−11.16	0.001
HNR	12.8	21.3	7.3	19	−11.44	0.001
Duration (in ms)	885	760	731	854	0.02	0.988

All acoustical parameters were adapted from McAleer et al. (2014) and extracted using Praat software (Boersma 2018). For F0 extraction, pitch ranges were set to 170–600 Hz for female and 100–370 Hz for male stimuli. F0 intonation = $F0_{max}$–$F0_{min}$; F0 glide = $F0_{End}$–$F0_{Start}$; Formant dispersion = ratio between consecutive formant means (from F1 to F4, maximum formant frequency set to 5 kHz, window length 0.025 s); alpha ratio (a measure of the spectral slope) = ratio of mean energy within low (0–1 kHz) and high frequencies (1–5 kHz), computed from the long-term average spectrum; HNR was extracted with the cross-correlation method (mean value; time step = 0.01 s; min pitch = 75 Hz; silence threshold = 0.1, periods per window = 1.0).
*Including both male and female speakers.
HNR, harmonics-to-noise ratio.

individual acoustic parameters along a fearful-angry morph continuum, while keeping the respective other acoustic parameters constant at an intermediate 50% morph level (ML). Thus, the relative effects of specific acoustic cues on the perception of vocal emotion expression could be quantified. In the F0 morph type condition, solely the parameter F0 was varied, while the other TANDEM-STRAIGHT acoustic parameters aperiodicity (AP), formant frequencies (FF), spectrum level (SL), and Time (T) were all kept at a 50% ML. Conversely, AP, FF, and SL were considered to reflect timbre, in line with previous research (Skuk & Schweinberger 2014) and were systematically varied in the morph type condition Timbre, while F0 and T were kept constant. In the morph type condition Time, we set F0, AP, FF, and SL at a 50% ML, while only T was varied. Note that the Time (T) parameter does not only reflect overall duration but rather interpolations of individual time anchor positions in the stimuli (for more details, cf. Skuk & Schweinberger 2014; Kawahara & Skuk 2019). Finally, in the morph type condition Full, all five parameters were varied. Morphed test voices were created at six MLs in steps of 20% from 0/100% (anger/fear; equivalent to a fearful voice) to 100/0% morph (equivalent to an angry voice). Please refer to Table 2 for stimulus characteristics of the continuum endpoints.

Altogether, 384 stimuli (four speakers × four pseudowords × four morph types × six MLs) were presented in the experiment. Mean duration was 808 ms (SD = 89 ms, range: 540 to 1.017 ms).

Experimental Setting

All participants performed the experiment using the same technical equipment. This included a Lenovo ThinkPad R500 notebook with a 32-bit operating system, an Intel Core Duo Mobile processor T5870 (2.0 GHz), 800 MHz FSB, 2 MB L2-Cache, and a 39.1 cm (15.4″) TFT display. Voice stimuli were presented binaurally at a presentation level of approximately 70 dB SPL, as measured with a Brüel and Kjær Precision Sound Level Meter Type 2206, using two Logitech loudspeakers (230 V ~ 50 Hz 40 mA). All participants were tested individually in a sound-attenuated room (~4 m²). They were sitting

on a comfortable chair 1 m away from the notebook monitor. Loudspeakers were placed next to the monitor.

Procedure

Experimental sessions lasted about 60 minutes for CI users and 30 minutes for NH individuals. Both groups filled in a number of paper-and-pencil questionnaires, including a written self-report questionnaire on demographic data. CI users further answered questions regarding their personal experience with their CIs and subjective causes of hearing loss. In addition, the CI users filled in a version of the 60-item Nijmegen Cochlear Implant Questionnaire* (NCIQ; Hinderink et al. 2000) to evaluate quality of life related to hearing loss.

Subsequently, the 384 voice stimuli were presented in a computer experiment programmed with E-Prime 2.0. Unilateral CI users with only one CI were asked to turn off or remove any hearing aids in the contralateral ear for the duration of this experiment to avoid the contribution of residual hearing. Note that bilaterally-implanted CI users were not tested with each CI independently but in the bilateral condition only. While performing the experiment, each CI user was using the processor he or she usually used in daily routine.

Experimental instructions were shown on the monitor at the beginning of the experiment to avoid possible interference from the experimenter's voice. Participants were asked to listen carefully to each voice and to decide as accurately and as fast as possible whether it sounded rather fearful or angry, using the keyboard ("D" for fearful and "L" for angry, German layout). Twenty initial practice trials were presented to ensure that all instructions were fully understood. After the experimenter had reassured that the participant did not have remaining questions, experimental trials were presented in six blocks of 64 trials each. All voices were presented once in random order. Self-paced breaks were allowed after each block. A trial started with a black screen with reminders of response labels ("fearful D," "angry L") in the upper left and right corners, respectively. After 500 ms, a green fixation cross appeared for 500 ms and was replaced by a green question mark, the onset of which coincided with the onset of a voice stimulus.

For practice trials, only unambiguous fearful or angry voices (i.e., ML 0% or ML 100%) were presented and participants received automatic feedback about the accuracy of their previous response. For experimental trials, all MLs were

---

* Note that we corrected for a coding issue for one item that was present in the original publication.

included and no feedback was given. In case participants failed to respond within 3000 ms following voice offset, the words "Zu langsam! Bitte reagieren Sie schneller!" ("Too Slow! Please respond faster!") appeared for 1000 ms. Mean duration of the computer experiment was approximately 28 minutes ($M = 27.85$ minutes, $SD = 8.60$ minutes) for CI users and 33 min ($M = 32.58$ minutes, $SD = 15.46$ minutes) for NH individuals. The study was approved by the Ethics Committee of Jena University Hospital (Reference Number 5282-10/17).

RESULTS

Here, we only report results that were of primary interest for the aim of this study. Further documents (including Supplemental Figures and Tables, Analysis Scripts, and Raw Data) can be found in the associated OSF Repository (https://osf.io/8pc4y/) (http://links.lww.com/EANDH/A984).

Statistical Analysis

Statistical analyses were performed using R (R Core Team 2020). Both errors of omission (no key press; 0.68% of experimental trials) and trials with individual reaction times < 200 ms (measured from voice onset; 0.04%) were excluded from analyses. We used Epsilon corrections for heterogeneity of covariances throughout where appropriate (Huynh & Feldt 1976) but did not otherwise test for distribution assumptions for performing analyses of variance (ANOVAs) due to the remarkable robustness of ANOVAs to violations from normality (cf. Schmider et al. 2010).

Vocal Emotion Recognition Performance Is Impaired in Cochlear Implant Users Compared to Normal-Hearing Individuals

We performed an initial 4 × 6 × 2 × 2 × 2 mixed ANOVA on the proportion of "angry"-responses, with within-subject factors morph type (MType: F0, Full, Timbre, Time), ML (0%, 20%, 40%, 60%, 80%, 100%), speaker sex (SpSex: female, male), and between-subject factors listener sex (LSex: female, male) and listener group (LGroup: CI, NH). As this ANOVA did not reveal any main effects or interactions involving LSex (all $ps \geq 0.063$), we collapsed data across LSex for all subsequent analyses. It is important to note that the ANOVA showed several two- and three-way interactions involving LGroup (cf. Table 3), revealing significant differences between the CI and the NH group. Relevant three-way interactions involving LGroup included LGroup × MType × ML, $F(15,720) = 8.432$,

$p < 0.001$, $\varepsilon_{HF} = 0.703$, $\eta_p^2 = 0.149$ (Fig. 1), and LGroup × ML × SpSex, $F(5,240) = 6.208$, $p < 0.001$, $\varepsilon_{HF} = 0.898$, $\eta_p^2 = 0.115$. Please note that, as we were not particularly interested in effects of SpSex, reports of effects and interactions involving this factor only appear in Table 3 and Supplemental Material 4.1.2, 4.2, and 5 (http://links.lww.com/EANDH/A984).

As expected, and as indicated by steeper gradients of classification performance across MLs, visual inspection of Figure 1 suggests that vocal emotion recognition performance generally is much more accurate in NH individuals than in CI users.

Cochlear Implant Users Make Disproportional Use of Timbre Information

Another observation from Figure 1 is that while F0 and timbre cues appear to make virtually identical contributions to performance in normal hearers, timbre cues appear to be more efficiently processed in CI users.

To follow-up significant interactions with LGroup (cf. Table 3), we performed subsequent statistical analyses separately for both groups. First, we assessed pairwise differences between morph types within the group of CI users at a global level by performing three separate 2 × 6 repeated-measures ANOVAs with the factors MType and ML. For these, we contrasted adjacent morph types (where "adjacent" was defined by decreasing degrees of performance levels between morph types) with each other, the three pairwise comparisons involved contrasts between Full and Timbre, Timbre and F0, and F0 and Time. For all these contrasts, the ANOVAs revealed significant main effects of ML, $Fs(5,120) \geq 7.566$, $ps < 0.001$, $\varepsilon_{HF} \leq 0.801$, $\eta_p^2 \geq 0.240$, but not for MType, $Fs(1,24) \leq 1.596$, $ps \geq 0.219$, $\eta_p^2 \leq 0.062$. Importantly, interactions of MType × ML were found for the contrast between Full and Timbre, $F(5,120) = 2.979$, $p = 0.014$, $\eta_p^2 = 0.110$, Timbre and F0, $F(5,120) = 4.200$, $p = 0.005$, $\varepsilon_{HF} = 0.725$, $\eta_p^2 = 0.149$, and F0 and Time, $F(5,120) = 3.039$, $p = 0.022$, $\varepsilon_{HF} = 0.789$, $\eta_p^2 = 0.112$. Accordingly, performance was best for Full morphs, and the results demonstrated a superior performance for Timbre morphs compared to F0 morphs in CI users. Time cues were less efficient than F0 cues to solve the task. Finally, further analyses (cf. Supplemental Material 4.1.1.1, http://links.lww.com/EANDH/A984) confirmed that while effects of morph type were significant for the more extreme MLs (0%, 20%, 80%, and 100%), they were not significant for the intermediate MLs, as expected.

Timbre and F0 Cues Are Equally Efficient in Normal-Hearing Individuals

Analogous to the analysis performed for the CI users, we first assessed pairwise differences between morph types within NH individuals by computing three separate 6 × 2 repeated-measures ANOVAs with the factors MType and ML. Again, we contrasted morph types Full and Timbre, Timbre and F0, and F0 and Time. For all these contrasts, significant main effects of ML, $Fs(5,120) \geq 72.162$, $ps < 0.001$, $\eta_p^2 \geq 0.750$, were found. For the contrast between Timbre and F0, we moreover found a significant main effect of MType, $F(1,24) = 4.712$, $p = 0.040$, $\eta_p^2 = 0.164$; there was no significant main effect of MType for the contrasts between Full and Timbre and between F0 and Time, $Fs(1,24) \leq 1.311$, $ps \geq 0.263$, $\eta_p^2 \leq 0.052$. Most importantly, the ANOVAs revealed interactions of MType x ML for the contrast between Full and Timbre, $F(5,120) = 27.120$,

TABLE 3. Results of the 4 × 6 × 2 × 2 ANOVA on the proportion of "angry"-responses with the factors MType, ML, SpSex, and LGroup

Main effects and interactions	F	df	p	η_p^2	ε_{HF}
ML	185.743	5, 240	< 0.001	0.795	0.478
SpSex	11.686	1, 48	0.001	0.196	
LGroup × ML	24.909	5, 240	< 0.001	0.342	0.478
MType × ML	54.647	15, 720	< 0.001	0.532	0.703
LGroup × MType × ML	8.432	15, 720	< 0.001	0.149	0.703
LGroup × ML × SpSex	6.208	5, 240	< 0.001	0.115	0.898
MType × ML × SpSex	2.802	15, 720	< 0.001	0.055	

ANOVA, analysis of variance; df, degrees of freedom; LGroup, listener group; ML, morph level; MType, morph type; SpSex, speaker sex.
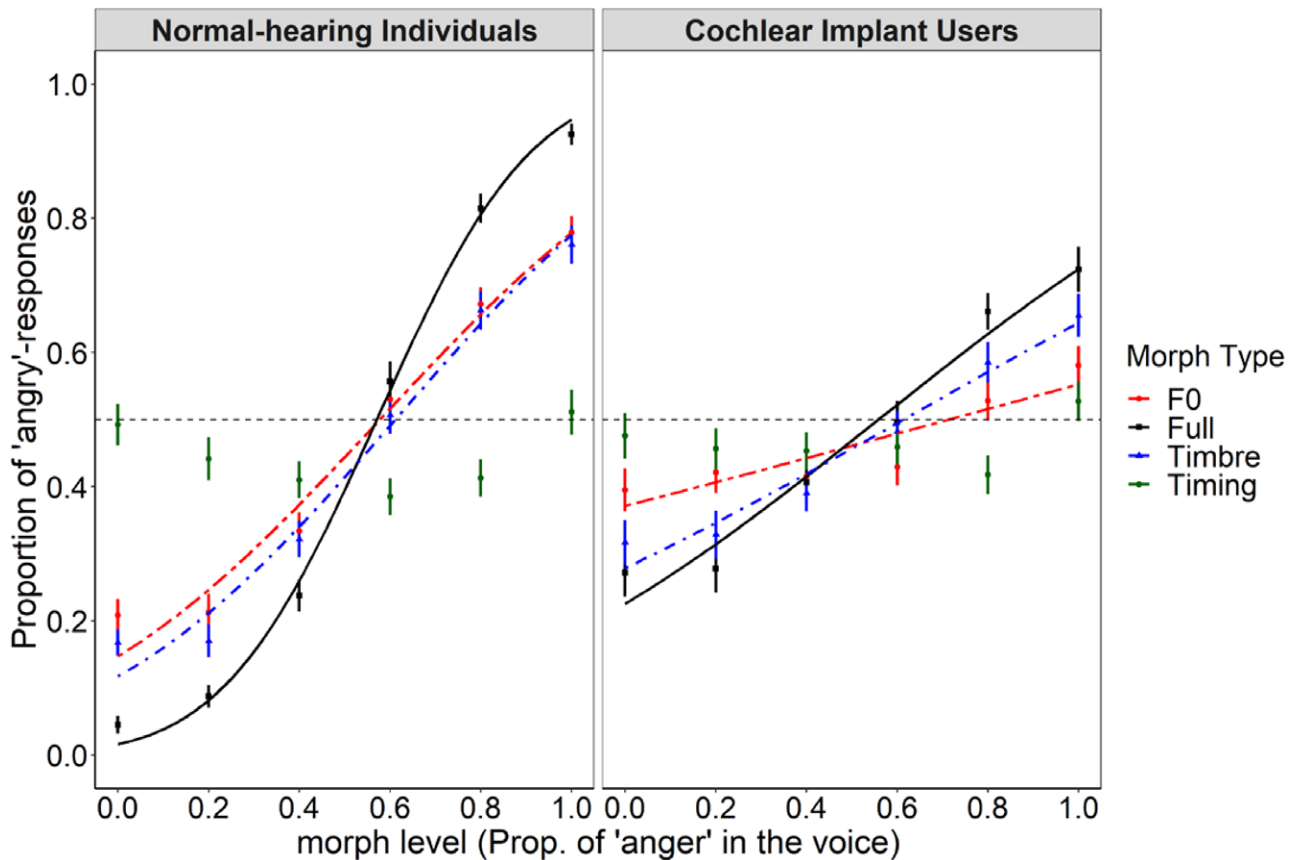
Fig. 1. The proportion of "angry"-responses for different morph levels and morph types used in the experiment, separately for normal-hearing individuals and CI users. Note that steeper slopes represent better performance. Error bars represent SEM. Best-fitting cumulative Gaussian functions are also shown. CI, cochlear implant; SEM, standard error of the mean.

$p < 0.001$, $\eta_p^2 = 0.531$, and between F0 and Time, $F(5,120) = 75.299$, $p < 0.001$, $\eta_p^2 = 0.758$, but not between Timbre and F0, $F(5,120) = 0.278$, $p = 0.924$, $\eta_p^2 = 0.011$. These results confirm the impression from Figure 1: while in CI users, timbre cues were more efficient than F0 cues to solve the task, NH individuals made equally efficient use of timbre and F0 cues. Finally, further analyses (cf. Supplemental Material 4.1.1.2, http://links.lww.com/EANDH/A984) confirmed that NH listeners´ performance for F0 and Timbre morphs did not differ significantly from each other at any ML. In addition and also at variance with the results for CI users, these analyses also indicated some sensitivity of NH listeners to Full morphs at the intermediate 40% and 60% MLs that should contain relatively ambiguous vocal emotional information only.

High-Performing Cochlear Implant Users Rely on Timbre Almost As Efficiently As Normal-Hearing Individuals Do, But Still Perform Lower When Having to Rely on F0

Since a visual inspection of the individual Gaussian fits on the proportion of "angry"-responses indicated considerable individual differences between CI users (see Supplemental Material 4.3.2.1, http://links.lww.com/EANDH/A984), the CI group was separated into two performance groups (PerfGroups) by using the median of the deviation of individual CI performance from the average performance of NH group DEV_{all} as a cutoff: the high-performing CI users (n = 13) and the low-performing CI users (n = 12).

The performance measure deviation (DEV) indicates how much a CI user's performance deviates from the average performance of the NH individuals. The smaller the DEV is for a given CI user, the more similar is her/his performance to the average performance of NH individuals. In that sense, smaller DEV scores indicate better performance (for a similar approach, see Fuller et al. 2014a; Skuk et al. 2020). For each CI user, we calculated DEV as follows: (1) For each stimulus of the experiment, we calculated how "angry" it was perceived on average across all NHs, that is, stimAngAVG. (2) Then, for each CI user and stimulus separately, we subtracted the performance of the CI user from the stimAngAVG and then took the absolute value of the result to get a difference measure for each stimulus independent of the polarity of the difference. (3) The DEV for a given CI user is then the absolute mean difference across all stimuli. We calculated DEV for all stimuli of all morph types together (DEV_{all}) and also separately for the stimuli of individual morph types (that is DEV_{Full}, DEV_{F0}, DEV_{Timbre}, DEV_{Time}).

A $4 \times 6 \times 2$ mixed ANOVA on the proportion of "angry"-responses with factors MType and ML and the between-subject factor PerfGroup (high-performing CI, low-performing CI) revealed main effects of PerfGroup, $F(1,23) = 9.316$, $p = 0.006$, $\eta_p^2 = 0.288$, and ML, $F(5,115) = 40.239$, $p < 0.001$, $\varepsilon_{HF} = 0.698$, $\eta_p^2 = 0.636$. They were qualified by several interactions (cf. Table 4) that were not post hoc tested any further, as the two CI PerfGroups were expected to differ significantly from one another.

TABLE 4. Two-way interactions of the 4 × 6 × 2 ANOVA on the proportion of "angry"-responses with the factors MType, ML, and PerfGroup, including both the high-performing and the low-performing CI users

Main effects and interactions	F	df	p	η_p^2	ε_{HF}
MType × ML	10.823	15, 345	< 0.001	0.320	0.738
PerfGroup × ML	15.066	5, 115	< 0.001	0.396	0.698

ANOVA, analysis of variance; CI, cochlear implant; df, degrees of freedom; ML, morph level; MType, morph type; PerfGroup, performance group.

However, visual inspection of Figure 2 suggests that the high-performing CI users exhibited a pattern of results similar to the NH individuals, while the low-performing CI users seemed to be close to guessing level.

To compare performance of the high-performing CI users with performance of NH individuals, we calculated a 4 × 6 × 2 × 2 mixed ANOVA on the proportion of "angry"-responses with the within-subject factors MType, ML, SpSex, and a between-subject factor PerfGroup (high-performing CI, NH). Only interested in the group differences here (refer to Table 5 for full display of interactions), we focused on the found interaction PerfGroup × MType × ML, $F(15,540) = 2.840$, $p = 0.001$, $\varepsilon_{HF} = 0.759$, $\eta_p^2 = 0.073$.

We post hoc tested this interaction by comparing each MType between high-performing CI users and NH individuals. Therefore, we separately calculated four ANOVAs, one per MType, with the factors ML and PerfGroup (PerfGroup: high-performing CI, NH). Importantly, we found no differences between high-performing CI users and NH individuals for timbre and timing (cf. Figure S8 in the Supplemental Material, http://links.lww.com/EANDH/A984), as all main effects and interactions involving PerfGroup were nonsignificant ($ps \geq 0.217$). However, some differences between high-performing CI users

and NH individuals were found for Full and F0 (cf. Figure S8 in the Supplemental Material, http://links.lww.com/EANDH/A984; refer to Table 6 for full statistics).

For F0, independent-sample two-tailed t tests revealed significant differences between high-performing CI and NH, reflecting better performance for NH listeners, for 20%, 60%, 80%, and 100% MLs, $|ts(74)| \geq 2.514$, $ps \leq 0.014$, but no differences were significant for 0% and 40% ML, $|ts(74)| \leq 1.473$, $ps \geq 0.145$. For Full, better performance for NH listeners was found for 0%, 80%, and 100% MLs, $|ts(74)| \geq 2.696$, $ps \leq 0.009$, with no significant differences for more ambiguous 20%, 40%, and 60% MLs, $|ts(74)| \leq 1.940$, $ps \geq 0.056$. In summary, the results indicate that timbre was used similarly efficiently by high-performing CI users as it was used by NH individuals. For Time, both high-performing CI users and NH individuals were close to guessing level. Full and F0, however, were used to a smaller extent by high-performing CI users than by NH individuals. Please also refer to Supplemental Material 4.3.1 (http://links.lww.com/EANDH/A984) for results of analyses of the cumulative Gaussian's slopes that reflected and, thus, supported these results.

The Cochlear Implant Users' Ability to Perceive Vocal Emotions Is Positively Correlated With Quality of Life

To explore relations between the CI users' ability to perceive vocal emotion and perceived quality of life, the performance measures DEV_{all}, DEV_{Full}, DEV_{F0}, DEV_{Timbre}, and DEV_{Time} were correlated with the scores of the NCIQ (i.e., the NCIQ total score and five subscores). Considering the directed hypothesis on the relationship between vocal emotion perception and quality of life (see Introduction), correlations were performed one-tailed. Because smaller DEV scores correspond to better performance, we expected negative correlations. As expected, DEV_{all} was negatively related to the total score of the NCIQ, $r_s = -0.390$, $p = 0.027$, $n = 25$, indicating that the CI users' ability
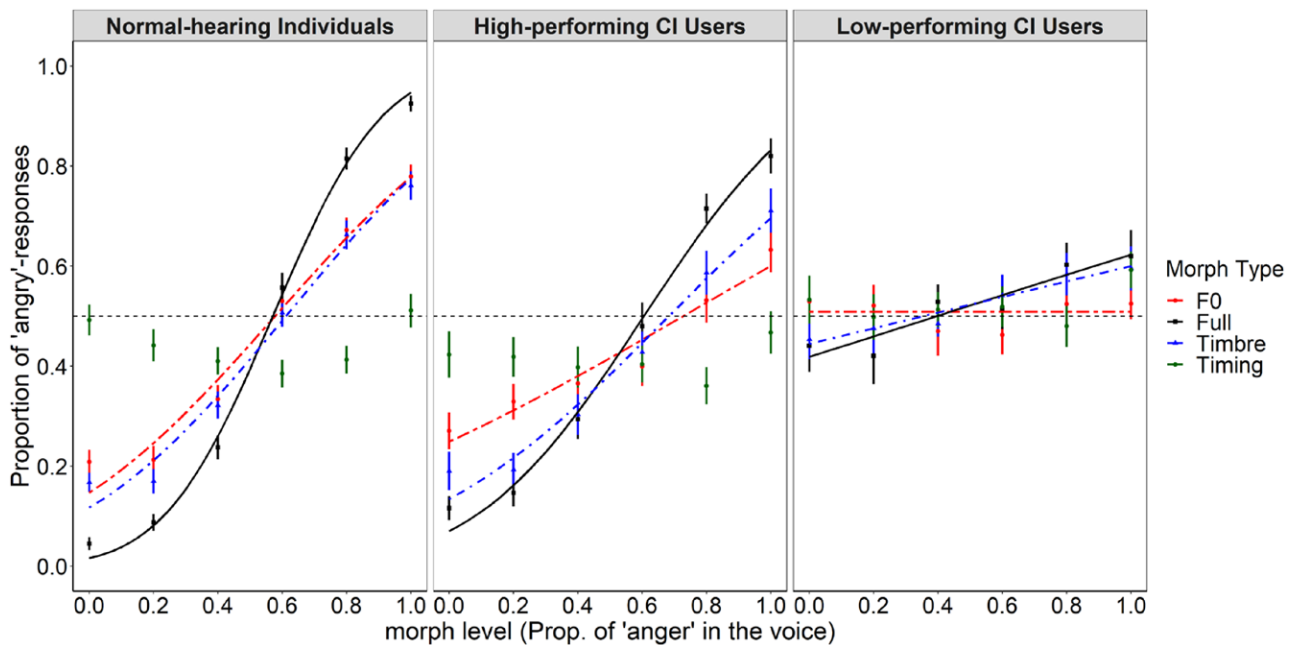

Fig. 2. The proportion of "angry"-responses for different morph levels and morph types used in the experiment, separately for the normal-hearing individuals, the high-performing CI users (n = 13), and the low-performing CI users (n = 12). Note that steeper slopes represent better performance. Error bars represent SEM. Best-fitting cumulative Gaussian functions are also shown. CI, cochlear implant; SEM, standard error of the mean.

TABLE 5. Results of the explorative 4 × 6 × 2 × 2 ANOVA on the proportion of "angry"-responses with the factors MType, ML, SpSex, and PerfGroup, including the high-performing CI users and the normal-hearing individuals

Main effects and interactions	F	df	p	η_p^2	ε_{HF}
ML	284.578	5, 180	< 0.001	0.888	0.637
SpSex	6.594	1, 36	0.015	0.155	
PerfGroup × ML	7.298	5, 180	< 0.001	0.169	0.637
MType × ML	69.354	15, 540	< 0.001	0.658	0.759
ML × SpSex	2.360	5, 180	0.042	0.062	
MType × SpSex	2.705	3, 108	0.049	0.070	
PerfGroup × MType × ML	2.840	15, 540	0.001	0.073	0.759
PerfGroup × ML × SpSex	3.651	5, 180	0.004	0.092	
MType × ML × SpSex	2.976	15, 540	0.001	0.076	0.796

ANOVA, analysis of variance; CI, cochlear implant; df, degrees of freedom; ML, morph level; MType, morph type; PerfGroup, performance group; SpSex, speaker sex.

to perceive the emotional expression in voices and perceived quality of life are related. Furthermore, DEV_{Full}, $r_s = -0.375$, $p = 0.032$, $n = 25$, and DEV_{Timbre}, $r_s = -0.404$, $p = 0.023$, $n = 25$, were also associated with the total score of NCIQ. Surprisingly, no significant relations between both DEV_{F0}, $r_s = -0.320$, $p = 0.059$, $n = 25$ and DEV_{Time}, $r_s = -0.319$, $p = 0.060$, $n = 25$ and the total score were revealed. We also observed significant relations between DEV scores and each of the NCIQ's subdomains except for self-esteem; at a descriptive level, relations tended to be most prominent for the subscales of advanced sound perception, speech production, and activity limitations (please refer to Supplemental Material 4.4.1, http://links.lww.com/EANDH/A984 for full details).

DISCUSSION

At a general level, our findings are in line with earlier reports that CI users perform lower in vocal emotion recognition than NH individuals (e.g., Luo et al. 2007; Schorr et al. 2009; Agrawal et al. 2013; See et al. 2013; Volkova et al. 2013; Wiefferink et al. 2013; Jiam et al. 2017; Kim & Yoon 2018; Paquette et al. 2018; Tinnemore et al. 2018; Waaramaa et al. 2018). However, at a more specific level, when considering contributions of fundamental frequency, timbre, and timing cues to the perception of vocal emotion in CI users,

TABLE 6. Results of the 6 x 2 ANOVAs for the morph types F0 and Full on the proportion of "angry"-responses with the factors ML and PerfGroup, including the high-performing CI users and the normal-hearing individuals

Main effects and interactions	F	df	p	η_p^2	ε_{HF}
Full					
ML	405.728	5, 180	< 0.001	0.919	0.646
PerfGroup × ML	6.684	5, 180	< 0.001	0.157	0.646
F0					
ML	80.518	5, 180	< 0.001	0.691	0.643
PerfGroup × ML	8.037	5, 180	< 0.001	0.183	0.643

ANOVA, analysis of variance; CI, cochlear implant; df, degrees of freedom; ML, morph level; PerfGroup, performance group.

the current findings represent an intriguing contrast to previous research: Our results suggest a greater contribution of timbre than F0, at variance with former reports—originating from gender perception tasks—that CI users cannot dependably use timbre (Fuller et al. 2014a). Fuller et al. (2014a) proposed two possible explanations for the poor usage of timbre. First, CI users could be unable to perceive timbre because its representation is not transferred to the auditory nerve due to large excitation fields of adjacent electrodes. Second, timbre representations—even when partially present in the neuronal code—may be too weak or too distorted to be reliably used. Our findings, however, suggest that some CI users can actually rely on timbre just as well as NH individuals can, at least under the conditions of the present experiment. This shows that CI systems, in principle, can efficiently transfer the acoustic parameters defining timbre (here, FF, spectral level information, and AP; cf. Skuk et al. 2020). In addition, recent research using harmonic complex tones suggests an interdependence between aspects of timbre processing (here, spectral slope) and F0 in both NH listeners and CI users (Luo et al. 2019). Concerning the present study on voices, the transmission of individual acoustic parameters defining voice timbre and their combined contributions to emotion perception will require further research. Moreover, since the neuronal representation of timbre will inevitably be distorted by the CI, our findings could suggest a remarkable degree of neuronal plasticity in the afferent pathway or the auditory cortex of a substantial proportion of CI users, enabling them to efficiently process timbre in emotional voices.

In the present experiment, vocal emotion recognition performance based on timing cues alone was virtually at chance levels. In our opinion, the most likely explanation for this finding was that timing cues were largely uninformative for the specific emotional contrast (fearful versus angry) we tested in this study (cf. Table 2). Crucially, timing cues were used neither by CI users nor NH listeners. As such, these results are not necessarily in contradiction with other studies proposing that CI users rely more on tempo-information than on other cues such as pitch (Kalathottukaren et al. 2015). Thus, timing cues may well be informative for CI users' recognition and discrimination of other emotional categories or contrasts (e.g., happy versus sad).

Several previous studies claim—using gender perception tasks—that fundamental frequency is the most robust and salient acoustic parameter for CI users or CI simulations in NH listeners (e.g., Fuller et al. 2014a; Fuller et al. 2014b). Our results, however, suggest difficulties in processing F0 information in some vocal emotions, as even the subgroup of high-performing CI users was significantly handicapped compared to NH individuals when judging emotions based on F0 cues alone. Of interest, in a vocal emotion task, Gilbers et al. (2015) reported a bias toward F0 *range* cues in CI users, whereas *mean* F0 constitutes a more salient cue for NH individuals. It is important to note that both mean F0 and F0 range were potentially diagnostic for the present emotion contrast (cf. Table 2). Several other studies investigated pitch perception irrespective of vocal emotions, thus only allowing restricted comparisons with the present study. Sucher and McDermott (2007) studied CI users' ability to perceive changes in pitch (with a range of 98 to 740 Hz) in complex musical stimuli and observed poor pitch change perception in CI users, broadly in line with the present findings. It is not entirely

clear how these results can be reconciled with other findings of relatively preserved pitch perception with a CI (e.g., Fuller et al. 2014b). One possibility is that F0 differences in the present emotion contrast were less salient compared to F0 differences for other emotions (Banse & Scherer 1996) or for other social signals. For instance, mean F0 in female voices (200 to 220 Hz) is almost twice as high than mean F0 in male voices (100 to 120 Hz; for a review, see Simpson 2009), and F0 alone is efficiently used by CI users in the perception of speaker gender (Skuk et al. 2020).

Overall, findings that inform about the degree to which F0 or timbre information can be perceived by CI users are somewhat inconsistent across studies. One way to account for such discrepancies is by considering the influence of other factors such as the nature of the auditory stimuli (e.g., vowels, words, sentences) or the type of social signal (e.g., speech comprehension, emotion perception, gender perception). For instance, Meister et al. (2016; see also Meister et al. 2020) compellingly argued that the ability of CI users to utilize timbre, while limited for brief stimuli, is relatively preserved for sentences with their larger phonetic variability and suprasegmental information. In our view, more research is also needed to determine the role of the social signal for voice perception in CI users (Schweinberger et al. 2020). The present data, for example, strongly suggest that the ability to use timbre information can be relatively preserved in a vocal emotion task, even for brief stimuli (bisyllabic pseudowords). The pattern of results from the present experiment also potentially forms a double dissociation relative to the pattern discovered by Skuk et al. (2020). Specifically, Skuk et al. (2020) found CI users´ ability to perceive speaker gender in brief bisyllabic stimuli to be exclusively based on F0, with minimal or no use of timbre, which is directly opposite to the present pattern for vocal emotion perception. Thus, it seems important to consider the type of social signal in tasks that assess nonverbal voice perception abilities in CI users. There is a need for more systematic evidence regarding interactive contributions of stimulus type and social signal to the use of F0 and timbre cues by CI users.

Overall, our data show that some CI users can efficiently process timbre in emotional voices beyond what would be expected based on earlier findings, and despite the fact that CIs degrade prosodic information (e.g., Nakata et al. 2012), probably partially due to their small number of electrodes. We are intrigued by recent evidence that rehabilitation programs may improve the perception of prosody in CI users (Vandali et al. 2015). In their review, Jiam et al. (2017) discuss the possibility that auditory trainings might potentially transfer to enhance vocal emotion recognition in CI users (e.g., Krull et al. 2012; for a review, see Nussbaum & Schweinberger 2021) but emphasize that much further research into the potential of such trainings is warranted. This seems particularly relevant because there is increasing evidence that vocal emotion recognition skills in CI users are positively linked to quality of life, both in children (Schorr et al. 2009) and adults, as shown in both the present study and a recent report (Luo et al. 2018).

As a perspective, further research on multimodal emotion perception in CI users also seems promising, especially when considering how current models of face and voice processing emphasize the multisensory nature of emotions (e.g., Young et al. 2020). Differences in CI technology alone may be insufficient to explain the present striking degree of individual differences. It seems more likely that the degree of reorganization triggered by the individual history of sensory deprivation (Ponton et al. 2000; Gordon et al. 2011) promotes speech-related facial processing through cross-modal plasticity, allowing more efficient audiovisual integration after cochlear implantation (e.g., Rouger et al. 2012). Last but not least, further research should aim at delineating the perceptual abilities and strategies that CI users employ when perceiving different types of (social) signals. Ultimately, a better understanding of possibilities and limitations of CI users to perceive different auditory cues and social signals might promote not only an improvement of CI design but also the development of tailor-made perceptual training programs. Together, such a focus on nonverbal aspects of the voice might further enhance social communication and, ultimately, quality of life for CI users.

In conclusion, when comparing vocal emotion perception in CI users and NH individuals using parameter-specific voice morphing (Skuk & Schweinberger 2014), CI users were far more efficient in using timbre than F0 information in the present experiment. We also observed an enormous degree of interindividual variability; a subgroup of high-performing CI users relied on timbre cues virtually as efficiently as NH individuals did while showing evidence for reduced usage of F0 information. Thus, in the context of the present vocal emotion task, CIs seem inefficient in conveying emotion based on F0 alone. Our results challenge many earlier findings by demonstrating that CI users actually can efficiently use timbre cues in some situations. Moreover, they form a potential double dissociation with a consistent previous pattern of results for voice gender perception, in which CI users exhibit efficient use of F0 but inefficient use of timbre. Accordingly, the current results could indicate that the type of social signal needs to be considered when assessing F0 and timbre perception skills in CI users. The ability to perceive vocal emotions was associated with quality of life. As a perspective, the present findings could inform both perceptual training interventions and improvements in CI technology and ultimately could contribute to enhancing CI users' social-emotional communication skills.

Address for correspondence: Celina I. von Eiff, Department for General Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University of Jena, Leutragraben 1, 07743 Jena, Germany. E-mail: celina.isabelle.von.eiff@uni-jena.de or Stefan R. Schweinberger, Department for General Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University of Jena, Am Steiger 3, 07743 Jena, Germany. E-mail: stefan.schweinberger@uni-jena.de

OPEN PRACTICES

REFERENCES

Agrawal, D., Thorne, J. D., Viola, F. C., Timm, L., Debener, S., Büchner, A., Dengler, R., Wittfoth, M. (2013). Electrophysiological responses to emotional prosody perception in cochlear implant users. *Neuroimage Clin, 2*, 229–238.

Agrawal, D., Timm, L., Viola, F. C., Debener, S., Büchner, A., Dengler, R., Wittfoth, M. (2012). ERP evidence for the recognition of emotional prosody through simulated cochlear implant strategies. *BMC Neurosci, 13*, 113.

Artières, F., Vieu, A., Mondain, M., Uziel, A., Venail, F. (2009). Impact of early cochlear implantation on the linguistic development of the deaf child. *Otol Neurotol, 30*, 736–742.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol, 70*, 614–636.

Boersma, P. (2018). Praat: Doing phonetics by computer [computer program]. version 6.0.46. Retrieved January 2020 from http://www.Praat.Org/

Brewer, R., Biotti, F., Catmur, C., Press, C., Happé, F., Cook, R., Bird, G. (2016). Can neurotypical individuals read autistic facial expressions? Atypical production of emotional facial expressions in autism spectrum disorders. *Autism Res, 9*, 262–271.

Chatterjee, M., Zion, D. J., Deroche, M. L., Burianek, B. A., Limb, C. J., Goren, A. P., Kulkarni, A. M., Christensen, J. A. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hear Res, 322*, 151–162.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychol Bull, 97*, 412-429.

Frühholz, S., Klaas, H. S., Patel, S., Grandjean, D. (2015). Talking in fury: The cortico-subcortical network underlying angry vocalizations. *Cereb Cortex, 25*, 2752–2762.

Frühholz, S., & Schweinberger, S. R. (2021). Nonverbal auditory communication - Evidence for integrated neural systems for voice signal production and perception. *Prog Neurobiol, 199*, 101948.

Fu, Q. J., Chinchilla, S., Galvin, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *J Assoc Res Otolaryngol, 5*, 253–260.

Fu, Q. J., Chinchilla, S., Nogaki, G., Galvin, J. J. 3rd. (2005). Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *J Acoust Soc Am, 118*, 1711–1718.

Fuller, C. D., Galvin, J. J. 3rd, Free, R. H., Başkent, D. (2014a). Musician effect in cochlear implant simulated gender categorization. *J Acoust Soc Am, 135*, EL159–EL165.

Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q. J., Free, R. H., Başkent, D. (2014b). Gender categorization is abnormal in cochlear implant users. *J Assoc Res Otolaryngol, 15*, 1037–1048.

GalvinIII, J. J., Fu, Q. J., Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners. *Ear Hear, 28*, 302-319.

Gaudrain, E., & Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear Hear, 39*, 226–237.

Gilbers, S., Fuller, C., Gilbers, D., Broersma, M., Goudbeek, M., Free, R., Başkent, D. (2015). Normal-hearing listeners' and cochlear implant users' perception of pitch cues in emotional speech. *Iperception, 6*, 0301006615599139.

Gordon, K. A., Wong, D. D., Valero, J., Jewell, S. F., Yoo, P., Papsin, B. C. (2011). Use it or lose it? Lessons learned from the developing brains of children who are deaf and use cochlear implants to hear. *Brain Topogr, 24*, 204–219.

Green, H., & Tobin, Y. (2009). Prosodic analysis is difficult... but worth it: A study in high functioning autism. *Int J Speech-Language Pathol, 11*, 308-315.

Guyatt, G. H., Feeny, D. H., Patrick, D. L. (1993). Measuring health-related quality of life. *Ann Intern Med, 118*, 622–629.

Hazrati, O., Ali, H., Hansen, J. H., Tobey, E. (2015). Evaluation and analysis of whispered speech for cochlear implant users: Gender identification and intelligibility. *J Acoust Soc Am, 138*, 74–79.

Hinderink, J. B., Krabbe, P. F., Van Den Broek, P. (2000). Development and application of a health-related quality-of-life instrument for adults with cochlear implants: The Nijmegen Cochlear Implant Questionnaire. *Otolaryngol Head Neck Surg, 123*, 756–765.

Huber, M. (2005). Health-related quality of life of Austrian children and adolescents with cochlear implants. *Int J Pediatr Otorhinolaryngol, 69*, 1089–1101.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *J Educ Stat, 1*, 69-82.

Jiam, N. T., Caldwell, M., Deroche, M. L., Chatterjee, M., Limb, C. J. (2017). Voice emotion perception and production in cochlear implant users. *Hear Res, 352*, 30–39.

Kalathottukaren, R. T., Purdy, S. C., Ballard, E. (2015). Prosody perception and musical pitch discrimination in adults using cochlear implants. *Int J Audiol, 54*, 444–452.

Kang, S. Y., Colesa, D. J., Swiderski, D. L., Su, G. L., Raphael, Y., Pfingst, B. E. (2010). Effects of hearing preservation on psychophysical responses to cochlear implant stimulation. *J Assoc Res Otolaryngol, 11*, 245–265.

Kawahara, H., Morise, M., Banno, H., Skuk, V. G. (2013). Temporally variable multi-aspect N-way morphing based on interference-free speech representations. In: 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference.

Kawahara, H., & Skuk, V. G. (2019). Voice morphing. In *The Oxford Handbook of Voice Perception* (pp. 685-706).

Kim, S. Y., Kim, H.-J., Park, E.-K., Joe, J., Sim, S., Choi, H. G. (2017). Severe hearing impairment and risk of depression: A national cohort study. *PLoS ONE, 12*, e0179973.

Kim, M. Y., & Yoon, M. S. (2018). Recognition of voice emotion in school aged children with cochlear implants. *Commun Sci Disorders, 23*, 1102-1110.

Kirtane, M., Mankekar, G., Mohandas, N., Patadia, R. (2010). Cochlear implants. *Int J Otorhinolaryngol Clin, 2*, 133-137.

Kong, Y. Y., Cruz, R., Jones, J. A., Zeng, F. G. (2004). Music perception with temporal cues in acoustic and electric hearing. *Ear Hear, 25*, 173–185.

Kovacić, D., & Balaban, E. (2009). Voice gender perception by cochlear implantees. *J Acoust Soc Am, 126*, 762–775.

Kovačić, D., & Balaban, E. (2010). Hearing history influences voice gender perceptual performance in cochlear implant users. *Ear Hear, 31*, 806–814.

Krull, V., Luo, X., Iler Kirk, K. (2012). Talker-identification training using simulations of binaurally combined electric and acoustic hearing: Generalization to speech and emotion recognition. *J Acoust Soc Am, 131*, 3069–3078.

Li, T., & Fu, Q. J. (2011). Voice gender discrimination provides a measure of more than pitch-related perception in cochlear implant users. *Int J Audiol, 50*, 498–502.

Limb, C. J., & Roy, A. T. (2014). Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hear Res, 308*, 13–26.

Luo, X., Fu, Q. J., Galvin, J. J. 3rd. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends Amplif, 11*, 301–315.

Luo, X., Kern, A., Pulling, K. R. (2018). Vocal emotion recognition performance predicts the quality of life in adult cochlear implant users. *J Acoust Soc Am, 144*, EL429.

Luo, X., Soslowsky, S., Pulling, K. R. (2019). Interaction between pitch and timbre perception in normal-hearing listeners and cochlear implant users. *J Assoc Res Otolaryngol, 20*, 57–72.

Massida, Z., Marx, M., Belin, P., James, C., Fraysse, B., Barone, P., Deguine, O. (2013). Gender categorization in cochlear implant users. *J Speech Lang Hear Res, 56*, 1389–1401.

McAleer, P., Todorov, A., Belin, P. (2014). How do you say 'hello'? Personality impressions from brief novel voices. *PLoS One, 9*, Article e90779.

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. *J Speech Lang Hear Res, 59*, 546–556.

Meister, H., Landwehr, M., Pyschny, V., Walger, M., von Wedel, H. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *Int J Audiol, 48*, 38–48.

Meister, H., Walger, M., Lang-Roth, R., Müller, V. (2020). Voice fundamental frequency differences and speech recognition with noise and speech maskers in cochlear implant recipients. *J Acoust Soc Am, 147*, EL19.

Moore, D. R., & Shannon, R. V. (2009). Beyond cochlear implants: Awakening the deafened brain. *Nat Neurosci*, *12*, 686–691.

Nakata, T., Trehub, S. E., Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. *J Acoust Soc Am*, *131*, 1307–1314.

Nussbaum, C., & Schweinberger, S. R. (2021). Links between musicality and vocal emotion perception. *Emotion Rev*, *13*, 211-224.

Paquette, S., Ahmed, G. D., Goffi-Gomez, M. V., Hoshino, A. C. H., Peretz, I., Lehmann, A. (2018). Musical and vocal emotion perception for cochlear implants users. *Hear Res*, *370*, 272–282.

Peterson, N. R., Pisoni, D. B., Miyamoto, R. T. (2010). Cochlear implants and spoken language processing abilities: Review and assessment of the literature. *Restor Neurol Neurosci*, *28*, 237–250.

Ponton, C. W., Eggermont, J. J., Don, M., Waring, M. D., Kwong, B., Cunningham, J., Trautwein, P. (2000). Maturation of the mismatch negativity: Effects of profound deafness and cochlear implant use. *Audiol Neurootol*, *5*, 167–185.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/.

Ray, J., Popli, G., Fell, G. (2018). Association of cognition and age-related hearing impairment in the English longitudinal study of ageing. *JAMA Otolaryngol Head Neck Surg*, *144*, 876–882.

Rouger, J., Lagleyre, S., Démonet, J. F., Fraysse, B., Deguine, O., Barone, P. (2012). Evolution of crossmodal reorganization of the voice area in cochlear-implanted deaf patients. *Hum Brain Mapp*, *33*, 1929–1940.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychol Bull*, *99*, 143–165.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., Buehner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodol Eur J Res Methods Behav Soc Sci*, *6*, 147-151.

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. *J Speech Lang Hear Res*, *52*, 141–152.

Schweinberger, S. R., von Eiff, C. I., Kirchen, L., Oberhoffner, T., Guntinas-Lichius, O., Dobel, C., Nussbaum, C., Zäske, R., Skuk, V. G. (2020). The role of stimulus type and social signal for voice perception in cochlear implant users: Response to the letter by Meister H et al. *J Speech Lang Hear Res, 63*, 4327–4328.

See, R. L., Driscoll, V. D., Gfeller, K., Kliethermes, S., Oleson, J. (2013). Speech intonation and melodic contour recognition in children with cochlear implants and with normal hearing. *Otol Neurotol*, *34*, 490–498.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Lang Linguistics Compass*, *3*, 621-640.

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. *J Speech Lang Hear Res*, *63*, 3155–3175.

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *J Speech Lang Hear Res*, *57*, 285–296.

Stickney, G. S., Zeng, F. G., Litovsky, R., Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *J Acoust Soc Am*, *116*, 1081–1091.

Sucher, C. M., & McDermott, H. J. (2007). Pitch ranking of complex tones by normally hearing subjects and cochlear implant users. *Hear Res*, *230*, 80–87.

Thomas, M., & Tripathi, S. (2014). Music perception and cochlear implants—A review of major breakthroughs. In: 2014 IEEE International Conference on Computational Intelligence and Computing Research (pp. 1-8).

Tinnemore, A. R., Zion, D. J., Kulkarni, A. M., Chatterjee, M. (2018). Children's recognition of emotional prosody in spectrally degraded speech is predicted by their age and cognitive status. *Ear Hear*, *39*, 874–880.

Trainor, L. J., Austin, C. M., Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychol Sci*, *11*, 188–195.

Vandali, A., Sly, D., Cowan, R., van Hoesel, R. (2015). Training of cochlear implant users to improve pitch perception in the presence of competing place cues. *Ear Hear*, *36*, e1–e13.

Volkova, A., Trehub, S. E., Schellenberg, E. G., Papsin, B. C., Gordon, K. A. (2013). Children with bilateral cochlear implants identify emotion in speech and music. *Cochlear Implants Int*, *14*, 80–91.

Waaramaa, T., Kukkonen, T., Mykkänen, S., Geneid, A. (2018). Vocal emotion identification by children using cochlear implants, relations to voice quality, and musical interests. *J Speech Lang Hear Res*, *61*, 973–985.

World Health Organization. (2021). *Deafness and Hearing Loss*. https://www.who.int/health-topics/hearing-loss#tab=tab_1.

Wiefferink, C. H., Rieffe, C., Ketelaar, L., De Raeve, L., Frijns, J. H. (2013). Emotion understanding in deaf children with a cochlear implant. *J Deaf Stud Deaf Educ*, *18*, 175–186.

Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: A remarkable past and a brilliant future. *Hear Res*, *242*, 3–21.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., Rabinowitz, W. M. (1991). Better speech recognition with cochlear implants. *Nature*, *352*, 236–238.

Xu, L., Zhou, N., Chen, X., Li, Y., Schultz, H. M., Zhao, X., Han, D. (2009). Vocal singing by prelingually-deafened children with cochlear implants. *Hearing Res*, *255*, 129-134.

Young, A. W., Frühholz, S., Schweinberger, S. R. (2020). Face and voice perception: Understanding commonalities and differences. *Trends Cogn Sci*, *24*, 398–410.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *Am Psychol*, *35*, 151-175.