Coding of Electrical Stimulation Patterns for Binaural Sound Coding Strategies for Cochlear Implants

Reemt Hinrichs[1], Tom Gajęcki[2], Jörn Ostermann[1], Waldo Nogueira[2]

***Abstract*— Binaural sound coding strategies can improve speech intelligibility for cochlear implant (CI) users. These require a signal transmission between two CIs. As power consumption needs to be kept low in CIs, efficient coding or bit-rate reduction of the signals is necessary. In this work, it is proposed to code the electrical signals or excitation patterns (EP) of the CI instead of the audio signals captured by the microphones. For this purpose we designed a differential pulse code modulation based codec with zero algorithmic delay to code the EP of the advanced combination encoder (ACE) sound coding strategy for CIs. Our EP codec was compared to the G.722 64 kbit/s audio codec using the signal-to-noise ratio (SNR) as objective measure of quality. On two audio-sets the mean SNR was 0.5 to 13.9 dB higher when coding the EP with the proposed coding method while achieving a mean bit-rate between 34.1 and 40.3 kbit/s.**

***Index Terms*— cochlear implants, excitation patterns, advanced combination encoder, differential pulse code modulation, binaural sound coding strategy**

I. Introduction

Today speech comprehension in quiet listening conditions can be considered good for cochlear implant (CI) users. However, their speech recognition performance decreases quickly as the level of background noise increases. Several studies show a deterioration in speech reception threshold (SRT), a measure which estimates the signal-to-noise ratio (SNR) needed by a subject to attain 50 % speech understanding. For CI users this value can go up to 11.4 dB and 0.75 dB for normal hearing (NH) listeners [1], [2], [3]. Users of CIs report speech understanding in noisy background as the main shortcoming of the device [4]. For this reason, several techniques were developed over the years to overcome this issue including beamformers, cross-devices and binaural sound coding strategies (BSCS) [5], [6].

Recently, bilateral implantation (BiCI), meaning a CI in both ears, has become more common, with demonstrated benefits in understanding speech in noisy situations when compared with single CI use [8]. Building on the BiCI-users BSCS were proposed. In these, signal information from the two CIs is used to generally improve speech perception.

Lopez-Poveda et al. presented in [9] a BSCS which mimics the contralateral medial olivocochlear (MOC) reflex of NH humans. In this strategy, the compression is adjusted for every frequency band according to the mean energy in the corresponding frequency band on the contralateral ear. The energy was estimated based on the CI's excitation patterns (EP), which are the current values applied to each electrode over time by applying a loudness growth function (LGF) on the envelope bands. The LGF in CI's is the mapping function between the acoustic amplitudes and the current amplitudes applied to the cochlea. With this approach, Lopez-Poveda et al. were able to achieve a 3 dB improvement in SRT. To make the strategy work, the information about the EP had to be known at the other ear. Another BSCS proposed in [10] synchronizes the bands selected for stimulation on the left and the right speech processor to improve the SNR. Generally, for real-life BSCS the information has to be transmitted between the two sides. As CIs are battery powered, their energy supply is limited and thus, the information has to be transmitted efficiently. From Shannon's classical paper we know that the energy needed for a reliable communication at fixed noise and bandwidth levels is directly linked to the capacity of the channel which itself is limited by the bit-rate needed for the desired transmissions across the channel [11]. Therefore the bit-rate of the information about the EP should be as little as possible. Additionally, the latency of the applied coding strategy should be as small as possible accounting for any additional delay due to transmission and delay from the particular BSCS. One way to send the LGF data, the EP, to the other ear would be to encode a slice of audio using a low delay audio codec and to use that information to calculate the EP by means of the CI sound coding strategy. In this work, we investigate a different method. As bit-rate reduction of audio and speech signals can be obtained through the reduction of redundant information and discarding of irrelevant information [12], we propose not to code and transmit the audio data captured by the microphones directly but instead to code the EP of the CI, as suggested by [13]. In the EP some irrelevancy for the CI user has been removed which would be coded by an audio codec. This approach allows easy selection of the information of interest (e.g. specific frequency bands) at just the quality needed and also avoids using the CI's sound coding strategy one more time on the receiving ear saving processing power [7]. In this paper, we present a codec for the EP of the ACE strategy which demonstrates the potential of our approach. The proposed codec is compared using the SNR of the EP to the alternative method of using a low delay audio codec. For comparison, the sub-band adaptive differential pulse code modulation (adaptive DPCM) based G.722 developed by the International Telecommunication Union (ITU) is used.

The structure of this paper is as follows: in Section II the CI sound coding strategy used in this work is described.

Thanks to Nina Aldag for the G.722 evaluation.

[1] Institut für Informationsverarbeitung, Leibniz Universität Hannover, Appelstraße 9a, 30167 Hannover

[2] Department of Otolaryngology, Medical University Hannover, Karl-Wiechert-Allee 3, 30625 Hannover

Afterwards, in Section III, the structure of the proposed EP codec is explained. Section IV presents the audio test sets employed as well as the experiment that has been performed. Next, in Section V, the results of the experiment are presented followed by a discussion of these results in Section VI. Finally, the paper is concluded with a summary of the work presented.

II. Advanced Combination Encoder

The reference sound coding strategy used in this work is the advanced combination encoder (ACE). Its main components are a filter bank, subsequent frequency band/channel selection and an acoustic to current level mapping block [14]. The term *channel* in this work is used synonymously with the frequency bands of the ACE. The mapping block determines the current level from the envelope magnitude and the channel characteristics. This is done using the logarithmically-shaped LGF that maps the acoustic envelope amplitude $a(k)$ of channel k to an electrical magnitude according to

$$\text{P(k)} = \begin{cases} \dfrac{\log(1+\rho((a(k)-s)/(m-s)))}{\log(1+\rho)}, & s \le a(k) \le m \\ 0, & a(k) < s \\ 1, & a(k) \ge m \end{cases} . \quad (1)$$

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 the base-level s is mapped to an output at threshold level, 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. The parameter ρ controls the steepness of the LGF [15]. For all experiments the default settings were used. These set $\rho = 0.8$, $s = 4/256$ and $m = 150/256$. The channel stimulation rate (CSR) was fixed at 1200 pulses per second. This results in the same number of values $P(k)$ per second per channel.

III. EP Codec

The EP codec (called *ElectroCodec* (EC)) is based on a DPCM and context-adaptive binary arithmetic coding (CABAC). The DPCM decreases a signal's variance by means of prediction and reduces the data-rate by follow up quantization. Predictive coding as known from speech coding most often uses linear prediction techniques. The theory of linear prediction usually assumes zero mean processes [16]. Due to the LGF function according to (1) this is not true for the EP. Assuming a slowly varying mean $\mathrm{E}(P(k,n))$, where $P(k,n)$ is the electrical magnitude of channel k according to (1) at discrete time n, the quantity $P(k,n) - P(k,n-1)$ is approximately (appr.) zero mean. This technique to stabilize the mean is also known from the autoregressive integrated moving average model [17]. Therefore, the structure of the DPCM was expanded to first transform the input signal into an appr. zero mean signal and then apply the classic DPCM structure onto that nearly zero mean signal. The resulting DPCM encoder structure implemented in every channel is depicted in Figure 1. From this, the transfer function $\frac{e(z)}{P(z)}$ can be derived (ignoring quantization) as

$$\frac{e(z)}{P(z)} = \frac{1-H(z)}{1-z^{-1}}, \quad (2)$$

with the predictor's transfer function $H(z)$, and the Z-transforms $P(z)$ and $e(z)$ of the input signal $P(n)$ and the error signal $e(n)$ respectively. The channel index k was dropped for clarity. The transfer function according to (2) has a pole on the unit circle and thus is only marginal stable. It could be made stable by moving the pole slightly from the unit circle towards zero while only marginally decreasing the prediction gain. This did not prove to be an issue for the coding in any of the experiments. As $1-H(z)$ can be shown (within the domain of linear prediction [16]) to always have zeros within the unit circle the inverse system (the DPCM decoder) is always stable. The predictor is adapted at every sample but solely on past samples which makes the codec have zero algorithmic delay.

Besides the values of the active channels, the activation state, represented by the *activity maps* (AMs), of every frame has to be encoded to allow decoding. A frame is understood to be the output currents for each electrode during one stimulation cycle. A stimulation cycle corresponds with the inverse of the stimulation rate on each channel. The AMs are arrays of the binary symbols 0 and 1 which denote whether a channel is inactive or active, respectively. It allows to map the coded EP to the correct channels when decoding the bitstream. Uncompressed, these maps have a length of M bits corresponding to the M channels of the CI. If all electrodes are used by the CI user M is equal to 22. This was the case for all experiments performed. Channels, for which the acoustic envelope amplitude is below base-level s, generate no output. We only have to indicate the channels that actually show an output. These have a value of 1 in the AM. All others are set to 0. Thus, separately, we encode the AM by means of CABAC. For CABAC the context-dependent probability $p(B_N|B_{N-1},\ldots,B_{N-m})$ for the Nth bit B_N of the AM for context-size (CS) m were derived based on histograms of the SQAM-set (compare Section IV-B). The context $(B_{N-1},\ldots,B_{N-m})$ is given by the previously encoded bits of the current frame. For $m >= N$ the CS is reduced to

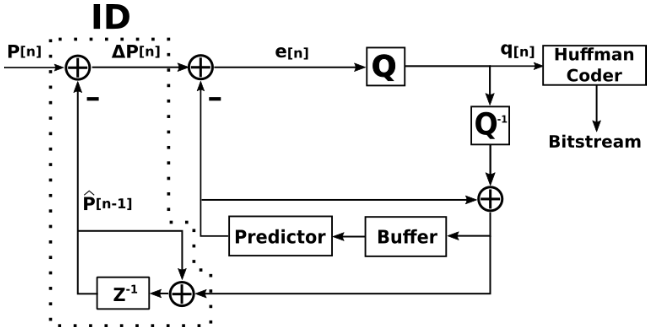

Fig. 1. Encoder structure of the DPCM used in every channel. The classic structure is expanded to add an initial differentiation (ID) stage resulting in an appr. zero mean input signal.

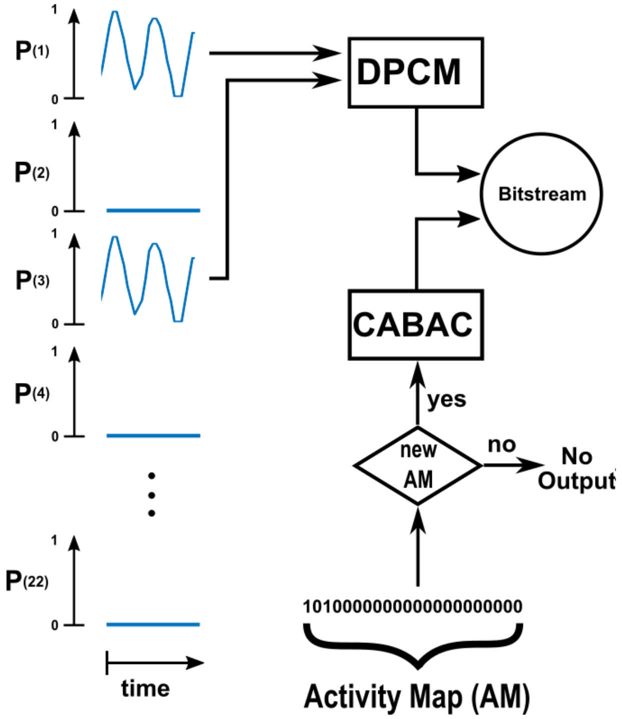

Fig. 2. In the depicted example of the encoding process of the EC only channel 1 and 3 show activity indicated by the sine like pulses. The other channels are inactive indicated by flat lines. The active bands of ACE are coded using a DPCM. Additionally, the activity state of the frames is stored as a bitmap, that is subsequently encoded using CABAC coding but only if the AM changed. In this example only two channels are active throughout the shown time frame resulting in the AM depicted in the Figure.

$N-1$. The process of creating the bitstream of a frame is shown in Figure 2. Because the AS of the channels may be constant throughout consecutive frames we do not always have to transmit the current activity map. To take advantage of this fact, an additional bitflag was added to the bitstream to indicate whether the activation changed from the previous frame or not. This allows to code silent sections, where the AM remains zero for all channels, with just 1 bit. Note that, because the number of channels M is fixed and known, no side information about the number of encoded symbols has to be encoded and no end of stream symbol is needed for the arithmetic coding.

IV. Experiment

This section first describes the audio codec for the alternative method of coding the audio instead of the EP and then describes the audio material and the experiment performed.

A. Audio Codec - G.722

Because to the best knowledge of the authors, no comparable approach exists, our codec can only be compared to the alternative method of encoding the captured audio using low delay, low bit-rate audio codecs. ITU's G.722 sub-band ADPCM, applied using FFmpeg, was chosen as a baseline which is a standard codec of cordless telephone and considered an HD-voice codec [18]. It codes 16 kHz audio at constant 64 kbit/s with an algorithmic delay of only 22 samples yielding only 1.375 ms of delay [19]. It was chosen for its combination of wideband audio, ultra-low delay, and moderate bit-rate. Other codecs with lower algorithmic delay exist, e.g. [20], but these either are designed for different sampling rates or higher bit-rates.

B. Method and Material

Two different audio-sets were used to test the codecs performances: the Hochmair, Schulz and Moser sentence-test (HSM) and the Sound Quality Assessment Material (SQAM). The HSM consists of 30 lists of 20 (German) everyday sentences and is commonly used to assess speech intelligibility in CI users [21]. All sentences are spoken once by a male speaker yielding 600 speech files and once by a female speaker yielding another 600 speech files. The SQAM contains high quality recordings of human speech and other signals [22]. Only the six speech samples were used. Because the input sampling rate of ACE is typically set to 16 kHz all recordings were resampled using Matlab's resample function before applying any signal processing. To compare the EC with the G.722, the speech recordings from the HSM- and SQAM-set were encoded and decoded using the G.722. Afterwards, the decoded signal was processed by ACE and the resulting EP compared to the clean signal's EP using the SNR as a measure of quality which was shown in [23] and [24] to accurately assess speech perception in noise for CI users outperforming other well known metrics. The generation of the EPs is depicted in Figure 3. The EP were calculated for the clean signal, the G.722 and the EC resulting in the signals $\mathrm{P}_{Clean}(k,n)$, $\mathrm{P}_{G.722}(k,n)$ and $\mathrm{P}_{EC}(k,n)$. The SNR of channel k was then calculated according to

$$\mathrm{SNR}_C(k) = 10\cdot\log_{10}\left(\frac{\sum_n P^2_{Clean}(k,n)}{\sum_n (P_{Clean}(k,n)-P_C(k,n))^2}\right), \qquad (3)$$

with $C \in \{\mathrm{EC}, \mathrm{G.722}\}$.

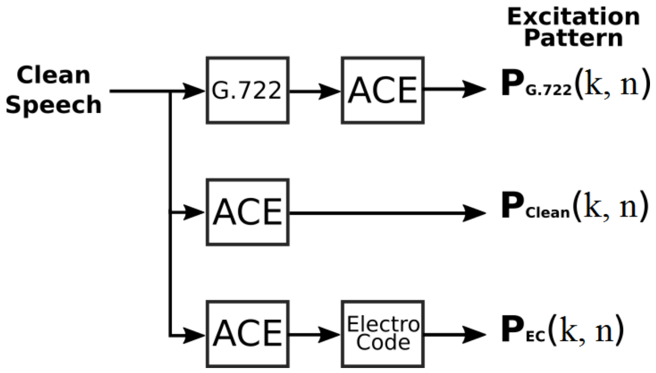

Fig. 3. Diagram of the signal calculation of the experiment performed. The SNR is calculated according to (3). To simulate a transmission between the ears, for the G.722 the audio signal is first encoded and decoded and then processed by the ACE to generate the EP. For the unprocessed reference, the audio is processed immediately by the ACE generating the reference EP. For our proposed EC, the audio is first processed by the ACE and then encoded and decoded using the proposed coding method. The resulting EP are then used for comparison to the other signals.

V. Results

Figure 4 shows the mean SNR for each individual channel for the EC and the G.722. For these SNRs, the mean and peak bit-rate of the EC for speech is depicted in Figure 5 for context-sizes of the CABAC between 0 and 12. The peak bit-rate is the highest bit-rate of any consecutive block of frames of 1-second length. The bit-rate in silent sections is 1.2 kb/s. The proposed EC for coding of the EP is able to outperform the G.722 on every channel in terms of SNR with an SNR advantage between 0.5 and 13.9 dB depending on channel and audio-set. This is achieved while using a lower mean bit-rate of 40.3 kbit/s down to 34.1 kbit/s at a context-size of 12 used in CABAC compared to 64 kbit/s of the G.722. Mean SNR across electrodes of the EC is 32.36 dB and 31.86 dB for the HSM- and SQAM-set, respectively. Mean SNR across electrodes of the G.722 is 27.98 dB and 26.14 dB for the HSM- and SQAM-set, respectively. The total SNR improvement across channels is thus 6.22 dB and 3.89 dB for the HSM- and SQAM-set, respectively.

VI. Discussion

For both sets the first order CABAC yields a significant decrease in bit-rate of between 4.7 and 7.3 kbit/s while higher orders only decrease it minorly. Interestingly, while the probabilities and quantization levels were learned from the SQAM-set, they generalized well onto the HSM-set indicated by the very similar decrease in bit-rate with increased context-size and very similar mean SNR per channel. Our results show that coding the EP directly with well known methods and zero algorithmic delay can outperform ultra-low delay audio codecs at least in low noise level scenarios. One reason why our approach works is the innate thresholding in the LGF and the N of M selection of ACE. Another reason is other irrelevant information (to the CI user) that is contained in the source audio data. ACE, in the standard configuration, does not make use of frequencies below 125 Hz, but these are coded by the G.722. We know from Qazi et al. [25] that CI users are sensitive to distortion of the band-selection but not sensitive to distortion of the signal envelopes, i.e. CI users can tolerate large distortions in speech segments given that the band-selection remains the same. As the band-selection is not corrupted by our coding approach and the SNR of the EP has shown to be a good indicator of speech intelligibility in noise for CI users in [23], it is to be expected that speech perception will be better with our proposed codec as well.

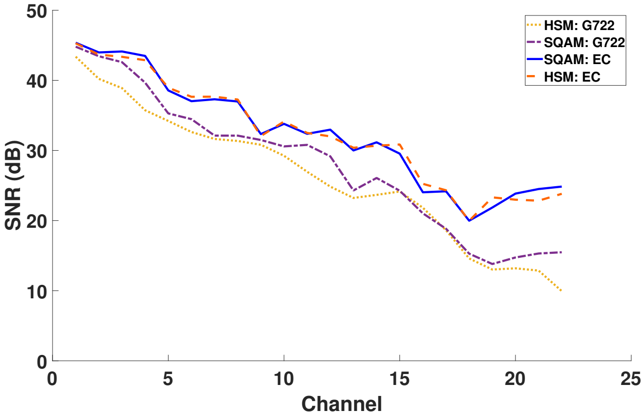

Fig. 4. Mean SNR of the EC and G.722 per channel for the audio-sets HSM and SQAM. The performance of the EC is very similiar for both audio-sets while the G.722 exhibits greater variability in coding quality.

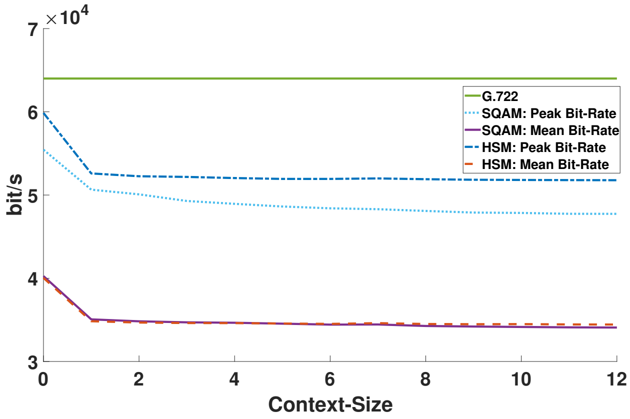

Fig. 5. Mean and peak bit-rate of the EC compared to the G.722 for the HSM and SQAM. The bit-rate is shown for context-sizes of the CABAC between 0 and 12. For comparison, the bit-rate of the G.722 is depicted as well.

VII. Conclusion

This work introduces a novel approach to signal coding for binaural sound coding strategies. Instead of the captured audio signal, the EP generated by the signal processor of the CI are coded. For this purpose this work introduces a novel coding scheme for the EP generated by the ACE sound coding strategy based on DPCM and CABAC. The performance of the new coding approach was compared to the standard G.722 64 kbit/s audio codec. This was done by comparing the resulting EP of the codecs in regards to SNR to clean EP. An SNR differential between 0.5 and 13.9 dB depending on channel in favor of the proposed coding method was obtained with a mean bit-rate between 34.1 and 40.3 kbit/s. These results show for the first time that the EP can be coded with higher SNR and lower bit-rate compared to a standard audio codec. With our approach, it is possible to potentially save computing and battery power for the communication needed in BSCS. The next step is to be tested in subjects and noisy conditions to evaluate the subjective quality of the codec.

VIII. Acknowledgements

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

References

[1] R. H. Gifford; J. K. Shallop; T. A. Peterson, *Speech recognition materials and ceiling effects: considerations for cochlear implant programs*, Audiology and Neurotology Vol. 13, 2008.

[2] D. M. Zeitler; M. A Kessler; V. Tenishkin et al., *Speech perception benefits of sequential bilateral cochlear implantation in children and adults: a retrospective analysis*, Audiology and Neurotology Vol. 13, 2008.

[3] R. H. Wilson; R. A. McArdle; S. L. Smith, *An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss*, Journal of Speech, Language, and Hearing Research, Vol. 50, 2007.

[4] F. Zhao; S. Stephens; S. W. Sim; R. Meredith, *The use of qualitative questionnaires in patients having and being considered for cochlear implants*, Clinical Otolaryngology Allied Sciences, Vol. 22, 1997.

[5] C. Honeder; R. Liepins; C. Arnoldner; H. Sinkovec; A. Kaider; E. Vyskocil et al., *Fixed and adaptive beamforming improves speech perception in noise in cochlear implant recipients equipped with the MED-EL SONNET audio processor*, PLoS ONE, Vol. 13, No. 1, 2018.

[6] N. Guevara et al., *The Voice Track multiband single-channel modified Wiener-filter noise reduction system for cochlear implants: patients outcomes and subjective appraisal*, International Journal of Audiology, 2016.

[7] W. Nogueira; J. Abel; T. Fingscheidt, *Artificial speech bandwidth extension improves telephone speech intelligibility and quality in cochlear implant users*, The Journal of the Acoustical Society of America, Vol. 145, No. 3, 2019.

[8] R. J. M. van Hoesel, *Speech perception, localization, and lateralization with bilateral cochlear implants*, The Journal of the Acoustical Society of America Vol. 113, 2003.

[9] E. A. Lopez-Poveda; A. Estaquio-Martin; J. S. Stohl; R. D. Wolford, *Binaural Cochlear Implant Sound Coding Strategy Inspired by the Contralateral Medial Olivocochlear Reflex*, Ear and Hearing, 2016.

[10] T. Gajecki; W. Nogueira, *A Synchronized Binaural N-of-M Sound Coding Strategy for Bilateral Cochlear Implant Users*, ITG Speech Communication, 2018.

[11] C. E. Shannon *Communication in the Presence of Noise*, Proceedings of the IEEE, Vol. 86, No. 2, 1998.

[12] T. Painet and A. Spanias, *Perceptual Coding of Digital Audio*, Proceedings of the IEEE, Vol. 88, No. 4, April 2000.

[13] B. Edler; A. Büchner; W. Nogueira; F. Klefenz, *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, Publication Number: WO/2007/033762, 2006-2007.

[14] W. Nogueira; T. Harczos; B. Edler; J. Ostermann; A. Büchner, *Automatic Speech Recognition with a Cochlear Implant Front-End*, Interspeech, 2007.

[15] W. Nogueira; A. Büchner; T. Lenarz; B. Edler, *A Psychoacoustic NofM-Type Speech Coding Strategy for Cochlear Implants*, EURASIP Journal on Applied Signal Processing, Vol. 127, 2005.

[16] P. P. Vaidyanathan, *The Theory of Linear Prediction*, Synthesis Lectures on Signal Processing, Vol. 2, No. 1, 2007.

[17] G. E. P. Box; D. A. Pierce, *Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models*, Journal of the American Statistical Association, Vol. 65, No. 332, 1970.

[18] B. Kovesi; S. Ragot; C. Lamblin; L. Miao; Z. Liu; C. Hu, *Re-engineering ITU-T G.722: Low delay and complexity superwideband coding at 64 kbit/s with G.722 bitstream watermarking*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011.

[19] International Telecommunication Union, *Recommendation ITU-T G.722*, Series G: Transmission Systems and Media, Digital Systems and Networks, 2012.

[20] S. Preihs; J. Ostermann, *Error Robust Low Delay Audio Coding Based on Subband ADPCM*, AES Convention: 131, October 2011.

[21] I.J. Hochmair-Desoyer; E. E. Schulz; L. H. Moser; M. Schmidt, *The HSM sentence test as a tool for evaluating the speech understanding in noise of cochlear implant users*, The American Journal of Otology, November 1997.

[22] European Broadcasting Union, *EBU SQAM CD - Sound Quality Assessment Material Recordings for Subjective Tests*, https://tech.ebu.ch/publications/sqamcd, 2008, online - last access 11.10.2018.

[23] G. D. Watkins; B. A. Swanson; G. J. Suaning, *An Evaluation of Output Signal to Noise Ratio as a Predictor of Cochlear Implant Speech Intelligibility* Ear and Hearing, Vol. 39, No. 5, September 2018.

[24] W. Nogueira; T. Rode; A. Büchner, *Spectral contrast enhancement improves speech intelligibility in noise for cochlear implants* The Journal of the Acoustical Society of America, Vol. 139, No. 2, 2016.

[25] O. R. Qazi; B. Dijk; M. Moonen; J. Wouters *Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility*, Hearing Research, Vol 299, May 2013.