Abstract—Mimicking the human ear on the basis of auditory models has become a viable approach in many applications by now. However, only a few attempts have been made to extend the scope of physiological ear models to be employed in cochlear implants (CI). Contemporary CI systems rely on much simpler filter banks and simulate the natural signal processing of a healthy cochlea to only a very limited extent. When looking at rehabilitation outcomes, current systems seem to have reached their peak potential, which signals the need for better algorithms and/or technologies. In this paper, we present a novel sound processing strategy, SAM (Stimulation based on Auditory Modeling), that is based on neuro-physiological models of the human ear and can be employed in auditory prostheses. It incorporates active cochlear filtering (basilar membrane and outer hair cells) along with the mechanoelectrical transduction of the inner hair cells, so that several psychoacoustic phenomena are accounted for inherently. Although possible, current implementation does not make use of parallel stimulation of the electrodes, which matches state-of-the-art CI hardware. This paper elaborates on SAM’s signal processing and provides a computational evaluation of the strategy. Results show that aspects of normal cochlear processing that are missing in common strategies can be replicated by SAM. This is supposed to improve overall CI user performance, which we have at least partly proven in a pilot study with implantees.

Index Terms—Auditory models, cochlear implants, coding strategies, signal processing.

I. INTRODUCTION

TECHNOLOGIES for the Hearing Impaired are becoming highly sophisticated and increasingly accessible. Hearing aids, bone-anchored hearing systems, cochlear implants (CI), and even brainstem implants are FDA approved and technologically mature enough to reliably support everyday life of their users. In the prevalent case of sensorineural hearing loss, a cochlear implant is likely to be the best choice for rehabilitation.

While CIs are inevitably the most successful neural prostheses to date, they are still far from perfect: The majority of the recipients claim not to enjoy listening to music and most CI users are also not capable of carrying on a conversation in noisy or reverberative environments. These deficiencies are at least partly due to the signal processing algorithms (also called strategies) residing in the CI processors. These strategies do not faithfully mimic the functionality of the intact human cochlea, but rely on linear filter banks and general purpose preprocessing methods.

There have been several attempts to enhance CI strategies to closer mimic the normal human cochlea to improve perceived quality and speech recognition rate. Most of these concepts try to incorporate one specific psychoacoustic property of the human ear. In that way, each of them addresses only one specific issue like missing cochlear delays [1], improper compression [2], lack of adaptation [4] or the complete absence of temporal fine structure [5]. Yet, by fixing any of these deficiencies alone, only small increases in overall user performance could be shown.

A different approach is not to extend a conventional CI signal processing strategy, but to completely replace it with an auditory model (AM) and a suitable stimulus coder. Although deemed as a good idea (see e.g., pp. 255–257 in [6]), this approach has been attempted in only a few published studies.

Wilson and his colleagues [8] have incorporated the dual resonance nonlinear (DRNL) basilar membrane model [7] with and without the Meddis inner hair cell model [17] into a series of CI processors. Even though they could not show outstanding user performance benefits, they rated their results as encouraging. Kim et al. have also conducted studies with an adapted version of the DRNL strategy [3]. Through hearing experiments with acoustic simulation they have shown improved syllable identification in normal hearing subjects in the presence of speech-spectrum-shaped noise.

Another study [29] has shown that only exchanging the FFT filter bank of the ACE strategy for an auditory filter bank does not automatically improve perceived quality or speech recognition rate. On one hand, the applied auditory filter bank contained a superseded inner ear model. On the other hand, the temporal fine structure preserved by the filter bank was destroyed by the stimulus coder, which was an unmodified ACE coder.

Considering all the above facts, we wanted to create a speech coding strategy (1) with an active cochlea model, (2) which would resemble human measurement data in terms of cochlear delays, compression, adaptation and cues for phase-locking, (3) processed further by a custom-tailored coder that would
be able to map temporal information from the model output onto the available slots of the sparse (sequential) CI stimulation pattern. We hypothesize that by combining the AM with an appropriate coder, all relevant psychoacoustic properties will implicitly show up in the stimulation pattern. This is expected to improve CI user performance (at the expense of computational complexity).

This paper presents a new CI strategy called SAM (Stimulation based on Auditory Modeling). The AM used in SAM includes sub-models of the peripheral ear, of the nonlinear mechanical filtering (basilar membrane and outer hair cells), and of the mecanoelectrical transduction (inner hair cells), as presented throughout Section II.A. The ensemble of these sub-models is responsible for the psychoacoustic features like realistic cochlear delays, adaptation, compression, and phase-locking, which we will elaborate on in Section III.A.

Electrical stimulus generation in SAM is completely based on the output of the real-time simulation of the above-mentioned auditory model. The coder is designed to preserve psychoacoustic properties of the auditory model. Unlike in common strategies, it is not restricted by a pre-defined channel stimulation rate and it activates stimulating electrodes in a stochastic manner. The coder is described in Section II.B.

Contemporary auditory prostheses use various sound processing strategies. Yet, the most wide-spread are CIS (Continuous Interleaved Sampling), ACE (Advanced Combination Encoder) and HiRes (“HiResolution”) as well as their variants (see e.g., pp. 381–432 in [24] for an overview). Even though it would be much fairer to compare SAM with other experimental CI strategies, due to their non-availability, we will use ACE in comparison to SAM throughout this study. Considering that most CI processors worldwide currently employ the ACE strategy, this comparison is still meaningful.

II. SYSTEM

SAM consists of two main parts: auditory model and coder, as shown in Fig. 1. The auditory model is generalized and mimics human auditory processing up to the level of neurotransmitter vesicle release of the inner hair cells. It can be seen as a bio-inspired filter bank. The coder, on the other hand, is designed to be fitted for each individual CI user.

A. Auditory Model

The auditory model, which is the cornerstone of the SAM strategy, is described in this section. Its general structure is shown on the left-hand side of Fig. 1. The peripheral ear is represented by a model of outer and middle ear (OME) filtering. It is connected to a model of nonlinear mechanical filtering (NLM) simulating the passive cochlear hydromechanics enhanced by the active outer hair cells (OHCs). The output of this stage is processed further by the model of mecanoelectrical transduction, which mimics the physiological properties of inner hair cells including the simulation of receptor potential, calcium kinetics and transmitter dynamics.

Most model components can be specified by differential equations, which are internally represented by equivalent electrical circuits using voltage-velocity and force-current analogies. All electrical networks are simulated in the time domain using wave digital filters, because of their excellent stability properties and efficiency [11].

Peripheral Ear: Outer and middle ear filtering, i.e., the transformation of incoming sound into the velocity of the oval window, is modeled by a linear filter. The resonant frequency of the filter is 3 kHz, while the damping is increasing for frequencies below, see right inset of Fig. 2. For frequencies above, constant magnitude response is assumed. This functionality can be described by an electrical network [12] as illustrated in the left inset of Fig. 2. Here, sound pressure at the outer ear is given by the current $i_{QO}$. The resonance is realized by a series RLC circuit, whereas the damping of lower frequencies is incorporated by $R_{ML}$ and $L_{ML}$. The velocity of the oval window is given by the output voltage $u_{OW}$.

Nonlinear Mechanical Filtering: The oval window velocity ($u_{OW}$) is the input to the nonlinear filtering, which models cochlear mechanics, i.e., the propagation of the traveling wave along the cochlear partition. For this purpose a model developed by Zwicker and Peiße [10] with modifications from Baumgarte (see [9] and [12]) has been employed (for the statement of reasons see Section IV.A). It is a one-dimensional macromechanical model of the cochlea, in which the unrolled cochlear duct is divided into sections of equal length. Each section mimics the passive cochlear hydromechanics and the active nonlinear effects of the outer hair cells of the corresponding cochlear region. This system can be described by differential equations, which can be represented by an electrical circuit as shown in Fig. 3.

Each section of the passive hydromechanical model consists of a parallel resonant circuit. RLC parameters express the di-
TABLE I

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>Symbol</th>
<th>Value(s)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Middle ear model resistor 1</td>
<td>$R_{m1}$</td>
<td>12.7 Ω</td>
<td>[12]</td>
</tr>
<tr>
<td>Middle ear model resistor 2</td>
<td>$R_{m2}$</td>
<td>31.75 Ω</td>
<td></td>
</tr>
<tr>
<td>Middle ear model inductor</td>
<td>$L_{m}$</td>
<td>3.175 mH</td>
<td></td>
</tr>
<tr>
<td>Outer ear model resistor</td>
<td>$R_{o}$</td>
<td>6.35 Ω</td>
<td></td>
</tr>
<tr>
<td>Outer ear model capacitor</td>
<td>$C_{o}$</td>
<td>1.575 μF</td>
<td></td>
</tr>
<tr>
<td>Outer ear model inductor</td>
<td>$L_{o}$</td>
<td>1.7875 mH</td>
<td></td>
</tr>
<tr>
<td>Number of NLM sections</td>
<td>$N_{NLM}$</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>Resolution of the NLM</td>
<td>$A_{r}$</td>
<td>0.25 Bark</td>
<td></td>
</tr>
<tr>
<td>Data rate of the auditory model</td>
<td>$f_{d}$</td>
<td>44100 s⁻¹</td>
<td></td>
</tr>
<tr>
<td>NLM friction loss resistance</td>
<td>$R_{f}$</td>
<td>400 kΩ</td>
<td>[12]</td>
</tr>
<tr>
<td>NLM endolymph mass capacitor</td>
<td>$C_{p}$</td>
<td>18.75 mF</td>
<td></td>
</tr>
<tr>
<td>NLM section mass capacitor</td>
<td>$C_{n}$</td>
<td>(4.2π + 310) mF</td>
<td></td>
</tr>
<tr>
<td>NLM helicotrema resistance</td>
<td>$R_{H}$</td>
<td>281.1 kΩ</td>
<td>[12]</td>
</tr>
<tr>
<td>NLM helicotrema capacitor</td>
<td>$C_{H}$</td>
<td>56.6 mF</td>
<td></td>
</tr>
<tr>
<td>NLM amplification factor 1</td>
<td>$g_{1}$</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>NLM amplification factor 2</td>
<td>$g_{2}$</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>NLM to IHC cilia coupling gain</td>
<td>$C_{c,a}$</td>
<td>-65 dB</td>
<td></td>
</tr>
<tr>
<td>NLM to IHC cilia time constant</td>
<td>$τ_{c}$</td>
<td>2.13 ms</td>
<td>[22]</td>
</tr>
<tr>
<td>Max. mechanical conductance</td>
<td>$C_{max}$</td>
<td>8 nS</td>
<td></td>
</tr>
<tr>
<td>Apical resting conductance</td>
<td>$C_{r}$</td>
<td>1.974 nS</td>
<td></td>
</tr>
<tr>
<td>Displacement sensitivity</td>
<td>$s_{0}$</td>
<td>85 e-9 m⁻¹</td>
<td></td>
</tr>
<tr>
<td>Displacement offset</td>
<td>$u_{0}$</td>
<td>7 nM</td>
<td></td>
</tr>
<tr>
<td>Displacement offset</td>
<td>$u_{1}$</td>
<td>5 e-9 m⁻¹</td>
<td></td>
</tr>
<tr>
<td>Displacement offset</td>
<td>$u_{2}$</td>
<td>7 nM</td>
<td></td>
</tr>
<tr>
<td>Cell capacitance</td>
<td>$C_{u}$</td>
<td>15 μF</td>
<td>[23]</td>
</tr>
<tr>
<td>Endocochlear potential</td>
<td>$E_{c}$</td>
<td>18 mS</td>
<td>[22]</td>
</tr>
<tr>
<td>Potassium conductance</td>
<td>$E_{k}$</td>
<td>100 mV</td>
<td></td>
</tr>
<tr>
<td>Potassium reversal potential</td>
<td>$E_{r}$</td>
<td>-70.45 mV</td>
<td></td>
</tr>
<tr>
<td>Resistance of the supporting cell</td>
<td>$R_{p} + R_{s}$</td>
<td>40 mΩ</td>
<td></td>
</tr>
<tr>
<td>Reversal potential (for calcium)</td>
<td>$E_{Ca}$</td>
<td>66 mV</td>
<td></td>
</tr>
<tr>
<td>Maximum calcium conductance</td>
<td>$C_{max}$</td>
<td>7.2 nS</td>
<td></td>
</tr>
<tr>
<td>Calcium constant 1</td>
<td>$β_{Ca}$</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>Calcium constant 2</td>
<td>$ρ_{Ca}$</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>Calcium time constant</td>
<td>$τ_{Ca}$</td>
<td>100 μs</td>
<td></td>
</tr>
<tr>
<td>Calcium diffusion time constant</td>
<td>$τ_{Ca}$</td>
<td>100 μs</td>
<td></td>
</tr>
<tr>
<td>Calcium concentration threshold</td>
<td>$[Ca^{2+}]_{o}$</td>
<td>0 mol</td>
<td></td>
</tr>
<tr>
<td>Release rate conversion scalar</td>
<td>$z$</td>
<td>2 e32</td>
<td>[15]</td>
</tr>
<tr>
<td>NT loss rate</td>
<td>$l$</td>
<td>2580 s⁻¹</td>
<td>[22]</td>
</tr>
<tr>
<td>NT recovery rate</td>
<td>$r$</td>
<td>6580 s⁻¹</td>
<td>[22]</td>
</tr>
<tr>
<td>NT reprocessing rate</td>
<td>$x$</td>
<td>66.3 s⁻¹</td>
<td>[15]</td>
</tr>
<tr>
<td>NT replenishment rate</td>
<td>$y$</td>
<td>10 s⁻¹</td>
<td>[15]</td>
</tr>
<tr>
<td>Maximum free NT quanta</td>
<td>$M$</td>
<td>10</td>
<td>[22]</td>
</tr>
</tbody>
</table>

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

...dimensions and material properties of the fluid filled cochlear duct. $L_{m}$, $C_{m}$ and $R_{m}$ represent the compliance (flexibility), the mass (including that of the axially moving endolymph) and the friction loss of the given section $n$, respectively. The resulting resonance frequency of the parallel resonant circuit represents the characteristic frequency of the cochlear partition at the given position. The additional capacitor $C_{q}$ and resistor $R_{q}$ represent the mass of the longitudinally moving endolymph and the resulting friction loss, respectively. The helicotrema is modeled by the parallel connection of a resistor ($R_{H}$) and a capacitor ($C_{H}$), as on the bottom of Fig. 3. Model parameters are based on physiological data from the literature.

As already stated above, the passive model is extended by a model of the outer hair cells. As shown in Fig. 3, the velocity of the cochlear partition is amplified within a nonlinear feedback loop using a voltage controlled voltage source ($f_{IHC}$) and a symmetrical saturating function (labeled SAT). These mimic the nonlinear amplification introduced by the outer hair cells in the physiological cochlea. The output voltage $v_{l_{OHC}}$ of the amplifier is fed back to the passive model and in addition, neighboring OHC models are coupled via resistors ($R_{a}$ and $R_{b}$). These correspond to the natural OHC to BM feedback and the interactions between neighboring OHCs, respectively. Furthermore, there is a second amplification stage (SAS) outside the feedback loop. It consists of a current controlled current source ($g_{2} l_{OHC}$) and of a parallel resonance circuit (labeled PRLC), which causes a frequency dependent voltage drop. The SAS was implemented to ensure the high amplification featured by the OHCs of the human ear while preserving stability of the model [12]. The output $v_{l_{BM,n}}$ of the SAS corresponds to the modeled basilar membrane (BM) velocity.

To better adapt to up-to-date measurement data, we have tuned relevant model parameters, as presented in Table I. First, the saturating characteristic of the amplifier stage has been calibrated so that the nonlinear gain of the model best fits experimental data published in [14]. Second, we have adjusted $C_{TH}$ and $C_{q}$ of the hydromechanical model in a way that traveling wave delays (as given in [13]) can be replicated appropriately by the model. These parameters were based on outdated measurements of the human cochlear geometrics provided by Zwislocki [34]. Zwicker and Peisl derived all other RLC values from the desired vibration properties of the given NLM section [10]. Therefore, these parameters cannot become obsolete and do not need to be updated.
We have also incorporated some changes to increase simulation speed, which is crucial for applicability in cochlear implant stimulation. One important parameter of the model is the sampling rate, \( f_s \). Because of inherent nonlinearities, the sampling rate should be at least quadruple the maximum signal frequency to avoid aliasing artifacts. Baumgarte [12] used 100000 S/s (samples per second) for applications in music coding, but since cochlear implants usually do not deal with frequencies above 8 kHz (16000 S/s input sampling rate), we have chosen a sampling rate of 44100 S/s for this application. Furthermore, we have reduced the number of sections, \( N_{BM} \), from 251 to 101, so that the frequency resolution of the down-sized model is \( \Delta f = 0.25 \) Bark, which makes a mean distance of 0.34 mm between adjacent sections along the modeled cochlea. Spatial discretization in this case is still fine enough to avoid reflections at section boundaries, which would lead to undesired artifacts by standing waves.

**Mechanoelectrical Transduction:** The calculated basilar membrane velocity along the cochlear partition forms the input to the model of mechnoelectrical transduction. To mimic the functionality of the inner hair cells (IHCs) the model described by Sumner et al. [16] was adapted and employed. Sumner et al. extended an earlier model of Meddis [17] by combining the neurotransmitter dynamics with a modified biophysical model of the receptor potential [18] and with calcium driven mechanism for the neurotransmitter release (see e.g., [19] and [20]) as shown in Fig. 4.

In the biophysical model (upper part of Fig. 4) of the hair cell, the vibration of the cochlear partition is first translated into the displacement of the stereocilia \( u(t) \), which has an effect on the number of open ion channels. This in turn alters the apical conductance \( G(u) \). The resulting change in the receptor potential \( V(t) \) can be modeled as a passive electrical circuit.

The release of neurotransmitter, which is dependent on the receptor potential, is mediated by calcium ions (middle part of Fig. 4). Calcium ion movements are modeled as a three-step process. First, the membrane depolarization opens calcium ion channels, so that the calcium current \( I_{Ca}(t) \) can be described as a function of the receptor potential. As calcium ions enter the cell, they accumulate in the vicinity of the synapse. The resulting calcium concentration \([Ca^{2+}]\) is modeled as a low-pass filtered function of the calcium current. Finally, the local calcium concentration determines the release probability, which is then expressed as the transmitter release rate \( k(t) \).

The transmitter release rate, on the other hand, drives a model of transmitter dynamics (lower part of Fig. 4). Instead of the stochastic version described by Sumner [16] we use the original Meddis transmitter recycling model [17] and incorporate random noise at a later stage, in the stimulus coder. This way, the magnitude of the stochastic variation can be customized more easily, and at a lower computational cost.

Neurotransmitters circulate between three reservoirs: the immediate store \( q(t) \), the cleft \( c(t) \) and the reprocessing store \( w(t) \). A certain quantity of available transmitter is kept in the immediate store. These are released across the membrane into the synaptic cleft at the momentary transmitter release rate. There they disperse and some are lost from the cleft (at a loss rate of \( \ell \)). The remaining transmitter is taken back (at a rate of \( \eta \)) into the reprocessing store and returned again into the immediate store (at a rate of \( \varphi \)). Neurotransmitter loss is counterbalanced by the factory, which produces new transmitter within the cell (at a rate of \( \gamma \)). For any time step \( i \) the discrete time simulation of the auditory model, \( C_n[i] \) denotes the neurotransmitter substance concentration near to the cell membrane exocytosed by the \( n^{th} \) IHC.

The modeled mechnoelectrical transduction can be tuned to reproduce rate intensity functions of low-, medium- and high-spontaneous rate fibers. The current implementation uses only high-spontaneous rate (HSR) fibers, because, on one hand, HSR is the most common fiber type in the human ear, and, on the other hand, sensitivity of HSR fibers fits the input dynamic range of cochlear implants (typically 25–65 dB SPL) best (see [16] and [21]).

Functional characteristics of synaptic adaptation are simulated by the model of transmitter dynamics. In addition, the model also reproduces phase-locking which is restricted to low frequencies. We will elaborate on the psychoacoustic features of the auditory model throughout Section III.A.

**B. Coding of Electrical Stimuli**

While the auditory model is generalized to mimic an average human ear, the coder must produce stimuli that uniquely fit the needs of each CI user. In addition, the large amount of output data from the auditory model must be decimated in a meaningful way to produce the sparse CI stimulation pattern. This section describes how the stimulation pattern gets calculated in SAM.
An overview of the processing steps is shown on the right-hand side of Fig. 1. Coder parameters are summarized in Table II.

**Interface:** First, the multi-channel output signal of the auditory model is resampled to the intended total rate of the stimulation. Considering that the auditory model includes low-pass filtering, a maximum-preserving decimation can also be used for typical target rates (like 9000 pps) at much lower computational costs than a full-featured resampling. Equation (1) explains the decimation for a one-second piece of data. In (1) \( \hat{C}_n \) denotes the \( n \)th channel of the auditory model output, \( \hat{C}_n \) is the decimated data and \( TPR \) is the total pulse rate.

\[
\hat{C}_n[i] = \max(\hat{C}_{n}[j], j = j_1, j_1 + 1, \ldots, j_2),
\]
\[
\hat{C}_n[i] = \left\lfloor f_s \cdot (i - 1)/TPR + 1 \right\rfloor,
\]
\[
\hat{C}_n[i] = \left\lfloor f_s \cdot i/TPR \right\rfloor
\]  

(1)

**Frequency Mapping:** Desired electrode frequencies may be different from user to user. One possible cause is the varying insertion depth of the electrode array. To match frequencies, model channels have to be mapped to the electrodes. In SAM mapping is fairly straightforward, because the auditory model provides enough channels to choose from. These have large bandwidths (see Fig. 5, bottom), which allow all frequencies to be present even after discarding approx. 80% of the channels. Moreover, since the position of the electrode array inside the cochlea is typically not exactly registered (see e.g., pp. 612–618 in [24]), the mapping does not need to be exact to the Hertz either. The characteristic frequency (CF) of each auditory model channel is known for any input level. Our algorithm looks for the channels with CFs nearest to the desired electrode frequencies at the input level of 55 dB SPL in the initialization phase and uses this mapping further on. The output of this stage is denoted by \( \hat{A}_m, m = 1, 2, \ldots, M_{CT} \), where \( M_{CT} \) is the number of CI electrodes. \( \hat{A}_m \) has the same data rate as \( \hat{C}_n \).

**Loudness Mapping:** To fit the huge dynamic range of the modeled ear into the operational range of the cochlear implant a loudness mapping has to be done. \( A_m \) is fitted to the input dynamic range (IDR) of the CI in a way that very soft signal parts are dropped and very loud parts get limited. IDR is the difference between the so-called T-SPL (threshold input intensity that results in electrical stimulation) and C-SPL (comfortable listening level), which are typically set to 25 dB SPL and 65 dB SPL, respectively [27]. For this step, the algorithm determines model output levels for each channel for pure tones at the level of T-SPL and C-SPL in its initialization phase, and uses these magnitudes further on for comparison. This is necessary, since the model is nonlinear and channels may vary quite a lot in terms of dynamics. Then, the data range of IDR is linearly mapped to the output range of \([0.0, 1.0] \). The result is denoted by \( \hat{A}_m \). Frequency responses of the ACE and SAM filter banks are presented for comparison in Fig. 5.

SAM does not use fixed channel stimulation rates by design (see also Section IV.B in the discussion), but it provides a mechanism to control repeated stimulation through the same electrode. This can be seen as a basic refractory model and the present calculation stage offers a very simple and efficient way to implement it. The mechanism is called repetition penalty and it means that every time an electrode would be activated in two consecutive cycles, the corresponding model channel gets attenuated by \( R_m \). If \( R_m \) is low, the channel may still remain attractive for the coder. If \( R_m \) is high, another channel may be picked to determine the current stimulus. Obviously, \( R_m = 0 \) turns the repetition penalty off, while \( R_m \rightarrow \infty \) does not allow any electrode to be stimulated more frequently than \( TPR/2 \). The output of this stage is denoted by \( \hat{A}_m \).

Finally, the loudness growth function (LGF) is applied. LGF was introduced in cochlear implants with linear filter banks to better mimic the nonlinearities of loudness perception (see e.g., pp. 303–304 and 345–346 in [24]). Since our strategy models parts of the ear, the use of an LGF is only required if the loudness growth should resemble that of another strategy. LGF is applied as shown in (2).

\[
L_m[i] = \frac{\log(1 + c_m \cdot \hat{A}_m[i])}{\log(1 + c_m)}
\]

(2)

In (2) \( c_m \) is the curve shaping factor, with which the steepness of the loudness growth can be varied. For \( c_m \rightarrow 0 \) the mapping...
is linear. For $c_m > 0$ or $-1 < c_m < 0$ the effect is compression or expansion, respectively.

Since we intended to test SAM with CI users already experienced with ACE, the default SAM configuration uses a $c_m$ value of 1.95, with which the loudness growth is very similar to that of ACE (see also Fig. 8).

Stimulus Rendering: The presented implementation of SAM is not designated for parallel electrode stimulation. Hence, in each cycle, the algorithm has to select an electrode to be activated. The decision is based on the momentary channel magnitudes. Given that $M_1$, $M_2$ and $M_3$ denote indexes of channels with the 1st, 2nd and 3rd highest momentary magnitudes, respectively, the probability of any of the three channels to be selected is given by (3).

$$
\begin{align*}
P(M_1) &= 1 - \xi_p, \\
P(M_2) &= \xi_p \cdot (1 - \xi_p), \\
P(M_3) &= \xi_p^2
\end{align*}
$$

(3)

The operation is deterministic if $\xi_p = 0$ and stochastic if $\xi_p > 0$. For example, $\xi_p = 0.1$ would mean that $M_1$, $M_2$ and $M_3$ are selected with 90%, 9% and 1% probability, which would allow less-dominant frequencies to be present in the stimulation pattern. This effect is even stronger if $H_m > 0$.

Next, the magnitude of the selected channel is translated into a current level, $Y$, as described in (4). $Y$ has no physical unit; it is used as an index to look up implant-typical currents in µA.

$$
Y = [0.5 + T H L_m + L_m[v] \cdot v \cdot (M C L_m - T H L_m)], \\
\quad m \in \{M_1, M_2, M_3\}
$$

(4)

In (4) $T H L_m$, $M C L_m$ and $v$ denote the threshold level, most comfortable level and global volume ($0 \leq v \leq 1$), respectively. $T H L_m$ and $M C L_m$ are determined for each channel for each CI user during the fitting sessions. The fitting procedure will be explained in a separate paper describing also test results with CI users.

As an example for the implant driving currents, the conversion from current level to µA with the Cochlear™ Nucleus® Freedom™ with Contour Advance™ electrode is given by (5).

$$
J[v] = 17.5 \cdot 100^Y[v] / 255 \, \mu A
$$

(5)

As a final step, other parameters like phase width and phase gap (i.e., pause between positive and negative parts of a biphasic stimulus) can be set for each stimulus in the pattern. Since there is evidence that jitter can enhance perception [25] and sound source localization abilities [26], we have introduced $\xi_j$, with which the possibility is given to vary the phase gap duration of a biphasic stimulus in a stochastic manner, as shown in (6). This random variation adds some irregularity to the stimulation signal, which reduces periodic characteristic, while preserving fine temporal structures.

$$
t_{\text{gap}} \in U((1 - \xi_j) \cdot T_{pg}, (1 + \xi_j) \cdot T_{pg})
$$

(6)

In (6) $T_{pg}$ is the standard phase gap value (typically around 8 µs), which is always used when $\xi_j = 0$. The phase gap duration used in the current stimulation cycle is denoted by $t_{\text{gap}}$. This value may also be limited by the CI hardware used.

In each cycle, the output of the coder is comprised of the following data: the identifier of the electrode to be activated, the amplitude of the stimulus that should be applied, and the duration of the phase gap. This set of information is denoted by $P$.

The presented stimulus rendering has similarities with both the CIS and n-of-m strategies: The presented coder also works in an interleaved manner (no parallel stimulation of electrodes) and at a high total pulse rate (typically used in CIS strategies). In contrast to CIS, but as is usual in n-of-m strategies, it selects the electrodes to be stimulated in its stimulation cycles. As opposed to ACE, however, each cycle in SAM consists of only one stimulus. This way, the destination of each stimulating pulse is determined by the SAM coder primarily based on the momentary output of the auditory model, complemented by the information about the previously stimulated electrode and by the repetition penalty and randomization settings.

III. COMPUTATIONAL EVALUATION

Prior to a clinical pilot study, we conducted verification and computational evaluation of SAM. The most important evaluation outcomes regarding psychoacoustic properties, influence of key parameters, and computational requirements of the new strategy are presented throughout this section. Where reasonable, a SAM vs. ACE comparison is presented. In all calculations, unless noted otherwise, the default parameter values listed in Tables I and II are used with SAM, and 900 pps channel stimulation rate combined with 10 selected channels per stimulation cycle are used with ACE.

A. Psychoacoustic Features

This section elaborates on the psychoacoustics properties of SAM. These are introduced by the auditory model and are meant to be preserved by the stimulus coder.

Cochlear Delays: SAM stimulation patterns show realistic hyperbolic delay characteristic, which is caused by frequency-specific group delays along the cochlear partition. Delays range from 0.3 ms (most basal electrode) to 6.5 ms (most apical electrode) in the presented 22-electrode setup. A side effect of these delays is to draw spectrally rich components of the input sound apart, so that they can be mapped better to the sequential stimulation pattern [1].

We have tuned the hydromechanical model so that it can replicate delays most appropriately (as given in [13]). Cochlear delays for speech signal can be observed e.g., in Fig. 9. To clearly demonstrate the difference between SAM and ACE, the response to a pulse input with both strategies was captured. A pulse with a duration of 200 µs and an intensity of 65 dB SPL was used. Fig. 6 illustrates the resulting stimulation patterns obtained using the two strategies.

The hyperbolic delay curves of SAM can well be seen, even though the stochastic coder adds a large amount of variation. The resulting delay characteristic of ACE is linear, with a sequential stimulation order from high to low frequencies.
Adaptation: Another psychoacoustic property of SAM is adaptation. It is induced by the saturating dynamics of the neurotransmitter pools of the simulated IHCs and can be observed as dynamic variation in response to constant-intensity input. The effect is most pronounced upon a sudden change of the input signal intensity. To demonstrate this property, the amplitude of a 500 Hz pure tone was scaled in a way that four current levels are used, which translates to a curve shaping factor of approximately 416.2.

Compression: In cochlear implants, compression of acoustic amplitudes is necessary to fit within the electrical dynamic range. Equation (2) on page 5 describes the loudness growth function, which is inherently used for this purpose. SAM inherently includes compression due to the nonlinearities of the cochlear mechanics, and uses (2) only to yield a loudness growth similar to that of ACE. The achieved similarity varies slightly across frequencies and is presented for 1000 Hz in Fig. 8. The plotted data was determined by finding the maximum stimulation level on the corresponding electrode for pure tones (with 50 ms fade-in) presented at various sound pressure levels.

Phase-Locking: The property of the auditory nerve fibers (ANFs) to tend to fire at a particular phase of the stimulating tone is called phase-locking. In normal hearing, it plays an important role in pitch perception and in fine discrimination of frequencies.

It is known that in electrical hearing ANFs are capable of firing in phase with stimulation pulses. Therefore, frequency information can be transmitted by varying the rate of stimulation according to the pitch cues provided by the temporal fine structure.

However, most contemporary CI speech processing strategies (including ACE) use constant rate pulse carriers that contain no temporal information [see panel (h) of Fig. 9]. By contrast, SAM is not restricted to a specific channel stimulation rate, hence its coder is able to transmit temporal fine structure well. As illustrated in panel (g) of Fig. 9 the distance between the pulse bursts on each apical electrode represents the fundamental frequency or its harmonics of the input signal.

A measure used in neurobiological studies to assess how well ANFs synchronize to stimuli is the synchronization index (SI). It can be calculated from the spike trains using FFT [28]. We have adapted this term to measure how well the fundamental frequency of the electrical stimuli represents the frequency of a pure tone input. A value near 1 indicates high synchronization to the input frequency, whereas a value of 0 means no synchronization.

SI can be calculated based on the stimulation pattern for a pure tone with known frequency as follows. First, the power spectrum of the zero-mean stimulation pattern is estimated via FFT. The value of SI is then defined as the sum of the normalized absolute values of the frequency bins that correspond to whole-number multiples of the pure tone frequency. Since the harmonics of the fundamental frequency are included in the SI calculation, this measure is independent of the envelope of the
Fig. 9. Example showing the waveform (top row), auditory model output (second row), SAM coder output (third row), and ACE coder output (bottom row) for the word “choice”. Left column shows data for the whole duration of the utterance, while the right column visualizes data for only 50 ms. While both SAM and ACE feature place coding, the temporal code is only present with SAM.

Fig. 10. Synchronization index trends for SAM and ACE. Higher values indicate better synchronization to the frequency of the pure tone input. While ACE cannot temporally code frequencies above 200 Hz, SAM features noticeable temporal coding up to frequencies well above 1 kHz depending on configuration. Dashed line (SAM) shows results with default parameters, while dashed-dotted line (SAM*) illustrates the case when peak selection randomness and repetition penalty are turned off.

B. Effect of Changing Key Parameter Values

In addition to the individually fitted physiologically induced parameters like $TH_{L_m}$ and $MCL_m$ there are several parameters in SAM that significantly affect coder behavior and may, thereby, affect perception. The next two paragraphs illustrate how the stimulation pattern changes as an effect of varying three of those parameters, namely $TPR$, $\xi_p$, and $R_m$.

Pulse Rate: One of the key parameters in SAM is the total pulse rate. Lower rates allow for longer phase widths, through which the same charges can be delivered employing lower electrical currents. On the contrary, higher rates enable the coder to better map temporally high-resolution parts of the auditory model output to the CI electrodes. CI users seem to have their preferred stimulation rate for any given strategy, which may have various physiological causes. A typical total pulse rate with the ACE strategy is 9000 pps, which we have adopted as default for SAM. In Fig. 11, a direct comparison of SAM and ACE stimulation patterns for lower (7200 pps) and higher (14400 pps) rates is given. During the ACE calculations with various total pulse rates, the number of peaks selected in each stimulation cycle was kept at 10. Note that panels (g) and (h) of Fig. 9 can be used as a reference with default parameter values.

As shown in Fig. 10, synchronization index always falls consistently with increasing input frequency. SAM locks to stimulus phase for frequencies up to about 2 kHz, which complies with knowledge about phase-locking [15]. This limit extends to approx. 4 kHz with peak selection randomness and repetition penalty turned off. In ACE, however, no considerable encoding of phase can be observed for over 100 Hz.
With SAM, cochlear delays and phase-locking effects are preserved over various total pulse rates. The amount of temporal information provided by ACE does not remarkably increases with higher rates.

![Fig. 11. Comparison of coder outputs (first row: SAM, second row: ACE) for a short snippet of the word “choice” for total pulse rates of (a) 7200 pps and (b) 14400 pps. For a clearer view only the 11 most apical electrodes are shown. With SAM, cochlear delays and phase-locking effects are preserved over various total pulse rates. The amount of temporal information provided by ACE does not remarkably increases with higher rates.](image1)

**Fig. 12.** Comparison of SAM output for a short snippet of the word “choice” for different peak selection randomness and repetition penalty settings. (a) $\xi_p = 0$ and $R_m = 0$. (b) $\xi_p = 0.2$ and $R_m = 40$ dB. For a clearer view only the 11 most apical electrodes are shown. In (a) the electrodes are often selected in consecutive stimulation cycles, while in (b) no electrode is active in two successive cycles and the variance of the stimulation is higher.

Peak Selection and Repetition: The other two parameters fine-shaping the stimulation pattern are peak selection randomness and repetition penalty. The former can enable less dominant spectral components to be coded, while the latter controls the likelihood of stimulating through the same electrode in two consecutive cycles. The two parameters are connected to some degree, since the peak selection can be more random if repetition is disallowed.

Fig. 12 shows a comparison for two distinct settings. $\xi_p = 0$ and $R_m = 0$ [Fig. 12(a)], and $\xi_p = 0.2$ and $R_m = 40$ dB [Fig. 12(b)] represent deterministic and stochastic operational modes of SAM, respectively. Panel (g) of Fig. 9 can, again, be used as a reference with default parameter values. Note that with $R_m = 0$ electrodes are often selected in consecutive stimulation cycles (highest momentary magnitude always wins). In the other case no electrode is active in two successive cycles and the variance of the stimulation is much larger.

**C. Computational Requirements**

We have used C/C++ to implement SAM. Current version (v0.93) of the core software consists of approximately 12 500 lines of source code (auditory model: 3000 lines, coder: 2500 lines, interfacing with OS and CI-hardware: 7000 lines). Since this software is designed to be used in clinical tests, a significant part of the software serves the safety of the test subjects, which slightly decreases calculation speed.

On a contemporary PC, the computational power required to process a given piece of audio data with the current implementation of SAM is about 30 times higher than that of ACE. Still, both strategies have linear time complexity and can be run in real-time. A comparison is presented in Fig. 13. Calculations were done on the same computer (Dell OptiPlex 760, Intel® Core™ 2 Duo processor E8500) under the same conditions. The two implementations (SAM and ACE) were optimized at about the same level for the target PC.

**IV. DISCUSSION**

With SAM we are presenting a new way of CI stimulus calculation that is based on an active cochlea model. When compared to wide-spread strategies like ACE, there are quite a few differences: A complex and nonlinear auditory model is used, the filter bank operates in the time-domain, the coder defaults to stochastic behavior and ignores the concept of “channel stimulation rate”. Critics say we have changed too much at once, so that causes for benefits (if any) can not be identified unambiguously. While this argumentation is true, this study has deliberately aimed towards a paradigm shift disregarding many traditions and limitations of previous speech processing strategies. In the next sections we will discuss some key points regarding the design of SAM.

**A. Auditory Model**

Today, a large number of auditory models are available [6]. They can for example be classified into phenomenological and mechanical models. The former reproduce only specific psychophysical properties of cochlear processing, while the latter recreate hydromechanical mechanisms of the cochlea. Consequently, mechanical models implicitly include psychophysical properties of the cochlea and they can be termed as more realistic (at the same time being computationally more intensive).

Unlike Wilson et al. [8] and Kim et al. [3], we have focused on detailed physical modeling and less on keeping computational efforts minimal. Therefore, we have chosen the hydromechanical Baumgarte model [9] instead of the phenomenological DRNL filter bank [7]. Furthermore, through parameter tuning, we have adapted the Baumgarte model to more recent human measurement data (see Table I). All these differences to DRNL...
could lead to more exact representation of psychoacoustic properties. In addition, with respect to Kim et al., we have included a more recent version of the Meddis IHC model [15]. However, the drawback of our compound auditory model utilized in SAM is its complexity.

Taft et al. recently published a study on a modified ACE strategy [1] that included “across-frequency delays” (as referred to by the authors). Both this approach and that of Kim et al. [3] have the advantage of being fast enough to be possibly real-time on contemporary CI processors. Current implementation of SAM is computationally more expensive, because of the time-domain simulation of the BM model and due to the presence of simulated inner hair cells. The former can clearly be identified as the computational bottleneck, and is therefore currently being investigated for possible optimizations. However, the ultimate objective is to turn SAM into a hardware implementation and make use of highly parallel structures to maximize speed. A good basis for this task is provided by the work of Hamilton et al. [30] and Wen and Boahen [31].

It should also be noted that the time complexity of the current implementation of SAM is only minimally influenced by its parameterization. On the contrary, the calculation speed of ACE is directly affected by its settings: For example, the number of FFT operations in a second depends on the value of \( N \) (number of electrode channels selected in a stimulation cycle).

### B. Coder

In SAM, we wanted to break away from defining a fixed channel stimulation rate (CSR). Instead, stimuli are distributed among electrode channels as required. We assumed that salient features of the auditory model output can be better represented in the stimulation pattern in this way. As a side effect, CSR vary over time and might (using the default configuration) yield several thousand pulses per second. Refractoriness of auditory nerve fibers would not justify the use of such high channel stimulation rate (see e.g., pp. 212–216 in [24]). This has been shown several times in combination with common strategies, see e.g., [32]. The situation may, however, be different with algorithms making use of the temporal fine structure. We hypothesize that even if auditory nerve fibers near to the place of stimulation are in the refractory state, distant ANFs may still react to the stimuli due to the current spread in the cochlea. Since SAM provides a lot of phase information, these “distant stimulus pick-ups” may still be useful to contribute to speech perception and pitch discrimination. Future work should investigate this assumption.

Another interesting question is whether to aim at simultaneous stimulation. Today’s common CI systems do not allow all electrodes to be activated at once. In fact, most systems allow for only one active electrode at a time. This way, spectrally rich portions of the input sound can typically not be mapped to the stimulation pattern sufficiently well. The stochastic coder of SAM is designed to represent also less dominant parts of the spectrum. Nevertheless, the auditory model could only take full effect if a way to shape intra-cochlear electrical fields would be given, which requires simultaneous activation of electrodes. A possible concept based on phased array stimulation is given in [33] and will be evaluated for viable application in future versions of SAM.

### C. Psychoacoustic Features

With SAM, aspects of normal cochlear processing that are missing in common strategies can be replicated by the auditory model and preserved by the stimulus coder. To prove this, we conducted a computational evaluation elaborating on specific psychoacoustic properties. Our results demonstrated the existence of all surveyed psychoacoustic features in SAM stimulation patterns.

As already mentioned, there have been several attempts to extend existing strategies by including only one specific psychoacoustic property. Even those approaches could show small increases in speech recognition rate. Therefore, we expect SAM, including all the different features at once, to allow for even greater improvements in CI user performance.

### D. Further Evaluation

Beyond the evaluation work presented in Chapter III, additional aspects of SAM have already been studied and published. These include simulations to explore horizontal-plane sound source localization ability with SAM [35], and model-based estimation of speech reception threshold and that of the amount of pitch cues present in the stimulation pattern [36]. Furthermore, based on a proprietary but universal vocoder approach, SAM has been evaluated by means of hearing tests with normal hearing listeners [37].

All above studies have been carried out as comparative evaluation trials (SAM vs. ACE), and each has shown considerable advantages in using SAM.

### E. Evaluation With Cochlear Implant Users

We have also tested SAM with CI users in a pilot study. Five post-lingually deafened German-speaking adult CI users participated in the tests. All testees had had a Nucleus® Freedom™ Contour Advance™ implant together with a Freedom™ sound processor (at least on the side, where SAM-tests took place) programmed with the ACE strategy. Subjects have had at least 2 years of CI experience at the commencement of the study, and it could be assumed that they have reached a plateau in their learning curve with ACE.

A broad range of tests were performed within the available 5 sessions (2 hours each) per subject. Most of them were direct comparisons against ACE, designed as blind tests. They included: (1) recognition of monosyllables, numbers and consonants, (2) recognition of speech in clean, noisy and reverberative environments, (3) pitch-discrimination with pure tones and sung vowels, and (4) subjective quality rating.

An in-depth description of methods along with the detailed results will be published in a separate paper. A brief overview of the most important outcomes is presented below.

Even without much time to habituate to the new signal processing strategy, subjects had been able to hear and understand with SAM. Tests have shown that CI users having poor speech recognition performance in noisy environments with ACE can achieve considerable benefit (in average, 2.4 dB lower speech reception threshold in speech shaped noise) when switching to SAM. Discrimination of several consonants has also gotten better with SAM. Reverberated environments seem to remain challenging (no statistically relevant differences between SAM
and ACE). However, we assume that results would get better after a longer habituation period. Great benefits have been proven with SAM in pitch discrimination tests, which can be important for perceiving prosody, listening to music, and for understanding tonal languages. The just noticeable difference of 3.9 semitones (STs) dropped to 2.2 STs for pure tones and from 7.9 STs to 5.1 STs for sung vowels, on average, when switching from ACE to the SAM strategy. Tests have also shown that no subject was stressed or disturbed by SAM.

V. CONCLUSION

In this paper we have presented a new cochlear implant speech coding strategy, SAM, which is intended to make use of the spectro-temporal representation produced by its nonlinear auditory filter bank. We have shown that the coder preserves psychoacoustic properties, which are also likely to be useful in electric hearing. A pilot study has delivered promising results, and we are currently running a more comprehensive study with CI users to identify strengths and weaknesses of SAM. Future work will investigate the influence of each single psychoacoustic property and that of the main processing steps. Furthermore, the implementation needs to be optimized in terms of computational complexity, which is crucial in applicability in CI processors.

ACKNOWLEDGMENT

The authors would like to thank the staff of the Cochlear-Implant Rehabilitationszentrum Thüringen in Erfurt, Germany, for their faithful support.

REFERENCES


Tamas Harczos (S’05) was born in Hungary in 1980. He received the M.Sc. degree in computer science from the University of Veszprém, Veszprém, Hungary, in 2004. During work on his thesis, he developed a novel low bit-rate low-complexity speech compression algorithm and its real-time DSP implementation. After graduation, he worked as an Embedded Programmer in the automotive industry before joining the Pázmány Péter Catholic University (PPCU), Budapest, Hungary, for further studies. In the Multidisciplinary Doctoral School of the Faculty of Information Technology at PPCU he studied neurobiology, neurophysiology and cellular-neural-networks, and participated in teaching undergraduate students. In 2006, he moved to Ilmenau, Germany, and joined the Medical Acoustics Group (now Bio-inspired Computing) of the Fraunhofer Institute for Digital Media Technology (Fraunhofer IDMT). Since 2007, he has been working toward the Ph.D. degree at Ilmenau University of Technology, Ilmenau, Germany. His current research interests focus on the use of auditory models in cochlear implants (CI), and the algorithmic aspects of CI speech processors and that of sound source localization with bilateral CI systems.

Anja Chilian was born in Germany in 1987. She received the B.Sc. degree in biomedical engineering from the Ilmenau University of Technology, Ilmenau, Germany, in 2010. In her thesis, she developed tools for evaluating cochlear implant speech processing strategies. She is currently working toward the M.Sc. degree in biomedical engineering (specialization electromedical engineering) at the Ilmenau University of Technology. She is also a Young Scientist at the Fraunhofer Institute for Digital Media Technology (Fraunhofer IDMT). Her current research interests include signal processing for cochlear implants and algorithmic and statistical aspects of biomedical signal analysis.

Peter Husar (SM’98) was born in Zilina, Slovakia, in 1956. He received the Dipl.-Ing. degree, the Dr.-Ing. degree (Ph.D.) and the habilitation degree (postdoc thesis) from the Ilmenau University of Technology (TUI), Ilmenau, Germany, in 1980, 1986, and 1999, respectively. From 1986 to 1989, he was with the Research Institute for Computer Engineering, Zilina, Slovakia, and with Siemens, Erlangen, Bavaria, Germany. From 1992 to 2005, he was a Researcher and Assistant Professor in the field of biosignal processing, measurement technology, and sensor development at TUI. From 2005 to 2006, he was with the University Clinics of Saarland, Germany, as the Head of the Medical Technology Center. From 2006–2007, he was a Professor of biomedical engineering at the University of Applied Sciences of Anhalt, Germany. Since 2007, he has been Head of the Department of Biosignal Processing at TUI. Since 2010, he also has been the Head of the Bio-inspired Processing Group at the Fraunhofer Institute for Digital Media Technology (Fraunhofer IDMT), Ilmenau, Germany. He has authored numerous publications in his research fields as well as several patents. Dr. Husar is a member of several expert groups and a senior member of the IEEE Engineering in Medicine and Biology Society.