A subjective and objective evaluation of a codec for the electrical stimulation patterns of cochlear implants

Reemt Hinrichs, Tom Gajecki, Jörn Ostermann, and Waldo Nogueira

Citation: The Journal of the Acoustical Society of America **149**, 1324 (2021); doi: 10.1121/10.0003571
View online: https://doi.org/10.1121/10.0003571
View Table of Contents: https://asa.scitation.org/toc/jas/149/2
Published by the Acoustical Society of America

JASA ARTICLE

A subjective and objective evaluation of a codec for the electrical stimulation patterns of cochlear implants

Reemt Hinrichs,[1,a)] Tom Gajecki,[2] Jörn Ostermann,[1] and Waldo Nogueira[2]
[1]*Institut für Informationsverarbeitung, Leibniz Universität Hannover, Appelstraße 9a, 30167 Hannover, Germany*
[2]*Department of Otolaryngology, Medical University Hannover, Karl-Wiechert-Allee 3, 30625 Hannover, Germany*

ABSTRACT:
Wireless transmission of audio from or to signal processors of cochlear implants (CIs) is used to improve speech understanding of CI users. This transmission requires wireless communication to exchange the necessary data. Because they are battery powered devices, energy consumption needs to be kept low in CIs, therefore making bitrate reduction of the audio signals necessary. Additionally, low latency is essential. Previously, a codec for the electrodograms of CIs, called the Electrocodec, was proposed. In this work, a subjective evaluation of the Electrocodec is presented, which investigates the impact of the codec on monaural speech performance. The Electrocodec is evaluated with respect to speech recognition and quality in ten CI users and compared to the Opus audio codec. Opus is a low latency and low bitrate audio codec that best met the CI requirements in terms of bandwidth, bitrate, and latency. Achieving equal speech recognition and quality as Opus, the Electrocodec achieves lower mean bitrates than Opus. Actual rates vary from 24.3 up to 53.5 kbit/s, depending on the codec settings. While Opus has a minimum algorithmic latency of 5 ms, the Electrocodec has an algorithmic latency of 0 ms.
 https://doi.org/10.1121/10.0003571

(Received 7 July 2020; revised 5 January 2021; accepted 3 February 2021; published online 24 February 2021)
[Editor: James F. Lynch] Pages: 1324–1337

I. INTRODUCTION

A cochlear implant (CI) is a surgically implanted device that can improve or restore hearing to people ranging from moderate to severe hearing loss. Its main components are an electrode array that is surgically implanted into a person's ear and a microphone and signal processor that are placed outside of the recipient's head. Currently, CI users achieve good speech understanding in quiet listening conditions. However, their speech recognition performance decreases quickly as the level of background noise increases (Gifford *et al.*, 2008; Wilson *et al.*, 2007; Zeitler *et al.*, 2008). To improve speech understanding, aside from noise reduction algorithms and directional microphones (Kokkinakis *et al.*, 2012), modern CIs offer wireless audio streaming from external devices (Boddy and Datta, 2018; Ceulaer *et al.*, 2015; Wolfe *et al.*, 2016b). Wireless audio streaming is performed through the head in contralateral routing of signals (CROS) and bilateral communication between two CIs. Wireless audio streaming is performed through the air from external devices such as smartphones or remote microphones to a CI (Ernst *et al.*, 2019; Mehrkian *et al.*, 2019).

This work investigates the impact of compressing the audio information for wireless transmission on speech understanding in CI users.

In binaural sound coding strategies, signal information from two CIs is combined to improve speech understanding and sound localization (Gajecki and Nogueira, 2020). A wireless communication link is required by binaural sound coding strategies and CROS-devices. Binaural sound coding strategies have been proposed (Gajecki and Nogueira, 2018; Kan, 2018; Lopez-Poveda *et al.*, 2016) to improve speech understanding of bilaterally implanted CI users. CROS-devices, applied to unilateral CI users, have a microphone placed on the non-implanted ear. Signals captured by the microphone are transmitted to the implanted side, improving speech understanding when a sound source is located away from the implanted ear (Weder *et al.*, 2015). External devices like Phonak's Roger Pen (Ceulaer *et al.*, 2015) or Cochlear's Mini Microphone (Boddy and Datta, 2018) improve speech understanding of CI users in difficult listening conditions, such as a conference or a class room, through wireless audio streaming from a remote microphone to a CI (Wolfe *et al.*, 2015). Bimodal CI users, i.e., CI users with an additional hearing aid in the non-implanted ear, can benefit from streaming phone calls to both devices (Wolfe *et al.*, 2016a).

The latency of a wireless communication can be critical for CI users. Especially in everyday situations with both visual and auditory cues, difficulties can arise. In face-to-face interactions with other humans, an end-to-end audio latency of less than 10 ms is required (European Telecommunications Standards Institute, 2013). In the case of unilateral CI users with one normal hearing ear, if an audio signal is presented through two pathways, even delays as low as 5 ms can impact the sound quality (Galster, 2010). Through the combination of the direct audio path and the wireless link, perceivable echo effects can arise that degrade the sound quality. However, for bimodal CI users, an additional delay of the

[a)]Electronic mail: hinrichs@tnt.uni-hannover.de

audio of the CI in comparison to the audio of the hearing aid could improve speech performance (Zirn *et al.*, 2015). While to the best knowledge of the authors no research exists on the impact of latency on binaural sound coding strategies or contralateral routing of signals, generally, any delay can be assumed to be undesirable. These latency constraints minimize the number of audio codecs applicable.

Generally, there is a trade-off between latency, audio bandwidth, and bitrate in audio coding (Allamanche *et al.*, 1999). Low latency at low bitrates is achieved by reducing the audio bandwidth, which decreases speech understanding. Low latency without reducing the audio bandwidth is achieved by increasing the bitrate of the applied audio coding. On the other hand, an increased audio bandwidth is accomplished at the cost of increased latency or increased bitrate.

However, the power consumption of a wireless link or channel is directly related to the capacity of the channel (Shannon, 1949). This capacity cannot be reduced below the bitrate of the information that is supposed to be sent through the wireless channel in a given timeframe. Therefore, in the context of audio streaming, the bitrate of the audio signals is the lower bound of the capacity of the wireless channel. Because of this, the bitrate of the applied audio coding algorithm determines the minimum capacity and therefore the minimum power consumption of the wireless communication. Therefore, the applied audio coding has a big impact on the CI's battery life.

To reduce the bitrate of audio signals, current wireless solutions for CIs (Boddy and Datta, 2018; Ceulaer *et al.*, 2015) apply audio coding algorithms like the predictive subband codec G.722 on an audio signal prior to transmission through the wireless link from an external device to a CI, often realized through or including Bluetooth (Wolfe *et al.*, 2015). For audio transmissions using Bluetooth, several well known audio codecs exist, such as the low complexity subband codec (SBC) (Hoene and Hyder, 2010), aptX by Qualcomm (Qualcomm, 2020), or in the near future the low complexity communication codec (LC3) (European Telecommunications Standards Institute, 2018). Many such codecs suffer from a rather high latency like SBC of more than 20 or even 40 ms. Other codecs, while achieving latencies well below 10 ms, like aptX, the ultra low delay codec (Kramer *et al.*, 2004) or predictive subband codecs (Preihs *et al.*, 2016), exhibit a rather high bitrate around or above 96 kbit/s at a sampling rate of 32 kHz and above. However, speech focused applications, such as most CI applications, can be coded at low latency with significantly less than 96 kbit/s (Böhmler *et al.*, 2010). This can be achieved by taking advantage of the typical audio bandwidth transmitted by CIs of around 8–10 kHz.

However, wireless communication through the head, either between two CIs or between a CI and a CROS-device, cannot be efficiently realized using Bluetooth (Edelmann and Ussmueller, 2018). The 2.4 GHz frequency band of Bluetooth is drastically attenuated by body tissue, rendering it impractical for through-the-head communication (Pal and Kant, 2019). A technology used for wireless transmission of audio through the head (Oticon, 2019; Phonak, 2016) is near-field magnetic induction, which, unlike Bluetooth, transmits information through modulation of the magnetic field, requiring significantly lower power (Pal and Kant, 2019). Currently, scientific literature regarding the application of near-field magnetic induction in CIs is very limited, and it is only used by CROS systems (e.g., Oticon, 2019; Phonak, 2016) or for the bilateral communication between two CIs of binaural beamformers (Phonak, 2012). A maximum data-rate of up to 424 kbit/s is reported (Pal and Kant, 2019). So far, no standard near-field magnetic induction audio codecs exist, but in principle, any codec used for Bluetooth could be applied in near-field magnetic induction based communication, too.

Audio codecs are typically evaluated in normal hearing listeners and therefore are not optimized to transmit audio through wireless transmission to CIs. Likewise, objective instrumental measures, such as the perceptual evaluation of speech quality (Khalifeh *et al.*, 2017; Kressner *et al.*, 2011), were designed to model the perception of normal hearing listeners.

Coding algorithms specifically aimed at wireless transmission of audio between conventional hearing aids have been proposed before (Li and Kleijn, 2007; Ostergaard *et al.*, 2009; Roy and Vetterli, 2007). These either consider the limited algorithmic complexity of hearing aids, consider their specific delay constraints, or make use of the spatial proximity and the signal correlations that come with it to reduce bitrate.

None of the aforementioned algorithms, however, considers perception differences between normal and assisted hearing, as they exist in CIs, and no audio codec exists specifically designed for wireless streaming to or between CIs.

To minimize both bitrate and latency, while not reducing the audio bandwidth, we proposed (Hinrichs *et al.*, 2019) to take advantage of the sound coding strategy of a CI, which computes the electrodograms, and to code these electrodograms as suggested by Edler *et al.* (2007). In sound coding strategies that perform a channel or band selection like the advanced combinational encoder (ACE) (Wouters *et al.*, 2015), which the Electrocodec was specifically designed for, in every stimulation cycle, only smaller segments of the full audio input spectrum are presented as electrical pulses to the cochlea. While the full audio input spectrum of CIs usually covers about 8–10 kHz, corresponding to a sampling rate of 16 kHz, the total frequency range of the segments selected by ACE covers between about 1 and 5 kHz. This frequency range is determined by the subbands that are selected in a given stimulation cycle. Additionally, a sound coding strategy like ACE does not necessarily select connected segments of the audio spectrum, resulting in "holes" in the audio spectrum that is presented to the cochlea. However, audio codecs usually cover a frequency range starting at 0 Hz up to some maximum frequency f_{max}. This would decrease speech understanding if f_{max} was set too low and in general transmit unnecessary

spectral information, as the encoded frequency range of an audio codec usually is connected, i.e., is an interval of the form $[0\ \text{Hz}, f_{\max}]$.

Therefore, coding the electrodograms instead of the corresponding audio signal should allow us to achieve lower bitrates and/or latencies at a given level of speech understanding. For this purpose, we proposed the Electrocodec in Hinrichs *et al.* (2019). The Electrocodec could be applied for audio streaming to CIs, e.g., in remote microphones, telephone call streaming from smartphones, or CROS-devices. Its only limitation is in cases where the electrodograms do not contain the required audio cue, e.g., phase information. The Electrocodec reduces bitrate, and therefore the possible power consumption of wireless transmissions, by coding the electrodograms. This coding introduces distortions in the electrodograms. To investigate the impact of these distortions on speech perception, subjective listening tests have to be performed.

Previously, a preliminary implementation of the Electrocodec was compared to the G.722 audio codec in quiet listening conditions, using the signal-to-distortion ratio as an objective measure of quality (Hinrichs *et al.*, 2019). The signal-to-distortion ratio is the ratio of the power of a signal and the power of the reconstruction error introduced by the coding algorithm. The result showed a benefit for the Electrocodec, which achieved a higher signal-to-distortion ratio at lower bitrates than the G.722.

In this work, an evaluation in CI-subjects of an optimized Electrocodec is presented. The Electrocodec is assessed monaurally, because the impact of the coding distortion on speech understanding is unknown. A bilateral evaluation would introduce further unknowns and should be investigated in the future after establishing the Electrocodec's monaural performance. Because this study aims at investigating the isolated impact of these distortions, no delay due to a wireless transmission is considered in this work.

The main research question of the current study is to investigate the impact of the signal distortion introduced by the Electrocodec on speech recognition and speech quality in noisy, monaural listening conditions. This impact is compared to a standard method to transmit audio signals between CIs or to a CI based on a low delay audio codec. For this purpose, the Electrocodec is compared to the Opus audio codec (IETF Codec Working Group, 2018a) in noisy, monaural listening conditions at different bitrates. The impact of the two codecs on both speech recognition and speech quality was evaluated in ten CI-subjects and compared to an unprocessed reference condition. The hypothesis was that the direct compression of the electrodograms as performed by our codec is able to achieve the same or better speech recognition and quality at lower bitrates than an audio codec can achieve with similar algorithmic latency.

Furthermore, the mean bitrate across signal-to-background-noise ratios (SNRs) of the Electrocodec is assessed, and an objective instrumental evaluation of speech performance using the short-time objective intelligibility measure (STOI) (Taal *et al.*, 2010) is given. It compares audio waveforms synthesized from electrodograms by a vocoder to an original unprocessed reference signal, and speech understanding is assessed by the STOI. If the scores obtained from the STOI agree well with the observed speech understanding of the subjective listening tests, it could be used to optimize the Electrocodec in the future.

The structure of this paper is as follows. In Sec. II, the sound coding strategy used for our study is described as well as the structure of the Electrocodec. Furthermore, the baseline Opus audio codec, the speech material, the test conditions, and the signal generation and testing procedure are described. In Sec. III, the results of the speech recognition and speech quality test are presented as well as the objective instrumental evaluation results of the two codecs, which includes an objective assessment of the speech recognition of the two codecs. The results are subsequently discussed in Sec. IV. The paper is concluded in Sec. V.

II. MATERIALS AND METHODS

A. Advanced combination encoder

The sound coding strategy used in this work is the advanced combination encoder. The block diagram of the research implementation of ACE is shown in Fig. 1.

ACE belongs to the class of so called N of M sound coding strategies, where at discrete time n only a subset of N electrodes out of the total M electrodes of the CI are selected. The main components of ACE are a filter bank, which splits the input audio waveform into M subbands; an envelope detection block that estimates the envelopes in each subband; subsequent frequency subband selection; and an acoustic to current level mapping block consisting of the loudness growth function (LGF) and a current mapping

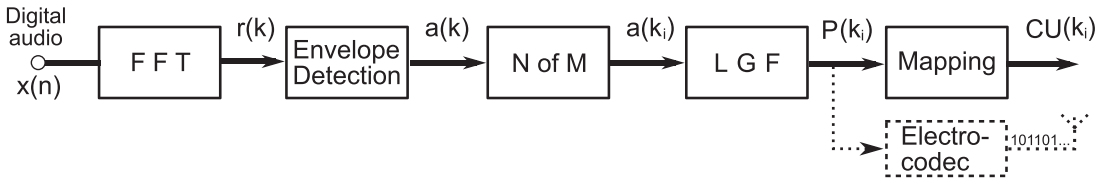

FIG. 1. Block diagram of the research implementation of the ACE sound coding strategy. The digital audio input signal $x(n)$ is separated and transformed into the M subbands of the CI by a fast Fourier transform (FFT) filter bank. From the output signal $r(k)$ of the filter bank, the acoustic envelopes $a(k)$ are calculated, and then the N largest acoustic envelope amplitudes out of the M subbands are selected. This yields the signals $a(k_i)$, where k_i refers to the selected subbands. Then the LGF according to Eq. (1) is applied, resulting in the electrodograms $P(k_i)$. Finally, the electrodograms are mapped to clinical units $CU(k_i)$. In a real implementation, the output signal of the Electrocodec is wirelessly transmitted, indicated by the antenna symbol, to another CI.

https://doi.org/10.1121/10.0003571

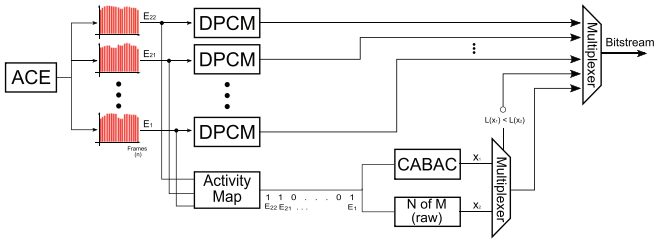

FIG. 2. Example of the encoding process of the Electrocodec. Only the most apical electrodes 21 and 22 as well as the most basal electrode 1 are selected, indicated by the red pulses. The other electrodes are not selected. This results in the activity map, a bit string representing the band selection, depicted in the figure. A value of "1" indicates an electrode that is selected, and a value of "0" indicates an electrode that is not selected. The signals of the selected electrodes of ACE are coded using DPCM. Additionally, the activity map is either encoded using CABAC or encoded without additional entropy coding by only using the N of M property. CABAC is used if the length $L(x_1)$ of the output bitstring x_1 of CABAC is smaller than the output length $L(x_2)$ of the output bitstring x_2 if only the N of M property is used [denoted N of M (raw)]. This decision is put into the total bitstream as a bit flag.

block. In ACE, the subband selection is performed on the basis of the largest magnitudes, i.e., the N subbands with the largest magnitudes are selected and processed further, and the other M–N subbands produce no output. For a detailed description, refer to Nogueira *et al.* (2005). The transformation from the acoustic to the electric domain is performed by the LGF that maps the acoustic envelope amplitude $a(k)$ of subband k to an electrical magnitude $P(k)$:

$$P(k) = \begin{cases} \dfrac{\log(1+\rho((a(k)-s)/(m-s)))}{\log(1+\rho)}, & s \le a(k) \le m \\ 1, & a(k) \ge m \\ \text{no output}, & a(k) < s. \end{cases} \quad \text{LGF.} \tag{1}$$

The magnitude $P(k)$, also called the electrodogram, is a fraction in the range from 0 to 1 that represents the proportion of the output current range from the threshold level to the most comfortable level. The parameters s, m, and ρ are described in Nogueira *et al.* (2005). For all experiments, $\rho = 416.2063$, $s = 4/256$, and $m = 150/256$ were used, which are the default values of the research implementation. The channel stimulation rate, which is the number of pulses in each band per second, was fixed at 900 pulses per second (pps), while the number of selected subbands was fixed at $N = 8$. This results in the same number of values $P(k)$ per second per band. $P(k)$ is the signal that is coded by the Electrocodec.

B. Electrocodec

1. Basic structure

The Electrocodec uses differential pulse-code modulation (DPCM) with backward adaptive linear prediction and context-adaptive binary arithmetic coding (CABAC) for signal compression. Through this design, it achieves zero algorithmic latency. Its basic structure and design is explained in Hinrichs *et al.* (2019). Figure 2 shows a detailed diagram of the construction of the bitstream of the Electrocodec. In contrast to Hinrichs *et al.* (2019), the Electrocodec uses quantizers with equally large codebooks in all subbands. The exact size of the codebooks depends on the tested condition; see Table I. Furthermore, the use of the additional bit flag has changed. Previously, a bit flag indicated whether the band selection has changed from the previous time step to the current one. If no change occurred, the band selection was not encoded to save bitrate. Now, this additional bit flag indicates whether the band selection was encoded without using entropy coding or using CABAC. This was found to be more robust with respect to bitrate, as this allows one to limit the maximum frame size. Because we wanted to avoid a bias in our coding approach, the probabilities used in CABAC were learned on clean speech and using the sound quality assessment material recordings (European Broadcasting Union, 2008). As a consequence, these probabilities do not fit perfectly when applying the codec in noisy conditions, and frames with very large numbers of bits can occur. In these cases, due to the bit flag, the codec switches to encode the band selection without CABAC, considering only the fact that the band selection is fully determined if N selected subbands have been encoded. This N of M property, also denoted as N of M (raw), allows one to represent the band selection often with significantly less than M bits. We could have learned the context probabilities from noisy

TABLE I. Overview of the conditions used in the study. In total, four conditions for the Electrocodec and three conditions for the Opus codec were selected for evaluation. A reference condition, corresponding to the original sound without applying a codec, was included as well. The specified average bitrate is achieved at 0 dB SNR. The latency specified is the algorithmic latency. For both the Electrocodec and Opus, mean bitrates across SNRs are depicted in Fig. 5.

Label	Description	Latency (ms)
REF	Reference condition. Original sound.	
EC2	Electrocodec with 4 quantization levels of the DPCM (2 bit). Mean bitrate: 24.3 kbit/s	0
EC3	Electrocodec with 8 quantization levels of the DPCM (3 bit). Mean bitrate: 30.6 kbit/s	0
EC4	Electrocodec with 16 quantization levels of the DPCM (4 bit). Mean bitrate: 37.6 kbit/s	0
EC7	Electrocodec with 128 quantization levels of the DPCM (7 bit). Mean bitrate: 53.5 kbit/s	0
Opus16c	Opus codec set to constant 16 kbit/s. Mean bitrate: 16 kbit/s	5
Opus16v	Opus codec set to variable 16 kbit/s. Mean bitrate: 31 kbit/s	5
Opus52v	Opus codec set to variable 52 kbit/s. Mean bitrate: 57.9 kbit/s	5

speech or applied an adaptive approach and certainly reduced the bitrate. But this would have made the comparison to an audio codec biased. Therefore, we decided to stick with the described approach. Figure 2 shows a detailed diagram of the construction of the bitstream of the Electrocodec.

2. Error resilience

The presented study is concerned with the impact on speech recognition and quality introduced by the signal coding of the Electrocodec. A wireless communication free of errors was assumed. In real applications, a packet of bits received through a wireless channel can be corrupted by interfering noise. Therefore, the error resilience of the Electrocodec is discussed briefly. For N of M sound coding strategies, such as ACE, the Electrocodec exhibits some resilience to packet loss, which is a transmission error that can be introduced in wireless communication (Korhonen and Wang, 2005). Because ACE selects the N subbands with the largest envelopes at every time step, the subband selection often changes from frame to frame. Because of this, it occurs repeatedly that subbands remain unselected for some adjacent frames. In these cases, when a subband is not selected in two adjacent frames, the Electrocodec resets the encoding of that specific subband. When that subband eventually is newly selected by ACE, no prediction is performed. Only quantization is applied for the first sample of that subband. Because of this reset of the subbands, once two consecutive frames occur in which a subband was not selected, the Electrocodec automatically resynchronizes the decoding. This allows recovery from packet losses that result in the loss of the information of one frame. Additionally, it is known (Qazi *et al.*, 2013) that the precise current level applied to the cochlear has little to no impact on speech recognition as long as the band selection remains unaffected. But the band selection of the decoded electrodograms always remains correct or uncorrupted after a packet loss, because the Electrocodec encodes the activity map independently from previous frames.

C. Baseline audio codec

As no other codec for the electrodograms exists, an audio codec had to be selected as a baseline algorithm. The Electrocodec was compared to this baseline approach to audio streaming.

The baseline audio codec had to (i) code frequencies up to 8 kHz (wideband), (ii) code with very low latency, and (iii) offer bitrate settings flexible enough for the requirements of our study. A wideband codec was necessary as baseline because of the 16 kHz sampling rate used by ACE. Furthermore, the Electrocodec encodes the whole spectrum (by using all electrodes) of the CI. The baseline audio codec had to do the same, especially as narrowband signals already significantly decrease speech recognition (Nogueira *et al.*, 2019). Low latency coding was required for a fair comparison to the Electrocodec, which has an algorithmic latency of 0 ms. Furthermore, to allow for lower and higher bitrates while using the same coding algorithm, an audio codec with variable bitrates was necessary. Lower bitrates were necessary to include a condition at which speech understanding was certainly reduced. Together, these requirements left only the Opus codec as a candidate.

The Opus codec is able to code audio with sampling rates from 4 up to 48 kHz, at algorithmic latencies between 5 and 60 ms, with specifiable bitrates ranging between 6 and 510 kbit/s (Valin *et al.*, 2013). Several studies showed Opus's performance to be equal to or better than other state of the art audio codecs while being very flexible with respect to algorithmic delay and bitrate (Jokisch *et al.*, 2016; Rämö and Toukomaa, 2011).

Opus uses variable bitrate coding and, unless constant bitrate is specified, will in general code at a bitrate at least slightly different from the specified one. In our study, the algorithmic latency of Opus was always 5 ms.

D. Test conditions

In total, eight different conditions were tested. These are summarized in Table I. The **REF** condition consisted of the electrodograms generated by mixing speech and noise signals without any further processing of the signals. The **EC2** condition was the Electrocodec with a quantization resolution of 2 bit used by the quantizer of every band's DPCM. Analogously, the **EC3** condition was the Electrocodec with a quantization resolution of 3 bit used by the quantizer of every band's DPCM. Accordingly, the **EC4** and **EC7** conditions used a 4 bit and 7 bit quantizer, respectively. The **EC2** to **EC4** conditions were introduced to investigate the finer dependencies between speech understanding and resolution of the quantizer. The **EC7** condition was introduced as a backup condition. Had the **EC2** to **EC4** conditions all shown poor speech performances, the performance of the **EC7** condition could have been used to determine if our approach could work at all. Additionally, the bitrate of the **EC7** condition is similar to the **Opus52v** condition, allowing for a better comparison.

The **Opus16c** condition was introduced to investigate a codec setting that could be expected to cause a decrease in speech understanding. To achieve 16 kbit/s at an algorithmic latency of 5 ms, constant bitrate was enforced for this condition. Otherwise, because of Opus's variable bitrate coding scheme, the actual bitrate was significantly higher than the specified one.

The **Opus52v** condition, due to variable bitrate coding, obtained a mean bitrate for the speech material of our study ranging from 58 to 60 kbit/s, depending on the level of background noise. It was expected that this condition would achieve transparent results, meaning no perceivable deterioration in intelligibility and quality.

The **Opus16v** condition attained a mean bitrate of about 31 kbit/s and was only tested in subjects ID5 to ID10 (see Table II). This codec was introduced because in the first four subjects, the **Opus16c** condition obtained very poor

https://doi.org/10.1121/10.0003571

TABLE II. Demographics of the CI-subjects as well as the SNR used throughout their speech recognition test. The tested side was always the better ear. Minimum, maximum, and median values of the dynamic ranges in clinical current units (CU) of the subjects' maps are specified as well.

ID	Participant's gender (age)	Tested side	Electrode type	Number of active electrodes	Dynamic range (CU) (min/max/median)	SNR (dB)
ID01	M (82)	Right	CI512	22	31/48/44	20
ID02	M (66)	Right	CI24R (CA)	20	57/77/70.5	8
ID03	M (76)	Left	CI522	22	30/76/57	1
ID04	M (73)	Right	CI24RE	21	36/84/69	3
ID05	M (72)	Right	CI24RE	20	43/54/52	5
ID06	F (71)	Right	CI24RE	20	40/64/51.5	5
ID07	M (50)	Right	CI24RE	22	61/74/69.5	0
ID08	M (78)	Right	CI512	19	52/66/65	6
ID09	F (49)	Right	CI522	20	51/71/65.5	3
ID10	F (76)	Right	CI522	20	37/64/54	0

speech performance, with many subjects understanding only a few words and many performing below 20% word recognition score. The only difference between the **Opus16v** condition and the **Opus16c** condition was that the **Opus16v** condition was set to 16 kbit/s with variable bitrate, allowing Opus to increase its bitrate above the nominal value of 16 kbit/s. It was expected that this condition would perform significantly better than the **Opus16c** condition.

E. Subjects

In total, ten subjects participated in the study, of which seven were males and three females. Mean age of the participants was 69.3 yrs. Except for one subject, the better ear was always the right ear. Detailed information about the participants is listed in Table II. All subjects gave informed consent to the project as approved by the Medical University Hannover Institutional Review Board.

F. Stimuli

The stimuli presented to the participants were created using the behind-the-ear head-related-transfer function from Denk *et al.* (2018), which simulates the impact of the sound propagation from the source to the microphones on the ears, including the head. In this simulated acoustic scenario, the virtual listener was positioned at a distance of 0.8 m from the speech source, which was located in front of the speaker at 0° azimuth. The noise source was positioned at a distance of 0.8 m at +90° (–90°) azimuth if the better ear was on the left (right) side such that the noise source was always on the opposite side of the better ear side. That way, the noise in the presented stimuli was always shaped by the head-related-transfer function. With the behind-the-ear head-related-transfer function, the speech and noise signals were mixed at the better ear. This mixture will be referred to as source audio signal. The generation of the electrodograms used in the study is depicted in Fig. 3. All stimuli for each listener were created prior to the testing session.

The source audio signal was processed by ACE without any additional coding method applied to create the reference electrodograms, labeled as the **REF** condition.

The source audio signal was encoded and decoded by Opus, as it would occur in a bilateral communication scenario, and then processed by ACE, resulting in the **Opus16c** to **Opus52v** conditions, depending on the setting of Opus used for the signal creation as listed in Table I. The source audio signal was processed by ACE, and the resulting electrodograms were then encoded and decoded by the Electrocodec to create the **EC2** to **EC7** conditions. These electrodograms were then used in the study and presented monaurally to the better CI-side of the subjects by streaming the electrodograms to the subject's CI processor. The streaming was performed using the nucleus implant communicator in a laboratory setting. The channel stimulation rate was fixed at 900 pps with eight subbands selected ($N = 8$). Phase duration was set to 25 μs, which all subjects also used in their clinical maps. All signal processing, including the streaming of the electrodograms (in combination with the nucleus implant communicator) was performed with

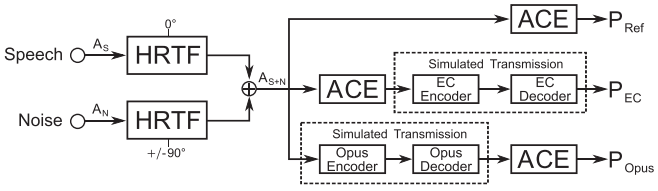

FIG. 3. Block diagram of the generation of the test signals presented in the study. First, speech and noise signals, denoted by A_S and A_N, respectively, were processed by head-related transfer functions (HRTF). +90° was used when the better ear was on the left, and –90° was used when it was on the right. Then the output signals were mixed. The values next to the HRTF boxes denote the angle at which the respective source was positioned. The resulting audio signal A_{S+N} was then processed in three different processing paths. The reference electrodograms P_{ref} were generated by processing A_{S+N} directly by ACE as defined in Fig. 1. For the Electrocodec (EC) conditions, the electrodograms P_{EC} were generated by first processing A_{S+N} by ACE (refer to Fig. 1) and then encoding and decoding the generated electrodograms by the Electrocodec. This simulated the processing chain necessary in a communication scenario between the ears if the Electrocodec was used. For the Opus conditions, the electrodograms P_{Opus} are generated by first encoding and decoding the audio signal A_{S+N}. This simulated the processing chain of a communication scenario between the ears if the Opus codec was used. The resulting audio signal was then processed by ACE, creating the electrodograms of Opus. For the speech quality test, the signal creation was identical, except that no noise was applied at the input. The respective electrodograms were then streamed to the subject's CI processor.

MATLAB, except for the creation of the audio signals coded by Opus. For these, Opus-tools 0.2 were used, which incorporate Opus 1.3 (IETF Codec Working Group, 2018b).

G. Speech material

The test material used was the Hochmair–Schulz–Moser sentence test (HSM) (Hochmair-Desoyer *et al.*, 1997). It consists of 30 lists, each consisting of a total of 106 words in 20 everyday sentences ranging in length between three and eight words. The background noise used in this study was the Consultatif International Téléphonique et Télégraphique (CCITT) noise according to Rec. G.227 (International Telecommuncation Union, 1993). The speech and noise material were mixed as described in Sec. II F.

H. Test procedure

Two experiments were performed with every subject. The first experiment consisted of a speech recognition test, and the second experiment was a speech quality test. The speech quality test was performed based on the multiple stimuli with hidden reference and anchor (MUSHRA) test (International Telecommuncation Union, 2015). All tests were performed monaurally, using the best performing ear in the case of bilateral CIs. If a subject used a different channel stimulation rate from the one selected for our study, first a fitting of the current levels with the new channel stimulation rate was performed. In the fitting procedure, the threshold level and most comfortable level were first decreased by a constant value and then gradually increased in steps of 2%. Before every increase, the subject was presented an example sentence from the HSM, and the subject was asked to report the perceived loudness. The perceived loudness was categorized using a clinical categorical loudness scale ranging from 0 (silence) to 10 (extremely loud) points, where it was aimed to achieve approximately 6 points, which corresponds to a loud but comfortable level where the subject had no difficulty understanding the speech presented. After the fitting procedure, the newly found values of the threshold level and most comfortable level remained fixed and unchanged throughout all experiments performed. Only the subjects ID01 and ID05, as specified in Table II, required a new fitting procedure as described in this section. The reason was a channel stimulation rate in their clinical map that differed from the 900 pps used in our study.

1. Speech recognition test

The word recognition score for each condition described in Sec. II D was tested to evaluate the impact on speech recognition of the respective codecs in noise. The word recognition score measures speech recognition by counting the correctly identified words from presented speech stimuli.

The participants were first trained ("warm-up") in the **REF** condition using the first few sentence lists of the HSM. Then the SNR was gradually decreased to identify the noise level at which the subjects understood around 70% of the words to avoid floor and ceiling effects. This SNR was used throughout the speech recognition test for all conditions. The lists used in this procedure were excluded from the actual speech recognition test. For each condition, two lists of the HSM were used. The lists were presented in random order, without the participant or the experiment conductors knowing the presented condition ("double blind").

2. Speech quality test

A MUSHRA test was performed to evaluate the impact of the different codecs on speech quality. In the MUSHRA test, the uncoded original is presented together with several encoded versions of the same signal. The listener is asked to rate the coded signals on a scale of 0–100 MUSHRA points, while no indication is given which signal belongs to which version. The listener can switch between all signals and listen to them repeatedly. The difference between the coded signals and the original should be evaluated. Among the signals to be evaluated are another copy of the uncoded original (the hidden reference) as well as several anchor signals.

The MUSHRA test was performed with noiseless speech material. The test was done without background noise, because a deterioration of quality is difficult to estimate if significant background noise is present as well. For the MUSHRA test, six sentences from the HSM were used, which were not presented during the speech recognition test. In total, eight conditions were tested: the four conditions of the Electrocodec, the **Opus16c** and **Opus52v** conditions of the Opus codec, one for the hidden reference, and one for the hidden anchor.

The **ANCHOR** condition was created using the Electrocodec with two quantization levels (1 bit) and subsequent deactivation of all subbands encoding frequencies of 850 Hz and above. This ensured very poor quality of the anchor. The MUSHRA test was repeated twice for each subject. Every repetition consisted of six sentences from the HSM for which all conditions were presented. The subjects then had to rate the speech quality of the conditions with respect to the reference condition. Combined, these two repetitions yielded 12 ratings per condition for each subject.

I. Objective intelligibility measure

As subjective listening tests are time and cost intensive, several algorithms have been proposed to objectively estimate the intelligibility of speech signals. While usually these metrics were designed for clean speech in additive background noise scenarios, some also work well for the assessment of the intelligibility of noisy speech processed by a CI. In these cases, the electrical stimulation patterns of a CI are resynthesized using a vocoder to create an audio waveform. These vocoded audio waveforms are then compared to the original unprocessed speech signals using any of the objective measures known from literature (Chen and Loizou, 2011). For our study, we selected the STOI (Taal *et al.*, 2010), which applied to vocoded speech files is labeled vocoder short-time objective intelligibility measure

https://doi.org/10.1121/10.0003571

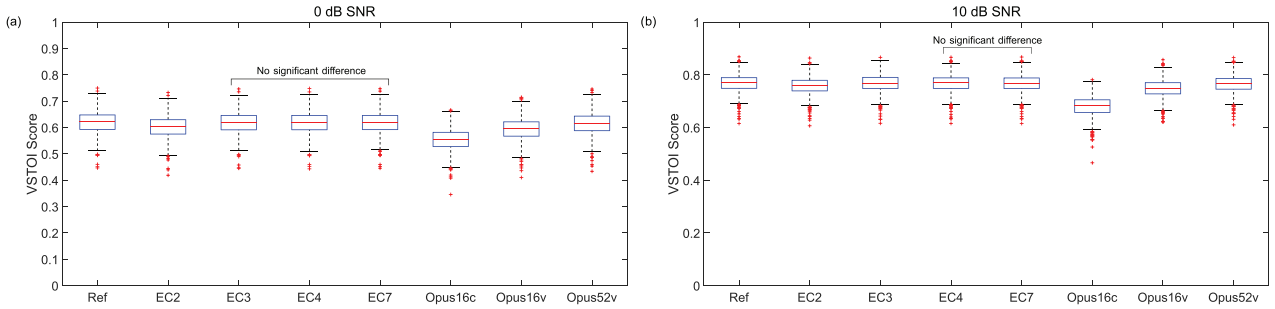

FIG. 4. Box-whisker plots of the VSTOI score of the entire HSM across all testing conditions for a SNR of (a) 0 dB and (b) 10 dB. At all other SNRs tested, the relative order between the conditions was the same, only the absolute value increased. At 0 dB, there was no significant difference between the **EC3**, **EC4**, and **EC7** conditions. At 10 dB, applying Bonferroni's correction, there was no significant difference between the **EC4** and **EC7** conditions. In all other cases, every condition was significantly different from all others. The red horizontal bars indicate the median result across the entire HSM. The bottom and top edge of the boxes indicate the 25% and 75% percentiles. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using red crosses.

(VSTOI). While STOI was originally developed for normal hearing listeners, it performs well for CI users too (Falk *et al.*, 2015).

To obtain the objective results of this study, the electrodograms of the test conditions, generated according to Fig. 3, were resynthesized using a sine vocoder with a threshold level of 80 and a most comfortable level of 150. The original noiseless, clean speech signals served as reference signals. Then the resynthesized audio waveforms were compared to the reference signals using STOI, yielding the respective VSTOI scores as done by Watkins *et al.* (2018). For all conditions, all 600 sentences of the HSM were used, resulting in 600 individual VSTOI scores per condition.

III. RESULTS

First, the results of the objective instrumental evaluation using VSTOI and the bitrate across SNRs for all tested codecs are presented. Second, the results of the subjective listening tests are presented.

A. Objective results

1. VSTOI

Results of the objective evaluation of the intelligibility of the test conditions are depicted in Fig. 4. Shown are the results for 0 and 10 dB SNRs for the entire 600 recordings of the HSM. These results were obtained using the ACE configuration as described in Sec. II A. For all other SNRs between 0 and 10 dB, the relative scores were the same, only the absolute values increased with increasing SNR. The median scores VSTOI scores are listed in Table III. Additionally, to give a rough idea of a corresponding word recognition, the median VSTOI scores mapped to word recognition scores are listed as well. The mapped word recognition scores were obtained by applying a logistic function to the median VSTOI scores as described in Taal *et al.* (2011).

A one-way analysis of variance (ANOVA) was performed as well as a Wilcoxon signed-rank test to investigate the VSTOI scores of the study conditions. The one-way ANOVA found a significant effect of testing condition with $F(7, 4792) = 159.02$ and $p < 0.001$ for 0 and 10 dB.

28 signed-rank tests were performed to investigate median differences between each pair of the tested conditions. After applying Bonferroni's correction to the significance level of $p < 0.05$, the new threshold of significance was $p/28 = 0.0018$. The signed-rank test revealed that all conditions achieved significantly different VSTOI scores except for the **EC3**, **EC4**, and **EC7** conditions at 0 dB SNR and the **EC4** and **EC7** conditions at 10 dB. These conditions are marked in Fig. 4.

2. Bitrates across signal-to-background-noise ratios

Figure 5 shows the mean bitrates of the different conditions across SNRs ranging from 0 to 10 dB for all sentences of the HSM. The **Opus16c** condition is not shown, as it uses a constant bitrate of 16 kbit/s independently of the audio signal. This range covers all SNRs tested in the study except that of subject one (20 dB). At 0 dB SNR, the **EC3** condition achieved a bitrate of 30.6 kbit/s, while the **Opus16v** condition achieved a bitrate of 31 kbit/s. All conditions, except

TABLE III. Median VSTOI scores for all conditions across the entire 600 recordings of the HSM for 0 dB SNR and 10 dB SNR. The VSTOI scores were mapped to word recognition scores (WRS) using a logistic function as described in Taal *et al.* (2011). To obtain the parameters of this logicistic function, the word recognition scores of the CI-subjects obtained for the **REF** condition were used.

	0 dB		10 dB	
Condition	Median VSTOI score	Mapped WRS (%)	Median VSTOI score	Mapped WRS (%)
REF	0.621	66.8	0.77	81.1
EC2	0.603	64.8	0.759	80.2
EC3	0.62	66.7	0.768	81.0
EC4	0.619	66.6	0.769	81.0
EC7	0.619	66.6	0.769	81.0
Opus16c	0.555	59.0	0.682	73.3
Opus16v	0.595	63.8	0.748	79.3
Opus52v	0.617	66.4	0.765	80.7

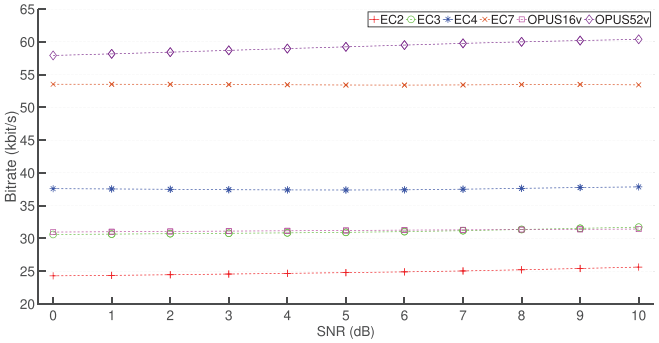

FIG. 5. Mean bitrates for the HSM of the **EC2** to **EC7** conditions as well as the **Opus16v** and **Opus52v** conditions. The **Opus16c** condition is not shown, as it codes at a constant bitrate of 16 kbit/s. The **EC3** and **Opus16v** conditions are almost coinciding. The bitrates are shown for a SNR ranging from 0 to 10 dB. This range covers all SNRs used in the study except for the SNR of the first subject.

for the **EC4** and **EC7** conditions, showed slightly increasing bitrate with increasing SNR. The calculation of the bitrate for the Electrocodec assumed the presence of 22 electrodes. For the Opus codec, the raw bitrate was used. This is the bitrate required solely for the coding of the signal excluding any additional packet information.

B. Evaluation in CI users

The results of the speech recognition test are shown in Fig. 6. Because the **Opus16v** condition was introduced after four subjects had been assessed, it was only evaluated in six subjects. The results of the speech quality test are shown in Fig. 7.

1. Speech recognition in CI users

The median word recognition score of the **REF** condition was 73%. For the Electrocodec, the median word recognition scores of the **EC2** to **EC7** conditions were 69, 71.5, 69, and 72%, respectively. The median word recognition scores

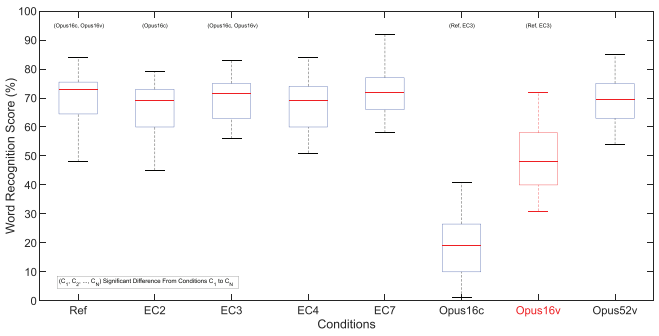

FIG. 6. Results of the speech recognition test for all conditions evaluated. The **Opus16v** condition was introduced after evaluating the results of the first four subjects and evaluated in six subjects only. All other conditions were evaluated in ten subjects. The labels above the box plots denote significant differences from other conditions for which a Wilcoxon signed-rank test was performed. The results are shown as box-whisker plots. The red bar indicates the median across all subjects. The bottom and top edge of the boxes indicate the 25% and 75% percentiles. The whiskers extend to the most extreme data points not considered outliers. No outliers occurred.

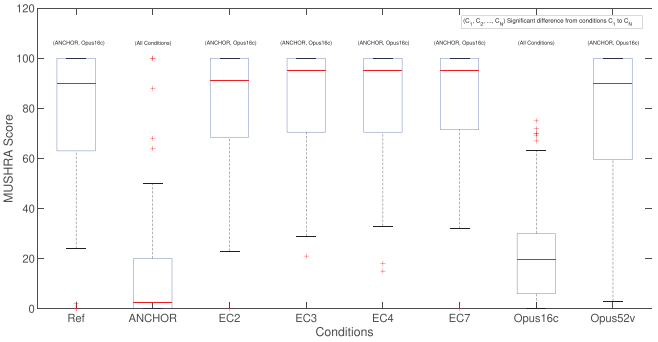

FIG. 7. Results of the speech quality test for all conditions evaluated. All conditions were evaluated in ten subjects. Note that speech quality was not evaluated for the **Opus16v** condition. The labels above the box plots denote significant differences from other conditions for which a Wilcoxon signed-rank test was performed. The red horizontal bars indicate the median results across all ten subjects. The bottom and top edge of the boxes indicate the 25% and 75% percentiles. The whiskers extent to the most extreme data points not considered outliers, and the outliers are plotted individually using red crosses.

for Opus conditions **Opus16c**, **Opus16v**, and **Opus52v** were 19%, 48%, and 69.5%, respectively. Individual word recognition scores in percent are given in Table IV. The IDs correspond to the IDs given in Table II. The **Opus16v** condition was not tested in the first four subjects. Except for subject ID08, the **EC2** condition always achieved a higher word recognition score than the **Opus16v** condition.

A one-way repeated measures ANOVA was performed as well as a Wilcoxon signed-rank test to investigate the results of the speech recognition test. The ANOVA found a significant effect of testing condition with $F(7, 63) = 71.98$ and $p < 0.001$. Fourteen signed-rank tests were performed to investigate median differences between each pair of interest of the tested conditions. The results are shown in Table V. Values in boldface show significant differences after applying Bonferroni's correction to the significance level of $p < 0.05$. The new threshold of significance after applying Bonferroni's correction was p/14 = 0.003 57. No

TABLE IV. Word recognition scores in percent across conditions for each tested subject. The listed IDs correspond to the subject IDs given in Table II. The **Opus16v** condition was tested in six subjects only. For each subject and each condition, the average word recognition scores across the two tested sentence lists are reported. N/A, not applicable.

	Conditions							
ID	REF	EC2	EC3	EC4	EC7	Opus16c	Opus16v	Opus52v
ID01	60.5	76.5	58	64.5	72	31.5	N/A	73
ID02	80	70	66.5	74	71	30	N/A	70.5
ID03	73.5	62	65	61.5	69.5	19	N/A	64
ID04	63	48	77	67.5	72	21.5	N/A	65.5
ID05	69	68	73	61	81	4.5	34.5	77.5
ID06	68.5	60	63	61	75	27	44	65
ID07	67	73.5	79	69	71	9.5	57.5	67.5
ID08	78	60.5	67.5	79	68.5	20	67.5	82.5
ID09	74	70.5	77	67	74	27.5	51.5	68.5
ID10	75	70.5	72.5	65	72	7	38	63

https://doi.org/10.1121/10.0003571

TABLE V. Results for the Wilcoxon signed-rank test for the results of the speech recognition test. Shown are the compared conditions in the columns labeled with A and B and the calculated p values in the column labeled with p. Values in boldface show significant values after applying Bonferroni's correction to the significance level of $p < 0.05$. The **Opus16v** condition was only tested in six subjects. The Wilcoxon signed-rank tests involving this condition were performed based on the results of the subset of these six subjects.

Conditions		
A	B	p
EC2	**REF**	0.048
EC3	**REF**	0.575
EC4	**REF**	0.198
EC7	**REF**	0.614
Opus16c	**REF**	**<0.001**
Opus16v	**REF**	**<0.001**
Opus52v	**REF**	0.433
EC2	**Opus16c**	**<0.001**
EC2	**Opus16v**	0.004
EC3	**Opus16v**	**0.002**
EC2	**Opus52v**	0.211
EC3	**Opus52v**	0.809
EC4	**Opus52v**	0.279
EC7	**Opus52v**	0.239

further comparisons were made to avoid further reducing the power of the statistical analysis.

2. Speech quality in CI users

The median MUSHRA score of the **REF** condition was 90. The **EC2** condition achieved a slightly higher median MUSHRA score of 91. The **EC3**, **EC4**, and **EC7** conditions achieved identical median MUSHRA scores of 95. The **Opus52v** condition achieved a median MUSHRA score of 90, identical to the **REF** condition. In contrast, the **Opus16c** condition achieved a significantly lower median MUHSRA score of 19.5, which is close to the **ANCHOR** condition, which achieved a median MUSHRA score of 2.5. For all conditions, a few subjects gave ratings either significantly below or significantly above the median MUSHRA scores of the respective conditions. But for all conditions, except for the **ANCHOR** and the **Opus16c**, about 75% of the ratings of the subjects were above a MUSHRA score of 60.

For the speech quality test, a one-way repeated measures ANOVA found an effect of tested condition $F(7, 63) = 264.1389$ with $p < 0.001$. The Wilcoxon signed-rank test revealed significant differences between the **REF** and the **Opus16c** and the **ANCHOR** condition ($p < 0.001$) as well as the **Opus16c** and the **ANCHOR** condition ($p < 0.001$). No significant difference was found between the **EC2** and the **REF** condition ($p > 0.05$). The mean results for the **EC2** to **EC7** conditions suggest no perceived reduction of speech quality. For Opus, the **Opus16c** condition performed slightly better than the **ANCHOR** condition. Both showed a significant difference from the reference condition ($p < 0.001$). The **Opus52v** condition was not significantly different from the reference condition ($p > 0.05$), achieving equal speech quality.

IV. DISCUSSION

In this study, we investigated the impact of the signal distortion introduced by the Electrocodec on speech recognition and speech quality. For this purpose, the Electrocodec was tested in ten CI-subjects and compared to the Opus audio codec. Additionally, the Electrocodec was evaluated using the objective instrumental measure VSTOI. The Electrocodec surpassed or matched the Opus audio codec with respect to word recognition score while achieving lower bitrate and lower latency.

A. Objective instrumental measures

The results of the objective instrumental evaluation of speech recognition using VSTOI as shown in Fig. 4 appear to qualitatively agree with the findings of the speech performance in CI users presented in this paper. However, the statistical analysis revealed significant differences between, e.g., the **REF** condition and the **EC3** condition and the **REF** condition and the **Opus52v** condition, which is not in agreement with the speech recognition test. This was caused by the large number of sentences assessed that covered the entire HSM-set. Due to this, the corresponding confidence intervals shrink until almost any median difference is significant. However, such small differences in VSTOI scores are unlikely to be observed in speech recognition tests. The mapping of the VSTOI scores to corresponding word recognition scores as shown in Table III, while useful to give a general idea of the meaning of the scores, is not going to be generally accurate. A word recognition score of, e.g., 59% for the **Opus16c** condition at 0 dB is in disagreement with the observed performance in the subjective tests. The reason is the speech recognition of the CI-subjects in the **REF** condition, which was aimed to be around 70% for all subjects. Therefore, no very high (>90%) or low (<40%) word recognition scores were measured in the **REF** condition. This made a reasonable fit of the parameters of the mapping function that maps the VSTOI scores to word recognition scores difficult.

Leaving aside the discussed statistical analysis, the usefulness of VSTOI for an optimization of the Electrocodec relies on VSTOI following the general trends of the subjective evaluations. The VSTOI score of condition **EC2** suggests a minor decrease in speech recognition compared to the other conditions of the Electrocodec, which, while not significant, is in accordance with the evaluation in CI users. The VSTOI score suggests a slightly poor performance of the **Opus16v** condition compared to the **EC2** condition, albeit the difference is very small. The VSTOI scores rank the **Opus16c** to **Opus52v** conditions as observed in the speech recognition test, i.e., rating the **Opus16c** condition significantly below the **Opus16v** and the **Opus52v** condition with a VSTOI score virtually equal to the **REF** condition. These observations suggest that VSTOI can be useful to

assess performance differences within the same type of algorithm but also for comparing different types of algorithms with each other. Therefore, the Electrocodec could be optimized or improved with respect to speech understanding by maximizing the VSTOI score of its coded signals.

B. Speech recognition

The results indicate that there is no significant reduction in speech recognition and quality for either of the tested conditions of the Electrocodec ranging from bitrates of about 24 kbit/s (**EC2**) to 55 kbit/s (**EC7**). All conditions of the Electrocodec showed significantly higher speech recognition scores than the **Opus16c** and the **Opus16v** condition, except for the **EC2** condition, which after applying Bonferroni's correction showed almost significantly better results than the **Opus16v** condition. Although the **Opus16v** was only evaluated in a subset of six subjects, only once and only on a single sentence list a subject (ID08) obtained a better performance with the **Opus16v** condition than with the **EC2** condition. In all other cases, the **EC2** condition performed consistently better, even on a list per list basis and at a lower mean bitrate. The **Opus16v** condition obtained a significantly higher mean word recognition score compared to the **Opus16c** condition but showed a significant decrease ($p < 0.001$) in word recognition score compared to the **REF** condition. No significant difference in speech recognition was observed between the **EC3** and **EC4** conditions ($p = 0.341$). The **EC7** condition achieved transparent performance ($p = 0.614$). The **Opus16c** condition achieved a very low word recognition score, with some subjects understanding only isolated words, and achieved a significantly ($p < 0.001$) worse result than the **REF** condition. The **Opus52v** condition showed transparent performance ($p = 0.4327$). All conditions tested for the Electrocodec outperformed the **Opus52v** condition in terms of bitrate while showing no significant decrease in terms of both speech recognition and speech quality compared to the **REF** condition.

C. Speech quality

The speech quality test showed that all conditions except the **Opus16c** and the **ANCHOR** condition performed transparently, i.e., without significant difference between the conditions. The **ANCHOR** and **Opus16c** conditions were consistently identified by all CI users and obtained the lowest rating.

The performance of the Electrocodec in both subjective listening tests supports the study of Qazi *et al.* (2013) that found CI users can tolerate large distortion of the subband envelopes and are far more sensitive to distortions of the band selection, which is not distorted by the Electrocodec, i.e., encoded losslessly. The potential small decrease in speech recognition of the **EC2** condition, though not significant, might actually stem from an increased noise level (due to the coding noise) in the speech gaps, which was shown to have significantly stronger detrimental effects on speech understanding for CI users than distortions of the envelope signals (Kressner *et al.*, 2019; Qazi *et al.*, 2013).

D. Bitrate

Figure 5 shows that all conditions except **EC4** and **EC7** exhibit a slight increase in bitrate with an increase in SNR. In terms of bitrate, the **EC2** condition considerably outperforms the **Opus16v** using about 7 kbit/s less while achieving equal or better speech recognition at a lower latency. At 0 dB SNR, the bitrate of the **EC3** and the **Opus16v** condition were virtually identical, with 30.6 kbit/s for the **EC3** and 31.0 kbit/s for the **Opus16v** condition, respectively. Both conditions also achieve similar bitrates at 10 dB SNR, with 31.7 kbit/s for the **EC3** and 31.4 kbit/s for the **Opus16v** condition.

For the Electrocodec, the slight increase in bitrate of the **EC2**, **EC3**, and **EC4** with increasing SNR is due to the increased information content of the band selection, which increases the mean word length of the lossless coding part of the Electrocodec (see Sec. II B). For the **EC7** condition, this effect is counteracted by the decrease in the mean word length of the lossless coding of the quantization indices of the DPCM. When the SNR increases, the acoustic scenario gets more similar to the noiseless training scenario used to learn the quantizer codebooks, and the lossless coding becomes more effective. Because in the **EC7** significantly more quantizer levels are used in the **EC2** and **EC3** conditions, this effect becomes more prominent.

E. Comparison to other codecs for hearing devices

This paper and the proposed Electrocodec deal with bitrate reduction in the context of audio streaming from external devices to CIs or between CIs. In a previous work (Roy and Vetterli, 2007), the contralateral audio signal of bilateral hearing devices was reconstructed by only coding and transmitting interaural level differences and interaural time differences. This method achieved a bitrate of 8 kbit/s, but at an algorithmic latency of 28 ms, which is too high, and the speech quality was only informally assessed. Furthermore, the approach is applicable in simple acoustic scenarios only. In Li and Kleijn (2007), an audio codec based on predictive coding combined with entropy coding similar to the Electrocodec, with an algorithmic latency of only 0.25 ms and a bitrate of 32 kbit/s for wideband speech, was proposed for audio streaming in hearing devices. While outperforming the G.722 audio codec, it was only assessed in clean speech, and its bitrate is significantly larger than the bitrate of the **EC2** condition. Another approach to bitrate reduction could be the combination of narrowband audio codecs like the BroadVoice16 (Chen and Thyssen, 2007), which exhibits an algorithmic latency of 5 ms, in combination with artificial bandwidth extension algorithms (Nogueira *et al.*, 2019) to achieve lower bitrates at improved speech recognition despite removing part of the speech information by reducing the bandwidth from 8 kHz down to approximately 3.4 kHz. This approach could achieve lower

bitrates than the Electrocodec, albeit at a higher latency and a decreased speech recognition, as the bandwidth extension algorithm of (Nogueira *et al.*, 2019) does not achieve the performance of the reference fullband condition.

F. Future work

The current study investigated the impact of some coding algorithms on speech recognition and speech quality in CIs in only one condition using CCITT noise. While CCITT noise is speech-shaped, in real scenarios the power spectrum of the background noise will not be stationary and can differ significantly from the CCITT. Furthermore, a single clinical speech set was used. Real speaker's prosodies vary so that the impact of, for example, differences in speech gaps could be vastly different from the performed speech recognition test of the presented study. Additionally, scenarios with more realistic environmental conditions, e.g., some level of reverberation, background noises from several positions, or the impact of nonideal wireless channels (Kozma-Spytek *et al.*, 2019), are yet to be investigated. Further bitrate reductions of the Electrocodec could be achieved by combining it with sound coding strategies such as the psychoacoustic advanced combination encoder (PACE) (Nogueira *et al.*, 2005). PACE improves speech understanding at a decreased number of selected subbands through an improved subband selection method. This could allow one to decrease the bitrate of the Electrocodec further by decreasing the number of selected subbands that need to be encoded. Finally, the Electrocodec will be evaluated in combination with the binaural sound coding strategy of Gajecki and Nogueira (2018) to investigate its benefits for binaural sound coding.

V. CONCLUSION

In this work, a subjective and objective evaluation of a source coding algorithm for the electrical stimulation patterns of CIs, called Electrocodec, was presented. The Electrocodec was compared to the Opus audio codec using a speech recognition, and a speech quality test was performed in ten CI users. Results from the study show no statistically significant reduction of speech recognition for either of the evaluated bitrates of the Electrocodec. For the Opus codec, except for the highest bitrate setting tested of about 58 kbit/s, a statistically significant reduction of the speech recognition was observed. For the speech quality test, the results indicated no difference between the reference signal and the coded signals of any of the codecs tested, except for the lowest bitrate setting of the Opus codec. The results show that the Electrocodec exhibits no reduction in speech recognition and speech quality at 24.3 kbit/s, while achieving an algorithmic latency of 0 ms compared to 5 ms for the Opus codec.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2177/1-Project ID 390895286.

Allamanche, E., Geiger, R., Herre, J., and Sporer, T. (**1999**). "MPEG-4 low delay audio coding based on the AAC codec," in *Proceedings of the 106th AES Convention*, 1 May 1999, Munich, Germany.

Boddy, C., and Datta, G. (**2018**). "The use of the cochlear mini microphone (MM) as a personal radio system (FM) with young children who are deaf," Cochlear Implants Int. **19**(6), 330–338.

Böhmler, E., Freudenberger, J., and Müller, M. (**2010**). "Comparison of SBC and G.722 speech codecs for Bluetooth wideband speech transmission," in *Proceedings of ITG Symposium Speech Communication*, 8 October 2010, Bochum, Germany.

Ceulaer, G., Bestel, J., Mülder, H., Goldbeck, F., Janssens de Varebeke, S., and Govaerts, P. (**2016**). "Speech understanding in noise with the Roger Pen, Naida CI Q70 processor, and integrated Roger 17 receiver in a multitalker network," Eur. Arch. Otorhinolaryngol. **273**, 1107–1114.

Chen, F., and Loizou, P. (**2011**). "Predicting the intelligibility of vocoded speech," Ear Hear. **32**(3), 331–338.

Chen, J.-H., and Thyssen, J. (**2007**). "The broadvoice speech coding algorithm," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 15–20, Honolulu, HI, Vol. 4, pp. IV–537–IV–540.

Denk, F., Ernst, S., Ewert, S., and Kollmeier, B. (**2018**). "Adapting hearing devices to the individual ear acoustics: Database and target response correction functions for various device styles," Trends Hear. **22**, 233121651877931.

Edelmann, J.-C., and Ussmueller, T. (**2018**). "Can you hear me now?: Challenges and benefits for connectivity of hearing aids and implants," IEEE Microwave Mag. **19**(7), 30–42.

Edler, B., Büchner, A., Nogueira, W., and Klefenz, F. (**2007**). "Cochlear implant, device for generating a control signal for a cochlear implant, device for generating a combination signal and combination signal and corresponding methods," International patent WO/2007/033762 (2006-2007).

Ernst, A., Baumgaertel, R., Diez, A., and Battmer, R.-D. (**2019**). "Evaluation of a wireless contralateral routing of signal (CROS) device with the advanced bionics Naída CI Q90 sound processor," Cochlear Implants Int. **20**(4), 182–189.

European Broadcasting Union (**2008**). "EBU SQAM CD—Sound Quality Assessment Material Recordings for Subjective Tests," https://tech.ebu.ch/publications/sqamcd (Last viewed 11.10.2018).

European Telecommunications Standards Institute (**2013**). "Electromagnetic compatibility and radio spectrum matters (ERM); system reference document; short range devices (SRD); technical characteristics of wireless aids for hearing impaired people operating in the VHF and UHF frequency range," Technical Report ETSI TR 102 791 V1.2.1 (2013-08), https://www.etsi.org/deliver/etsi_tr/102700_102799/102791/01.02.01_60/tr_102791v010201p.pdf (Last viewed 12/21/2020).

European Telecommunications Standards Institute (**2018**). "Digital enhanced cordless telecommunications (DECT); study of super wideband codec in DECT for narrowband, wideband and super-wideband audio communication including options of low delay audio connections ($\leq$ 10 ms framing)," Technical Report ETSI TR 103 590 V1.1.1, https://www.etsi.org/deliver/etsi_tr/103500_103599/103590/01.01.01_60/tr_103590v010101p.pdf (Last viewed 11.06.2020).

Falk, T., Parsa, V., Santos, J., Arehart, K., Hazrati, O., Huber, R., Kates, J., and Scollie, S. (**2015**). "Objective quality and intelligibility prediction for users of assistive listening devices," IEEE Signal Process. Mag. **32**(2), 114–124.

Gajecki, T., and Nogueira, W. (**2018**). "A synchronized binaural N-of-M sound coding strategy for bilateral cochlear implant users," in *Proceedings of the 13th ITG Symposium on Speech Communication, VDE*, October 10–12, Oldenburg, Germany, pp. 1–5.

Gajecki, T., and Nogueira, W. (**2020**). "The effect of synchronized linked band selection on speech intelligibility of bilateral cochlear implant users," Hear. Res. **396**, 108051.

Galster, J. (**2010**). "A new method for wireless connectivity in hearing aids," Hear. J. **63**(36), 38–39.

Gifford, R. H., Shallop, J. K., and Peterson, T. A. (**2008**). "Speech recognition materials and ceiling effects: Considerations for cochlear implant programs," Audiol. Neurotol. **13**(3), 193–205.
Hinrichs, R., Gajęcki, T., Ostermann, J., and Nogueira, W. (**2019**). "Coding of electrical stimulation patterns for binaural sound coding strategies for cochlear implants," in *Proceedings of the 41st IEEE Engineering in Medicine and Biology Society (EMBC)*, July 23–27, Berlin, pp. 4168–4172.
Hochmair-Desoyer, I., Schulz, E., Moser, L., and Schmidt, M. (**1997**). "The HSM sentence test as a tool for evaluating the speech understanding in noise of cochlear implant users," Am. J. Otol. **18**(6 Suppl), S83.
Hoene, C., and Hyder, M. (**2010**). "Optimally using the Bluetooth subband codec," in *Proceedings of the IEEE Local Computer Network Conference*, October 10–14, Denver, CO, pp. 356–359.
IETF Codec Working Group (**2018a**). "Opus 1.3," https://opus-codec.org/ (Last viewed 15.11.2019).
IETF Codec Working Group (**2018b**). "Opus-Tools 0.2," https://opus-codec.org/release/dev/2018/09/18/opus-tools-0_2.html (Last viewed 10.09.2019).
International Telecommuncation Union (**1993**). "ITU Recommendation G.227," https://www.itu.int/rec/T-REC-G.227-198811-I/en (Last viewed 10.09.2019).
International Telecommuncation Union (**2015**). "ITU-RBS.1534-0 (Method for the subjective assessment of intermediate quality levels of coding systems)," https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1534-1-200301-S!!PDF-E.pdf (Last viewed 01.12.2019).
Jokisch, O., Maruschke, M., Meszaros, M., and Iaroshenko, V. (**2016**). "Audio and speech quality survey of the Opus codec in web real-time communication," in *Proceedings of the 27th Conference on Electronic Speech Signal Processing (ESSV)*, March, Leipzig, Germany, pp. 254–262.
Kan, A. (**2018**). "Improving speech recognition in bilateral cochlear implant users by listening with the better ear," Trends Hear. **22**, 2331216518772963.
Khalifeh, A. F., Al-Tamimi, A., and Darabkh, K. A. (**2017**). "Perceptual evaluation of audio quality under lossy networks," in *Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, March 22–24, Chennai, India, pp. 939–943.
Kokkinakis, K., Azimi, B., Hu, Y., and Friedland, D. R. (**2012**). "Single and multiple microphone noise reduction strategies in cochlear implants," Trends Amplif. **16**(2), 102–116.
Korhonen, J., and Wang, Y. (**2005**). "Effect of packet size on loss rate and delay in wireless links," in *Proceedings of the IEEE Wireless Communications and Networking Conference, WCNC*, March 13–17, New Orleans, LA, Vol. 3, pp. 1608–1613.
Kozma-Spytek, L., Tucker, P., and Vogler, C. (**2019**). "Voice telephony for individuals with hearing loss: The effects of audio bandwidth, bit rate and packet loss," in *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 3–15.
Kramer, U., Schuller, G., Wabnik, S., Klier, J., and Hirschfeld, J. (**2004**). "Ultra low delay audio coding with constant bit rate," in *117th Audio Engineering Society Convention*, October 28–31, San Francisco.
Kressner, A. A., Anderson, D. V., and Rozell, C. J. (**2011**). "Robustness of the hearing aid speech quality index (HASQI)," in *Proceedings of the 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 209–212.
Kressner, A., May, T., and Dau, T. (**2019**). "Effect of noise reduction gain errors on simulated cochlear implant speech intelligibility," Trends Hear. **23**, 233121651982593.
Li, M., and Kleijn, W. (**2007**). "A low-delay audio coder with constrained-entropy quantization," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 21–24, New Paltz, NY, pp. 191–194.
Lopez-Poveda, E., Eustaquio-Martin, A., Stohl, J., Wolford, R., Schatzer, R., and Wilson, B. S. (**2016**). "A binaural cochlear implant sound coding strategy inspired by the contralateral medial olivocochlear reflex," Ear Hear. **37**(3), 138–148.
Mehrkian, S., Bayat, Z., Javanbakht, M., Emamdjomeh, H., and Bakhshi, E. (**2019**). "Effect of wireless remote microphone application on speech discrimination in noise in children with cochlear implants," Int. J. Pediatr. Otorhinolaryngol. **125**, 192–195.
Nogueira, W., Abel, J., and Fingscheidt, T. (**2019**). "Artificial speech bandwidth extension improves telephone speech intelligibility and quality in cochlear implant users," J. Acoust. Soc. Am. **145**(3), 1640–1649.
Nogueira, W., Büchner, A., Lenarz, T., and Edler, B. (**2005**). "A psychoacoustic 'NofM'-type speech coding strategy for cochlear implants," EURASIP J. Appl. Signal Process. **2005**, 3044–3059.
Ostergaard, J., Quevedo, D., and Jensen, J. (**2009**). "Low delay moving-horizon multiple-description audio coding for wireless hearing aids," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, April 19–24, Taipei, Taiwan, pp. 21–24.
Oticon (**2019**). "Oticon CROS" https://wdh01.azureedge.net/-/media/oticon/main/pdf/germany/cros/pbr/65949de_pbr_promotional_sheet_oticon_cros.pdf?&la=de-de&rev=03B9&hash=AEA390E8EA49233F01CB28085C65F1AB (Last viewed 21.12.2020).
Pal, A., and Kant, K. (**2019**). "NFMI: Connectivity for short-range IoT applications," Computer **52**(2), 63–67.
Phonak (**2012**). "Binaural voicestream technology," https://pdfs.semanticscholar.org/b12a/16d5808e841d0a1cf41be9051ce40edc7cda.pdf (Last viewed 31.03.2020).
Phonak (**2016**). Technical information Naída Link CROS, https://advancedbionics.com/content/dam/advancedbionics/Documents/Regional/DK/Datasheets/Datasheet_Phonak_Naida_Link_CROS_210x297_GB_V1.00.pdf (Last viewed 06.05.2020).
Preihs, S., Lamprecht, T., and Ostermann, J. (**2016**). "Error robust low delay audio coding using spherical logarithmic quantization," in *Proceedings of the 24th European Signal Processing Conference (EUSIPCO)*, pp. 1970–1974.
Qazi, O., van Dijk, B., Moonen, M., and Wouters, J. (**2013**). "Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility," Hear. Res. **299**, 79–87.
Qualcomm (**2020**). "aptX audio codec," https://www.aptx.com/ (Last viewed 31.03.2020).
Rämö, A., and Toukomaa, H. (**2011**). "Voice quality characterization of IETF Opus codec," in *Proceedings of the 12th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, August, Florence, Italy, pp. 2541–2544.
Roy, O., and Vetterli, M. (**2007**). "Distributed spatial audio coding in wireless hearing aids," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 21–24, New Paltz, NY, pp. 227–230.
Shannon, C. E. (**1949**). "Communication in the presence of noise," Proc. IRE **37**(1), 10–21.
Taal, C., Hendriks, R., Heusdens, R., and Jensen, J. (**2010**). "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, March 14–19, Dallas, TX, pp. 4214–4217.
Taal, C. H., Hendriks, R. C., Heusdens, R., and Jensen, J. (**2011**). "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," IEEE Trans. Audio Speech Lang. Process. **19**(7), 2125–2136.
Valin, J.-M., Maxwell, G., Terriberry, T., and Vos, K. (**2013**). "High-quality, low-delay music coding in the Opus codec," *Proceedings of the 135th Audio Engineering Society Convention 2013*, October 17–20, New York, pp. 73–82.
Watkins, G., Swanson, B., and Suaning, G. (**2018**). "An evaluation of output signal to noise ratio as a predictor of cochlear implant speech intelligibility," Ear Hear. **39**(5), 958–968.
Weder, S., Kompis, M., Caversaccio, M., and Stieger, C. (**2015**). "Benefit of a contralateral routing of signal device for unilateral cochlear implant users," Audiol. Neurotol. **20**(2), 73–80.
Wilson, R., McArdle, R., and Smith, S. (**2007**). "An evaluation of the BKB-SIN, HINT, QUICKSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss," J. Speech Lang. Hear. Res. **50**(4), 844–856.
Wolfe, J., Duke, M., and Schafer, E. (**2016a**). "Speech recognition of bimodal cochlear implant recipients using a wireless audio streaming accessory for the telephone," Otol. Neurotol. **37**, e20–e25.
Wolfe, J., Duke, M., Schafer, E., Cire, G., Menapace, C., and O'Neil, L. (**2016b**). "Evaluation of a wireless audio streaming accessory to improve mobile telephone performance of cochlear implant users," Int. J. Audiol. **55**(2), 75–82.

Wolfe, J., Morais, M., and Schafer, E. (**2015**). "Improving hearing performance for cochlear implant recipients with use of a digital, wireless, remote-microphone, audio-streaming accessory," J. Am. Acad. Audiol. **26**(6), 532–539.
Wouters, J., McDermott, H. J., and Francart, T. (**2015**). "Sound coding in cochlear implants: From electric pulses to hearing," IEEE Signal Process. Mag. **32**(2), 67–80.
Zeitler, D. M., Kessler, M. A., Terushkin, V., Roland, T., Jr., Svirsky, M. A., Lalwani, A. K., and Waltzman, S. (**2008**). "Speech perception benefits of sequential bilateral cochlear implantation in children and adults: A retrospective analysis," Audiol. Neurotol. **13**(3), 314–325.
Zirn, S., Arndt, S., Aschendorff, A., and Wesarg, T. (**2015**). "Interaural stimulation timing in single sided deaf cochlear implant users," Hear. Res. **328**, 148–156.