OPEN ACCESS

Short Report

Perception of emotional expression in cochlear implant users

Wahrnehmung des emotionalen Ausdrucks durch Cochlea-Implantat-Nutzer

Abstract

Cochlear implants (CIs) have significantly improved speech comprehension in individuals with profound hearing loss. However, the perception of vocal emotions – critical for successful interpersonal interactions and quality of life – remains underexplored in CI users. This short report summarizes findings from a series of experiments, focusing on how CI users perceive vocal emotions, including the weighting of acoustic parameters in their perception, and the potential for perceptual training to improve performance. It also examines the role of audiovisual (AV) integration in emotion recognition. Across experiments, CI users demonstrated reduced emotion recognition accuracy, with substantial inter-individual differences. Whereas normal-hearing (NH) individuals utilized timbre and fundamental frequency (F0) information to equivalent degrees when recognizing vocal emotions, CI users were considerably more efficient in using timbre (compared to fundamental frequency) information. Importantly, targeted perceptual training using vocal caricatures improved recognition performance. In addition, audiovisual integration enhanced emotion recognition, with a larger benefit in CI users than in NH controls. Together, these findings highlight the importance of vocal emotion perception in CI outcomes and suggest that enhancing this perceptual ability in CI users could play a vital role in improving their social communication and quality of life.

Keywords: cochlear implants, hearing loss, vocal emotion perception, quality of life, psychosocial functioning

Zusammenfassung

Cochlea-Implantate (CIs) haben das Sprachverstehen bei Menschen mit hochgradigem Hörverlust deutlich verbessert. Die Wahrnehmung von stimmlichen Emotionen – entscheidend für erfolgreiche zwischenmenschliche Interaktion und Lebensqualität – ist jedoch bei CI-Nutzern bislang wenig untersucht. Der vorliegende Kurzbeitrag fasst einige Experimente zusammen, in denen untersucht wurde, wie CI-Nutzer stimmliche Emotionen wahrnehmen, einschließlich der Gewichtung akustischer Parameter in ihrer Wahrnehmung und der Möglichkeiten einer Verbesserung der stimmlichen Emotionserkennung durch perzeptuelles Training. Darüber hinaus wird die Rolle der audiovisuellen Integration bei der Emotionswahrnehmung analysiert. CI-Nutzer wiesen in allen Experimenten deutlich geringere Erkennungsraten für Emotionen auf, mit großen interindividuellen Unterschieden. Während normalhörende (NH) Personen Timbre und Fundamentalfrequenz (F0) gleichermaßen bei der Erkennung stimmlicher Emotionen nutzten, waren CI-Nutzer deutlich effizienter in der Nutzung von Timbre (im Vergleich zur Grundfrequenz). Wichtig ist, dass gezieltes perzeptuelles Training mit Karikaturen von stimmlichen Emotionen die Erkennungsleistung verbesserte. Zusätzlich steigerte die audiovisuelle Integration die Emotionswahrnehmung, wobei CI-Nutzer stärker profitierten als NH-Kontrollen.

Celina I. von Eiff[1,2,3]
Stefan R. Schweinberger[1,2,3,4]

1 Department for General Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University Jena, Germany

2 Voice Research Unit, Institute of Psychology, Friedrich Schiller University Jena, Germany

3 German Research Foundation Priority Programme SPP 2392 Visual Communication (ViCom), Frankfurt am Main, Germany

4 German Center for Mental Health (DZPG), Site Jena-Magdeburg-Halle, Germany

Zusammengenommen unterstreichen diese Ergebnisse die Bedeutung der Wahrnehmung stimmlicher Emotionen für den Erfolg einer CI-Versorgung und deuten darauf hin, dass die Förderung dieser Wahrnehmungsfähigkeit die soziale Kommunikation und Lebensqualität von CI-Nutzern entscheidend verbessern könnte.

Schlüsselwörter: Cochlea-Implantate, Hörverlust, stimmliche Emotionswahrnehmung, Lebensqualität, psychosoziale Funktionsfähigkeit

Introduction

Cochlear implants (CIs) have significantly improved auditory capabilities in individuals with profound hearing loss, primarily enhancing speech comprehension. However, CI users often remain impaired in perceiving non-verbal vocal signals, such as vocal emotions [1]. Emotions are essential for successful interpersonal communication, empathy, and social bonding. Vocal emotions refer to the emotional tone of speech and are conveyed through multiple acoustic cues, including pitch (fundamental frequency, F0), timbre, temporal characteristics, and intensity. While previous research suggested surprisingly weak relationships between speech recognition and perceived quality of life (QoL) in CI users [2], studies reported consistent correlations between vocal emotion perception and QoL in the same group (e.g., [3]). Despite its relevance for psychosocial functioning, vocal emotion perception is still not routinely assessed in clinical practice, and its underlying mechanisms remain insufficiently understood.

The present short report synthesizes findings from different experimental studies included in a publication-based doctoral dissertation comprising four journal publications [4]. Specifically, the experiments investigated how CI users perceive vocal emotions compared to normal-hearing (NH) individuals, in terms of the weighting of acoustic parameters, potential for perceptual training, and benefits of audiovisual (AV) integration. Results have significant implications for both clinical assessment of CI outcomes and development of rehabilitation strategies.

Material and methods

This short report summarizes a series of experimental studies reported in detail in [4]. Across experiments, between 15 and 26 adult CI users participated (depending on the experiment), each matched with an equal number of age- and gender-matched NH controls. Close matching between groups was considered essential because previous research has demonstrated age- and gender-related differences in emotion perception. Thus, matching procedures ensured that observed group differences were primarily attributable to CI use rather than demographic confounds.

Sample sizes were determined via a priori power analyses in all but one study. Notably, in the acoustic-parameter weighting experiment, effect size estimates were unavailable; therefore, sample size was guided by previous CI studies with similar paradigms.

All participants were native German speakers without neurological or psychiatric diagnoses. CI users reported no other otologic disorders and had either bilateral or unilateral implants with severe to profound hearing loss in the non-implanted ear. Device manufacturers included Cochlear, MED-EL, and Advanced Bionics.

The studies were approved by the Ethics Committee of Jena University Hospital (Reference Number 5282-10/17 and Reference Number 2019-1606_1-B0). At the beginning of the experiments, all participants gave written informed consent after being carefully informed about the procedure and the aim of the studies, that all data were rendered pseudonymized, that results of the studies might be published in a scientific journal, and that participation was voluntary and could be discontinued at any time if they wished so.

Across studies, we used emotional pseudoword stimuli spoken by multiple speakers, which we manipulated with state-of-the-art voice morphing based on TANDEM-STRAIGHT [5] – a speech analysis, modification, and resynthesis framework which enables precise control of acoustic parameters in stimuli while preserving natural voice quality. This approach not only allowed independent, systematic manipulation of timbre and F0 information but also the creation of so-called “vocal caricatures”, which exaggerate acoustic parameters beyond the original emotional recording.

Importantly, most experiments used emotional pseudowords that were produced using an emotion induction technique rather than instructed emotional acting. This distinction is relevant because acted vocal emotions differ from authentic, non-acted vocal emotions – potentially introducing both quantitative and qualitative biases in experimental findings. The use of non-acted vocal emotions therefore increases the ecological validity of the experimental stimuli.

We additionally developed a novel perceptual training. Both CI users and NH individuals underwent 30-day online training sessions consisting of 64 trials per session (approximately 7 minutes). In these sessions, participants classified caricatured and original vocal emotions and received trial-by-trial feedback, with the aim of enhancing emotion recognition accuracy.

To explore AV emotion perception, we created a substantial AV database for emotional voice and dynamic face stimuli (with voices varying in emotional intensity via different morph levels, a feature allowing adaptive testing

and calibration of task difficulty) [6]. CI and NH participants were tested on their ability to recognize vocal emotions, with or without precisely synchronized facial information.

Results

Statistical analyses were performed using the software R and conducted separately for each experiment using methods appropriate to the respective experimental design, including analyses of variance (ANOVAs), t-tests, Welch tests, correlational analyses, and modelling response data with cumulative Gaussian functions. Details are reported in the original publications summarized in the dissertation [4].

Across experiments, CI users exhibited significantly reduced accuracy in recognizing vocal emotions compared to NH individuals, both with and without facial information (e.g., in an auditory-only experiment: listener group × morphtype × morphlevel, $F(15,720)=8.432$, $p<0.001$, $\varepsilon_{HF}=0.703$, $\eta_p^2=0.149$). However, accuracy varied considerably across CI users, with some individuals performing close to chance levels while others approached NH performance levels.

Analysis of acoustic cues used by participants revealed that NH individuals utilized timbre and fundamental frequency (F0) information to similar degrees when recognizing vocal emotions ($F(5,120)=0.278$, $p=0.924$, $\eta_p^2=0.011$), while CI users were considerably more efficient in using timbre (compared to F0) for the same task ($F(5,120)=4.200$, $p=0.005$, $\varepsilon_{HF}=0.725$, $\eta_p^2=0.149$). Notably, high-performing CI users relied on timbre similarly efficiently as NH individuals ($ps\geq0.217$), whereas impairments remained particularly pronounced for F0-based emotion information (for 20%, 60%, 80%, and 100% morphlevels, $|ts(74)|\geq2.514$, $ps\leq0.014$; however, note that for 0% and 40% morphlevels, $|ts(74)|\leq1.473$, $ps\geq0.145$). These findings suggest that current CI devices can transmit timbre-related emotional signals relatively effectively, while emotional information conveyed by F0 remains comparatively difficult to process.

Perceptual training with vocal caricatures improved vocal emotion recognition in CI users, with training effects remaining detectable approximately six weeks after training completion. Importantly, these improvements generalized at least partially to untrained speakers, suggesting that training enhanced perceptual processing of emotional voice information rather than merely stimulus-specific learning. Vocal caricaturing furthermore enhanced vocal emotion recognition performance directly in another experiment, both in CI users ($t(24)=4.144$, $p<0.001$, $t(24)=5.643$, $p<0.001$, and $t(13)=3.708$, $p=0.001$ for auditory-only, AV congruent, and AV incongruent stimuli; $M_{CI}\pm SEM=0.061\pm0.015$, 0.079 ± 0.014, and 0.135 ± 0.036, respectively) and NH individuals ($|ts(24)|\geq8.352$, $ps<0.001$, for auditory-only, AV congruent, and AV incongruent; $M_{NH}\pm SEM=0.146\pm0.010$, 0.119 ± 0.011, and 0.136 ± 0.016, respectively), demonstrating the potential of exaggerated diagnostic acoustic information for facilitating socio-emotional communication.

Participants also showed substantial benefits from multisensory stimuli: AV integration enhanced emotion recognition in both CI and NH groups; however, CI users exhibited stronger benefits to vocal emotion perception from time-synchronized congruent facial emotional information than NH individuals ($t(32.031)=4.580$, $p<0.001$) – and these larger crossmodal benefits were maintained even at equal auditory-only performance levels ($M_{CI}=71.2\pm1.79$ vs. $M_{NH}=78.3\pm3.50$; $t(18)=-1.808$, $p=0.044$). Importantly, this suggests that these AV benefits result from deafness-related compensatory plasticity rather than merely degraded acoustic representations. Together, these findings highlight the importance of AV mechanisms for emotional communication in CI users and suggest promising perspectives for rehabilitation approaches incorporating multisensory emotional information.

Finally, the results confirmed previous findings demonstrating positive associations between vocal emotion recognition abilities and QoL ratings in CI users (e.g., $r_s=-0.390$, $p=0.027$, $n=25$; note that negative correlations were expected, as smaller scores corresponded to better performance here).

Discussion and implications

Vocal emotion perception as a clinically relevant CI outcome

Traditional assessments of CI outcomes focus on speech comprehension, overlooking the role of non-verbal cues like vocal emotions. This research suggests that vocal emotions play a critical role in social communication and well-being. By integrating emotion recognition into outcome measures, clinicians might achieve a more holistic evaluation of CI users' functional abilities, better reflecting their capacity to engage in meaningful social interactions.

Implications for CI rehabilitation and perceptual training

The results suggest that targeted perceptual training can improve emotion recognition in CI users. Refined training programs incorporating vocal caricatures could become an effective tool in CI rehabilitation, offering a tailored approach to enhancing emotional competence.

Audiovisual integration and crossmodal plasticity

The investigation highlights the potential for multisensory integration to compensate for sensory deprivation. CI users benefit more from AV cues than NH individuals (even at equal auditory-only performance levels), implicating crossmodal plasticity – CI users' brains may have re-

organized to integrate visual and auditory information more efficiently. Rehabilitation programs could leverage this crossmodal benefit to enhance communication skills in CI users.

Social and psychological well-being

Positive correlations between vocal emotion recognition and QoL in CI users imply the importance of this perceptual ability for well-being and social inclusion. Improving emotion recognition might lead to better interpersonal relationships, emotional resilience, and overall life satisfaction. The present findings suggest that emotional competence is not a supplementary ability but a core component of social participation and health.

Broader implications for auditory research

The morphing-based stimulus design used here offers a tool for investigating vocal emotion perception, with or without precisely synchronized facial information. This technique opens up new possibilities for understanding the mechanisms of both auditory-only and AV emotion recognition in hearing-impaired people.

Conclusion

The present short report summarizes recent findings on the perception of vocal emotions in CI users. Across experiments, CI users exhibited reduced recognition of vocal emotions and showed different weighting of acoustic parameters in their perception, compared to NH individuals. Importantly, perceptual training can improve CI users' emotion recognition. The findings also demonstrate the benefits of AV integration, implying its potential for rehabilitation. Together, these results have important implications for the clinical assessment of CI outcomes, the development of rehabilitation programs, and the understanding of neuroplasticity in hearing-impaired individuals. Extending CI outcome measures beyond speech comprehension to include vocal emotion perception may therefore provide a more ecologically valid evaluation of communicative functioning and quality of life. Furthermore, rehabilitation approaches specifically targeting emotional communication may help enhance social participation, emotional well-being, and overall quality of life in CI users.

Notes

Conference presentation

This contribution was presented at the 27th Annual Conference of the German Society of Audiology and published as an abstract [7].

Funding

CIvE was funded by a fellowship from the Studienstiftung des deutschen Volkes. This research was also funded by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft (DFG), grant reference: Schw 511/25-1) within the DFG Priority Programme SPP 2392 Visual Communication.

Competing interests

The authors declare that they have no competing interests.

References

1. Jiam NT, Caldwell M, Deroche ML, Chatterjee M, Limb CJ. Voice emotion perception and production in cochlear implant users. Hear Res. 2017 Sep;352:30-9. DOI: 10.1016/j.heares.2017.01.006
2. Huber M. Health-related quality of life of Austrian children and adolescents with cochlear implants. Int J Pediatr Otorhinolaryngol. 2005 Aug;69(8):1089-101. DOI: 10.1016/j.ijporl.2005.02.018
3. Luo X, Kern A, Pulling KR. Vocal emotion recognition performance predicts the quality of life in adult cochlear implant users. J Acoust Soc Am. 2018 Nov;144(5):EL429. DOI: 10.1121/1.5079575
4. von Eiff CI. Perception of emotional expression in cochlear implant users [doctoral dissertation]. Jena: Friedrich Schiller University Jena; 2024.
5. Kawahara H, Morise M, Takahashi T, Nisimura, R, Irino T, Banno H. Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE Xplore; 2008. p. 3933-6. DOI: 10.1109/ICASSP.2008.4518514
6. von Eiff CI, Kauk J, Schweinberger SR. The Jena Audiovisual Stimuli of Morphed Emotional Pseudospeech (JAVMEPS): A database for emotional auditory-only, visual-only, and congruent and incongruent audiovisual voice and dynamic face stimuli with varying voice intensities. Behav Res Methods. 2024 Aug;56(5):5103-15. DOI: 10.3758/s13428-023-02249-4
7. von Eiff CI. Perception of emotional expression in cochlear implant users. In: Deutsche Gesellschaft für Audiologie e. V.; ADANO, editors. 27. Jahrestagung der Deutschen Gesellschaft für Audiologie und Arbeitstagung der Arbeitsgemeinschaft Deutschsprachiger Audiologen, Neurootologen und Otologen. Göttingen, 19.-21.03.2025. Düsseldorf: German Medical Science GMS Publishing House; 2025. Doc217. DOI: 10.3205/25dga217

Corresponding author:
Celina I. von Eiff
Department for General Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University Jena, Am Steiger 3, Haus 1, 07743 Jena, Germany
celina.isabelle.von.eiff@uni-jena.de

Please cite as
von Eiff CI, Schweinberger SR. Perception of emotional expression in cochlear implant users. GMS Z Audiol (Audiol Acoust). 2026;8:Doc21. DOI: 10.3205/zaud000098, URN: urn:nbn:de:0183-zaud0000983

This article is freely available from
https://doi.org/10.3205/zaud000098

Published: *2026-09-09*