# Lossless Compression at Zero Delay of the Electrical Stimulation Patterns of Cochlear Implants for Wireless Streaming of Audio Using Artificial Neural Networks

Reemt Hinrichs
*Institut für Informationsverarbeitung*
*L3S Research Center*
*Leibniz University Hannover*
Hannover, Germany
hinrichs@tnt.uni-hannover.de

Lukas Ehmann
*Institut für Informationsverarbeitung*
*Leibniz University Hannover*
Hannover, Germany

Hendrik Heise
*Institut für Informationsverarbeitung*
*Leibniz University Hannover*
Hannover, Germany

Jörn Ostermann
*Institut für Informationsverarbeitung*
*Leibniz University Hannover*
Hannover, Germany

***Abstract*—Cochlear implants (CIs) are battery-powered, surgically implanted hearing-aids capable of restoring a sense of hearing in people suffering from moderate to profound hearing loss. To achieve this, audio signals captured by the microphone of the CI are processed by its signal processor and converted into electrical pulses, the stimulation patterns, which then excite certain areas of the cochlear. Nowadays wireless transmission of audio from external devices, like remote microphones and smartphones, is used to improve speech understanding and localization or for the convenience of the CI user. To conserve energy or channel capacity in this wireless transmission, data compression is commonly applied. In this work, zero delay lossless compression of the so called clinical units of the CIs is proposed and a zero delay lossless codec (ZDLLC) based on artificial neural networks is investigated for this purpose. The ZDLLC is compared to the lossless compression algorithms PAQ and PPM as well as the lossy Opus audio codec. On the TIMIT speech corpus and various acoustic scenarios the ZDLLC achieved a mean bitrate of 28.6 kbit/s at zero algorithmic latency compared to 33.6 kbit/s to 35.2 kbit/s for the Opus audio codec at 5 ms to 7.5 ms algorithmic latency. In contrast, at very high latency, PPM achieved a mean bitrate of 37.3 kbit/s and PAQ achieved a mean bitrate of 25.1 kbit/s. It was found that lossless compression of the stimulation patterns could be useful for wireless streaming of audio.***

***keywords*—lossless compression, cochlear implants, neural networks**

## I. Introduction

Cochlear implants (CIs) are battery-powered, surgically implanted hearing-aids capable of restoring a sense of hearing in people suffering from moderate to profound hearing loss. While good speech understanding is achieved in high speech-to-background noise environments, more challenging environments as encountered in common social situations like a restaurant setting still pose a problem [1]. Wireless streaming of audio can be used to improve speech understanding of CI users in these situations. Methods like wireless streaming of audio from smartphones, remote microphones [1], contralateral routing of signals or binaural sound coding strategies [2] require transmission of audio from external devices to a CI processor. To either save power or bandwidth in this wireless transmission, signal compression or coding is commonly applied to reduce the bitrate of the audio signal before transmission. This coding usually introduces an additional delay in the processing chain and thus has to be kept as small as possible, as hearing aid users cannot tolerate delays above the range of $5-10$ ms without affecting their speech perception [3]. The selection of source coding algorithms is severly limited due to this delay constraint. In CIs, an audio input is mapped to the electrical domain by the sound coding strategy of the signal processor. In this processing some irrelevant information is removed, and thus compression in the electrical domain allows to reduce bitrate. Previously, we proposed [4], [5] the Electrocodec to code and transmit the electrical stimulation patterns generated by the sound coding strategy of the CI. We proposed [4] a combination of differential pulse-code modulation (DPCM) and arithmetic coding to compress the current magnitudes and the band selection of the electrical stimulation patterns generated by the advanced combinational encoder (ACE) sound coding strategy. Using this approach, we achieved [5] lower bitrates and zero latency at equal or better speech understanding than the state-of-the-art Opus audio codec. However, in [4], [5] we coded the output of the so called loudness growth function (LGF), the logarithmic function that is used in CIs to map from the acoustic to the electrical

TABLE I

SPEECH AND NOISE AZIMUTHS, SIGNAL-TO-NOISE RATIOS (SNRS), NOISE TYPES AND ACOUSTIC SCENARIOS CONSIDERED IN THIS WORK. A:B:C DENOTES THE SET $\{A, A+B, A+2B, \dots, C\}$. BFR IS RESTAURANT NOISE, CCITT IS SPEECH-SHAPED NOISE.

| Label | Speech Azi. [°] | Noise Azi. [°] | SNR [dB] | Noise | Scenario |
|---|---|---|---|---|---|
| Train | -90:15:90 | -90:15:90 | -5:5:20 30 50 | BFR, Bus, CCITT, Office | Anechoic, Office |
| Test | -90:5:90 | -90:5:90 | -2.5:2.5:10 20 40 | BFR, Bus, CCITT, Office | Anechoic, Office, Cafeteria |

domain. But, another processing step is performed to compute the actual currents applied to the electrodes of the CIs. This step covers the mapping to the so called current levels, usually measured in clinical units, integers which correspond one to one to current values in microampere. The precise mapping from the output of the LGF to clinical units is set individually for each CI user by an audiologist and involves setting the dynamic range in the electrical domain for each subband of the CI. These clinical levels apriori contain only the relevant information for the CI user. While the advantage of coding the output of the LGF as done in [4], [5] is some independence of the individual parameters of the CI user, the downside are some irrelevancies still left in the LGF data. Thus, compression of the clinical units could be useful for wireless transmission of audio for CIs. Therefore, in this work, lossless compression of the current levels at zero latency is investigated. The general application in wireless streaming of audio for CIs is depicted in Fig. 1. Conventionally, an audio codec compresses the audio signal of an external device and transmits it to the CI processor, where it is decompressed and processed by the sound coding strategy. In our approach, the audio signal is first converted into stimulation patterns, which are then compressed by our codec and transmitted to the CI processor. While lossy compression allows for lower bitrates, it can generally lead to undesired distortions of a signal or certain signals.

Lossless compression does not distort a signal and thus guarantees unaffected audio quality and intelligibility. Additionally, CI users could be very sensitive to distortions of the clinical units requiring extensive listening tests to develope a lossy codec based on the clinical units. For this purpose, a simple lossless codec is investigated which mixes several contexts using a feedforward network. The codec is compared on the TIMIT speech corpus using several acoustic scenarios

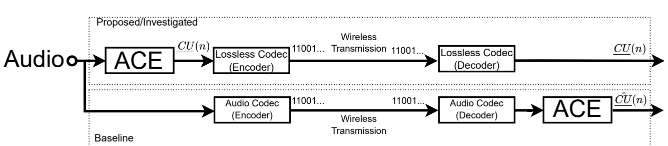


Fig. 1. Two methods to wireless transmission of audio for cochlear implants (CIs). Conventionally, the audio signal would be encoded by an audio codec, transmitted to the signal processor of the CIs where the audio is decoded by the same audio codec. In the investigated approach, the audio signal is first processed by the sound coding strategy of the CI, in our case the advanced combinational encoder (ACE), and then compressed and decompressed before and after transmission by a lossless codec. Due to that, the current levels, specified in clinical units (CU), are reconstructed without distortion on the receiver side, unlike for the (usually lossy) audio codec.

to the well known lossless algorithms PAQ and prediction by partial matching (PPM) as well as the lossy state-of-the-art audio codec Opus. Our Electrocodec is included as a baseline as well. Section II elaborates on the required details of the sound coding strategy used to generate the data as well as the dataset and baselines used. Additionally, the lossless codec investigated in this work is explained. In Section III, the performance of the lossless codec is investigated and compared to baseline compression algorithms. In Section IV, the results are discussed and compared to previous works and the manuscript concludes in Section V.

## II. METHODS AND MATERIALS

### A. Advanced Combinational Encoder

The sound coding strategy used in this work is the advanced combinational encoder (ACE). 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 or subbands 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, subsequent frequency subband selection and an acoustic to current level mapping block consisting of the loudness growth function (LGF) and the actual current mapping. 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. [6]. The mapping block determines the current level from the envelope magnitude and the band characteristics. This is done using the logarithmically-shaped LGF that maps the acoustic envelope amplitude $a(k)$ of subband $k$ to an electrical magnitude $P(k)$ according to

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

The magnitude $P(k)$ 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 comfort level). An input at base-level $s$ is mapped to an output at threshold level (THR), and no output is produced for an input of lower amplitude. The parameter $m$ is the input level at which the output saturates. Inputs at this level or above result in stimuli at comfort level (MCL). The parameter $\rho$ controls the amount of compression of the LGF [6]. For all experiments the default settings were used. These set $\rho = 416.2063$, $s = 4/256$ and $m = 150/256$. 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 $N = 8$ and the number of total subbands was fixed at $M = 22$. Finally, the clinical units $CU(k)$ delivered by the CI are derived from $P(k)$ by

$$CU(k) = \text{THR}(k) + [(\text{MCL}(k) - \text{THR}(k)) \cdot P(k)], \quad (2)$$

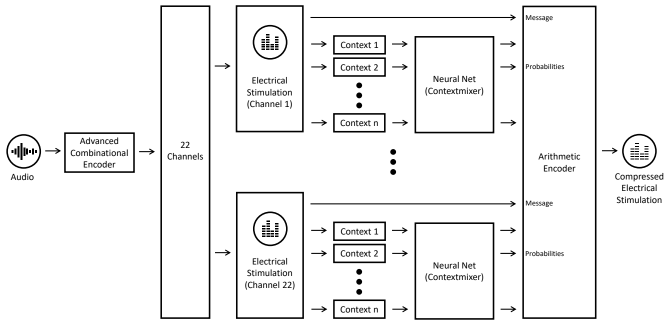


Fig. 2. General structure of the lossless codec. From an audio input the electrical stimulation patterns are generated by the advanced combinational encoder sound coding strategy. A feedforward neural network mixes the estimated probabilities of several contexts to compute (hopefully) improved symbol probabilities which are then used for the encoding of each subband symbol by the arithmetic encoder. The decoder uses the same structure.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| [-5, -8] 0.00 | [-5, -7] 0.00 | [-5, -6] 0.00 | [-5, -5] 0.01 | [-5, -4] 0.01 | [-5, -3] 0.01 | [-5, -2] 0.00 | [-5, -1 0.02 | [-5, 0] 0.01 |
| [-4, -8] 0.00 | [-4, -7] 0.00 | [-4, -6] 0.01 | [-4, -5] 0.00 | [-4, -4] 0.01 | [-4, -3] 0.01 | [-4, -2] 0.00 | [-4, -1 0.03 | [-4, 0] 0.01 |
| [-3, -8] 0.00 | [-3, -7] 0.00 | [-3, -6] 0.01 | [-3, -5] 0.01 | [-3, -4] 0.01 | [-3, -3] 0.01 | [-3, -2] 0.00 | [-3, -1 0.03 | [-3, 0] 0.03 |
| [-2, -8] 0.00 | [-2, -7] 0.01 | [-2, -6] 0.01 | [-2, -5] 0.01 | [-2, -4] 0.01 | [-2, -3] 0.01 | [-2, -2] 0.00 | [-2, -1 0.02 | [-2, 0] 0.04 |
| [-1, -8] 0.01 | [-1, -7] 0.01 | [-1, -6] 0.01 | [-1, -5] 0.01 | [-1, -4] 0.02 | [-1, -3] 0.02 | [-1, -2] 0.01 | [-1, -1 0.04 | [-1, 0] 0.33 |
| [0, -8] 0.05 | [0, -7] 0.08 | [0, -6] 0.08 | [0, -5] 0.09 | [0, -4] 0.11 | [0, -3] 0.09 | [0, -2] 0.04 | [0, -1] 0.68 | [0, 0] 1.00 |
| [1, -8] 0.00 | [1, -7] 0.01 | [1, -6] 0.01 | [1, -5] 0.01 | [1, -4] 0.02 | [1, -3] 0.02 | [1, -2] 0.01 | [1, -1] 0.04 | |
| [2, -8] 0.00 | [2, -7] 0.01 | [2, -6] 0.01 | [2, -5] 0.01 | [2, -4] 0.01 | [2, -3] 0.01 | [2, -2] 0.01 | [2, -1] 0.02 | |
| [3, -8] 0.00 | [3, -7] 0.01 | [3, -6] 0.01 | [3, -5] 0.01 | [3, -4] 0.01 | [3, -3] 0.01 | [3, -2] 0.00 | [3, -1] 0.03 | |
| [4, -8] 0.00 | [4, -7] 0.00 | [4, -6] 0.00 | [4, -5] 0.01 | [4, -4] 0.01 | [4, -3] 0.01 | [4, -2] 0.01 | [4, -1] 0.02 | |
| [5, -8] 0.00 | [5, -7] 0.00 | [5, -6] 0.00 | [5, -5] 0.00 | [5, -4] 0.01 | [5, -3] 0.01 | [5, -2] 0.00 | [5, -1] 0.03 | |

Reference
> 1/4
> 1/16
> 1/32
> 1/64
> 1/128
< 1/128

Fig. 3. Heatmap showing the partial correlation for the sixth subband. The most relevant contexts for prediction are the same subband at the previous time step and the previous subband at the same time step. The notation [$\Delta k$, $\Delta n$] is channel independent, e.g. assuming the subband $k$ at frame $n$ is to be encoded then [-1,-2] is the context corresponding to the subband $k-1$ and timestep $n-2$.

where $\mathrm{THR}(k)$ and $\mathrm{MCL}(k)$ are the threshold level and the most comfortable level, respectively, of subband k. The $[\cdot]$ operator represents the rounding operation and rounds towards the nearest integer. The dynamic range of subband $k$ is defined as $\mathrm{MCL}(k) - \mathrm{THR}(k)$. If a band is not selected or $a(k) < s$, then $CU(k)$ is assigned the value zero. Else it is in the range $CU(k) \in \{THR(k), THR(k)+1, \ldots, MCL(k)\}$. The quantity $CU(k)$ is mapped to a current value in microamperes by an exponential function which is then the actual current applied to the electrodes. In this work, it is proposed to compress $CU(k)$ losslessly. The dynamic range, which also depends on the respective subband, often lies between 5 bits to 7 bits [7] at least for devices by Cochlear Ltd. For the subjects of our study [5] the range of $CU(k)$ was mostly within the range of 5 bits to 6 bits. In general, there is a larger gap between the dynamic ranges of prelingually implanted CI users and postlingually implanted CI users. The latter exhibit a considerably smaller dynamic ranges [8]. This small range makes lossless compression attractive, which guarantees no loss of information.

## B. Datasets

To create realistic noisy speech signals, the TIMIT speech corpus was processed by the behind-the-ear head related transfer functions (HRTF) from [9]. These HRTFs allow to simulate speech in noise scenarios, where the azimuth of each source can be independently varied with respect to its incident azimuth in the range of $\pm 90°$ in steps of $5°$ except for certain acoustic environments. An azimuth of $-90°$ corresponds to a source located to the left of the listener and $+90°$ corresponds to a source located to the right of the listener, $0°$ corresponding to the front of the listener. Each speech recording of the training and test data of TIMIT was processed using signal-to-noise ratios (SNRs), speech and noise azimuths, acoustic environments and noise type from a list of values given in Tab. I. For each category (SNR, noise type, ...) and each speech file, values were selected and applied randomly. While combinations of conditions like bus noise in a cafeteria environment are certainly less realistic than others, these still give important information about the robustness and generalization capabilities of the lossless codec. The range of values for the SNR and other categories for the test set was chosen such that it allowed to assess the impact of out-of-group values, e.g. the impact of a speech azimuth not used in training like $5°$. As noise, we used Comité Consultatif International Téléphonique et Télégraphique (CCITT) noise, bus noise, office noise and restaurant noise. CCITT noise is speech-shaped noise often used in clinical research. The test data was mapped to clinical units using Eq. 2 with dynamic ranges of size $2^B - 1$ for all subbands, where $B = 3, \ldots, 6$ was investigated. Because subbands not selected are assigned the value zero, which is not included in the values according to Eq. 2, another symbol had to be the zero, and together in total $2^B$ symbols were used per subband. Additionally, to assess the compression using real dynamic ranges, the data was mapped to clinical units using mean dynamic ranges from CI users. For this purpose, values given in [8] for prelingually and postlingually implanted CI users were considered, which should yield a meaningful statistic.

## C. Lossless Codec

The block diagram of the investigated zero delay lossless codec (ZDLLC) including ACE is depicted in Fig. 2. The general structure is heavily inspired by PAQ [10]. The purpose of the contexts is to estimate symbol probabilities through relative frequencies, i.e. each context estimates a conditional probability $P(CU_n(k)|CU_{n_{i_1}}(k_{i_1}), \ldots, CU_{n_{i_N}}(k_{i_N}))$, where $N$ is the length or order of the context. $n_{i_s}$ and $k_{i_s}$ were such that only already encoded subbands were considered at each time step. Contexts of length one to three were considered. The purpose of the artificial neural networks is to mix the context probabilities to obtain a hopefully more accurate symbol probability, subsequently used in the artihmetic coding for the compression of the current frame. A frame is understood as the vector $\underline{CU}_n := (CU_n(1), CU_n(2), \ldots, CU_n(M))^T$ with $CU(k)$ as in Eq. 2 and $n$ denoting discrete time. The combination of a feedforward network and contexts allows to capture the dynamics of the subband signals while keeping a simple network structure. This is beneficial as recurrent networks tend to be more difficult to train. For an optimal algorithm each subband had its own neural network. Compression was

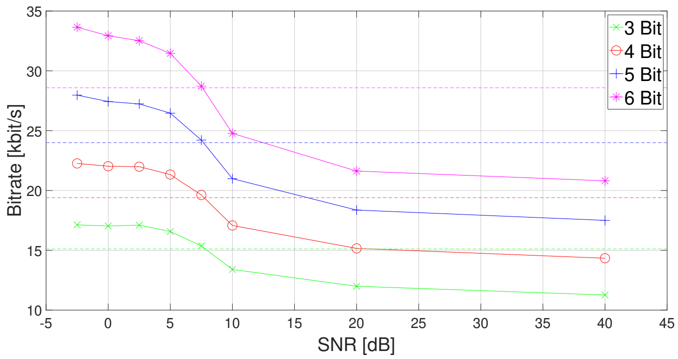


Fig. 4. Mean bitrate across signal-to-noise ratio (SNR) of the lossless codec on the test set. A considerable increase in bitrate with decreasing SNR was observed. The cause are more subbands that are selected at low SNRs. The dashed lines indicate the mean bitrate across the test set. These results were obtained using the complete set of contexts given in Table II.

performed from top to bottom, i.e. starting with $CU_n(1)$. The codec made use of the $N$ of $M$ selection of ACE to optimize the compression. At time step $n$, once $N$ selected subbands have been encoded the encoding process can stop as the other subbands cannot be selected which considerably reduces the required bitrate. As only one frame of data is compressed at a time, the algorithmic latency of the lossless codec is zero. To find an initial set of useful contexts, i.e. contexts that carry most information about the future course of the clinical units, the partial correlation coefficients were computed and the subbands with the largest partial correlation coefficient were initially considered for the compression algorithm. The computed partial correlation coefficients are depicted in Fig. 3. Training of the network was performed for 500 epochs using the cross-entropy loss and a batch size of 256. Early stopping based on the training loss was applied and the training was stopped if the loss decreased by less than $1\,\%$ over the course of ten epochs.

### D. Baselines

The well known lossless source coding algorithms prediction by partial matching (PPM) and PAQ [10] were used as baseline. For their application, the vector $\underline{CU}(n)$ was binarized for each $n$ and the resulting binary vectors concatenated. This sequence of binary vectors was then compressed. While PAQ and PPM do not make a fair comparison due to their algorithmic latency - they compress the entire bitstream at once - they provide a useful estimate of the upper bound for the achievable compression. For both algorithms, bitrates are given, computed using the compressed size and the total duration of the stimulation patterns, instead of compression ratios, as bitrate is the more common measure for source coding algorithms related to data transmission. Additionally, the Opus audio codec [11] served as baseline to assess the bitrates achieved by a state-of-the-art audio codec. Opus can code at almost any bitrate between 6 kbit/s and 520 kbit/s and at algorithmic latencies between 5 ms and 60 ms. For the comparison between Opus and the lossless codecs, the audio signals were compressed and decompressed using Opus as required in a wireless transmission, and the decompressed signal was processed by ACE. The resulting stimulation patterns were then resynthesized using a vocoder, and the resulting audio signal compared to the original audio signal using the short-time objective intelligibility measure (STOI) [12], a common algorithm to assess the intelligibility of speech signals which has found application in CI research [13]. This process is described in detail in [5]. Using STOI, Opus was applied at algorithmic latencies of 5 ms and 7.5 ms and several bitrate settings to find the minimal mean bitrate at which no reduction in speech intelligibility could be observed. This process was also guided by the results of [5]. This bitrate was then used for the comparison for each latency setting. Finally, to contrast the lossless compression of the stimluation patterns with lossy compression, the Electrocodec, proposed by us in [4], [5], was used as reference as well. In contrast to the investigated ZDLLC, the Electrocodec is a lossy codec which compresses the quantity $P(k)$, defined in Eq. 1, at zero delay. It was applied using quantizer resolutions of two and three bits per subband, accordingly labeled EC2 and EC3.

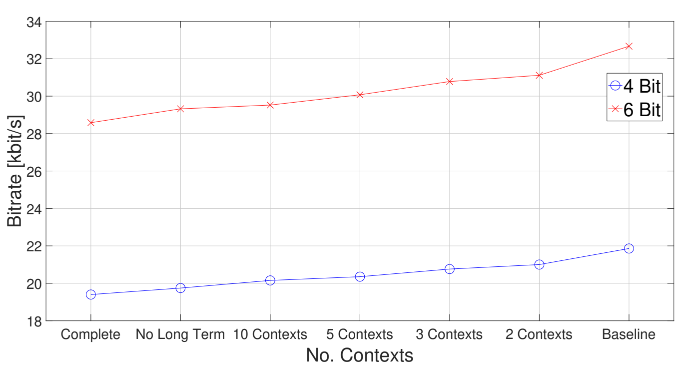


Fig. 5. Bitrate across the number of considered contexts per subband of the lossless codec. The single context was [0, -1], i.e. the same subband but the previous time step, and served as reference to assess the benefit of context mixing. Invalid contexts like [-1,0] in the first subband were ignored.

TABLE II
THE CONTEXTS OF ORDER ONE TO THREE CONSIDERED FOR THE LOSSLESS COMPRESSION. THE NOTATION $[\Delta k, \Delta n]$ IS CHANNEL INDEPENDENT, E.G. ASSUMING THE SUBBAND $k$ AT FRAME $n$ IS TO BE ENCODED, THEN [-1,-2] IS THE CONTEXT CORRESPONDING TO THE SUBBAND $k-1$ AND TIMESTEP $n-2$.

| | Contexts | | | | | |
|---|---|---|---|---|---|---|
| 1st Order | [0, -1] | [-1, 0] | [0, -2] | [-2, 0] | [1, -1] | [-1, -1] |
| 2nd Order | [[0, -1], [-1, 0]] | | [0, -1], [0, -2] | | [0, -1], [1, -1] | |
| | [0, -1], [-1, -1] | | [-1, 0], [-1, -1] | | [-1, 0], [-2, 0] | |
| | [-1, -1], [-2, 0] | | [-1, -1], [0, -2] | | [1, -1], [0, -2] | |
| 3rd Order | [0, -1], [-1, 0], [-1, -1] | | | [0, -1], [-1, 0], [1, -1] | | |
| | [0, -1], [-1, 0], [0, -2] | | | [0, -1], [-1, 0], [-2, 0] | | |
| | [0, -1], [0, -2], [-1, -1] | | | [0, -1], [0, -2], [1, -1] | | |
| | [0, -1], [-1, -1], [1, -1] | | | [0, -1], [-1, -1], [-2, 0] | | |
| | [-1, 0], [-1, -1], [-2, 0] | | | | | |
| Longterm | [-3,0] | [0,-3] | [0,4] | [0,-5] | [0,-6] | [0,-7] |
| | [0,-8] | [0, -6], [0, -7], [0, -8] | [0, -2], [0, -3] | [0, -3], [0, -4] | [0, -4], [0, -5] | [0, -5], [0, -6] |
| | [0, -6], [0, -7] | [0, -7], [0, -8] | | | | |

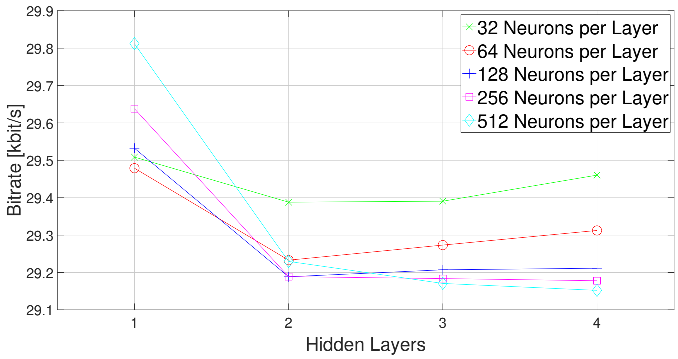


Fig. 6. Bitrate across number of neurons and layers of the artificial neural networks. Increasing the number of layers above two did not yield substantial benefits and the number of layers could be kept at two to reduce the computational complexity. A resolution of 6 bit/symbol was used to generate the depicted results.

## III. Results

Mean bitrates of the ZDLLC across the test set and SNRs for 3 bit to 6 bit are depicted in Fig. 4. Starting at about 20 dB the bitrate considerably increased from about 21.6 kbit/s for 6 bit at 20 dB to about 33.7 kbit/s for 6 bit at -2.5 dB. Similar but smaller increases occured for lower number of bits/symbol. The cause is the number of selected bands often being less than $N$ for SNRs greater than about 7.5 dB, which makes compression considerable easier. At and below an SNR of about 5 dB for most frames the maximum number $N$ out of the $M$ subbands are selected, thus making compression more difficult. The effectiveness of the contexts is shown in Fig. 5, where the bitrate is plotted across the number of contexts considered by the ZDLLC. A single context served as reference which did not use a neural network. Increasing the number of contexts from one to two allowed to decrease the bitrate by about 1.5 kbit/s for the 6 bit test set and 1 kbit/s for the 4 bit test set. Adding further contexts yielded a decrease of the bitrate by about 2 kbit/s for the 6 bit and by about 1.3 kbit/s for the 4 bit dataset. The impact of the neural network on the bitrate is shown in Fig. 6, where the bitrate across number of layers is depicted for several number of neurons per layer. A substantial reduction in bitrate by 0.3 kbit/s to 0.6 kbit/s was observed when two or more layers were used compared to using only a single hidden layer. Mean bitrates across the test set using 6 bit data resolution of PPM, PAQ, Opus, the Electrocodec and the ZDLLC are given in Table V. Both, PAQ and the ZDLLC, outperformed Opus at either latency with respect to bitrate. However, the Electrocodec achieved the smallest bitrate of 20.1 kbit/s when using 2 bit quantization per subband, rising to only 22.7 kbit/s for an SNR $\leq$ 5 dB. PPM achieved considerably worse bitrates more than 10 kbit/s above those of PAQ and the ZDLLC. PAQ achieved about 2 kbit/s to 3 kbit/s lower bitrate than the ZDLLC, however, at very high latency. Mean bitrates of the ZDLLC, PPM and PAQ using dynamic ranges of prelingually and postlingually implanted CI users taken from [8] across the entire test set are given in Table III. For postlingually implanted CI users, a substantial decrease in bitrate by about 3 kbit/s compared to prelingually implanted CI users was observed, which is in accordance with the results shown in Fig. 4. PAQ and the investigated ZDLLC both achieve bitrates around or below 30 kbit/s, outperforming Opus at either investigated latency.

TABLE III
Mean bitrates in kbit/s on the test set and the subset with a signal to noise ratio $\leq$ 5 dB of the investigated zero delay lossless codec (ZDLLC) as well as prediction by partial matching (PPM) and PAQ, using the dynamic ranges of prelingual and postlingual implanted CI users as listed in Table IV.

| **Dataset\Condition** | **Prelingual** | | | **Postlingual** | | |
|---|---|---|---|---|---|---|
| | ZDLLC | PAQ | PPM | ZDLLC | PAQ | PPM |
| Test Set | 27.4 | 24.9 | 36.7 | 24.6 | 21.9 | 33.8 |
| Test Set ($\leq$ 5dB) | 31.4 | 30.0 | 43.5 | 28.1 | 26.3 | 40.0 |

## IV. Discussion

In this work a zero delay lossless codec (ZDLLC) for the clinical units of cochlear implants was investigated. It consists of artificial neural networks and several contexts which estimated conditional symbol probabilities used in arithmetic coding. The lossless codec achieved similiar or lower bitrates than the Opus audio codec at zero latency for real dynamic ranges of CI users compared to 5 ms or more for Opus. The ZDLLC outperforming PPM and achieving worse but similiar bitrates/compression ratios as PAQ is comforting. PAQ outperforming the investigated ZDLLC despite being a general purpose algorithm is due to several reasons: PAQ compresses the entire bitstream and thus has a latency as long as the bitstream. It cannot be used as is for low latency applications. In contrast, the ZDLLC compresses frame by frame to achieve zero latency, thereby reducing the number of symbols per message, and thus its arithmetic coding will lose some of its effectiveness. This alone can explain a redundancy in our approach of up to 2 bit/frame, i.e. a redundancy of 1.8 kbit/s. However, this still would not entirely account for the gap in bitrate between PAQ and the ZDLLC. Secondly, PAQ adapts the mixing of its contexts on the fly, allowing to adapt better to changes in signal statistics unlike our ZDLLC which uses a static feed-forward network. Finally, PAQ uses many different kind of contexts which might allow to adapt to almost any kind of data despite its general purpose nature. The performance of the ZDLLC, which by no means used an optimal implementation, could be improved further, e.g. by increasing the size of the buffer used for the estimation of the symbol probabilities. Our Electrocodec outperforming all other codecs supports its design further and was expected due to its lossy coding of the stimulation patterns. While lossy compression of the output of the loudness growth function as done by the Electrocodec is beneficial for some applications, lossless compression of the clinical units could be advantagous for wireless streaming of audio to guarantee no loss of information, which cannot be guaranteed for all kind of signals by the Electrocodec. An interesting result is the achievable bitrate when using only a

TABLE IV

MEAN DYNAMIC RANGES OF EACH CHANNEL OR SUBBAND FOR PRELIGUNAL AND POSTLINGUAL IMPLANTED CI USERS TAKEN FROM [8].THE AVERAGE (AVG) OF THE DYNAMIC RANGES OF THE SUBBANDS IS GIVEN AS WELL.

| Subject\Subband | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | AVG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Prelingual | 56 | 58 | 57 | 58 | 57 | 57 | 57 | 57 | 57 | 58 | 58 | 57 | 55 | 56 | 57 | 59 | 58 | 58 | 58 | 58 | 59 | 59 | 58 |
| Postlingual | 38 | 38 | 39 | 40 | 39 | 40 | 41 | 40 | 41 | 40 | 40 | 40 | 39 | 38 | 38 | 37 | 36 | 35 | 34 | 32 | 30 | 29 | 37 |

TABLE V

MEAN BITRATES OF THE INVESTIGATED ZERO DELAY LOSSLESS CODEC (ZDLLC) AND THE BASELINE ALGORITHMS PAQ, PREDICTION BY PARTIAL MATCHING (PPM), THE OPUS AUDIO CODEC AT 5 MS AND 7.5 MS ALGORITHMIC LATENCY AND THE ELECTROCODEC AT TWO BITS (EC2) AND THREE BITS (EC3) PER SUBBAND ACROSS THE TEST SET USING A 6 BIT RESOLUTION AND THE SUBSET OF CONDITIONS WITH A SIGNAL TO NOISE RATIO $\leq 5$ DB. FOR OPUS NEGLIGIBLE VARIATION OF THE BITRATE ACROSS SNR WAS OBSERVED.

| Dataset\Codec | ZDLLC | PAQ | PPM | $Opus_{5ms}$ | $Opus_{7.5ms}$ | EC2 | EC3 |
|---|---|---|---|---|---|---|---|
| Test Set | 28.6 | 25.1 | 37.3 | 35.2 | 33.6 | 20.1 | 24.3 |
| Test Set ($\leq$5 dB) | 32.6 | 30.4 | 44.3 | 35.2 | 33.6 | 22.7 | 27.8 |

single context in each subband and the $N$ of $M$ selection of ACE. This way zero latency at bitrates between about 30 kbit/s to 40 kbit/s could be achieved without major computational complexity. While the Electrocodec could be used for binaural sound coding strategies like [14], this seems unlikely for the ZDLLC, as the information loss in the clinical units is considerably larger compared to the output of the loudness growth function. Nonetheless, the guaranteed unaffected quality and intelligibility of the speech signals, which cannot be in all cases guaranteed for the Electrocodec, could make it desireable for wireless streaming of audio for CIs.

## V. CONCLUSION

In this work a zero delay lossless codec (ZDLLC) for the compression of the current levels, given in clinical units, of cochlear implants (CIs) was investigated. The current levels contain only the relevant information for the CI user. The investigated ZDLLC used arithmetic coding supported by several contexts which estimated the conditional symbol probabilities and context mixing, which was implemented by an artificial neural network. The ZDLLC was compared to the lossless algorithms PAQ and PPM as well as the lossy audio codec Opus on the TIMIT speech corpus using several noisy conditions. At a data resolution of six bit, the ZDLLC achieved a mean bitrate of 28.6 kbit/s at zero latency compared to 33.6 kbit/s to 35.2 kbit/s and a latency of 5 ms to 7.5 ms for the Opus audio codec and, at very high latency, 25.1 kbit/s for PAQ and 37.3 kbit/s for PPM. The results suggest that lossless compression of the clinical units can be beneficial for wireless streaming of audio for CIs.


## ACKNOWLEDGEMENTS

This work was supported by the DFG Cluster of Excellence EXC 1077/1 Hearing4all and funded by the German Research Foundation (DFG) - Project number: 381895691


## REFERENCES


[1] K. I. Bieńkowska, K. Babula, M. Lewczuk-Siliwoniuk, M. Demiańczuk, K. Kasica, B. Kasica, A. Kolanowska, M. B. Gawłowska, P. Mickiewicz, and A. Jedlińska, "Effectiveness of the use of mini-microphones for the treatment of children with cochlear implants - a multicenter study," *Szkoła Specjalna*, no. LXXXI(1), p. 37–46, 2020.

[2] T. Gajecki and W. Nogueira, "A synchronized binaural n-of-m sound coding strategy for bilateral cochlear implant users," in *Speech Communication; 13th ITG-Symposium*, 2018, pp. 1–5.

[3] M. Stone, B. Moore, K. Meisenbacher, and P. Derleth, "Tolerable hearing aid delays. v. estimation of limits for open canal fittings," *Ear and hearing*, vol. 29, pp. 601–17, 09 2008.

[4] R. Hinrichs, T. Gajęcki, J. Ostermann, and W. Nogueira, "Coding of electrical stimulation patterns for binaural sound coding strategies for cochlear implants," in *Proc. 41st IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 4168–72.

[5] R. Hinrichs, T. Gajecki, J. Ostermann, and W. Nogueira, "A subjective and objective evaluation of a codec for the electrical stimulation patterns of cochlear implants," *The Journal of the Acoustical Society of America*, vol. 149, no. 2, pp. 1324–1337, 2021. [Online]. Available: https://doi.org/10.1121/10.0003571

[6] W. Nogueira, A. Büchner, T. Lenarz, and B. Edler, "A psychoacoustic "nofm"-type speech coding strategy for cochlear implants," *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 3044–59, 11 2005.

[7] M. Ozbal Batuk, B. Cinar, E. Zeren, O. Bayulgen, I. Dusunmez, and G. Sennaroglu, "Evaluation of ecap thresholds, t and c levels in children with sequential bilateral cochlear implants," p. 1, 07 2019.

[8] A. Zarowski, A. Molisz, L. Coninck, A. Vermeiren, T. Theunen, L. Theuwis, T. Przewoźny, J. Siebert, and E. Offeciers, "Influence of the pre- or postlingual status of cochlear implant recipients on behavioural t/c-levels," *International Journal of Pediatric Otorhinolaryngology*, vol. 131, p. 109867, 01 2020.

[9] H. Kayser, S. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, p. 6, 12 2009.

[10] M. V. Mahoney, "Adaptive weighing of context models for lossless data compression," 2005.

[11] O. Jokisch, M. Maruschke, M. Meszaros, and V. Iaroshenko, "Audio and speech quality survey of the opus codec in web real-time communication," in *Proc. 27th Conference on Electronic Speech Signal Processing (ESSV)*, 03 2016, pp. 254–62.

[12] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 4214–4217.

[13] E. H.-H. Huang, C.-M. Wu, and H.-C. Lin, "Combination and comparison of sound coding strategies using cochlear implant simulation with mandarin speech," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 2407–2416, 2021.

[14] M. J. Fumero, A. Eustaquio-Martín, J. M. Gorospe, R. Polo López, M. A. Gutiérrez Revilla, L. Lassaletta, R. Schatzer, P. Nopp, J. S. Stohl, and E. A. Lopez-Poveda, "A state-of-the-art implementation of a binaural cochlear-implant sound coding strategy inspired by the medial olivocochlear reflex," *Hearing Research*, vol. 409, p. 108320, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378595521001544