An Auditory Model Based Strategy for Cochlear Implants

Waldo Nogueira, András Kátai, Tamás Harczos, Frank Klefenz, Andreas Buechner and Bernd Edler

***Abstract*— A physiological and computational model of the human auditory system has been fitted in a signal processing strategy for cochlear implants (CIs). The aim of the new strategy is to obtain more natural sound in CIs by better mimicking the human auditory system.**

The new strategy was built in three independent stages as proposed in [6]. First a basilar membrane motion model was substituted by the filterbank commonly used in commercial strategies. Second, an inner hair cell model was included in a commercial strategy while maintaining the original filterbank. Third, both the basilar membrane motion and the inner-hair cell model were included in the commercial strategy.

This paper analyses the properties and presents results obtained with CI recipients for each algorithm designed.

I. INTRODUCTION

Cochlear implants are accepted as an effective method for improving the auditory receptive abilities of people with profound hearing loss. Current cochlear implants consist of a microphone, a speech processor, a transmitter, a receiver and an electrode array which is positioned inside the cochlea [1]. The speech processor is responsible for decomposing the input audio signal into different frequency bands and delivering the most appropriate stimulation pattern to the electrodes.

Speech coding strategies play a very important role in maximizing the user's overall communicative potential. "N-of-M" strategies [2] such as Advanced Combinational Encoder (ACE), separate speech signals into M sub-bands and derive envelope information from each band signal. N bands with the largest amplitude are then selected for stimulation (N-of-M). The N selected bands are compressed to adapt for the narrower dynamic range of electrical evoked hearing.

These strategies represent only a very simple approximation to processing in the normal cochlea [6], [5]. A bank of linear bandpass filters is used instead of the nonlinear and coupled filters that would model normal auditory function. Furthermore, an instantaneous non-linear mapping is used to produce the whole compression that the normal system performs in several steps with large adaption effects[6], [5].

In [5] has been pointed the importance of phase relationships (delay trajectories) between the stimulation patterns in adjacent positions along the basilar membrane. These phase relationships can be used to extract important features of sound. Results from [10] have shown that it is possible to recognize speech from the phase relationships produced in the basilar membrane indeed in difficult noise situations. The filterbank used in the ACE strategy does not model this effect.

W. Nogueira and B. Edler are with the Laboratory of Information Technology, Lebniz University of Hannover, Germany nogueira@tnt.uni-hannover.de

A. Katai, T. Harczos and F. Klefenz are with Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany klz@idmt.fraunhofer.de

A. Buechner is with the Hoerzentrum Hannover, Medical University of Hannover, Germany buechner@hoerzentrum.de

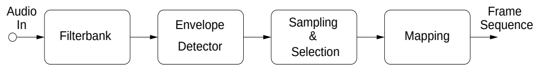

Fig. 1. *ACE Block Diagram.*

Therefore, different signal processing strategies have been designed to provide a closer mimicking of normal auditory functions. The new processor is based on a Extended Zwicker/Meddis-Poveda auditory model [10]. The new strategies have been included separately and together in the commercial ACE strategy.

Section 2 gives an introduction to the ACE strategy. Section 3 presents the auditory model. In section 4, the new strategies based on the auditory model are presented. Section 5 shows some objective experiments performed with the new strategies. Section 6 presents the results obtained with cochlear implant recipients. Finally, Section 7 outlines some conclusions.

II. THE ACE STRATEGY

The ACE (Figure 1) is an "NofM"-type strategy used with the Nucleus implant [3]. A signal sampled at 16 kHz is sent through a filterbank. The filterbank is implemented with an FFT (Fast Fourier Transform). The block update rate of the FFT is adapted to the rate of stimulation on a channel (i.e the total implant rate divided by the number of bands selected, N).

The FFT bins are combined by summing the powers to provide the required number of frequency bands M; the envelope in each spectral band is thus obtained. Each spectral band is allocated to one electrode and represents a single channel.

In the "Sampling and Selection" block, a subset of N ($N < M$) envelopes with the largest amplitude are selected for stimulation. It has been shown that by selecting a subset of bands speech performance can be improved [2].

Finally, the "Mapping" block, determines the current level from the envelope magnitude and the channel characteristics. This is done by using the Loudness Growth Function (LGF), which is a logarithmically-shaped function that maps the acoustic envelope amplitude $a(z_i)$ to an electrical

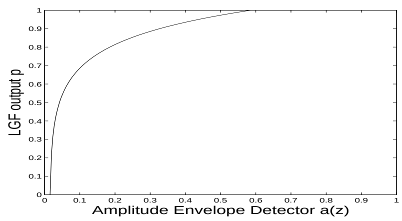

Fig. 2. *Loudness Growth Function.*

magnitude $p(z_i)$ as follows.

$$p(z_i) = \begin{cases} \frac{log\left(1+\rho\left(\frac{a(z_i)-s}{m-s}\right)\right)}{log(1+\rho)} & s \leq a(z_i) \leq m \\ 0 & a(z_i) < s \\ 1 & a(z_i) \geq m. \end{cases} \quad (1)$$

The magnitude $p(z_i)$ is a fraction in the range 0 to 1 that represents the proportion of the output range (from the threshold *T* to the comfort level *C*). 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 [3]

Finally, the channels z_i are stimulated with levels:

$$l_i = T + (C - T)p_i. \quad (2)$$

The set of l_i $(i = 1..N)$ form the frame sequence. A frame is generated at a rate defined by the channel stimulation rate.

III. The Auditory Model

The auditory model first computes the velocitiy of the basilar membrane (BM) excited by a time varying window using the Extended Zwicker (EZ) model [7]. The EZ model generates 251 -channel output data, where each channel has a different frequency, ranging from 5 Hz to 21 kHz.

The mechano-chemical coupling of the BM velocity is mediated by the forced movement of the stereociliae of the inner hair cells (IHC). The movement depolarizes the IHCs resulting in neurotransmitter vesicle releases. This process is modeled according to the rate kinetics equations as given by Meddis and collegues in [8].

The goal of this work is to fit the auditory model described above into a cochlear implant strategy. Compression, bandwidth, frequency and temporal resolution have to be conveyed to the limitations of the cochlear implant while maintaining the features of these models.

IV. A Cochlear Implant Speech Processing Strategy based on an Auditory Model

The fitting of the above described auditory model has been performed in three stages as suggested in [6]. First, the bank of linear bandpass filters of the ACE has been substituted by the Extended Zwicker model. This configuration has been termed EZ-ACE strategy.

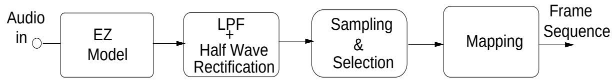

Fig. 3. *EZ-ACE Block Diagram.*

Second, the envelope detector of the ACE strategy has been substituted by the Meddis IHC model described in [9]. This configuration has been termed IHC-ACE strategy.

Finally, in the third configuration, both the bank of linear bandpass filters and the envelope detector have been substituted by the Extended Zwicker model and the Meddis IHC model. This strategy has been termed EZ-IHC-ACE.

The fitting of these three strategies is presented in more detail in the following subsections.

A. EZ-ACE Strategy

The block diagram of the EZ-ACE strategy is presented in Figure 3. An audio signal sampled at 100 kHz is decomposed into several bands or sections, each one corresponding to the movement of the basilar membrane in one position of the cochlea according to the EZ model. The EZ model produces 251 sections along 21 Bark bands. The sample rate at the output of each section is 100 kHz. From all these sections only M=22 are selected for stimulation as the total number of electrodes for the Nucleus implant is 22. The sections selected are always fixed and they correspond to the ones that are closer to the center frequencies of each filter band in the commercial ACE strategy.

The temporal resolution has been reduced using low-pass filtering and half-wave rectification in order to accommodate for the implant stimulation rate.

Afterwards, some of the bands (N-of.M) are selected for stimulation in order to reduce interaction between channels as it is done in the ACE strategy.

Finally, the parameters of the loudness growth function in the EZ-ACE strategy have been adapted to the new filterbank. The EZ model models the compression produced in the basilar membrane (Figure 4b). This issue is not modeled by the FFT and envelope detector used in the ACE (Figure 4a). For this reason, the steepness of the loudness growth function was reduced in the new strategy.

B. IHC-ACE Strategy

The block diagram of this strategy is presented in Figure 5. An audio signal sampled at 16 kHz is processed using the same FFT and envelope detector as the ACE strategy. An IHC model was incorporated at the output of each envelope detector using an interface. This interface interpolates each envelope over the time until 40 kHz. This temporal resolution was necessary to avoid negative quantities appearing in the reservoirs that describe the IHC Meddis Model.

The cleft contents signal obtained at the output of the Meddis model [9] was used as output of the IHC model as proposed in [6]. The cleft content signal was integrated during a time interval equal to the cochlear implant stimulation period.

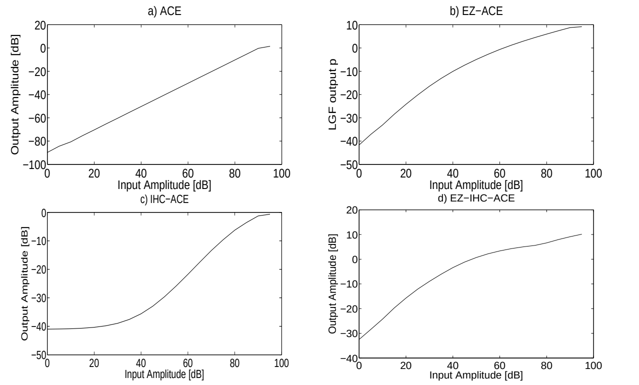

Fig. 4. *Compression obtained at the input of the LGF (8th electrode) for the a) ACE , b) EZ-ACE, c) IHC-ACE strategy and d) EZ-IHC-ACE. The compression was obtained by presenting a set of tone bursts of 1 kHz and 1 second length with amplitudes varying from 0 until 90 dB.*

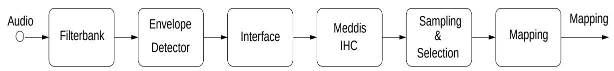

Fig. 5. *IHC-ACE Block Diagram*

The Loudness Growth Function was configured to be less steep than in the commercial ACE as the IHC model already performs a compression (Figure 4c).

C. EZ-IHC-ACE Strategy

The third strategy designed was termed EZ-IHC-ACE and its block-diagram is presented in Figure 6.

An audio signal sampled at 100 kHz is introduced in the EZ model. Only M=22 sections of the 251 were selected as in the EZ-ACE strategy. At the output of this stage each of the M sections are introduced into an independent IHC Meddis model. The output of the IHC model was the cleft content signal which was temporally integrated during a time interval equal to the cochlear implant stimulation period. Afterwards NofM bands were selected to further processing. The N amplitudes selected were then introduced into the LGF function. The compression produced by the EZ and the IHC model caused that no further compression in the LGF was necessary.

V. Objective Experiments

Objective experiments have been performed to test the adaption, phase locking and delay trajectories with each strategy.

A. Adaption

The adaption phenomena was evaluated using 250 ms, 1 kHz tone bursts of increasing amplitude from interspersed with long silent intervals. In each iteration the tone burst

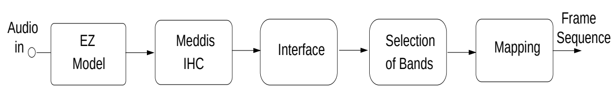

Fig. 6. *Block Diagram of the EZ-IHC-ACE strategy*

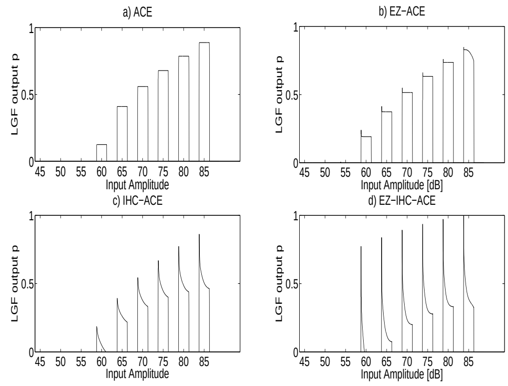

Fig. 7. *Adaption with a) ACE, b) EZ-ACE, c) IHC-ACE and d) EZ-IHC-ACE*

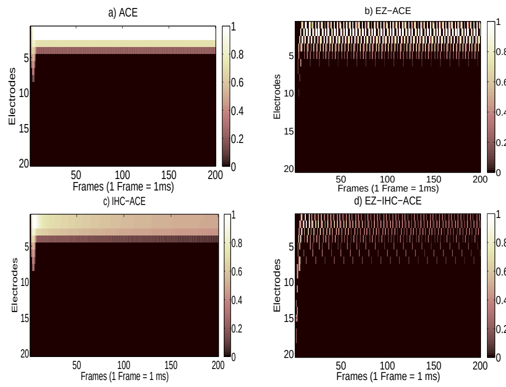

Fig. 8. *a) Stimulation pattern obtained for a tone burst of 250 Hz and 80 dB for the ACE, b) EZ-ACE, c) ACE-IHC and d) EZ-IHC-ACE*

amplitude was increased in 5 dB, in total the amplitude range covered from 45 dB until 100 dB. The stimulation patterns produced in the 8th electrode by each strategy are presented in Figures 7a, 7b, 7c and 7d. These results can be compared with the ones presented in [8].

B. Delay Trajectories and Phase Locking

A tone burst of 250 Hz and 80 dB has been used to analize the delay trajectories and phase locking produced by each strategy.

Figures 8b and 8d show that including the EZ model in the ACE causes the representation of the so called delay trajectories, which are not represented by the ACE filterbank (Figures 8a and 8c). Furthermore, the temporal structure of the sinusoid is better represented by the EZ model than by the filterbank used in ACE and IHC-ACE.

C. Stimulation Patterns

Finally, the stimulation patterns for a speech token, where “aka” is uttered by a man, are presented in Figures 9a, 9b, 9c and 9d.

VI. Subjective Experiments

The EZ-ACE, the IHC-ACE and the EZ-IHC-ACE strategies have been incorporated into a research ACE strategy

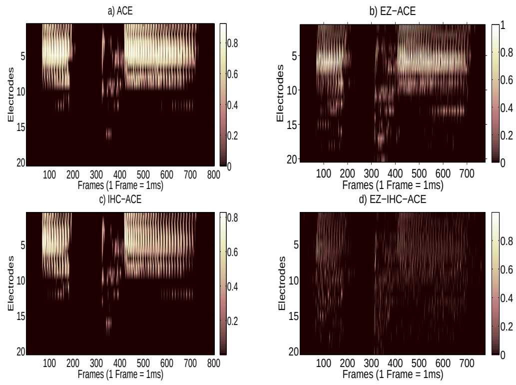

Fig. 9. *a) Stimulation pattern obtained for the speech token 'aka' for the a) ACE, b) EZ-ACE, c) IHC-ACE and d) EZ-IHC-ACE*

made available by Cochlear Corporation, termed NIC (Nucleus Implant Communicator). The NIC processes the audio signals on a personal computer (PC). A specially initialized clinical speech processor serves as a transmitter for the instructions from the PC to the subject's implant. The three strategies programmed within the NIC environment were tested on subjects using the Nucleus 24 implant.

id	**Age**	**Duration deafness (years)**	**Implant experience (years)**	**Rate** (Hz)
P1	62	0	5	900
P2	26	0	4	1440
P3	53	3.66	6	720
P4	68	0.75	9	1080
P5	52	7.75	3.66	1080
P6	39	15.33	7	720
P7	37	0	4	1200
P8	42	0	6	720

TABLE I

SUBJECT DEMOGRAPHICS

The test material was the HSM (Hochmair, Schulz, Moser) sentence test [4]. The signals were processed in noise, with a signal-to-noise ratio (SNR) of 15 dB. Furthermore, the test material had previously been pre-emphasized by a filter which mimics the frequency response of the microphone used in commercial cochlear implant systems. The test subjects (Table 1) spent some minutes listening to the processed material. 2 lists of 20 sentences were presented with the ACE, the EZ-ACE, the IHC-ACE and the EZ-IHC-ACE. The subjects had to repeat each sentence without knowing which strategy they were listening to. This procedure was carried out on three patients for each strategy. Figure 10 presents the averaged scores obtained by each test subject for the different strategies.

VII. DISCUSSION

This paper has presented three signal processing strategies for cochlear implants based on a physiological and computational model of the human auditory system. The improvement of these strategies can be caused by the better

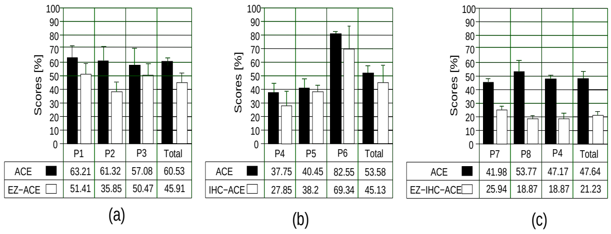

Fig. 10. Score by patient (average and standard deviation) in noise conditions (SNR=15 dB) for a) ACE vs EZ-ACE, b) ACE vs IHC-ACE and c) ACE vs EZ-IHC-ACE.

mimicking of the delay trajectories that occur in the basilar membrane, the adaption effect that occurs in the inner-hair cells, the compression produced by both models and the better modeling of phase-locking. All these features are not modeled in the commercial ACE strategy.

First results measuring speech intelligibility with cochlear implant recipients did not reveal an improvement respect to the ACE. However all patients were used to the ACE strategy and the stimulation patterns produced by the new strategies were significantly different to those produced by ACE. It is speculated that with a longer period of accommodation, the new processors could achieve better speech intelligibility performance.

VIII. ACKNOWLEDGMENTS

The authors would like to thank the eight patients that participated in this study.

REFERENCES

[1] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz, "Better speech recognition with cochlear implants,", Nature, vol. 352, no. 6332, pp. 236-238, 1991.

[2] W. Nogueira, A. Buechner, Th. Lenarz, B. Edler, " A Psychoacoustic "NofM" -Type Speech Coding Strategy for Cochlear Implants", EURASIP Journal on Applied Signal Processing, vol. 2005, no. 18, pp. 3044-3059, 2005.

[3] "Nucleus MATLAB Toolbox 2.11," Software User Manual, N95246 Isuue 1, Cochlear Corporation, October, 2001.

[4] I. Hochmair-Desoyer, E. Schulz, L. Moser, M. Schmidt, "The HSM sentence test as a tool for eavaluating the speech understanding in noise of cochlear implant users", J. Otol, 18 (Suppl. 6): S83, November 1997.

[5] B. C. J. Moore, "Coding of Sounds in the Auditory System and Its Relevance to Signal Processing and Coding in Cochlear Implants", Otology & Neurotology, 24:243-254, 2003.

[6] B. S. Wilson, R. Schatzer, E. A. Lopez-Poveda, X. Sun, S. T. Lawson and R. D. Wolford, "Two new Directions in Speech Processor Design for Cochlear Implants", Ear & Hearing, 2005.

[7] F. Baumgarte, "A Physiological Ear Model for the Emulation of Masking", Journal of Oto-Rhino-Laryngology and its Related Specialities, Vol. 61, No. 5, 1999.

[8] R. Meddis, M. J Hewitt and T. M. Shackleton, "Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse", J. Acoust. Soc. A, 87(4), April, 1990.

[9] R. Meddis,"Simulation of mechanical to neural transduction in the auditory receptor", J. Acoust. Soc. Am, 79(3), March, 1986.

[10] T. Harczos, G. Szepannek, A. Kátai and F. Klefenz, "An auditory model based vowel classification", in Proc. IEEE. Biomed. Circuits and Systems Conf. BioCAS, pp. 69-72, London, 2006.