COCHLEAR IMPLANT ELECTRODE STIMULATION STRATEGY BASED ON A HUMAN AUDITORY MODEL

Auf einem menschlichen Gehörmodell basierende
Elektrodenstimulationsstrategie für Cochleaimplantate

Dissertation

Zur Erlangung des akademischen Grades

Doktor-Ingenieur (Dr.-Ing.)

th.

vorgelegt an der Fakultät für Elektrotechnik und Informationstechnik

der Technischen Universität Ilmenau

von

Dipl.-Ing. Tamás Harczos

geboren am 9. Mai 1980 in Baja (Ungarn)

Édesapám emlékére

–

To my Father, in loving memory

Legal notices

Companies and trademarks

For the sake of readability, the cochlear implant manufacturers (and their corresponding subsidiaries) Cochlear Ltd. (Australia), MED-EL Elektromedizinische Geräte GmbH (Austria), Advanced Bionics LLC (USA), and Neurelec (an MXM Company, France) are referred to as Cochlear, MED-EL, Advanced Bionics, and Neurelec, respectively, all through this document. Product names, company names and logos are trademarks or registered trademarks of their respective holders.

Hyperlinks

This document includes hyperlink references. This is solely for the convenience of the reader. The author is not responsible for any data obtained from the linked internet sites. Access to them is done entirely at the reader's own risk.

Independence of research

No cochlear implant company influenced the composition of this thesis. Neither had they any control over the research design, methodology, analysis, or findings of the underlying research.

Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich oder unentgeltlich geholfen:

———

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt.

Ich bin darauf hingewiesen worden, dass die Unrichtigkeit der vorstehenden Erklärung als Täuschungsversuch bewertet wird und gemäß §7 Abs. 10 der Promotionsordnung den Abbruch des Promotionsverfahrens zur Folge hat.

Ilmenau, 12. Januar 2015

Abstract

Modern cochlear implants (CIs) combined with professional rehabilitation have enabled several hundreds of thousands of hearing-impaired individuals to re-enter the world of verbal communication. Cochlear implants are inevitably the most successful neural prostheses to date, but, when looking at rehabilitation outcomes, current systems seem to have reached their peak potential. The fact that even after years of rehabilitation, most CI recipients claim not to enjoy listening to music and are not capable of carrying on a conversation in noisy or reverberative environments shows that there is still room for improvement. To overcome the above problems, there have been several attempts to enhance CI signal processing strategies to closer mimic the intact human cochlea, but most of these attempts achieved only limited success.

This dissertation presents a new cochlear implant signal processing strategy called *Stimulation based on Auditory Modeling* (SAM), which is completely based on a computational model of the human peripheral auditory system. SAM has been developed from scratch to avoid the limiting technological heritage of former strategies. 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.

SAM has been evaluated through simplified models of CI listeners, with five cochlear implant users, and with 27 normal-hearing subjects using an acoustic model of CI perception. Results have always been compared to those acquired using *Advanced Combination Encoder* (ACE), which is today's most prevalent CI strategy. First simulations showed that speech intelligibility of CI users fitted with SAM should be just as good as that of CI listeners fitted with ACE. Furthermore, it has been shown that SAM provides more accurate binaural cues, which can potentially enhance the sound source localization ability of bilaterally fitted implantees. Simulations have also revealed an increased amount of temporal pitch information provided by SAM. These findings have guided the choice of properties to be tested in the subsequent pilot study. For the pilot study with five CI users, a fast, safe, and real-time capable PC implementation of SAM and several graphical software tools supporting the experiments were developed in C/C++ and MATLAB. The pilot study, which ran smoothly, revealed several benefits of using SAM. First, there was a significant improvement in pitch discrimination of pure tones and sung vowels. Second, CI users fitted with a contralateral hearing aid reported a more natural sound percept and substantially better quality of both speech and music. Third, all subjects were accustomed to SAM in a very short period of time (in the order of 10 to 30 minutes), which is particularly important given that a successful CI strategy change typically takes weeks to months. An additional test with 27 normal-hearing listeners using an acoustic model of CI perception delivered further evidence for improved pitch discrimination ability with SAM as compared to ACE.

Although SAM is not yet a market-ready alternative, it strives to pave the way for future strategies based on human auditory models and it is a promising candidate for further research and investigation.

Zusammenfassung

Moderne Cochleaimplantate (CI), verbunden mit einer professionellen Rehabilitation, haben mehreren hunderttausenden Hörgeschädigten den Wiedereintritt in die Welt der verbalen Kommunikation ermöglicht. Cochleaimplantate sind heute unbestritten die erfolgreichsten neuronalen Prothesen. Betrachtet man jedoch die Rehabilitationserfolge, so haben sie inzwischen ihre Grenzen erreicht. Die Tatsache, dass die meisten CI-Träger sogar noch mehrere Jahre nach der Rehabilitation nicht in der Lage sind, Musik zu genießen oder einer Konversation in geräuschvoller oder verhallter Umgebung folgen zu können, zeigt, dass es noch Raum für Verbesserungen gibt. Es gab bereits einige Versuche, CI-Signalverarbeitungsstrategien durch bessere Nachahmung der intakten menschlichen Cochlea zu verbessern, um so die genannten Probleme zu überwinden. Die meisten dieser Versuche waren jedoch nur begrenzt erfolgreich.

Diese Dissertation stellt die neue CI-Signalverarbeitungsstrategie *Stimulation based on Auditory Modeling* (SAM) vor, die vollständig auf einem Computermodell des menschlichen peripheren Hörsystems beruht. SAM wurde von Grund auf neu entwickelt, um technologische Begrenzungen früherer Strategien zu vermeiden. Die Strategie berücksichtigt die aktive cochleäre Filterung (Basilarmembran und äußere Haarzellen) sowie die mechano-elektrische Übertragung der inneren Haarzellen, sodass das System verschiedene psychoakustische Eigenschaften direkt beinhaltet.

Im Rahmen der vorliegenden Arbeit wurde die SAM Strategie dreifach evaluiert. Sie wurde einerseits mit vereinfachten Wahrnehmungsmodellen von CI-Nutzern geprüft. Andererseits wurden auch Tests mit fünf CI-Nutzern durchgeführt. Außerdem fand eine Evaluierung mit 27 Normalhörenden mittels eines akustischen Modells der CI-Wahrnehmung statt. Die Evaluationsergebnisse wurden stets mit Ergebnissen, die durch die Verwendung der *Advanced Combination Encoder* (ACE) Strategie ermittelt wurden, verglichen. Die ACE Strategie stellt die zurzeit verbreitetste Strategie dar. Erste Simulationen zeigten, dass die Sprachverständlichkeit der CI-Hörer, welche mit SAM versorgt wurden, genauso gut war, wie die der ACE-Versorgten. Weiterhin konnte gezeigt werden, dass die SAM Stimulationsmuster genauere binaurale Merkmale enthalten, was potentiell zu einer Verbesserung der Schallquellenlokalisierungfähigkeit bei bilateral versorgten CI-Trägern führen kann. Die Simulationen zeigten ebenfalls einen erhöhten Anteil an zeitlichen Pitchinformationen, welche von SAM bereitgestellt wurden. Diese Erkenntnisse haben dazu beigetragen, die in der nachfolgenden Pilotstudie zu testende Eigenschaften auszuwählen. Für die Pilotstudie mit fünf CI-Nutzern wurden eine schnelle, sichere und echtzeit-fähige PC-Implementierung von SAM und mehrere grafische Softwaretools zur Unterstützung der Experimente in C/C++ und MATLAB entwickelt. Die Ergebnisse der Pilotstudie zeigten mehrere Vorteile von SAM auf. Erstens war eine signifikante Verbesserung der Tonhöhenunterscheidung bei Sinustönen und gesungenen Vokalen zu erkennen. Zweitens bestätigten CI-Nutzer, die kontralateral mit einem Hörgerät versorgt waren, eine natürlichere Klangwahrnehmung und eine wesentlich bessere Qualität von Sprache und Musik. Als ein sehr bedeutender Vorteil stellte sich drittens heraus, dass sich alle Testpersonen in sehr kurzer Zeit (im Bereich von 10 bis 30 Minuten) an SAM gewöhnen konnten. Dies ist besonders wichtig, da ein erfolgreicher CI-Strategiewechsel typischerweise Wochen oder sogar Monate dauert. Ein

zusätzlicher Test mit 27 Normalhörenden, die mittels eines akustischen Modells der CI-Wahrnehmung stimuliert wurden, lieferten weitere Nachweise für die verbesserte Tonhöhenunterscheidung von SAM im Vergleich zu ACE.

Obwohl SAM noch keine marktreife Alternative ist, versucht sie den Weg für zukünftige Strategien, die auf menschlichen Gehörmodellen beruhen, zu ebnen und ist somit ein erfolgversprechender Kandidat für weitere Untersuchungen und Forschungsarbeiten.

Acknowledgments

This PhD thesis is supposed to be solely my own work, which –in a practical sense– it is. However, when thinking about how my naive wish of studying cochlear implants evolved into a firm definition of tasks and eventually led to the development of this thesis, I was surprised how many people helped and also influenced my work, in one way or another.

I would like to thank:

- *my Parents* for all their love, and for providing me financial support to make my studies available, paving the way for the knowledge and ground work for this thesis,

- *Hans* and *Gudrun Sachs* for teaching me that life is exactly what I make of it and for having encouraged me whenever needed,

- *András Rohonyi, Márton Péri, Jeff Preston* and all my pre-university language teachers for all the English and German knowledge I gathered from their excellent lessons;

- *Tamás Kiezer, Norbert Sárközi, Zoltán Szlávik, Balázs Gaál, Gyula Gottlieb* and *András Kutrovics* for turning learning at the university dormitory into a fun activity and also for all the inspiration I got from them;

- *Gyula Kovács* for supporting my wish to commence PhD studies while warning me of possible pitfalls;

- *Prof. György Karmos* for drawing my attention to cochlear implants, for his useful classes on electrophysiology, and for reviewing parts of this thesis;

- *Prof. Tamás Roska* for supporting my enthusiasm and for his help in making first contact with Fraunhofer IDMT;

- *Frank Klefenz* for making my internship placement at Fraunhofer IDMT possible;

- *Prof. Karlheinz Brandenburg* and *Prof. Peter Husar* for facilitating my stay and supervising my research at Fraunhofer IDMT;

- *Prof. Birger Kollmeier* for his mentorship and for answering my questions always in a very timely manner;

- *András Kátai* for teaching me all the basics needed for my work at Fraunhofer IDMT, for his help with all the cochlear implant related projects, and for being a friend and supporter of mine ever since my very first day in Ilmenau;

- *Anja Chilian* for her help with the development and evaluation of SAM, as well as for being a model student (and later a model colleague) of mine, always open to discussion and always ready to help;

- *Gero Szepannek* for all the discussions about the design and evaluation of experiments and for his patience in explaining statistics to me;

- *Ute Feuer* and *Peter Voigt* and all the colleagues at the Cochlear-Implant Rehabilitationszentrum Thüringen for their faithful help through all the years;

- *Thomas Vandahl* for his voluntary help with the implementation of the OLSA Tester;

- *Jan Rennies* and *Francois Xavier Nsabimana* from the Project Group Hearing, Speech and Audio Technology (HSA) of Fraunhofer IDMT, for providing me a tool for calculating speech transmission indices;

- *Nina Wardenga* and *Ralf Meyer* for the MATLAB implementation of the text reception threshold test, and to *Stefan Fredelake* for establishing the connection with them;

- *Prof. Waldo Nogueira* for all the fruitful discussions and joint research;

- the volunteers for the participation in the experiments related to SAM;

- all the former students of our project group for their help in the development and testing of SAM;

- *Horst Hessel, Thomas Hocke, Peter Schleich, Dirk Meister,* and *Dan Gnansia* for their contribution to clarifying features and parameters of commercial signal processing strategies for cochlear implants; and *Horst* once again for making this thesis more valuable through his comments and constructive criticism;

- *Richard Francis "Dick" Lyon* and *Roy Patterson* for the correspondence about the history of gammatone and gammachirp filter banks, and *Dick* once more for proofreading section 2.3;

- *Erzsike Dezsőné Boda, Julcsi Lászlóné Wohlráb, Erzsébet Keresztes, Helga Blúmné Erdős, Erika Baloghné Horváth, Ida Szabóné Kószó,* and *Mayor Zoltán Alszegi* for providing me free access to a quiet office in Vaskút to work on my thesis while I was in Hungary;

- all the coauthors and contributors of joint papers that have arisen during the work;

- the members of the AUDITORY mailing list helping me out with papers I could not find elsewhere.

Special thanks go to Grace Lim, who did a marvelous job in proofreading and commenting on the document in a very timely manner!

Finally, sincere thanks to my loving wife, Yvonne, for her patience and support throughout the writing of this thesis, and to our wonderful three-year-old son, Tom, for all the love and joy he brings us every day, and for allowing me to write whenever I explained that I needed to.

Contents

Glossary

Acronyms and abbreviations

ABI	Auditory Brainstem Implant
ACE	Advanced Combination Encoder (strategy)
AGC	Automatic Gain Control
ADRO	Adaptive Dynamic Range Optimization
AI	Auditory Image (representation)
AM	Auditory Model
AMI	Auditory Midbrain Implant
AN	Auditory Nerve
ANF	Auditory Nerve Fiber
ANOVA	Analysis of Variance
ANP	Auditory Nerve Population
AP	Action Potential
APGF	All-Pole Gammatone Filter
ART	Auditory Nerve Response Telemetry
ASIC	Application Specific Integrated Circuit
ASR	Automatic Speech Recognition (or Recognizer)
BAHA	Bone-Anchored Hearing Aid
BM	Basilar Membrane
BPF	Band-Pass Filter
BPNL	Band Pass Nonlinear (model)
BTE	Behind the Ear (processor)
BW LPF	Butterworth Low Pass Filter
CAP	Central Auditory Processing
CC	(Neurotransmitter) Concentration in the (synaptic) Cleft
CCITT	International Telegraph and Telephone Consultative Committee (French: Comité Consultatif International Téléphonique et Télégraphique)
CCns	Non-stochastically calculated CC
CF	Center (or Characteristic) Frequency
CGCF	Compressive Gammachirp Filter
CI	Cochlear Implant
CIS	Continuous Interleaved Sampling (strategy)

CIX   Cochlear Implant eXperiment(al software)

CNN   Cellular Neural Network

CNN-UM   CNN Universal Machine

CPW   Cochlear Pressure Wave

CSR   Channel Stimulation Rate

CSSS   Channel Specific Sampling Sequence

CUDA   Compute Unified Device Architecture

DCGCFB   Dynamic Compressive Gammachirp Filter Bank

DCT   Discrete Cosine Transform

DFT   Discrete Fourier Transform

DLL   Dynamic-Link Library

DRNL   Dual Resonance Nonlinear (filter bank)

DSP   Digital Signal Processor

DT   Delay Trajectory

EABR   (Electrically) Evoked Auditory Brainstem Response

EAS   Electric-Acoustic Stimulation

ECAP   (Electrically) Evoked Compound Action Potential

ESPC   Electrode Selection Probability Control (function)

EZ   Extended Zwicker (model)

EZB   Extended Zwicker-Baumgarte (model)

FDA   (U.S.) Food and Drug Administration

FFT   Fast Fourier Transform

FPGA   Field-Programmable Gate Array

FSP   Fine Structure Processing (strategy)

GCC   Generalized Cross-Correlation

GPU   Graphics Processing Unit

GT   Gammatone (filter bank)

GUI   Graphical User Interface

HDCIS   High Definition CIS (strategy)

HiRes   High Resolution (strategy)

HL   Hearing Loss

HMM   Hidden Markov Model

HPF   High Pass Filter

HSR   High Spontaneous Rate (nerve fiber)

HTK   Hidden Markov Model Toolkit

IDMT   (Fraunhofer) Institute for Digital Media Technology

IDR   Input Dynamic Range

IHC	Inner Hair Cell
INR	Internal Representation
IPA	International Phonetic Alphabet
IR	Impulse Response
ITD	Interaural Time Difference
LGF	Loudness Growth Function
LPC	Linear Predictive Coding
LPF	Low Pass Filter
LSR	Low Spontaneous Rate (nerve fiber)
MAP	Model of Auditory Processing
MBPNL	Multiple Band Pass Nonlinear (model)
MCL	Most Comfortable Level
MEI	Middle Ear Implant
MEM	Multi-channel Envelope Modulation (strategy)
MFCCs	Mel-Frequency Cepstral Coefficients
MPIS	Main Peak Interleaved Sampling (strategy)
MPS	Multiple Pulsatile Sampler (strategy)
MSR	Medium Spontaneous Rate (nerve fiber)
NIC	Nucleus Implant Communicator
NIH	National Institutes of Health
NLM	Nonlinear Mechanical (filter)
NRT	Neural Response Telemetry
NT	Neurotransmitter
OAE	Otoacoustic Emission
OHC	Outer Hair Cell
OLSA	Oldenburg sentence test (German: Oldenburger Satztest)
OME	Outer and Middle Ear (filter)
OZGF	One-Zero Gammatone Filter
PACE	Psychoacoustic ACE (strategy)
PAP	Peripheral Auditory Processing
RF	Radio Frequency
RP	(Vesicle) Release Probability
SAM	Stimulation based on Auditory Modeling (strategy)
SAS	Second Amplification Stage
SGC	Spiral Ganglion Cell
SI	Synchronization Index
SNR	Signal to Noise Ratio

SPEAK   Spectral Peak (strategy)

SpecRes   Spectral Resolution (strategy)

SRT   Speech Reception Threshold

ST   Semitone

STI   Speech Transmission Index

THL   Threshold Level

TRNL   Triple-Path Nonlinear (filter)

TPR   Total Pulse Rate

TRT   Text Reception Threshold

VCIS   Virtual Channel Interleaved Sampling (strategy)

VCV   Vowel-Consonant-Vowel

VLSI   Very-Large-Scale Integration

WDF   Wave Digital Filter

Mathematical notation, symbols and units

A_m	Result of the frequency mapping of $\hat{C}_n$
$\ddot{A}_m$	Result of the dynamic range mapping of A_m
$\hat{A}_m$	Result of applying repetition penalty to $\ddot{A}_m$
c	Neurotransmitter substance concentration near a simulated inner hair cell
$[Ca^{2+}]$	Calcium concentration inside a simulated inner hair cell
$\tilde{C}_n$	Neurotransmitter substance concentration near the nth simulated inner hair cell
$\hat{C}_n$	Result of the maximum-preserving decimation of $\tilde{C}_n$
dB FS	Decibel (relative to) full scale
dB HL	Decibel (relative to normal) hearing level
E_i	Electrode i
F_s	Sampling rate (see also sps)
G	Apical conductance of a simulated inner hair cell
I	Electrical current of a stimulus to be applied to a cochlear implant electrode
I_{Ca}	Inward calcium current of a simulated inner hair cell
i_{OE}	Sound pressure at the outer ear expressed as electrical current
J_n	Current density in the proximity of stimulating electrode n
k	Neurotransmitter release rate inside a simulated inner hair cell
λ_C	Variable of the spatial current function, i.e., range of current spread
L_m	Result of the amplitude mapping of $\hat{A}_m$
N_C	Number of simulated auditory nerve cells
P_S	Sound pressure level
$\underline{P}$	Result vector of the stimulus renderer
pps	Pulses per second
q	Amount of neurotransmitter substance in the immediate store of a simulated inner hair cell
σ_{int}	Multiplicative internal noise
sps	Samples per second (the SI equivalent is Hz, see also F_s)
t_{gap}	Duration of the phase gap in a biphasic stimulus
$\mathcal{U}(a,b)$	Uniform distribution with minimum and maximum values of a and b, respectively

u	Displacement of a stereocilia of a given inner hair cell
$u_{BM,n}$	Velocity of the nth basilar membrane section expressed as electrical voltage
u_{OW}	Velocity of the oval window expressed as electrical voltage
V	Receptor potential of a given inner hair cell
w	Amount of neurotransmitter substance in the reprocessing store of a simulated inner hair cell
$\overline{x}$	Complex conjugate of x
$\lfloor x \rfloor$	Largest integer less than or equal to x (floor function)
Y	Current level of a stimulus to be applied to a cochlear implant electrode

See also Table A.1 and Table A.2 starting on page 131 for further symbols and parameters used in this thesis.

Chapter 1

Introduction

1.1 Motivation and objective

Cochlear implants (CIs) are undisputedly the most successful neural prostheses to date. CIs can provide a way for individuals with severe or profound hearing loss to perceive sounds and speech. Users of current CI systems, however, tend to encounter difficulties with understanding speech in noisy or reverberated environments; they have trouble discriminating pitch and perceiving prosody (which is crucial in tonal languages), and the joy of music also remains beyond their reach. Furthermore, there is also room for improvement regarding the duration of the necessary rehabilitation after implantation.

This thesis presents a novel CI signal processing algorithm (also called strategy), that may inherently by its design, overcome these known (and possibly also further) issues. The new strategy is called SAM[1] and is based on detailed models of the human auditory system. It is meant to reproduce psychoacoustic properties that are likely to be useful also in electric hearing. SAM has been validated via computational evaluation tasks and its usefulness has been proven in a pilot study with cochlear implant users. This study has also demonstrated encouraging outcomes regarding the time needed to get accustomed to the new strategy.

As the full form of the abbreviation SAM already suggests, the main concept is to use auditory models to calculate CI stimulation patterns. I started developing the idea of combining models of the basilar membrane, the inner hair cells, and the auditory nerves for use in CIs in 2005. The work was triggered by a very powerful talk by Professor György Karmos, who gave a presentation about CIs at the Péter Pázmány Catholic University in Budapest, Hungary. He explained the inner workings of the ear and then that of the common speech processing strategies of (at that time) state of the art CIs. The gap between those two was startling.

A search on the internet showed that there were two approaches to fill this gap. In the first, existing strategies were augmented by algorithms that helped mimic specific properties of the ear, normally not included in the given strategy. In the second, large parts of existing strategies were exchanged with auditory models. Some literature research revealed that good computa-

---

[1] Stimulation based on Auditory Modeling

tional models of the human ear were already published that time. Furthermore, the CI technology (especially the electrode driving circuitry) seemed to be mature enough to transmit significantly more information, so that the idea of switching from DFT[2] to a more complex auditory filter bank was promising. The latter approach also suited my personal preference and it became the base of the design considerations for a novel speech processing strategy.

1.2   Outline of the thesis

The remainder of this thesis is structured as follows:

Chapter 2 provides the necessary background information and places the problem to be solved into the context of modern hearing research. The chapter first provides an overview of the physiology of normal hearing followed by a summary of technologies that can be useful for the Hearing Impaired. Since cochlear implants are the focus of the presented research, the state of the art regarding these neural prostheses is represented in detail. Next, computational models of the human ear (with emphasis on cochlea models) are reviewed and categorized.

Chapter 3 summarizes published work of other laboratories on similar concepts. The few existing CI strategies based on auditory models are compared and the state of the art is discussed.

Chapter 4 proposes SAM. First, the design considerations and requirements on a new CI speech processing strategy are discussed, which is followed by the detailed description of the algorithm. Every processing stage, starting from the auditory model based filter bank, through the stimulus coder, all the way to the electrode discharges, along with system parameters and their origins are described. Next, the chapter explains the software framework into which SAM has been embedded and presents some applications implemented for evaluating the strategy with CI recipients. The chapter is rounded off by showing a suggested way of fitting[3] the strategy.

In Chapter 5, validation and evaluation work on SAM is reviewed. First, the chapter elaborates on the requirements imposed in the previous chapter and validates SAM against them. One section is dedicated to the effects of changing coder parameters to see how those affect the fulfillment of the requirements. Parallel (and partly prior) to the pilot study with implantees, SAM was extensively tested and evaluated in computational studies. The outcomes of these latter are discussed in detail. Next, the chapter elaborates on the setup and results of the evaluation studies involving CI users and normal-hearing listeners, and draws conclusions.

---

[2] Discrete Fourier Transform (DFT) is used in the speech processing strategies of Cochlear. By contrast, MED-EL uses a combination of bandpass filters and Hilbert-transform in their filter banks.

[3] Customizing strategy parameters to fit the needs and preferences of a cochlear implant user.

Remaining open questions and limitations regarding both the development and the evaluation of SAM are discussed in Chapter 6.

Chapter 7 concludes this thesis by summarizing the main results and findings. It also points out directions for future research and brings up two possibly useful concepts being yet quite unexplored in the field of computational auditory modeling.

Chapter 2

Hearing, aided hearing, and models of hearing

2.1 Hearing

In case of a hypothetical decision on losing a sense, people generally prefer to keep their vision as opposed to their hearing. Though the amount of information provided by the visual and somatosensory systems can be much larger, hearing is essential to taking part in cultural and social life. Today, hearing might not be crucial for survival, as it was when nature started developing this highly efficient system, but it has become the key component of human communication, and it is beyond debate that it makes life easier and more enjoyable, too. The importance of the ears and the auditory system (as well as that of all other sensory organs) gets apparent, once they fail.

The following section gives an overview of the functional anatomy and physiology of the intact human auditory system. Subsequently, the most common causes and symptoms of hearing loss are reviewed, along with the technologies that can help in these cases.

2.1.1 Normal hearing

The audio signal processing of the ear consists of a chain of consecutive steps. The outer ear (auricle and ear canal) and the middle ear (eardrum and auditory ossicles) work like amplifying transducers towards the inner ear (cochlea). Their combined transfer function yields peak sensitivity in the range of about 1 to 3 kHz. The middle ear, however, can also enhance speech intelligibility at high loudness levels [BZ73] and protect the inner: the movement of the ossicles can be stiffened by two muscles, which results in reduced transmission gain. An overview of the human auditory periphery is shown in Figure 2.1. [Møl06a: 26-34, 187-190]

Incoming sound waves are transmitted via the hammer, anvil and stirrup, to the cochlea through the oval window. The cochlea contains three fluid-filled canals: the scala tympani, the scala media, and the scala vestibuli. (For a cross section of the cochlea, see Figure 2.3.) The scala media and the scala tympani are separated by the basilar membrane (BM). The oscillations of the oval window are transformed into propagating pressure waves within the scalae, which, in turn, cause a displacement of the basilar membrane. The BM is stiff and narrow near the base of the cochlea, and flexible and wide near the apex. Due to these properties and due to the

gradiently changing wavelength to canal depth ratio, the vibration of the BM at any position is dominated by only a small range of frequencies. This phenomenon is called tonotopy, and through this feature, the BM is able to spatially separate sounds based on their frequency components.
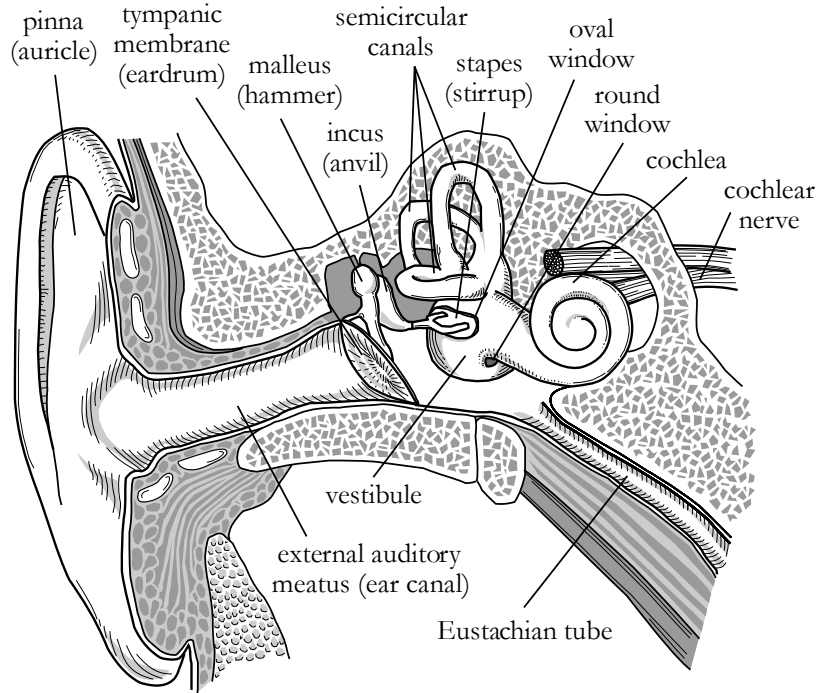
Figure 2.1: The human auditory periphery. Adapted from [Wat93].

The length of the human BM typically ranges from 31 to 35 mm [Gre90, Møl06a: 10]; an arbitrary position on the BM can be mapped to its approximated characteristic frequency by e.g. the Greenwood equation [Gre90] shown below.

$$f_c(p) = 165.4 \cdot \left(10^{0.06 \cdot p} - 1\right). \tag{2.1}$$

The function f_c approximates the characteristic frequency (in Hz) at the position p measured (in mm) from the apical end of the unrolled cochlea (see Figure 2.2).

The organ of Corti is located along the length of the BM, containing three to five (in humans typically three) rows of outer hair cells (OHCs) and one row of inner hair cells (IHCs), see Figure 2.3 and Figure 2.4. The OHCs are assumed to enhance perception by acting as tiny amplifiers, shaping the traveling wave by adding energy to the BM movement. OHCs receive a number of efferent nerve fibers from the central nervous system, which means that their operation is controlled by higher-level entities of the brain.
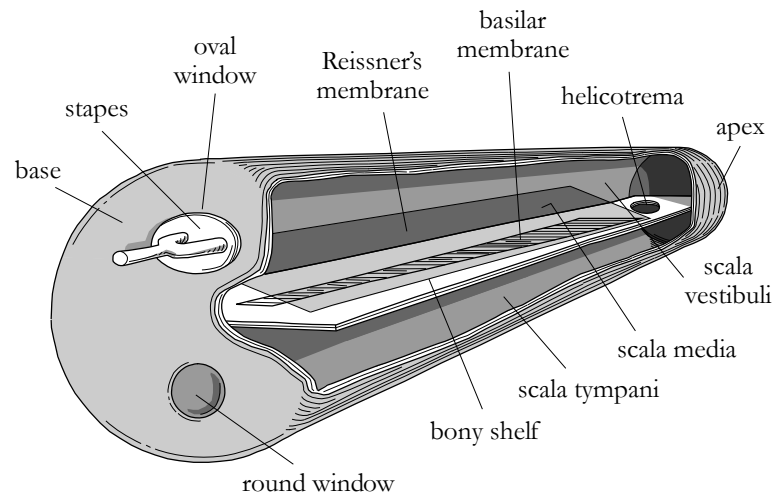
Figure 2.2: The unrolled cochlea, highly simplified. Adapted from [Wat93].

Inner hair cells, by contrast, seem not to have this strongly marked control, and they act as sensory transducers of the mechanical waves into bio-electrical potentials. In a human cochlea, a total of about 12000 OHCs and 3500 IHCs are located [Møl06a: 11].
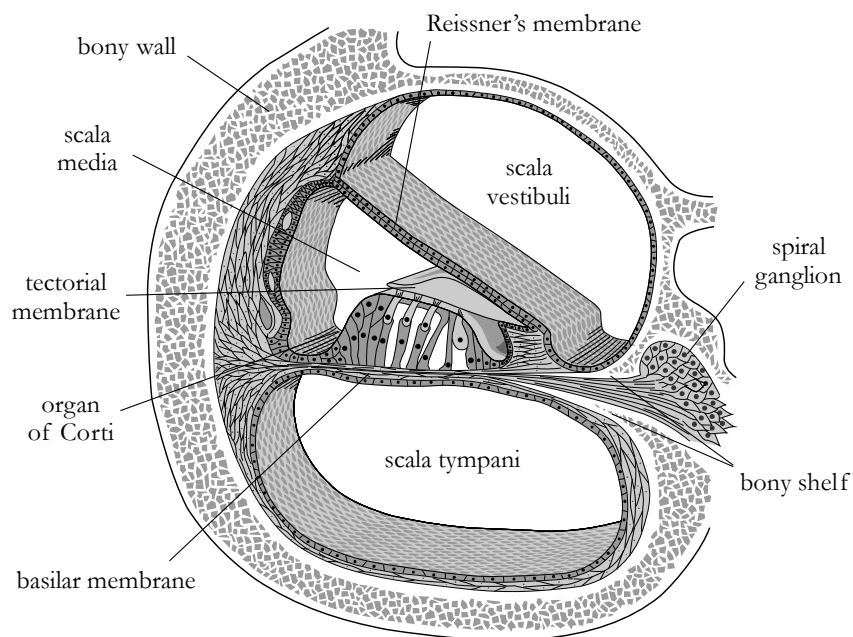
Figure 2.3: Cross section of the cochlea. Adapted from [Wat93].

The signal-transmission inside the IHCs is well described in both [Møl06a: 13-15, 48-52] and [SLOM02] and can be summarized as follows. On top of the IHCs, three rows of stereocilia follow the movement of the BM. Stereocilia deflection results in opening and closing of potassium-ion-channels. Influx of K^+ leads to depolarization of the IHC, resulting in half-wave rectification of the band-pass filtered sound wave. As a function of the IHC-membrane potential, Ca^{++}-ions enter the cell and evoke the release of neurotransmitter substance at the presynaptic end of the IHC.

The diffusion of the neurotransmitter across the synaptic cleft causes postsynaptic depolarization of the auditory nerve fibers (ANF). If the postsynaptic potential reaches a specific threshold, an action potential is generated. After that, the ANF underlies some refractory period, during which the threshold for firing is largely increased and thus firing is less probable.
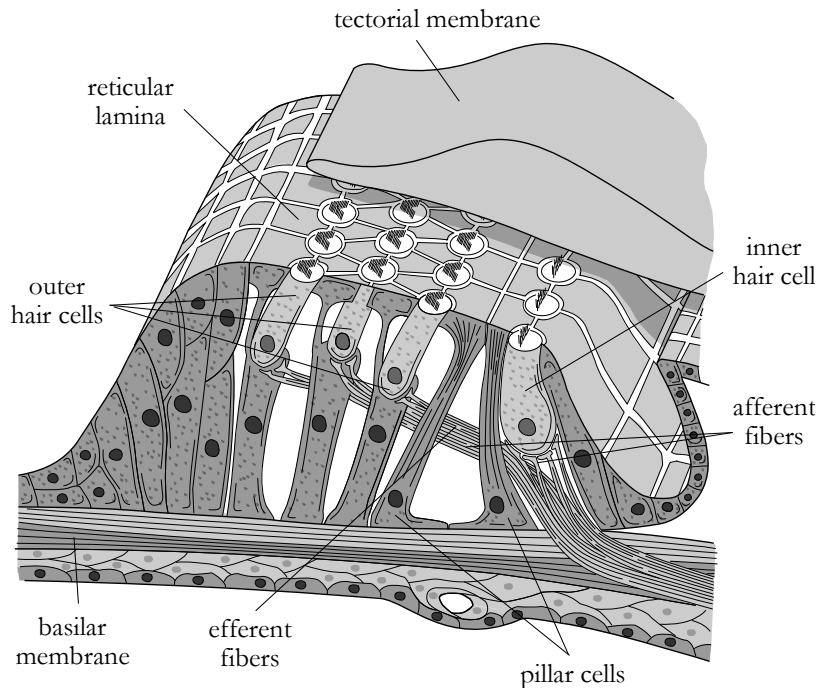
Figure 2.4: The organ of Corti. Adapted from [Wat93].

The pathway from the hair cells towards the brain is realized by the spiral ganglion cells (SGCs). One IHC has about ten such afferent connections. The human auditory nerve consists of about thirty thousand afferent nerve fibers in total [Møl06a: 13]. The information carried by the SGCs will start to be integrated at the next processing stage: in the cochlear nuclei of the brainstem. The subsequent stages of processing are the midbrain, the thalamus, and finally, the place of auditory perception: the auditory cortex. (It is to be noted, though, that speech understanding is achieved through the interaction of various regions of the brain.)

2.1.2 Impaired hearing

Disorders of the auditory system can be divided into two classes: decreased function and abnormal function. Decreased function is generally known as hearing impairment, which includes elevated hearing threshold and decreased speech discrimination. Examples of abnormal function are tinnitus and hyperacusis (over-sensitivity to sound). The following paragraphs will only address decreased auditory function, which is typically subdivided into conductive, sensorineural and neural hearing loss (HL). [Møl06a: 201-251, Coc10a]

2.1.2.1   Types and levels of hearing loss

In the case of *conductive HL*, parts of the outer and middle ear complex do not work properly. It can be congenital, but it may also emerge later in life due to head injury, middle ear infections, or other diseases. Pure tone audiometry, tympanometry and recordings of the acoustic middle-ear reflex response are typically used to diagnose conductive hearing loss.

Sensorineural HL indicates a problem within the inner ear. The most common problem is damage to the hair cells: missing amplification of the OHCs and/or impaired transduction function of the IHCs. This kind of hearing loss is also termed "sensory" or "cochlear" hearing loss and its most common causes are acoustic trauma, genetics, head injury and certain medications. However, a certain degradation of the hair cell functions is also part of the normal aging process (presbycusis). (Sometimes a conductive HL occurs in combination with a sensorineural HL, which is referred to as *mixed HL*.)

Neural HL refers to a problem with the auditory nervous system. Diagnostically, it is hard to differentiate from sensorineural HL, but the possible causes are different. Evoked auditory brainstem response audiometry (middle and late potentials) are often used to diagnose neural HL. Some forms of sensorineural and neural HL may be distinguished by psychoacoustic tests. For example, injuries to OHCs cause the cochlear filter to become broader, which may increase masking and impair temporal coding of broadband sounds such as vowels.

Hearing loss levels are generally categorized as mild (20-40 dB HL), moderate (40-60 dB HL), severe (60-90 dB HL), and profound (>90 dB HL) [MS07: 749]. The unit dB HL (or dB hearing level) is "the sound level in dB above the average threshold of young individuals without disorders of the ear" [Møl06a: 167].

2.1.2.2   Hearing technologies

There are various technologies that restore hearing or at least allow for a better auditory sensation. The right technology must be chosen for each individual, since it depends on both the location and the severity of the hearing impairment, see Table 2.1.

The most established method to restore hearing is the use of a *hearing aid* (HA). The main function of a HA is to amplify sound. The gain can be set for individual frequency bands, so that the hearing aid can be fitted to the specific hearing loss of the wearer of the HA. Hearing aids are typically worn behind the ear (BTE) or in the ear, but there are also body-worn HAs. A hearing aid can compensate for decreased function of the outer, middle and inner ear, as long as the impairment does not exceed a certain limit.

Should the degree of conductive hearing loss exceed about 50 dB HL or should the subject be unable to wear the earmold of a normal BTE hearing aid, a *bone conduction implant* (also cal-

led bone anchored hearing aid or BAHA) may be the right choice. However, the cochlea and auditory nerves are required to be intact. This technology completely bypasses the outer and middle ear and transmits sound to the inner ear through the skull bone. The result is a near-normal hearing experience. [DFW+11]

An alternative to BAHA is the middle ear implant (MEI). A MEI is implanted in the middle ear cavity and (for some manufacturers) partly in the mastoid area. The implant works by directly moving the auditory ossicles, or by vibrating the membrane window of the cochlea. MEIs are most often used for those with sensorineural HL. [Cha02]

Table 2.1: Hearing technologies as a function of location and severity of hearing loss.

Conductive hearing loss (Outer and middle ear)	Sensorineural hearing loss (Inner ear)	Neural hearing loss (Auditory nerve)	Technology that may help most
mild to moderate	mild to moderate	(intact)	hearing aid
moderate to single-sided deafness	(intact)	(intact)	bone conduction implant
mild to moderate	mild to severe	(intact)	middle ear implant
(n/a)	moderate to profound	(intact)	cochlear implant
(n/a)	moderate to profound in high-frequency range	(intact)	electric-acoustic hearing implant
(n/a)	(n/a)	severe	auditory brainstem implant

Hearing technologies listed in Table 2.1 support or substitute different parts of the human auditory system. A short explanation of each technology is given below.

If the damage is located in the cochlea, a *cochlear implant* can help most. A CI works by electrically stimulating the auditory nerve fibers inside the inner ear. The electrodes are typically placed in the scala tympani and their arrangement follows the tonotopic principle. Depending on the signal processing algorithm of the CI, the auditory sensation may be more or less similar to natural hearing. However, today's CIs definitely do not reproduce the richness of normal hearing. Cochlear implants are going to be discussed in more detail in section 2.2. [Møl06a: 268-277, Cla03]

An *electric-acoustic hearing implant* is a combination of a hearing aid and a cochlear implant in the same ear. Electric-acoustic stimulation (EAS) is the best choice if cochlear hearing loss is limited to the mid to high frequency range. A special surgical technique attempts to preserve re-

sidual hearing of the subject to the greatest possible extent, so that low frequency sound components can be acoustically amplified by the HA. High frequency parts, on the other hand, are transmitted by the CI. Fair understanding of speech in noise and good pitch discrimination abilities are evidence that the auditory system successfully merges these two kinds of stimuli. Most EAS subjects (unlike most CI users) do also appreciate music. [JFD+06]

Auditory brainstem implants (ABI) and *auditory midbrain implants* (AMI) are currently the last and riskiest choice to provide a sense of sound to a profoundly deaf person. They bypass the outer, middle and inner ear, plus the auditory nerves. The stimulating technology is quite similar to that of CIs, but instead of stimulating the cochlea, the stimulation is applied to the brainstem (more precisely to the cochlear nucleus of it) or to the midbrain to produce auditory sensation. Due to the nature of the implantation (i. e., brain surgery) the total number of ABI- and AMI-recipients worldwide is not more than several thousand. [Møl06a: 277-279]

2.2   Cochlear implants

A cochlear implant is the first choice in the case of moderate to profound sensorineural hearing loss. As stated above, a CI electrically stimulates the cochlea, which results in sound perception. This section provides a short historical overview of the evolution of cochlear implants, followed by the indications for and a functional description of current CIs. At the end, a short paragraph addresses the importance of the postoperative rehabilitation.

Table 2.2: Timeline of the cochlear implant research. Based on [Cla03: 6-57] and [ZRH+08].

1790s	A. Volta discovers that electrical stimulation of the ears creates auditory sensation.
1868	R. Brenner uses AC stimulation with varied rate, polarity and intensity.
1950	N. Lundberg directly stimulates the auditory nerve with electrical current.
1957	A. Djourno and C. Eyries perform a more detailed study on stimulating auditory nerves with inductive coils. They also do postoperative rehabilitation to assess speech development in a previously deaf person.
1961	W. House, John Doyle and James Doyle insert single-electrode implants into the scala tympani of three profoundly deaf patients for a test period of about three weeks.
1966	F. Simmons implants a bundle of six electrodes into the modiolus (central axis) of the cochlea and makes the first attempts to separate spectral components into frequency bands.
1972	W. House builds the first wearable signal processor for single-electrode implants.

1978	G. Clark and B. Pyman perform the first multi-electrode cochlear implant operation.
1984	FDA[4] approves first CI (from House/3M/Cochlear) for implantation in adults.
1985	FDA approves the Nucleus multi-channel implant from Cochlear to be used in adults.
1990	FDA approves the Nucleus implant to be used from the age of 2 years.
1996	FDA approves the Clarion multi-channel implant from Advanced Bionics.
2001	FDA approves the Combi40+ multi-channel implant from MED-EL.
2004	First successful combination of a hearing aid and a cochlear implant in the same ear.

2.2.1 Retrospection

In the late 1790s, the Italian physicist Alessandro Volta discovered that electrical stimulation of the ears created auditory sensation. Due to the unpleasantness of his self-experiments he did not follow up with his studies [Cla03: 1-3]. Nevertheless, his research was the start of the long evolution of CIs. The most important stages of CI research are summarized in Table 2.2. The presented timeline is only an excerpt and is not complete. The dashed line after the year 1978 indicates clinical maturity of cochlear implants. In the 1990s a new phase of CI evolution started [Cla03: 46], which had been dominated by patents (not listed here).

2.2.2 Indications

Indications for a cochlear implant have been dramatically expanded due to increasingly sophisticated design and surgical techniques. In the 1990s and early 2000s, the requirement for cochlear implantation was bilateral profound hearing loss. Today, a significant drop in speech perception may be sufficient, if it cannot be handled by hearing aids [DGG+07, SAO11]. Bilateral profound hearing loss (with severe auditory nerve damage) is currently a requirement for brainstem implantation.

Due to the expanded indications, the number of cochlear implantees is increasing rapidly. Today, the total number of CI users worldwide is estimated to be about two hundred thousand (USA: seventy thousand [NIH10], Germany: thirty thousand [Jut11], UK: twelve thousand [TEF10]).

The costs of an implantation (including pre-operational evaluation, hardware, surgery and rehabilitation) are in the range of 40000-120000 USD. The cost-benefit analysis of insurance companies shows, however, that cochlear implantation is a cost-saving alternative. This outcome results from the fact that adult implantees have higher productivity than non-implanted

---

[4] U. S. Food and Drug Administration. For a full listing of cochlear implant related approvals see [FDA12].

deaf adults, and that implanted children typically do not need to attend special kindergartens and schools. Huge benefits of binaural electrical hearing have also been shown, so that bilateral (CI in both ears) and bimodal (CI in one ear and hearing aid in the other, or hybrid implants, see EAS in section 2.1.2.2) systems are now strongly encouraged [Cla03: 438-442]. Since the cost-effectiveness of CIs can be measured quantitatively, costs are generally covered by health insurance in developed countries. [Cla03: 769-771]

In developing countries, cochlear implantation is financially less feasible. Intense discussions are currently in process to clarify if simpler and cheaper CI devices [WRZ+98, APJ+07] for those countries should be produced. The greatest concern, however, would probably not be the hardware, but limited audiological and rehabilitative resources. [SB11]

2.2.3   System components and functionality

A cochlear implant consists of external and internal parts as shown in Figure 2.5. External parts include the processor, the transmitting coil, and the cable connecting them. Internal parts are the ones surgically inserted inside the skull. They include the receiver coil and stimulator, the cochlear electrode array, reference (or ground) electrode(s), and the cable connecting all these parts together.
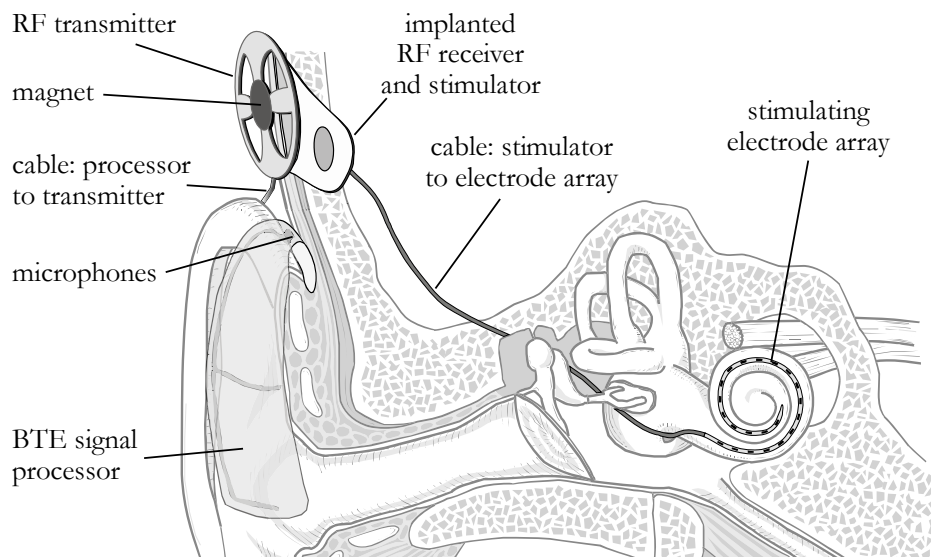
Figure 2.5: Components of a cochlear implant system. Based on [Wat93].

The processor is in most cases designed to be worn behind the ear, but there are also other types like on-the-body processors. The processor picks up the sounds by its built-in omnidirec-

tional and/or directional microphone(s). Then, it digitizes the signal and runs (optional) pre-processing algorithms like AGC[5] or ADRO[6] [Cla03: 432-433].

Next, the pre-processed audio signal is coded by the selected speech processing strategy (see section 2.2.4), for which the parameters are read from the CI processor's non-volatile memory. These parameters are determined by an audiologist during the fitting sessions. A BTE processor typically runs on one to three button cells, which last one to three days. The system inside is in most cases based on a low-consumption DSP[7] or ASIC[8].

The sound signal coded by the processor is streamed to the radio frequency (RF) transmitter, which consists of an antenna coil and a magnet. The counterpart of the magnet –together with the RF receiver– is embedded surgically into the skull bone and it ensures the transmitter to remain in place exactly over the receiver. The RF-link not only serves as the data connection, but it also supplies the implanted parts with energy. In modern CI systems, the RF link is bidirectional and it can provide the operating audiologist with intracochlear measurement data like impedances and electrically evoked compound action potential recordings. The latter is also termed neural or auditory nerve response telemetry (NRT or ART, respectively), see e.g. [Cla03: 683-684].

Next, the received stimulation data is interpreted and buffered by the stimulator, which is generally put inside a biocompatible enclosure together with the RF-receiver. Finally, the stimulator activates the intracochlear electrodes and delivers charges according to the requested order and magnitudes of stimulating pulses.

2.2.4   Signal processing strategies

Commercially available cochlear implant processors are capable of running various signal processing strategies. While early systems have used continuous analog signals to stimulate the cochlea [Cla03: 729-730], modern CIs employ principles from the channel vocoder [Dud39, Loi06] and apply pulsatile stimulation. Furthermore, all multi-electrode CI systems on the market make use of the tonotopy, i.e., low-to-high frequency signal parts are mapped to electrodes located apical-to-basal.

Current strategies mostly differ in terms of type and parameters (like number of bands) of the filter bank used, feature extraction methods (like tracking of the fundamental frequency or extracting linear predictive coding [LPC] parameters) implemented, rate and type (sequential or par-

---

[5] Automatic Gain Control

[6] Adaptive Dynamic Range Optimization

[7] Digital Signal Processor

[8] Application Specific Integrated Circuit

allel multi-channel) of the pulsatile stimulation, and the extent of preserving temporal fine struc-
tures of the input sound. A clear and unambiguous categorization related to the listed properties
is difficult, because access to specific pieces of information regarding CI strategies is often limited
by the manufacturers. Yet, some papers provide an overview of the most common strategies
[Cla03: 381-432, Møl06b, Loi06]. The signal processing strategies currently available in recent
commercial systems are listed in Table 2.3.

Since Cochlear's ACE is currently the most widely used signal processing strategy on the
market, it will be used as a baseline reference to evaluate SAM. A detailed description of ACE
including the step-by-step explanation of its signal processing is given in the following section.
More information on the flagship-strategies of the other manufacturers and recent advances in
CI signal processing follow in section 2.2.4.2.

Table 2.3: Signal processing strategies of current CI systems, listed by manufacturer.

Advanced Bionics:	**CIS** (Continuous Interleaved Sampling) [WFL+93],
	HiRes/HiRes Fidelity 120 (High Resolution strategy) [Nog08: 34-45],
	MPS (Multiple Pulsatile Sampler) [GFRM05].
Cochlear:	**ACE** (Advanced Combination Encoder) [VWPC00, Nog08: 29-34],
	CIS (see above),
	MP3000 (also known as the Psychoacoustic ACE, or PACE) [Nog08: 47-58],
	SPEAK (Spectral Peak strategy) [Loi98, Loi06].
MED-EL:	**HDCIS** (High Definition CIS) [LZOS10],
	FSP/FS4/FS4p (Fine Structure Processing) [VPV10, HNJ+06, Zie01].
Neurelec:	Crystalis/**MPIS** (Main Peak Interleaved Sampling) [LBL+10, DBP+10].

2.2.4.1 The advanced combination encoder

The advanced combination encoder is currently the world's most widespread CI strategy,
which was introduced by Cochlear with the 22-channel (electrode) Nucleus implants. ACE is
an N-of-M type strategy, which means that it only stimulates through $N < M$ electrodes within a
given period of time, where M is the total number of electrodes (in this case 22). Furthermore,
ACE belongs to the group of strategies operating based on the waveform representation of the
input signal, i. e., it is not doing feature extraction like F0 (fundamental frequency) tracking or
formant frequency estimation. "The principle underlying the use of this strategy is that speech
can be well understood even when only the peaks in the short-term spectrum are transmitted."
[Loi06: 122]

A simplified block diagram of ACE is shown in Figure 2.6. The block diagram, as well as the following explanation of ACE is based on [Nog08: 29-34].
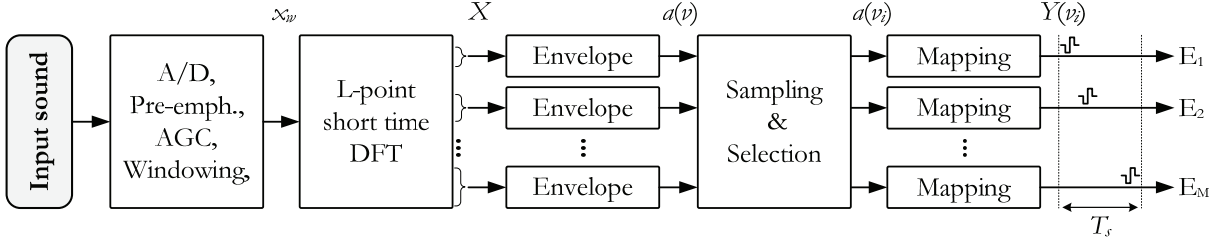
Figure 2.6: Block diagram of the ACE strategy. Based on [Nog08].

In ACE, the input signal is digitized at 16000 samples per second (sps) and the data is run through a pre-emphasis filter, which attenuates low frequency components. Then, an automatic gain control (AGC) is applied to reduce distortion for signal parts that are too loud. Next, the pre-processed signal is split into overlapping L-sample blocks by applying a Hann window [OSB99: 468-469]. Each block of windowed samples is sent through a filter bank that employs L-point Discrete Fourier Transform. The origins of consecutive windows are N samples apart. In the typical case that L=128 and N=8, windows are overlapped to an extent of 93.75%. For each sample block, the complex output of the DFT calculation is denoted $X(n)$ for each DFT bin n=1, 2, ..., L. The real-valued power of each bin can be calculated as shown in Equation 2.2, where $\overline{X(n)}$ is the complex conjugate of $X(n)$.

$$R^2(n) = X(n) \cdot \overline{X}(n). \qquad (2.2)$$

Now, from the L=128 bins (i.e., from the 65 bins due to real DFT input and Hermitian symmetry [OSB99: 630-633]) M=22 bands need to be composed. Since the ACE strategy only makes use of the envelope information of the band-filtered signals, composition of bands and extraction of their envelopes are handled together as shown below.

$$a(\nu) = \sqrt{\sum_{n=N_\nu^*}^{N_\nu^{**}} g_\nu \cdot R^2(n)}, \quad \nu = 1, 2, ..., M. \qquad (2.3)$$

In Equation 2.3, $a(\nu)$ is the envelope of band ν. Gains and limits for the summation for a typical ACE configuration are listed in Table 2.4.

Table 2.4: ACE settings for the default 22-electrode configuration. Based on [Nog08].

Band number v	1	2	3	4	5	6	7	8	9	10	11
First DFT-bin N_v^*	3	4	5	6	7	8	9	10	11	12	14
Last DFT-bin N_v^{**}	3	4	5	6	7	8	9	10	11	13	15
Gains g_v ($\cdot$1e^{-3})	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.68	0.68
Center freqs (Hz)	250	375	500	625	750	875	1000	1125	1250	1438	1688
Band number v	12	13	14	15	16	17	18	19	20	21	22
First DFT-bin N_v^*	16	18	20	23	26	30	34	39	44	50	57
Last DFT-bin N_v^{**}	17	19	22	25	29	33	38	43	49	56	64
Gains g_v ($\cdot$1e^{-3})	0.68	0.68	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65
Center freqs (Hz)	1938	2188	2500	2875	3313	3813	4375	5000	5688	6500	7438

Next, the "Sampling & Selection" block (see Figure 2.6) selects the N largest amplitude bands for stimulation. The selected bands are indexed v_i, $i=1,2,...,N$.

In the consecutive step, envelope magnitudes of the selected bands are mapped to electrical magnitudes using the so-called loudness growth function (LGF), see Equation 2.4. The most important purpose of the LGF is to apply compression, which is not included by the DFT-based linear filter bank.

$$p(v_i) = \begin{cases} \log\left(1 + \varrho \cdot \left(\dfrac{a(v_i) - s_{base}}{m_{sat} - s_{base}}\right)\right), & s_{base} \le a(v_i) \le m_{sat}; \\ 0, & a(v_i) < s_{base}; \\ 1, & a(v_i) \ge m_{sat}. \end{cases} \tag{2.4}$$

The calculation of the electrical magnitude $p(v_i)$ depends not only on the band envelopes, but also on the values of ϱ, s_{base} and m_{sat}. The parameter ϱ controls the steepness of the LGF. Input values of base level s_{base} are mapped to the electrical threshold level of the given electrode, $THL(v_i)$, while inputs at even lower amplitudes do not contribute to electrical stimulation. Input values exceeding the saturation level m_{sat} are mapped to the most comfortable level $MCL(v_i)$. The global volume setting, which CI users can normally adjust on their own, is typically implemented by reducing the dynamic range ($MCL - THL$) per band. (For example, $THL=100$ and $MCL=200$ would change to $THL=100$ and $MCL=150$ in the case of setting the global volume to 50%.)

Since *THL* and *MCL* are coded as 8-bit integers, the fraction $p(v_i)$ yields an integer value $Y(v_i)$ in the range of 0 to 255 via the mapping described in Equation 2.5. As an example for the implant driving currents, the conversion from $Y(v_i)$ to actual μA with the Cochlear Nucleus Freedom device with Contour Advance electrode is given in Equation 2.6.

$$Y(v_i) = \left\lfloor THL(v_i) + \left(MCL(v_i) - THL(v_i) \right) \cdot p(v_i) \right\rfloor. \tag{2.5}$$

$$I(v_i) = 17.5 \cdot 100^{Y(v_i)/255} \ \mu A. \tag{2.6}$$

Finally, the currents $I(v_i)$ are applied to the electrodes $v_i, i = 1, 2, ..., N$, sequentially, using biphasic pulses and in a stimulation order basal (high frequencies) to apical (low frequencies). Once the N selected electrodes have been activated, the input audio signal is shifted N samples and a new cycle of processing starts. An electrodogram showing the ACE output for a sample waveform is presented in Figure 2.7. (More electrodograms comparing ACE and SAM outputs are presented throughout Chapter 5.)
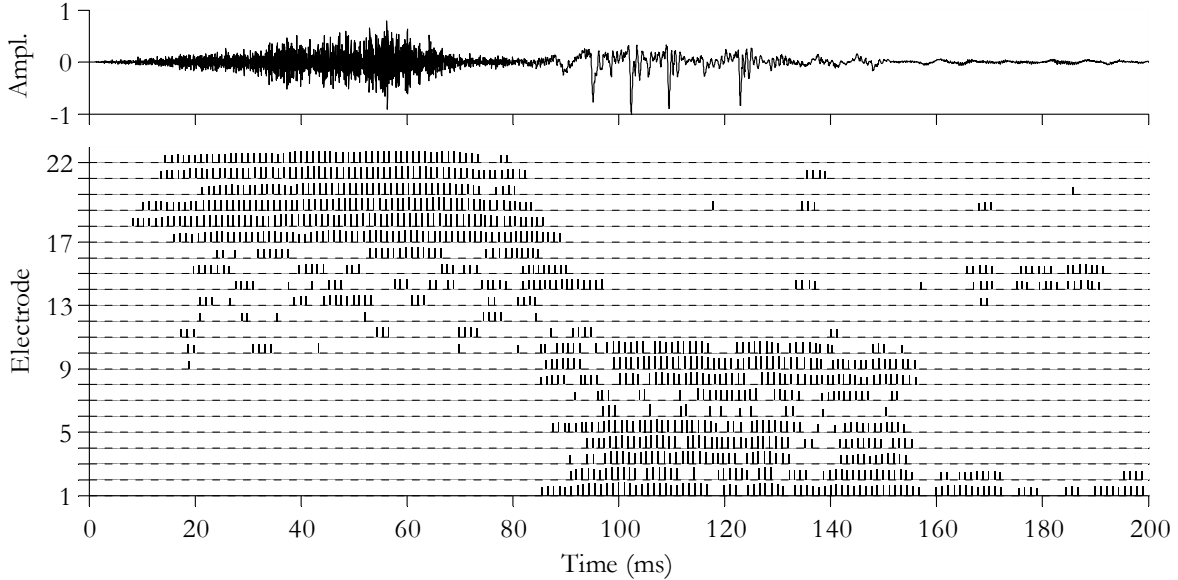
Figure 2.7: ACE electrodogram for the short utterance "sun". The strategy output was calculated with the standard ACE settings. In the figure, electrode no. 1 is the most apical one. Each stimulus pulse is shown as a vertical bar, the height of which represents the stimulus intensity. Maximum and minimum intensity corresponds to the *MCL* and *THL* of the given electrode channel, respectively.

Standard values used in the most Nucleus Freedom devices are ϱ=416.2, s_{base}=4/256, and m_{sat}=150/256. Threshold and most comfortable levels are fitted for each patient separately:

$THL(\nu)$ should represent the threshold of perception, while $MCL(\nu)$ should correspond to a "comfortably loud" level for each electrode ν. The audiologist may also decide to change AGC parameters, frequency mapping, LGF steepness ϱ, channel stimulation rate (*CSR*, 900 pps[9] by default), N, or even the pulse width (25 µs by default) of the biphasic stimuli.

It is important to see that ACE does not process phase information of the signal, so that it only provides place pitch cues to the CI user [LWM04, VST+05, Loi06: 127-133]. The lack of temporal pitch cues is also evident from Figure 2.7, where both low and high frequency electrodes use the same stimulation rate *CSR* without much variation over time.

In addition to the channel stimulation rate, the term *total pulse rate* (*TPR*) is also frequently used. It indicates the total number of discharges within a second. For the ACE strategy $TPR \cong N \cdot CSR$ holds.

2.2.4.2 Other common strategies and recent advances

To date, the ACE strategy is employed exclusively in Cochlear devices. A further developed version called MP3000 [Nog08: 47-58] is now making its way to an increasing number of CI users. MP3000 incorporates a psychoacoustic model into the "Sampling & Selection" block (see Figure 2.7) of the ACE strategy. This model determines a masking threshold based on current and previous data in each cycle of operation. Then, only non-masked spectral peaks are used to select the electrodes to be activated in the given cycle. While no outstanding increase in user performance could be shown with this strategy, it is beyond argument that MP3000 filters out many unneeded pulses. This contributes to energy conservation, which is a key point in implant technology. [BBS+11]

All the other manufacturers use CIS-related strategies (cf. Figure 3.2 on p. 35). CIS can be seen as a special parameterization ($N=M$) of ACE, in which all available electrodes are activated (in a pre-defined order) within each processing cycle. An advantage of CIS is the typically high stimulation rate associated with this strategy. High rates can potentially help provide temporal information to the auditory system [Loi06: 114-117].

MED-EL currently focuses on FSP [HNJ+06] and derived strategies in combination with their 12-electrode implants. FSP also comes with a heritage from CIS, but it uses channel specific sampling sequences (CSSS) [Zie01] in the low-frequency bands. In brief, it means that CSSS-channels do not work with single pulses (presented at the channel stimulation rate, as in CIS), but they employ pulse-bursts shaped with a cosine window like envelope. Furthermore, these bursts are synchronized with the zero-crossings of the band-pass filtered signal of the

---

[9] Pulses Per Second

given CSSS-channel. This implies that CI channels employing channel specific sampling sequences do also code temporal pitch as opposed to fixed-rate channels. In general, the two to three most apical electrode channels are enabled in FSP to be used with CSSS.

There are also enhanced versions of FSP. FS4, for example, uses up to six times higher rates on the CSSS-channels and the number of these channels can be increased to four. The latest variant (FS4p) features, in addition, parallel stimulation of the electrodes. This can help boost stimulation rate and give rise to virtual channels [WLZF92], which, in turn, can increase place pitch resolution. [Mei12]

Advanced Bionics' proprietary CIS-like strategy is called HiRes [Nog08: 34-35] and it makes use of the manufacturer's 16-electrode implants. Compared to other CIS implementations, HiRes calculates short-time averages of the half-wave rectified filter bank bands instead of using an envelope detector. HiRes runs with total pulse rates of up to 37 kpps. The current and enhanced version of this strategy is called HiRes Fidelity 120. It includes parallel stimulation and current steering, which are employed to create virtual channels, onto which the analyzed results of 120 spectral bands are mapped. This approach promises place pitch cues that are five times as accurate as those of ACE. A research version of this strategy was called SpecRes (Spectral Resolution strategy), see e. g. [Nog08: 35-45].

Neurelec works with 20 electrodes and their cochlear implant processor runs the strategy abbreviated as MPIS. It is also a CIS-derivative, but it can rather be thought of as a mixture of ACE and CIS. [LBL+10, Gna12]

2.2.5   The role of rehabilitation

Normal-hearing people tend to believe that speech perception of CI patients would be restored right after the surgical implantation. Unfortunately, it is not that simple.

After implantation there is a four to six weeks period of healing, during which the external processor is not connected to the implant. The journey from silence to sound begins only after that time. Turning on the processor for the very first time often does not result in any sound perception, and if it does, recipients describe the first sounds as being "unnatural", "robotic", or "weird". This signals that the cochlear implant processor must first be fitted to its user by an audiologist. CI users are unique, having their own history of deafness, so that a series of cochlear implant fitting sessions may be necessary. During these sessions, the parameters of the signal processing strategy (like stimulation rate, and threshold and comfort levels of the electrical currents, see section 2.2.4.1) are matched with the needs and preferences of the CI user. On the other hand, thanks to the brain's plasticity, the patient's auditory system will also adapt to the newly available sound information during the time between the fitting sessions. [Bos03]

The most important factors determining the number of required sessions are:

(1) *duration of deafness*: the longer one lives in silence the more unused neural structures will die away (i.e., the more time is needed for the auditory pathway to recover as much as possible);

(2) *brain health*: the readiness of the neural system (plasticity) to adapt to the new sensory input may vary a lot (e.g. according to the patient's age and fitness);

(3) *electrode position*: the exact location of the implanted electrode array is almost never known, so it is hard to set the correct center frequency for each electrode channel; (as a matter of fact, this is just another issue that plasticity must typically take care of)

(4) *signal processing*: current CI sound processing strategies deliver excitation patterns unfamiliar to the ear.

There now exist a large number of cochlear implant rehabilitation centers worldwide. Audiologists and speech (and other) therapists are there to provide specialized help to the CI recipients on a regular basis. Beyond technical aspects, this help includes logopedics and therapeutic pedagogy (and often medical and psychological care, too). Apart from some isolated cases, a number of sessions are needed to reach considerably good results in understanding speech. However, the duration also depends on whether "only" a rehabilitation (for post-lingually deafened CI users) or a habilitation (for pre-lingually deafened CI users) is needed.

Rouger et al. made a study in 2007 to investigate whether CI patients still can have benefits of multisensory integration [RLF+07]. They have found that, in fact, cochlear implant listeners are the better multisensory integrators. They have published a summary of an extensive experiment in which they have monitored speech-understanding abilities of CI users during the rehabilitation process and a long-term follow-up. Speech perception tests have been carried out with and without allowing visual cues, i.e., lip reading. See Figure 2.8 for an excerpt of the results based on 97 subjects.
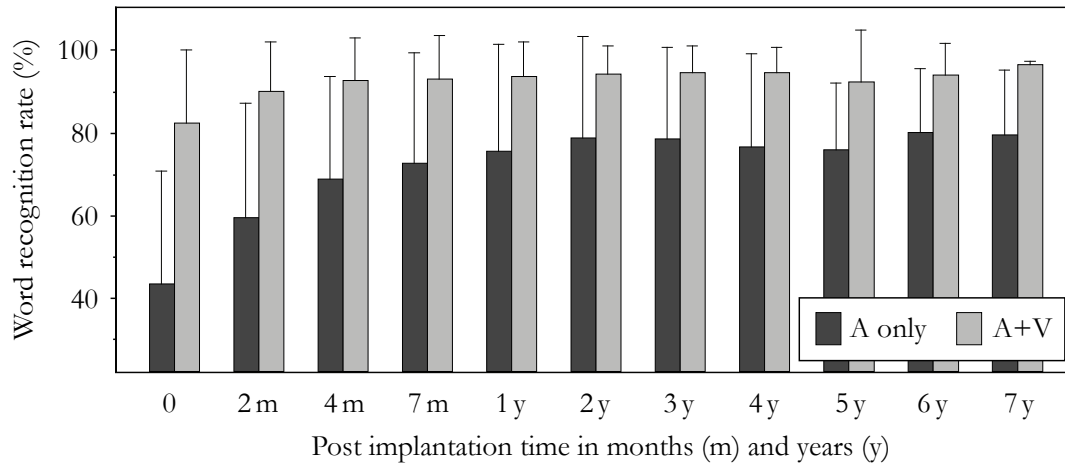
Figure 2.8: Average word recognition rates as a function of time after implantation using
different modalities. Auditory (A only) and audio-visual (A+V) situations were
tested. Adapted from [RLF+07].

Note that in Figure 2.8 the "audio only" recognition rate reaches approx. 80 % after 2 years of implantation. Patients at this rehabilitation stage are often able to communicate over the phone, once the topic of the conversation is well defined, i. e., the spectrum of words during the talk is limited.

Indeed, professional rehabilitation and follow-up trainings are crucial for success after cochlear implantation, and their importance should not be underestimated.

2.3   Computational models of hearing

Models of hearing (also called auditory models) date back to ancient times. Aristotle already explained hearing from an early point of view in his *De Anima* (book II, chapters 5-12) [Aris30] dated approximately 330 BC. He drew attention to analogy to other sensations and used metaphors to characterize the processes in the ear. This can be seen as at least a theoretical model of hearing. However, to design a new cochlear implant signal processing strategy that is based on an auditory model, one would need a computational model of hearing. This section surveys available models that can possibly be used in CIs.

The mechanisms of the outer and middle ear have been described early due to the accessibility of these parts of the ear [HC33, Zwi62, Møl65]. In the field of cochlear implants, these parts are typically modeled together as a simple linear filter reducing low-frequency (and in some implementations also high-frequency) energy before the signal is further encoded [Cla03: 432-433, 482] (see also pre-emphasis in the ACE strategy on p. 16).

Computational models of the inner ear and further stages of the auditory pathway are discussed throughout the next sections. Though this retrospection is not an essential part of the

presented work, it should prove that choosing the right models is not an easy task. The final choice of models for the CI stimulation strategy is discussed in sections 4.3.2 and 6.1.2.

2.3.1   Inner ear models

It was a long way from the observational hearing research to the anatomically correct functional description of the human in vivo inner ear. Some relevant early works were [Hel63, WL24, Bék28]. Computational models of the inner ear emerged only during the last third of the 20th century due to the increasing availability of computers. The well-documented findings of Békésy[10] [Bék60] also had a huge impact on the progress. Research groups around the world were driven by various aims and they concentrated on different entities within the cochlea. For the possible use in the CI stimulation method described in this thesis, two groups of models are most relevant: basilar membrane models, and models of the mechanoelectrical transduction (i. e., inner hair cell and auditory nerve complex). These are discussed in the next two sections.

2.3.1.1   Basilar membrane

The main purpose of the basilar membrane is the frequency decomposition of the input signal. Since this is also the most essential part of CI sound processing, this paragraph surveys possible basilar membrane models thoroughly.

Today, a large number of basilar membrane models exist [MFLP10]. They can be categorized according to various criteria [Chi10a]. One of them is detailedness, i. e., micro- vs. macromechanical models. In the latter, some structures within the cochlear partition are neglected. Furthermore, assumptions can be made to simplify the three dimensional cochlea and model it in only two dimensions or as a one-dimensional segment. A further decisive property is whether the effect of outer hair cells is included; if yes, the BM model is said to be active. If the OHCs are modeled with their nonlinear amplification, then it adds up to an active nonlinear BM model. Another categorization could be transmission line and point models of the basilar membrane. Transmission line models mimic the energy flow along the BM, while point models can reproduce the activity of the BM at any distinct position along the membrane, without the need to calculate more than that.

For practical reasons, the overview presented here is restricted to macromechanical one-dimensional models. Most of these can be associated with one of three well-differentiated tracks of development, as shown in Figure 2.9. They are all based on measurement and modeling work of earlier studies, but they all seek to prove different claims.

---

[10] I am going to use the original Hungarian name *György Békésy* instead of *Georg von Békésy* throughout this thesis.
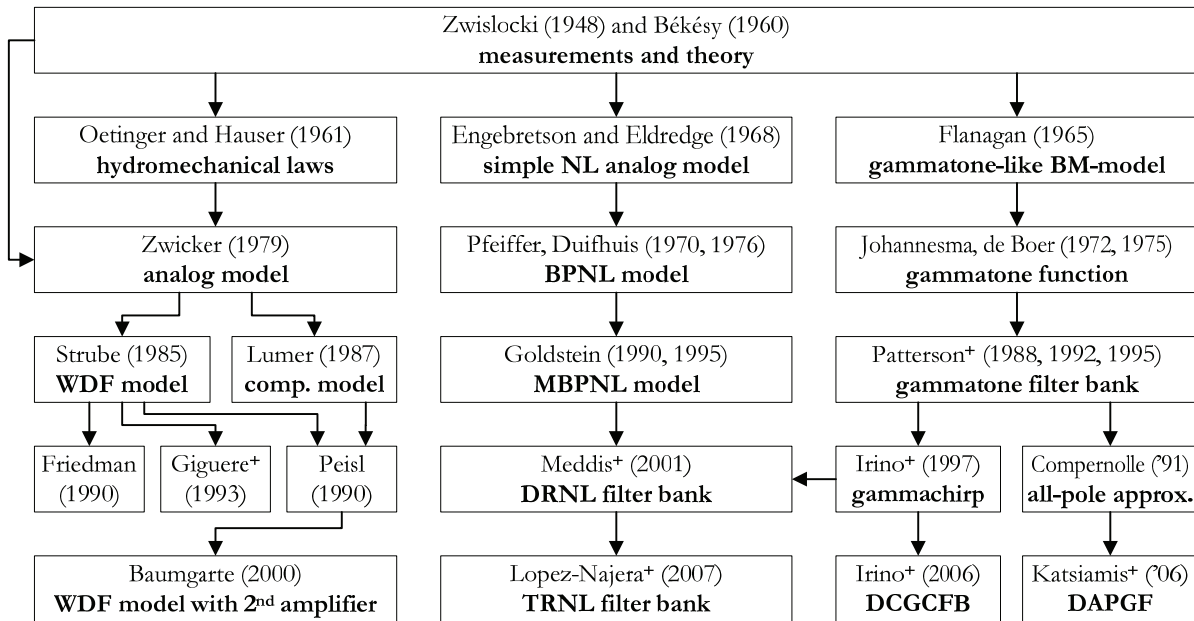
Figure 2.9: Historical tracks of basilar membrane model development. Highly simplified and incomplete representation. (Superscripted plus signs indicate more than one author.) Based on [Chi10a, LKD10, MFLP10].

Filter banks based on cochlear hydromechanics

One development track (the leftmost column in Figure 2.9) has aimed at modeling the human BM through its hydromechanical properties. Békésy [Bék47, Bék49, Bék52, Bék60] and Zwislocki [Zwi48] delivered compliance and mass data of the basilar membrane, respectively. However, they did their experiments with post mortem cochleae, so they could only capture the passive characteristics of the BM. Based on the published physical and geometric properties, Peterson and Bogert [PB50], Ranke [Ran50], and Oetinger and Hauser [OH61] progressively established the hydromechanical laws of the cochlea. Oetinger and Hauser described these by using two simultaneous partial differential equations. The equation system could be resolved through a lattice network after a series of approximations.

Zwicker noticed the necessity of active elements. He divided cochlear pre-processing into nonlinear pre-amplification with feedback (outer hair cells) and transformation into neural signals (inner hair cells). Eventually, he developed an analog cochlea model including active nonlinear effects [Zwi79, Zwi86]. This was transformed into a Wave Digital Filter (WDF) [Fet86] structure by Strube [Str85], with which a numerical solution could be achieved. (Strube's model with 100 sections covering the whole frequency range of human hearing and with a sampling rate of 48 ksps was employed for example in [HGH07].) Strube's model was extended by Friedman, who included another degree of freedom mimicking the vibration of

the tectorial membrane [Fri90]. The Strube-Zwicker model was revised and validated by Giguère and Woodland [GW94a, GW94b].

Lumer developed a complex computational model [Lum87a, Lum87b] to reproduce two-tone suppression effects (see e.g. [ML10:11]), which is an important aspect of masking in human listeners. His model included linear parts for pre-processing from Oetinger and Hauser, and nonlinear elements from the model of Zwicker. Peisl also developed a computational model based on Zwicker's system [Pei90, ZP90]. It included enhancements regarding the lateral coupling of the modeled BM sections. Finally, Baumgarte has further extended the model of Peisl by adding a second amplification stage [Bau99, Bau00] beyond the feedback loop of the OHC structures. [Chi10a]

All elements of this track are transmission line models of the cochlea, while the following two paragraphs only enumerate point models.

Filter banks connected to gammatone functions

Another aim in the history of BM model development was to mathematically describe the measured impulse responses of the cat cochlea (third column in Figure 2.9). Johannesma [Joh72] and de Boer [Boe75] published an analytic expression for the reverse correlation function to estimate the IR of the fibers concerned. A similar function was already employed by Flanagan in a simple basilar membrane model [Fla60, Fla65]. This function was later named *gammatone* [AJ80], which should refer to the fact that it was the product of gamma-distribution envelope and a sinusoidal tone[11]. Schofield confirmed that experimental data on masking could be explained by the amplitude frequency response of gammatones [Sch85]. Patterson and colleagues made great efforts to create a practical filter bank of gammatone filters, which could also be used as a standard (software) platform for further research [PNHR88, PRH+92, PA95].

In 1997, Irino and Patterson extended the gammatone function with a chirping phase: the *gammachirp* function was born [IP97]. Gammachirp now included the level-dependent asymmetric shapes of the cochlear filters (see e.g. [ML10:16]). In 2001, they modified their architecture and created the compressive gammachirp filter (CGCF), with which both physiological data and nonlinear masking effects could be explained [IP01]. The CGCF was fitted in 2003 to more recent data from masking experiments [PUI03]. Eventually, the CGCF approach was further extended by a new level-control path and named dynamic compressive gammachirp filter bank (DCGCFB) [IP06]. The DCGCFB also enabled to explain two-tone suppression. [Chi10a]

---

[11] Though not named gammatone, Eaglesfield [Eag45] and Tucker [Tuc46] already used a similar structure (cascade of N resonators) in the mid 1940s to analyze filters by their step response.

Based on the linear gammatone filter bank of Patterson et al. [PNHR88, PRH+92] Carney developed a nonlinear model to better simulate responses of the cat auditory nerves [Car93]. The nonlinearity was introduced by a feedback mechanism. A significant enhancement to this model was published by Zhang and colleagues [ZHBC01] in 2001. They exchanged the feedback mechanism with a broad-bandwidth control path, through which further physiological effects like two-tone suppression and asymmetrical growth of suppression above and below the characteristic frequency were included. A further extension included level-independent frequency glide and level-dependent compressive nonlinearity (see e.g. [ML10:21]), which was published by Tan and Carney [TC03] in 2003. [Chi10a]

Also based on the 1988 gammatone filter of Patterson and colleagues, efficient approximations by discarding the zeros from a pole-zero decomposition were published [Com91, Sla93, SL93]. The result was called all-pole gammatone filter (APGF). Lyon continued the work by introducing the one-zero gammatone filter (OZGF) [Lyo96] and the differentiated all-pole gammatone filter (DAPGF) [Lyo97]. Eriksson and Robert (inspired by the work of Carney [Car93]) turned the APGF into an active filter by adding distributed feedback loops [RE99]. Their model reproduced several observed phenomena including compression, two-tone suppression, and suppression of tones by noise. Katsiamis, Drakakis and Lyon released VLSI (very-large-scale integration) implementations of the DAPGF and OZGF models in 2006 [KDL06]. They published an overview and comparison of the models of this development path (including APGF, OZGF and DAPGF) in 2007 [KDL07]. Eaglesfield showed as early as 1945 that cascades of resonators could be used as a basis to build gammatone-like filters in hardware [Eag45]. In a recent paper, Lyon gave an overview of filter banks of this approach [Lyo11].

Filter banks based on band-pass and dual resonance nonlinear filters

Yet another development track of basilar membrane models originates from the work of Engebretson and Eldredge (second column in Figure 2.9). They were about to build a simple mathematical model describing the nonlinear behavior of the cochlea [EE68]. In 1970, Pfeiffer modified the network by Engebretson and Eldredge and built the band pass nonlinear (BPNL) model [Pfe70]. This was able to reproduce two-tone inhibition properties more similar to those observed experimentally. The explanation of the physiological basis and a thorough analysis of the BPNL model were published by Duifhuis in 1976 [Dui76]. He hypothesized that frequency selectivity would be enhanced by sensory hair cells, which were represented by the BPNL model. Goldstein added another signal path to the BPNL and called it multiple band pass nonlinear (MBPNL) model [Gol90, Gol95]. The restructured filter architecture was able to explain a number of cochlear filter properties such as compression, suppression, distortion, sim-

ple-tone interference [ML10:17-19], level-dependent tuning and best-frequency shifts (see e.g. [LBA07]).

Based on the concept described in [EE68] and [Pfe70], Kim and colleagues developed a system of nonlinear differential equations that was able to reproduce a large number of frequency-dependent nonlinear phenomena of the in vivo basilar membrane [KMP73]. Neely and Kim [NK83] and Davis [Dav83] proceeded by explicitly identifying the function of the outer hair cells in such a system. A further enhancement to account for otoacoustic emissions (OAEs, see e.g. [Møl06a]) was published by Neely in 1985 [Nee85].

In 2001, motivated by the MBPNL approach, Meddis et al. developed the dual resonance nonlinear (DRNL) filter [MOL01], which was also a combination of two signal paths: one linear and one nonlinear. One difference to MBPNL however, is, that the center frequencies of the two paths in DRNL are unmatched. This produces the shift in the best frequencies as the input signal intensity increases. It is important to mention that the band-pass filters of the DRNL employ gammatone functions (cf. Figure 3.4 on p. 36). A DRNL filter bank tuned with human cochlea parameters was published in the same year [LM01]. Plack and colleagues showed in 2002 that DRNL could reproduce the majority of known masking phenomena [POD02]. Lopez-Poveda extended previous work on DRNL by giving an approximate analytic transfer function that allowed fast calculation and analysis of the level-dependent frequency-domain response for at least tone inputs [Lop03]. (However, his transfer function did not reflect model nonlinearities in a proper way and was criticized e.g. by Duifhuis [Dui04, Lop04]). In 2003, Sumner et al. presented a DRNL-based complete nonlinear model of the guinea pig [SOLM03]. In the following year, Holmes and colleagues reported good agreement with physiological data after having tuned some of the parameters [HSOM04]. In 2007, Lopez-Najera et al. enhanced the DRNL by adding a third parallel signal path, thus creating the triple-path nonlinear (TRNL) model [LLM07]. The new path included a linear amplifier and an all pass filter. Using linear regression across the optimized parameters for seven different sites of the modeled BM, they could create a filter bank that simulated the chinchilla cochlea for tone inputs reasonably well.

2.3.1.2 Mechanoelectrical transduction

The mechanical energy (i.e., the movement of the stereocilia) is turned into electricity (i.e., into action potentials propagating towards the auditory cortex) by the complex of inner hair cells and spiral ganglion neurons. Though these embody separate stages of the peripheral auditory processing, they are often studied together. They play an essential role in a variety of psycho-acoustic properties like adaptation and temporal masking.

Since the CI signal processing strategy presented in this thesis is intended to mimic the parts of the ear that are bypassed by the cochlear implant, an appropriate model of the inner hair cells is needed in addition to a model of the basilar membrane.

A good neurophysiological and neuroanatomical overview with emphasis on the auditory nerve fibers and inner hair cells was given by Davis in the early 1960s [Dav62]. Shortly after, Weiss published a spiking auditory nerve model, which was tuned with spike activity patterns recorded in cats [Wei63]. In a further study, he also prepended a linear BM model and a simple IHC model to his system to create a complete computational model of the peripheral auditory system [Wei64, Wei66]. (This was found inaccurate for non-sinusoidal stimuli [Gei68].) In 1965, Siebert proposed a way to model the stochastic behavior of the primary auditory neurons better [Sie65]. Eggermont developed a computational model (and an analog electronic circuit) of the cochlear adaptation based on measurement data from frogs [Egg73] and extended it soon after with stochastic properties [Egg75]. In 1974, Schroeder and Hall published a computational model of the mechanoelectrical transduction [SH74]. A significant innovation was to introduce quanta of neurotransmitter to mimic vesicle releases. Ten years later, Lyon proposed a system consisting of computational models of neural auditory processing to support automatic speech recognition [Lyo84]. In 1985, Weiss and colleagues developed a multi-scale (macro- and micromechanical) model of the alligator lizard's ear, including a very detailed description of the biomechanics of the IHCs [WPR85, RPL+85, WL85]. One year later, Shamma et al. developed a complete biophysical model of the cochlear processing of the guinea pig [SCW+86].

A milestone in modeling the mechanoelectrical transduction was marked by the model of Meddis [Med86]. His system combined several aspects of previous models and added new features like the neurotransmitter reuptake. The "Meddis model" (as it is now generally referred to) produced realistic mammalian rate intensity functions and adaptation effects, among others. It was reparameterized and extended in 1988 [Med88] and 1990 [MHS90], and further revised in 1998 [LOM98] (e.g. quantal release of the neurotransmitter as seen in [SH74] was included). A good overview was compiled by Christiansen [Chr01], while a comparison with seven other IHC models was presented in [HM91]. VLSI implementations were developed and published by McEwan and Schaik [MS03, Sch03, MS04].

In 2002, Sumner et al. reworked the Meddis model by splitting it into neurophysiologically meaningful submodels, which could be tuned better with measurement data available at that time [SLOM02]. While the model was first published with a guinea pig parameter set, Wiegrebe and Meddis brought out the parameters for human settings in the appendix of [WM04]. Model

features like adaptation, forward masking, and first spike latency were thoroughly evaluated in 2003 [SLOM03], 2005 [MO05], and 2006 [Med06a, Kri06, Med06b], respectively.

Other sophisticated IHC models were published for example by Nam et al. [Nam05, NCG07] and by Lopez-Poveda and Eustaquio-Martín [LE06].

2.3.2 Cochlear nucleus and beyond

As mentioned in section 2.1.1, the next stages of auditory processing are the brainstem, midbrain, the thalamus, and finally, the auditory cortex. The structures towards the auditory cortex are known to get more and more complex, which also means that accurate and widely accepted models of these regions are rare. Typically, individual cell types are modeled fairly well (e.g. bushy, stellate, or chopper cells of the cochlear nucleus, see [VZ10, WHB10]), but the functional interconnections between them are not fully identified.

Since the auditory model in SAM only includes parts up to the IHC processing (see section 4.3.1), models of higher auditory entities are not discussed here. For the status quo in electrical stimulation of the brainstem, see [Møl06a]. For a succinct overview of the central auditory processing, refer to [Cla03: 83-93]. For a comprehensive summary of beyond-the-cochlea models, see [MFLP10].

Chapter 3

State of the art

As stated in section 1.1, there seem to be two approaches to better mimic the human auditory system in cochlear implant signal processing strategies. One is to augment existing strategies by algorithms that mimic specific properties of the ear. The other is to completely replace a large part of a given strategy with an auditory model that includes many features given in physiological hearing. This chapter reviews published work on both of the above concepts.

The following section discusses augmented strategies and gives ACE-related examples. Subsequently, section 3.2 summarizes published attempts to create auditory model based strategies. The chapter ends by summing up key issues.

3.1 Augmented strategies

One concept to enhance speech recognition of CI listeners is to incorporate psychoacoustic features of the human ear into the sound processing algorithm of cochlear implants. Most studies published so far have addressed only one specific issue like missing cochlear delays [TGB10], improper compression [FS98], lack of adaptation [GW99], or the absence of temporal pitch (or fine structure) information [VST+05, NSZ05]. By fixing any of these deficiencies alone, a small increase in user performance can typically be shown.

Since there are many ideas worldwide on how to add new features to existing CI signal processing strategies, it would be impossible to cover them all in this thesis. Instead, only two of them are presented here in more detail. Namely, incorporating the fundamental frequency and cochlear delays into the stimulation pattern. (The MP3000 strategy incorporating masking effects into ACE, as discussed in section 2.2.4.2, could be just another example.)

3.1.1 Fundamental frequency

Contemporary CI systems provide far fewer pitch cues than the intact peripheral ear does. On one hand, the spectral resolution (place pitch) is limited by the number of electrodes and by the current spread [CRSC03] within the cochlea. On the other hand, temporal pitch is not (or not sufficiently) supported by many of the signal processing strategies. Consequently, CI listeners

are not able to perceive several aspects of speech (e.g. prosody) and cannot typically identify melodies and instruments in music. Speakers of tonal languages, such as Mandarin, have even greater trouble with their CIs, since they may miss lexical meaning normally encoded in vocal pitch variations. Hence, efforts to improve the coding of F0 are crucial.

The use of virtual channels (see section 2.2.4.2) is an attempt to increase spectral resolution. However, including temporal pitch by varying the stimulation rate according to the fundamental frequency is presumably less hazardous from a technological point of view. An overview on schemes incorporating F0 into the CIS and ACE strategies is presented in [Loi06:127-133]. One of them is called Multi-channel Envelope Modulation (MEM) and is detailed by Vandali et al. in [VST+05]. In MEM, the fundamental frequency is extracted from the input signal by a series of low and high pass filters. The envelopes $a(v)$ estimated by the ACE filter bank (see section 2.2.4.1) are then modulated by the F0 envelope signal. Other stages of the ACE processing are unaltered. An illustration of the MEM stimulation pattern (termed F0-ACE) is presented in Figure 3.1.

Based on the above references, most of the F0 enhancement strategies (including MEM) have been shown to improve pitch perception. At the same time, they have not affected speech recognition in any significant way.

3.1.2   Cochlear delays

A more recent study by Taft et al. [TGB10] aimed to include traveling wave delays in ACE stimulation patterns. They hypothesized that this could contribute to better CI speech perception in noisy environments. In fact, they could show a small but significant improvement in an experiment with eight CI listeners.

In the implementation by Taft and colleagues, the envelopes $a(v)$ estimated by the ACE filter bank are delayed according to corresponding electrode locations. Knowing that the electrodes of Cochlear's Nucleus array are spaced 0.635 mm apart, the delays are estimated using the equation of Donaldson and Ruth [DR93] as shown below. In Equation 3.1, d is the distance from the stapes (in mm) and L is the approximated delay (in ms).

$$L(d) = 0.3631 \cdot e^{0.11324 \cdot d}. \tag{3.1}$$

Though the mechanism underlying the increased recognition rate with this strategy is not 100 % clarified, the introduction of frequency-dependent delays definitely has a positive effect on ACE: it helps separate components of spectrally rich input sound segments, which, in turn, manifests itself in better sound to stimuli mapping. For details, see also [Taf09, TGB09]. An il-

lustration of the stimulation pattern calculated with this strategy (termed TW-ACE) is also presented in Figure 3.1.
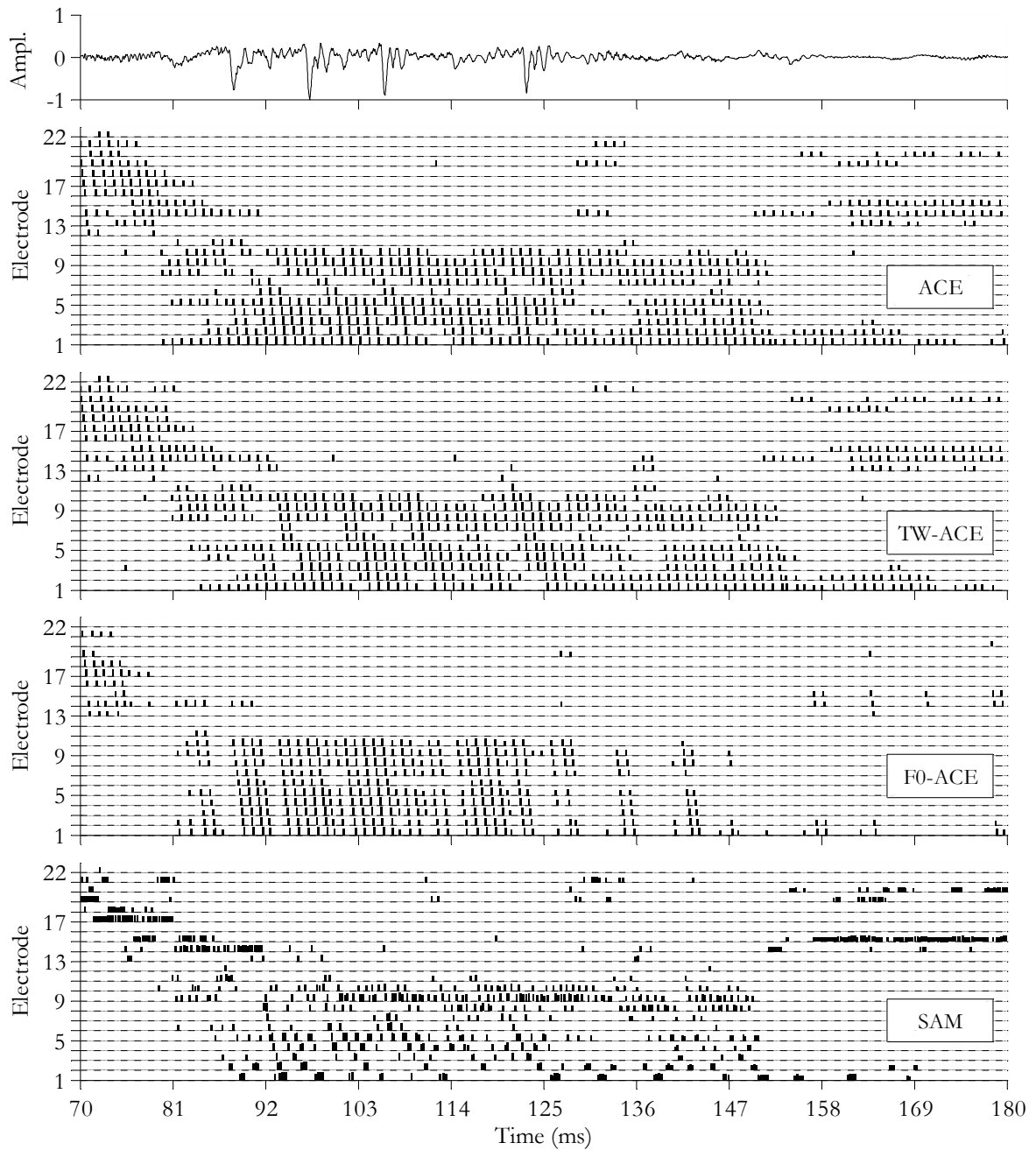
Figure 3.1: Comparison of electrodograms for the 110 ms middle part of the short utterance "sun". Using the same parameterization, the output of ACE is compared to those of the traveling wave ACE (TW-ACE) and F0-modulated ACE (F0-ACE) implementations. The electrodogram on the bottom shows the output of SAM and is only provided here as a matter of interest.

3.1.3   Remarks

The studies presented in this section have the advantage that they only tackle one single property missing from conventional CI strategies. The relatively small number of extra parameters makes them easier to evaluate. Furthermore, they can be implemented in current CI processors, because of their simplicity. At the same time, however, the use of these strategies does not promise huge differences in user performance.

3.2   Strategies based on auditory models

A completely different approach –with more potential, but also with more pitfalls– is to rely completely on auditory models. This way, important properties of the auditory system shall be included inherently.

This idea was first explored and documented by Schatzer and colleagues [SWWL03] in 2003, who were about to exchange the band-pass filter bank of the CIS strategy with the DRNL filter bank (cf. Figure 2.9 on p. 24). In 2007, Nogueira et al. [NKH+07] managed to incorporate the active nonlinear Baumgarte basilar membrane model along with the Meddis IHC model into the ACE strategy. In the same year, Kim and colleagues [KKK07] carried on the work of Schatzer et al. with the DRNL strategy. They relied on acoustic simulation for the evaluation part. Even though there are signs of similar activity in other laboratories (e.g. in that of Zachary M. Smith in Englewood, Colorado, USA or of Werner Hemmert in Munich, Germany), SAM completes the enumeration of the published attempts on combining auditory models with signal processing strategies (as of July 2014). The following paragraphs provide more details on each of the above studies.

3.2.1   Better mimicking with DRNL

In a progress report [SWWL03] of an NIH[12]-funded project, Reinhold Schatzer, Blake Wilson, Robert Wolford, and Dewey Lawson documented a novel signal processing strategy for "closer mimicking of normal auditory functions". The basis for their studies was a standard CIS processor design, as shown in Figure 3.2. Their aim was to introduce nonlinear processing elements resembling the human peripheral auditory system in the CI speech processor. They hypothesized that these nonlinear auditory models should provide a better approximation of the normal cochlear processing, which, in turn, should contribute to better speech reception in both quiet and noise.

---

[12] National Institutes of Health
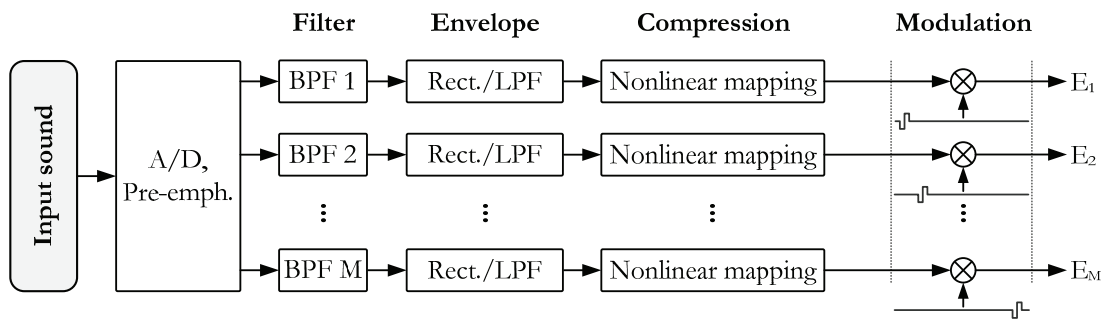
Figure 3.2: Standard CIS processor structure. (BPF: band-pass filter, Rect.: rectifica-
tion, LPF: low pass filter.) Adapted from [SWWL03].

The development of the new speech processor structure (shown in Figure 3.3) was divided
in three steps. First, the bank of linear band-pass filters of the CIS processor should be substi-
tuted with a DRNL filter bank. Second, the envelope detector and parts of the compression ta-
ble of CIS should be replaced with the Meddis IHC model. Third, steps 1 and 2 should be
combined carefully (paying attention to inter-stage gains and compression).

Schatzer et al. reported in [SWWL03] that they have accomplished the first step of the above
enumeration. (Unfortunately, it seems that they have not proceeded to the next steps since
then.) They modified the original DRNL structure [MOL01, LM01] slightly to work with a
normalized amplitude range and to preserve the necessary numeric precision in filter calcula-
tions. The modified structure is shown in Figure 3.4.
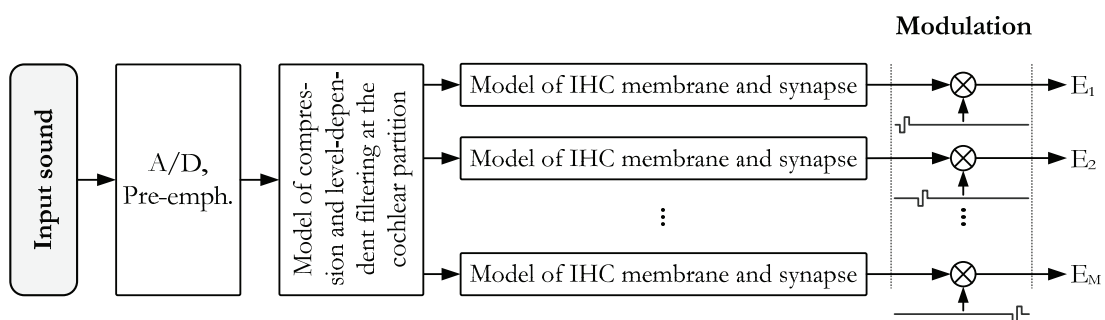
Figure 3.3: Processor structure for a closer mimicking of normal auditory functions as
proposed by Schatzer et al. Adapted from [SWWL03].

The new signal processing strategy was tested with eight CI listeners (the majority of which
was equipped bilaterally with MED-EL devices), using various parameterizations of the DRNL
filter bank. Speech reception was tested in a VCV[13] context, where the vowel was fixed to /a/,
and the consonant was varied. The utterances were made noisy using CCITT speech-spectrum

---

[13] Vowel-Consonant-Vowel

noise and the signal to noise ratios (SNRs) were selected to create challenging but not impossible listening situations.
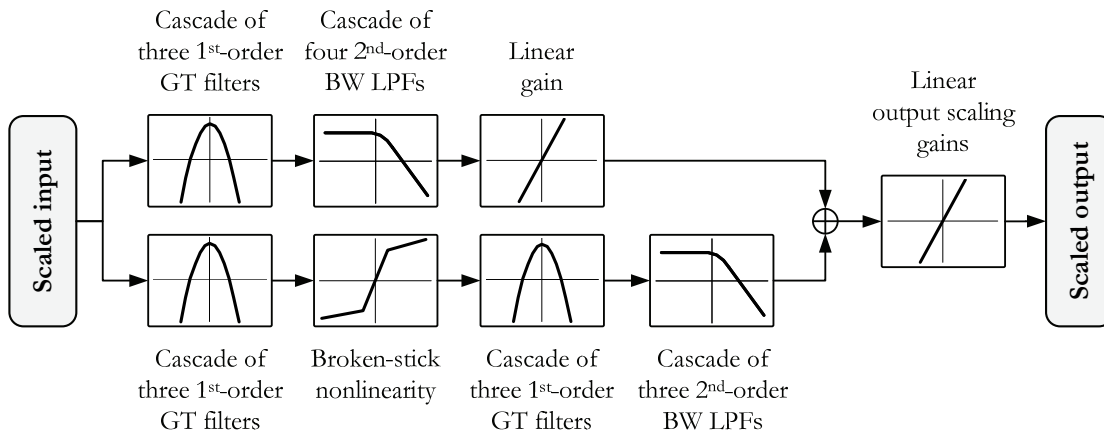
Figure 3.4: Modified DRNL filter structure as implemented by Schatzer et al. for processing digitized sound in the normalized amplitude range [0,1]. Adapted from [SWWL03].

The results of the listening tests are mixed. On average, the users performed better with the standard CIS strategy. However, there are cases (mainly in high SNR conditions) where the DRNL-based implementation slightly outperformed CIS. Significant improvements (with both the CIS and the DRNL strategy) could be observed when using 12 instead of just 8 electrodes.

In the next progress report [WWS+03] Wilson et al. combined the DRNL-CIS implementation with the idea of virtual channel interleaved sampling (VCIS) [WLZF92]. (In the published implementation, 21 virtual electrodes can be addressed for stimulation, while only having six physical electrodes. In VCIS, adjacent physical electrodes may be stimulated simultaneously to shift the perceived pitch in any direction, and hence, to create virtual electrodes.) The DRNL-VCIS strategy was tested with only one CI listener in a similar setup as discussed above. It could be shown that (a) in no test case did the standard CIS strategy perform significantly better than the DRNL-VCIS, and that (b) DRNL-VCIS scored significantly higher than CIS in all tests at 5 dB SNR.

Though these results are encouraging, they (unlike the idea itself) cannot be deemed as a breakthrough in CI speech processing. Further publications of Wilson, Schatzer, and colleagues tackling their above presented approach are e. g. [WSL+05, WSL06, WD08a, WD08b, WLS10].

3.2.2   Further investigations with DRNL

The idea of the DRNL-CIS signal processing strategy was carried on by Kim and colleagues [KKK07, KCKK09]. They further altered the DRNL structure (see Figure 3.5) already modi-

fied by Schatzer et al. and they claimed to have it implemented to run with minimal computational requirements. This should even allow real-time operation on an unmodified CI processor. In this study, the DRNL-CIS strategy was implemented and parameterized to be used in four-, eight- and twelve-channel cochlear implants.

Unfortunately, the evaluation did not involve CI listeners. Instead, spectral analysis (measures of preserving formant information and dominant frequency component analysis) and listening tests with normal-hearing listeners using acoustic simulation of CI strategies took place. Listening tests included syllable identifications tests similar to those in the study of Schatzer et al. There were two noise conditions: a) no noise, and b) 2 dB SNR speech-shaped noise.

Results show that formant information is more noise robust with the DRNL-CIS than with the standard CIS strategy. Syllable identification tests showed statistically significant benefits (about 20 % more correct) with DRNL-CIS over CIS with the eight- and twelve-channel implementations.
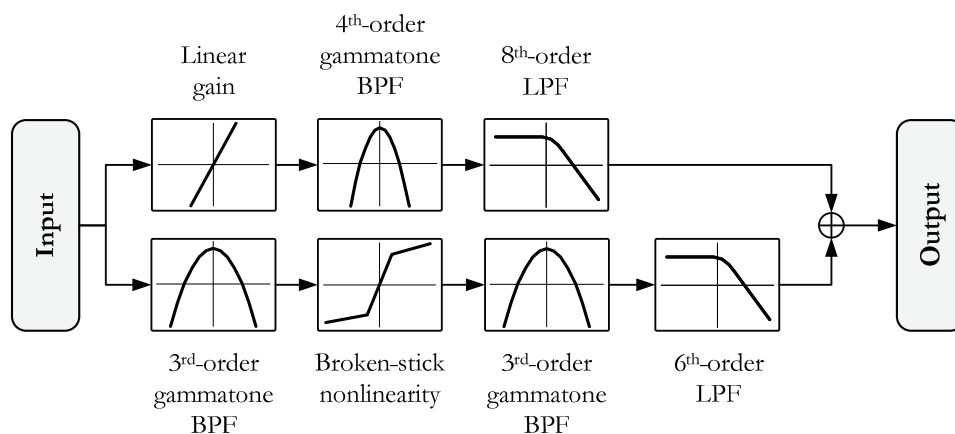
Figure 3.5: Modified DRNL filter structure as implemented by Kim et al. Adapted from [KCKK09].

The above results are quite impressive. However, an obvious shortcoming of this study is that the implemented strategy has never actually been tested with CI users.

3.2.3 First experiments with the Baumgarte and Meddis models

In 2007, Nogueira et al. [NKH+07] published the first results of a study in which auditory models were combined with the ACE strategy. This study was the first attempt worldwide to use a transmission line BM model (see section 2.3.1.1) as auditory filter bank. More specifically, the Baumgarte model (also referred to as the extended Zwicker model, see page 25) and the 1986 Meddis IHC model (see page 28) were chosen. A group of the Fraunhofer Institute for Digital Media Technology IDMT (including two of my colleagues and myself) participated in

the project by delivering the PC implementations of the selected auditory models and by providing easy-to-use interfaces to them.
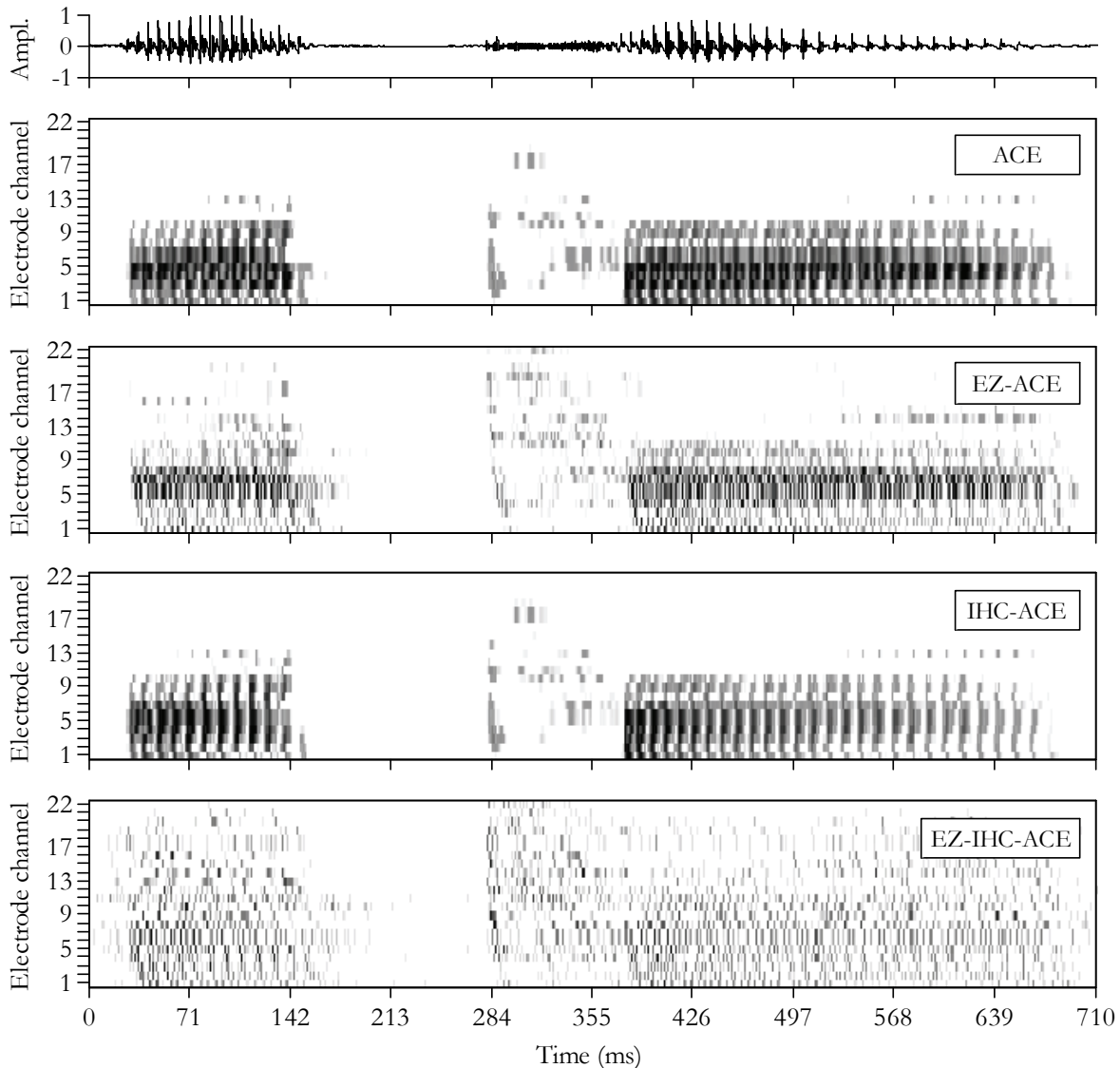
Figure 3.6: Comparison of envelope data for the short utterance "aka" calculated with the strategies ACE, EZ-ACE, IHC-ACE, and EZ-IHC-ACE. Adapted from [NKH+07].

As proposed by Schatzer et al. [SWWL03], the changes to the reference strategy were made in distinct steps. First, the DFT filter bank of ACE was replaced with an implementation of the extended Zwicker (EZ) model (the resulting strategy was termed EZ-ACE). Then, only the envelope extraction module of ACE was substituted with the Meddis IHC model (this implementation was called IHC-ACE). Finally, both the filter bank and the envelope extractor were replaced with their bio-inspired counterparts (termed EZ-IHC-ACE). A comparison of the envelope data $a(\nu)$ for the short utterance "aka" calculated via the four different strategies is pre-

sented in Figure 3.6. The effect of the IHC model alone seems to be subtle. However, when combined with the BM model, the envelope data becomes quite different from that of ACE.

The new strategies were tested with CI listeners using speech material from the HSM [HSMS97] sentence test with added speech-shaped noise (15 dB SNR). Each of the new strategies was tested on three subjects, the task of which was to repeat as much of each sentence as possible.

While the small number of subjects did not allow for detailed statistics, the main result was that every subject achieved higher scores with the commercial ACE strategy. Averaged over the three subjects per new strategy, EZ-ACE, IHC-ACE, and EZ-IHC-ACE performed about 15 %, 8 %, and 26 % worse than the reference strategy. The authors hypothesized that the disillusioning results originated from the fact that the tested CI recipients were already used to the ACE strategy. It was speculated that with a longer accommodation period the new algorithms could have scored better.

During the development of SAM, we have often recalled lessons learned with EZ-ACE, IHC-ACE, and EZ-IHC-ACE. In the aftermath, it seems we have disregarded some issues during the 2007 study. First of all, the resolution of the auditory model output was limited by the channel stimulation rate, which probably destroyed a considerable amount of temporal pitch cues and fine structure (see Figure 3.6). Second, the "Sampling & Selection" and "Mapping" parts of the ACE coder (cf. Figure 2.6 on p. 16) were left unmodified. As a consequence, temporal cues were further decimated. Finally, the parameters of the BM and IHC models could have been tuned more exhaustively.

3.2.4 Remarks

Until now, only a handful of projects have tried to integrate complex auditory models into cochlear implants. Despite some encouraging results, a real breakthrough has not been achieved. Notwithstanding, it is important to mention that no study based on this concept has allowed for a sufficiently long habituation period before evaluating user performance with a novel strategy. Ideally, all new processing schemes should be implemented in portable processors allowing CI users to gather experience through daily use before benchmarking.

3.3 Summary

Cochlear implants have become the most successful neural prostheses with more than two hundred thousand patients implanted worldwide. The success of CIs has also been supported by a large number of rehabilitation centers, in which implantees train to make the best of their system. It has been shown that the speech reception performance of CI listeners increases rap-

idly during the first year of CI use and it plateaus after the second to third year post-
implantation. Even then, most CI users have trouble understanding speech in noisy or rever-
berated environments and only very few of them (re-)develop a sense of music.

Quite a number of ideas have been published on how to improve cochlear implants. One
possibility is to enhance CI signal processing. The most straightforward approach is to augment
and further develop strategies that are already successful. A recent example could be the incor-
poration of psychoacoustic masking effects into ACE. Yet, the resulting MP3000 strategy has
not (yet) been proven to provide benefits to CI listeners (except for the extended battery life).
This example shows that changes that are more essential may be needed to achieve consider-
able improvements in CI user performance. One such possibility is the substitution of sophisti-
cated models of the human auditory system for the simple processing blocks of current CI
strategies.

The idea of using auditory models in auditory prostheses is as attractive as obvious. Even so,
up to today, there have only been a handful of attempts to embody such a system. This may be
due to the bottom-up way of thinking, which suggests that a problem of this complexity would
go far beyond the possibilities of current CI hardware. The studies so far have reported en-
couraging but no ground-breaking outcomes.

The SAM strategy, which is the focus of this thesis, ventures another try with auditory mod-
els. The major difference is that SAM is not based on an existing strategy, so that every part of
it is tailored to the chosen auditory models. Through that, SAM is expected to reproduce the
features of natural signal processing of the human ear more accurately than its predecessors do.

Chapter 4

Development of the SAM signal processing strategy

The systems and methods presented throughout the previous chapters indicate the huge advancement that cochlear implants have undergone since Volta's first experiment. Current CIs provide a considerable increase in quality of life when looking at the daily routine of a previously deaf patient. At the same time, current cochlear implant systems seem to have reached their peak potential (cf. Figure 2.8 on page 22), which signals the need for better algorithms and/or technologies.

There have been many attempts to enhance signal processing strategies (see section 2.2.4 and Chapter 3), but a real breakthrough is yet to be achieved. Having said that, the idea of designing a new strategy from scratch, having very little in common with contemporary strategies, may indeed be worth trying. Should the tests with cochlear implant users show remarkable benefits, their causes can be probed and conventional strategies may eventually be enhanced accordingly. Seen from this perspective, SAM is not meant to be a competitor to current strategies in the first place. It should rather be a proof of concept acting as a guide for further research and development.

Notwithstanding, we defined some important requirements at the beginning of the design process of SAM. These are reviewed in section 4.1. Section 4.2 presents the basic structure of SAM, whereas sections 4.3 and 4.4 describe the strategy in detail. Section 4.5 reveals some details about the reference implementation and about the safety measures that have been taken to make the pilot study safe for the implantees. Section 4.6 explains a possible way of fitting the SAM strategy to a CI user. Subsequently, the chapter ends with a summary.

Since I did not do the whole implementation work alone, I will use the personal pronoun *we* whenever referring to the collective work of my colleagues and I.

4.1 Requirements and design considerations

The two main concepts considered during the design process of SAM (beyond being safe for its users) were modularity and adaptability. The concept of modularity is meant to keep every part of the system exchangeable, so that, for example, parts of the auditory model can be re-

placed with newer or better ones. For that, a cornerstone is the separation of parts that are uniformly parameterized from those that are uniquely fitted for all CI users.

Adaptability, on the other hand, means that the strategy should be suited for all implant and electrode configurations, or at least be easily adapted to any new system, while maintaining error-free functioning. (The first cochlear implant system on which we intended to test SAM was a Nucleus Freedom with a Contour Advance electrode array from Cochlear. Details follow in the next sections.)

Beyond the main concepts, we have laid down some concrete requirements:

(R1) *Safety*: the core components of the software, which communicate with the implants, must not make mistakes. At the same time, they must handle any error or malfunction that can possibly be detected. Input data and configuration need to be checked for validity before use. Any output data must be double-checked for valid range before further processing. The stimulation must be interruptible at any time. Related to safety, the software must be as stable as possible (regardless of the activity on the user interface, the software must not terminate unexpectedly) and be responsive at all times (user must be informed about the progress of calculation, stimulation, etc.).

(R2) *Stability*: in no instance may any input sound corresponding to the normal hearing range of 0 to 120 dB SPL induce instability (i. e., make the software get into or stuck in an unexpected state). This is particularly important for the simulation of the auditory model, because it is a composition of more nonlinear systems, which may yield unexpected states with certain configurations for extreme input data.

(R3) *Features*: the auditory model currently used in SAM features many attributes (including psychoacoustic properties) of the intact human auditory system. To satisfy expectations, the CI stimulation patterns generated by the SAM strategy must also exhibit a number of these properties. SAM should at least be able to transmit a fair amount of F0 information and resemble human-specific traveling wave delays.

(R4) *Speed*: the implementation must be fast enough on a contemporary PC to run in real-time (require less time for calculation than the duration of the input sound). The implementation must also support processing of the input sound on a block-by-block basis, which is crucial for continuous operation.

4.2 System overview

In accordance with the design considerations presented in section 4.1, SAM can be broken up into two distinct parts: auditory model (also referred to as the auditory filter bank) and coder (or stimulus coder), as shown in Figure 4.1. The auditory model is generalized and it provides the bio-inspired spectro-temporal representation of the input signal. The coder, on the other hand, is designed to be custom-fitted for each individual CI user.

The signal path shown in Figure 4.1 can be summarized as follows. The outer and middle ear filter facilitates a pre-emphasis of the input signal. The model of the nonlinear mechanical filtering consists of the basilar membrane and the outer hair cells. The connected inner hair cells rectify the decomposed signal and introduce adaptation. The coder reduces the data rate, maps frequencies and amplitudes to fit the user's needs, and provides the stimuli with stochastic properties to reduce periodic characteristic. Symbols for input, output, and intermediate signals (as referred to in the rest of this thesis) are shown in italics.
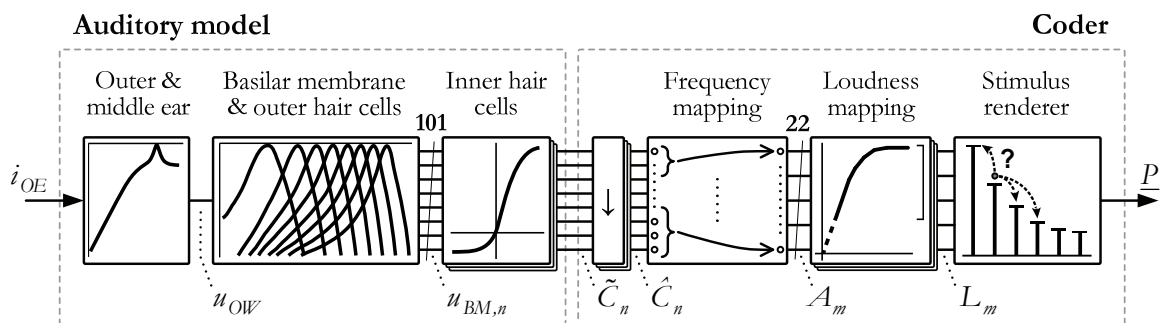
Figure 4.1: System overview and signal path in SAM.

4.3 Auditory filter bank

It was a key issue in the design of SAM to find the appropriate modeling level to be employed in the auditory filter bank. On one hand, modeling only the very first steps of the auditory processing could result in losing psychoacoustic properties. On the other hand, modeling too much could yield a representation of the data that is too abstract while wasting computational power. The following subsection elaborates on this problem and presents a simulation study that might help with the conclusion. The auditory model combined from several submodels and eventually used in the current implementation of SAM is presented in section 4.3.2.

Meddis et al. have published a note in their book on the use of auditory models: "Models serve many different purposes, and it is important to match the level of detail to the purpose in hand." [MFLP10: 10]. Driven by the same motivation and keeping the SAM design considera-

tions in mind, the auditory model has been customized to best meet requirements. Some note-worthy steps of customization and tuning are explained in section 4.3.3.

4.3.1   Adequate modeling level

In the electrically stimulated cochlea, the afferent dendrites of the spiral ganglion cells are stimulated directly by the electrical charges delivered by the implant electrodes. In the healthy cochlea, these afferents form a glutamate synapse with the inner hair cells: Action potentials arise whenever the neurotransmitter concentration in the synaptic cleft raises above a critical limit (unless the afferent fibers are in a refractory state). These facts lead to the recognition that the optimal CI filter bank should mimic the workings of the human ear up to the concentration of the released neurotransmitter substance in the synaptic cleft. Furthermore, the magnitude of the electrical stimulation should correlate with the calculated concentration values.

The SAM filter bank was built with the above theory kept in mind. In the preparation phase of the SAM development a computational study was carried out to verify if the proposed mod-eling level also meets practical requirements. That study is described in more detail in the fol-lowing section.

4.3.1.1   Comparative evaluation of successive cochlea modeling stages

This section recapitulates methods and results of the simulation study published in [HNSK07]. The aim of that study was to clarify which modeling depth of the auditory system would be most appropriate as an auditory filter bank for the SAM strategy from a practical point of view. Based on hidden Markov models (HMMs), an automatic speech recognition (ASR) framework was built. Using discrete cosine transform (DCT) the dimensionality of the auditory model output was reduced (251 channels to 13 DCT coefficients). Then, first and second order re-gression terms were calculated and appended to the DCT coefficients to form the final feature vectors. Finally, these vectors were routed into the HMM-based ASR framework. An overview of the whole system is presented in Figure 4.2.
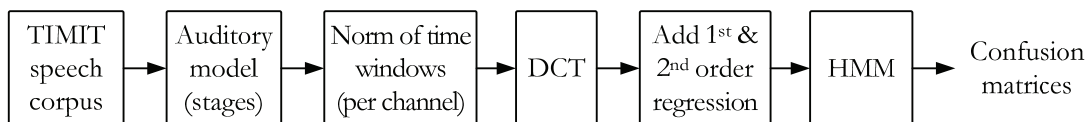
Figure 4.2: Overview of the system and methods used in the study.

Auditory model used in the preliminary study

The first step in the study was to construct a layered auditory model, which provided output from as many meaningful intermediate modeling stages as possible. The Baumgarte outer and

middle ear model along with his BM model [Bau00] was coupled with the models of the IHC-AN[14] complex of Meddis [Med86] and Sumner et al. [SLOM02]. An overview of the modeling stages is presented in Figure 4.3. The employed submodels are described in more detail in section 4.3.2.
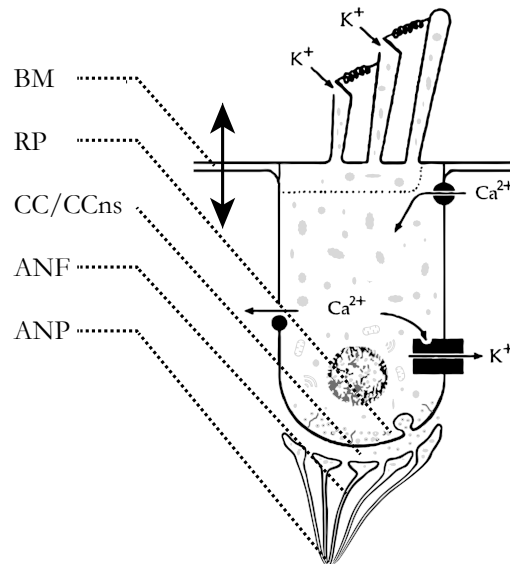
Figure 4.3: Cochlea modeling stages presented on a schematic view of the IHC-AN complex. The abbreviations are defined as follows. **BM**: basilar membrane motion (axial displacement of the depicted cell), **RP**: neurotransmitter release probability, **CC**: concentration of the neurotransmitter substance in the synaptic cleft, **CCns**: CC calculated in a non-stochastic way, **ANF**: spike activity of an auditory nerve fiber, **ANP**: summarized activity of a population of auditory nerve fibers. Adapted from [HNSK07], based on [Bau00, Zbo97].

Please note that the auditory nerve fibers of the ANF stage have been tuned as high spontaneous rate fibers. ANP was modeled as the mean activity of 12 auditory nerve fibers consisting of 6 HSR, 4 MSR, and 2 LSR (high/medium/low spontaneous rate, respectively) fibers. For details on differently tuned nerve fibers, see [SLOM02, SLOM03].
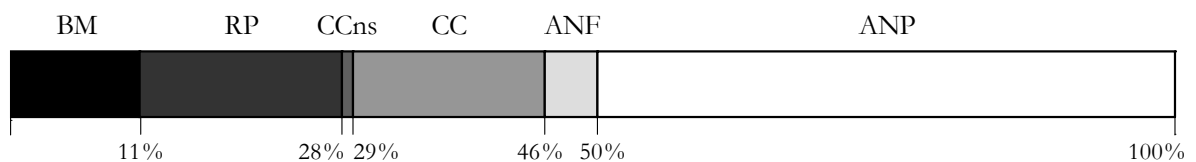
Figure 4.4: Relative computational times to perform simulation up to each modeling stage.

---

[14] Auditory Nerve

We created a computer program simulating the layered auditory model in the time domain. Figure 4.4 presents how the relative computational times for each cochlear modeling stage add up.

Speech corpus

Recognition experiments were conducted on a subset of the DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus [LKS86, ZSG90]. The subset included 768 sentences (576 training + 192 test sentences) and had a total duration of over 38 minutes. No speaker appeared in both the training and the test sets.

Design of the ASR experiments

The continuous speech recognition system was built with Cambridge University's Hidden Markov Model Toolkit (HTK) [You94]. For the hidden Markov modeling, left-to-right HMMs were trained for each of the 61 TIMIT monophones (see [GLF+93: 29-31]), consisting of three states each. States were modeled as Gaussians. Based on the calculated features and the given original phonetic transcriptions of the training material, the system was designed to carry out 19 rounds of re-estimation of the HMM parameters using HTK's embedded Baum-Welch algorithm. After the 6th, 9th and 12th iterations of the Baum-Welch re-estimation, the Gaussians were split into mixtures of two Gaussians yielding a total of eight Gaussian mixtures per state by the end of the training. State initialization was done by linearly segmenting the utterances (*flat init*). No advanced grammar model (such as *bi-gram*, or *tri-phone*) was used; recognition of phones was therefore solely based on the actual feature vectors.

All recognition tasks were run several times, simulating different noise conditions. The signal to noise ratios were 96 dB (no added noise), 24 dB, 18 dB, 12 dB, 9 dB, 6 dB, and 0 dB. The added noise was pseudorandom with a period long enough to avoid recurrence of noise patterns.

Each noisy setup was evaluated under three different training conditions. These were called clean, mixed, and dirty training. In the case of clean training the HMMs were only trained with clean (96 dB SNR) data, while during dirty training the training material only consisted of noisy data. In the mixed case, training was carried out using 50 % noisy and 50 % clean data. The total duration of the training material was invariant among the three conditions. Independently of the actual training variant, recognition tests were always accomplished exclusively on data corresponding to the actual SNR. Only the main outcomes related to the development of the SAM strategy are presented in this thesis. For the full results, please see [HNSK07].

Results and conclusions

The purpose of this study was not to build an ultimate ASR system, but to show effects of employing consecutive levels of auditory modeling. The effect of noise was analyzed systematically. We have found that higher-level auditory modeling can indeed be an advantage in terms of noise-robustness. Furthermore, the auditory modeling stage of CCns (non-stochastically calculated neurotransmitter concentration in the synaptic cleft) seems to fit well as a base for an auditory filter bank for cochlear implants.

The most important trends in word recognition rate vs. SNR are presented in Figure 4.5 and Figure 4.6. In the HTK terminology, a word is said to be recognized *correctly* if all of its phonemes have been recognized in the appropriate order. Insertion errors (i.e., additional phonemes between two correctly recognized ones) are ignored by this metric. (By contrast, word recognition *accuracy* does not allow insertion errors.)

Figure 4.5 shows percentage of correctly recognized words as a function of the signal to noise ratio for various cochlear modeling stages, after clean training of the recognizer. The latter means that the recognizer has never encountered noisy utterances prior to testing. With other words, these recognition results are hypothesized to correlate to the performance of a listener in an unexpected noise condition. Even though the BM modeling stage outperforms CCns in the 96 dB SNR condition, CCns performs best at all other SNRs.

Figure 4.6 shows word recognition rates after mixed training. This shall correlate to the situation where a listener is used to listening in a specific type of noise. Here, the BM stage would suffice to yield good results in all cases. However, it is most unlikely that a listener would be used to every kind of noise that may appear. Since the CCns stage yields approximately the same results as BM for any noisy condition, it remains attractive as a final choice.

Taking the required computational power (see Figure 4.4) into account, and considering the theory about the electrode-tissue interface (presented in section 4.3.1 on p. 44), the CCns stage seems to be the best alternative for the auditory filter bank of the SAM strategy.

A final note on the relatively low magnitude of the overall recognition rates presented in the figures below: please keep in mind that a large phonetic set (61 phones) was used without grammar in an open-set recognition task.
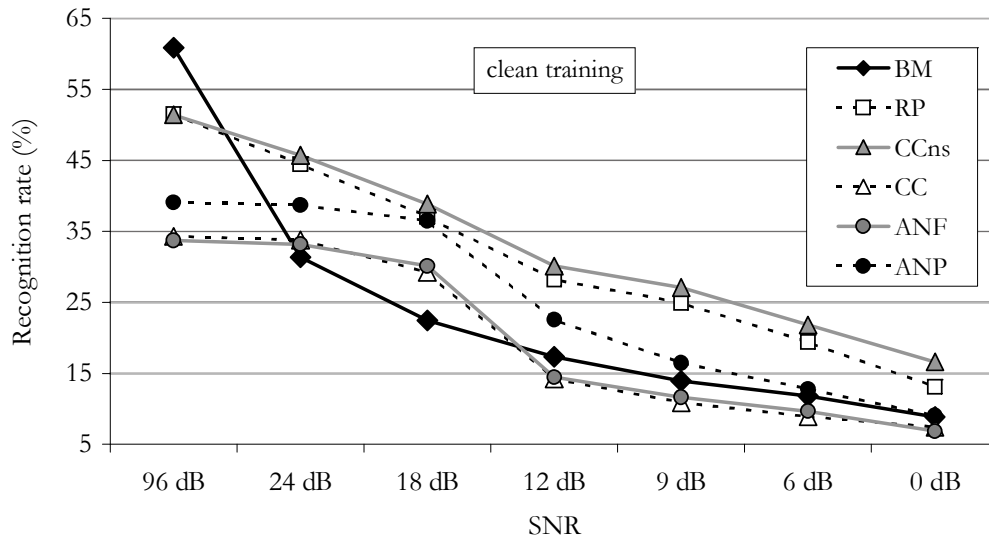
Figure 4.5: Percentage of correctly recognized words as a function of the signal to noise ratio for various cochlear modeling stages, after clean training of the recognizer. Adapted from [HNSK07].
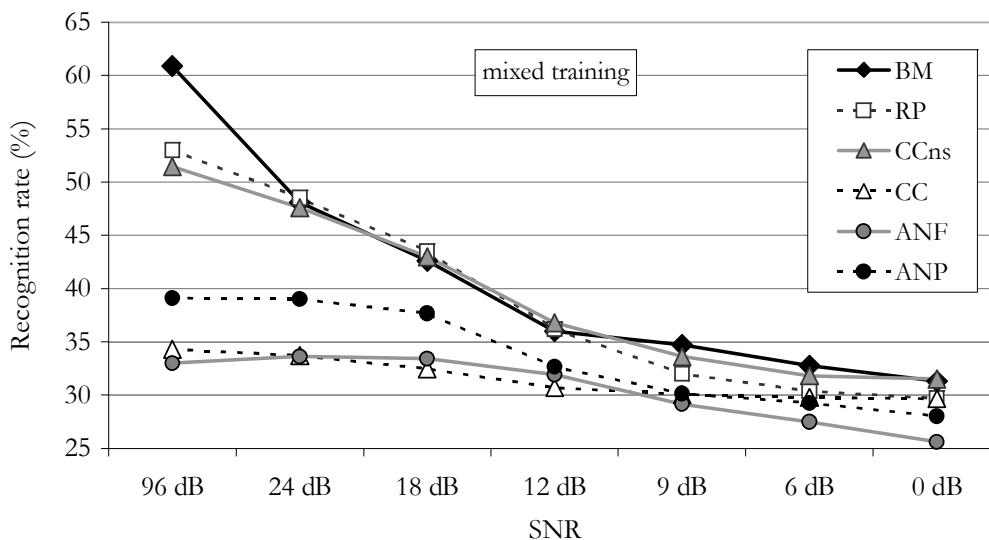
Figure 4.6: Percentage of correctly recognized words as a function of the signal to noise ratio for various cochlear modeling stages, after mixed training of the recognizer. Adapted from [HNSK07]. Note that the scaling of the Y axis has been changed (in contrast to Figure 4.5) to help distuingish symbols.

4.3.2 The compound auditory model

This section describes the auditory model (AM) of SAM, which is the cornerstone of the presented strategy. A large part of the contents shown here has already been published in [HCH12]. The reuse was permitted by the coauthors.

The general structure of the AM is shown on the left-hand side of Figure 4.1. 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. The output of this stage is processed further by the model of mechanoelectrical transduction, which mimics the physiological properties of inner hair cells including the simulation of receptor potential, calcium kinetics and transmitter dynamics. Each band of the NLM, i.e., each simulated BM section, is connected to one simulated inner hair cell.

Most model components can be specified by differential equations, which are internally represented by equivalent electrical circuits using velocity-voltage 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 [Fet86].

4.3.2.1 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 $3\,\mathrm{kHz}$, while the damping is increasing for frequencies below that, see Figure 4.7b. For frequencies above $3\,\mathrm{kHz}$, constant magnitude response is assumed. This functionality can be described by an electrical network [Bau00] as illustrated in Figure 4.7a. Here, sound pressure at the outer ear is given by the current i_{OE}.
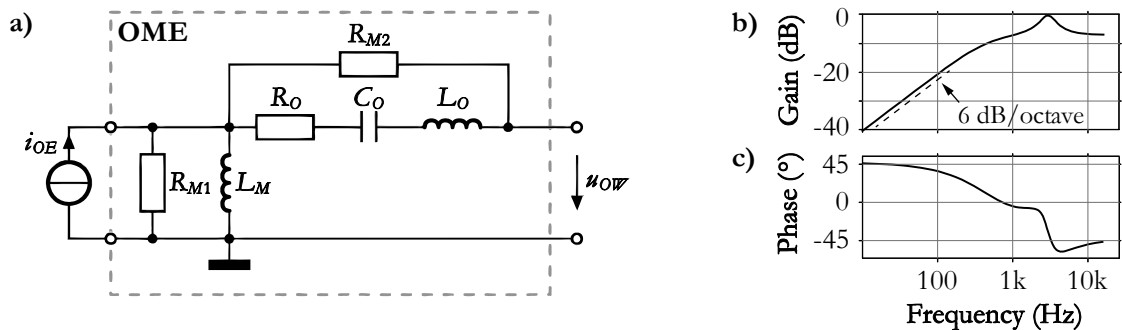
Figure 4.7: Electrical model of the outer and middle ear (a) and its estimated amplitude (b) and phase characteristics (c). Adapted from [Bau00].

The resonance is realized by a series RLC circuit, whereas the damping of lower frequencies is incorporated by R_{M1} and L_M. The velocity of the oval window is given by the output voltage u_{OW}.

Baumgarte states in his dissertation [Bau00:39] that the phase characteristic (shown in Figure 4.7c) was not taken into account during the modeling process. Since model stages of the nonlinear mechanical filtering (see section 4.3.2.2) also introduce frequency-dependent delays, it is not necessary to have a linear phase response here, but it is favorable to calibrate the phase characteristic of the full system (as described in section 4.3.3.3).

4.3.2.2 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 Peisl [ZP90] with modifications from Baumgarte (see [Bau99] and [Bau00]) has been employed (for the statement of reasons see section 6.1.2 of the discussion). 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, in turn, can be represented by an electrical circuit as shown in Figure 4.8.

Each section of the passive hydromechanical model consists of a parallel resonant circuit. RLC parameters express the dimensions and material properties of the fluid-filled cochlear duct. L_n, C_n and R_n 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 Figure 4.8. Model parameters are based on physiological data from the scientific literature.

As already stated above, the passive model is extended by a model of the outer hair cells. As shown in Figure 4.8, the velocity of the cochlear partition is amplified within a nonlinear feedback loop using a voltage controlled voltage source ($g_1 u_{c,n}$) 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 $u_{c,n}$ of the amplifier is fed back to the passive model and in addition, neighboring OHC models are coupled via resistors (R_{la} and R_{lb}). 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 ($g_2 i_{c,n}$) and of a parallel resonance circuit (labeled PRLC), which causes a frequency dependent voltage drop. The SAS was implemented by Baumgarte to ensure the high amplification featured by the OHCs of the human ear while preserving stability of the model [Bau00]. The output ($u_{BM,n}$) of the SAS corresponds to the modeled basilar membrane velocity.
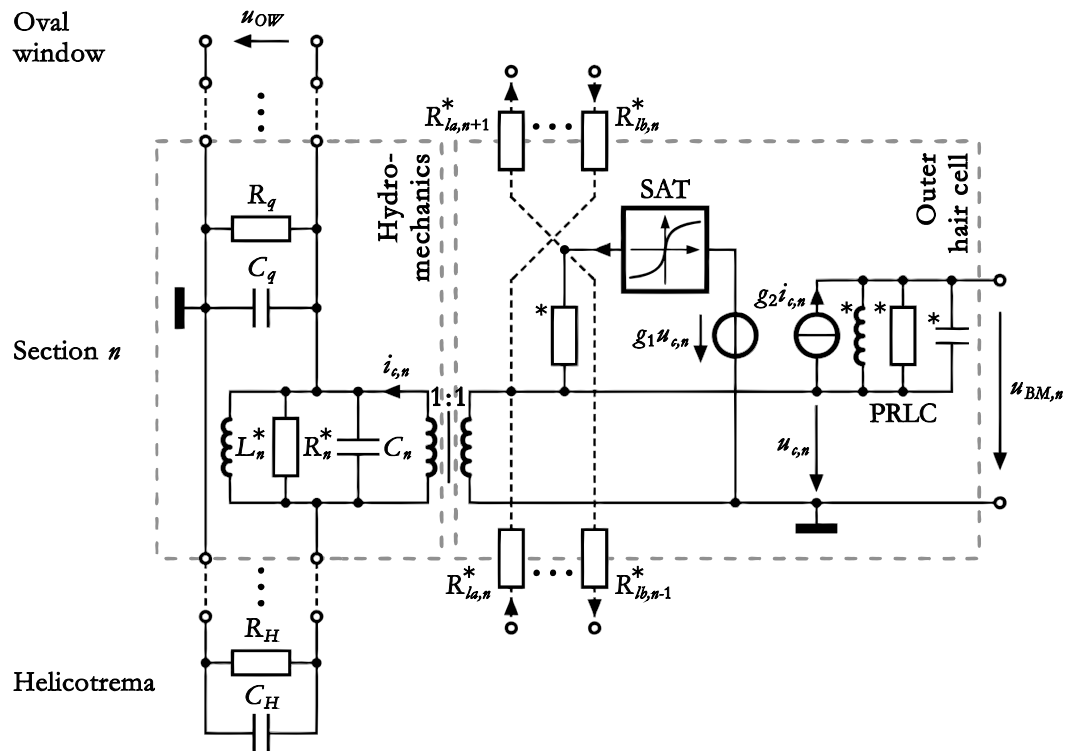
Figure 4.8: Electrical model of the nonlinear mechanical filtering. Adapted from [Bau99]. The values of electrical components, except for those marked with an asterisk (*), are listed in Table A.1 starting on page 131. The values of the marked components can be calculated for each cochlear section via the equations given in [Bau00].

4.3.2.3 Mechanoelectrical transduction

The calculated basilar membrane velocity along the cochlear partition forms the input to the model of mechanoelectrical transduction. To mimic the functionality of the inner hair cells, the model described by Sumner et al. [SLOM02] was adapted and employed. Sumner et al. extended an earlier model of Meddis [Med86] by combining the neurotransmitter dynamics with a modified biophysical model of the receptor potential [SCW+86] and with calcium driven mechanism for the neurotransmitter release (see [HL88] and [KW90]) as shown in Figure 4.9.

In the biophysical model (upper part of Figure 4.9) 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 Figure 4.9). 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 en-

ter 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 $k(t)$.
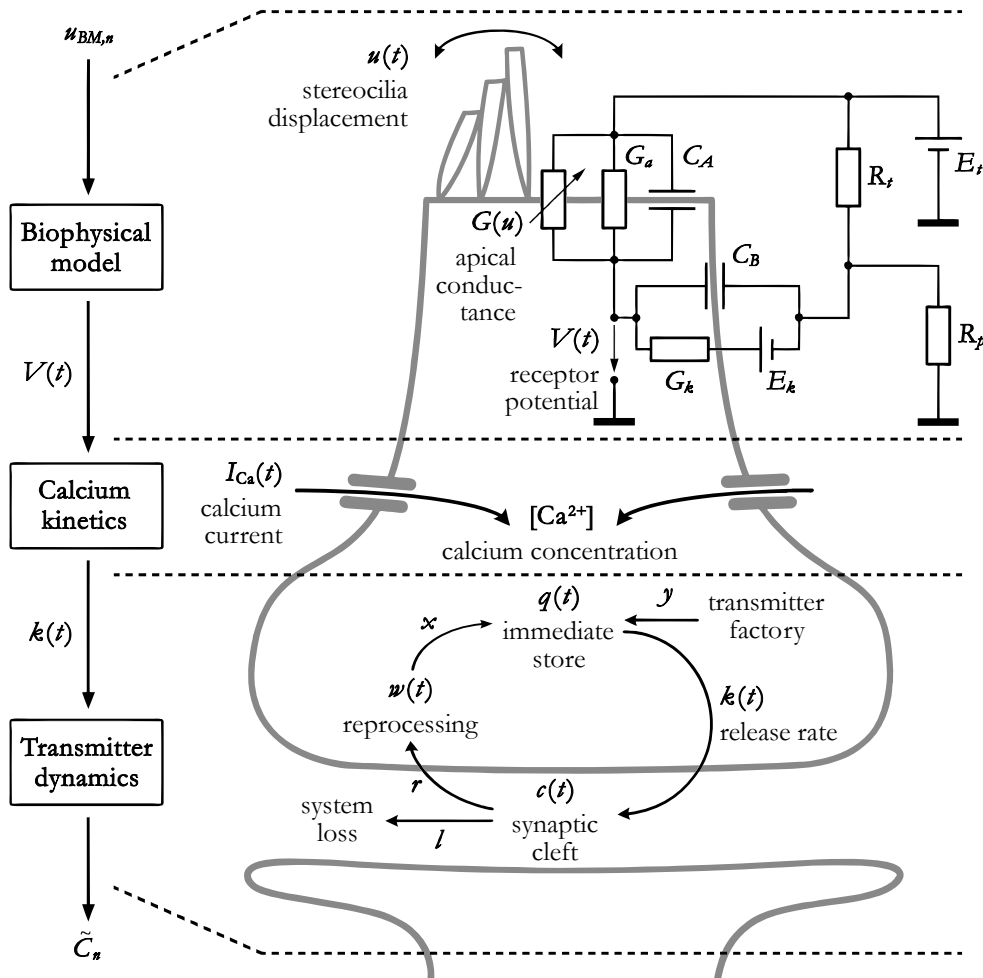
Figure 4.9: Overview of the mechanoelectrical transduction. Left: simplified signal path. Right: more detailed schematical view of the operation of an inner hair cell. Adapted from [SLOM02]. The values of electrical components are listed in Table A.1 starting on page 131.

The transmitter release rate, on the other hand, drives a model of transmitter dynamics (lower part of Figure 4.9). Instead of the stochastic version described by Sumner [SLOM02] we use the original Meddis transmitter recycling model [Med86] 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 l). The remaining transmitter is taken back (at a rate of r) into the reprocessing store and returned into the immediate store (at a rate of x). Neurotransmitter loss is counterbalanced by the factory, which produces new transmitter within the cell (at a rate of y). For any time step i of the discrete time simulation of the auditory model, $\tilde{C}_n[i]$ denotes the neurotransmitter substance concentration near to the cell membrane exocytosed by the nth IHC.

The modeled mechanoelectrical 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 [SLOM03] and [YWR90]).

Functional characteristics of synaptic adaptation are simulated by the model of transmitter dynamics. In addition, the model also reproduces phase-locking[15], which is restricted to low frequencies. A review of psychoacoustic features of the auditory model is presented in section 5.1.3.

4.3.3   Customization and tuning

The compound auditory model was customized and tuned prior to the use with the new CI signal processing strategy. Customization work manifested itself in (down)scaling the model without losing important properties and without adding artifacts. Details follow in the next section. Tuning consisted of the following issues: (a) finding coupling factors between model stages to ensure correct operating ranges of the submodels, (b) tweaking the magnitude response nonlinearity of the modeled active basilar membrane, and (c) revising group delays. Tuning work is explained through sections 4.3.3.2 and 4.3.3.3.

4.3.3.1 Practical considerations

One key question at the beginning of the development of SAM was if the simulation of the auditory model could be reasonably fast. Reasonably, in the first place, meant that the calculation should not take longer than the duration of the input sound. This was not given with the reference implementation provided by Baumgarte along with his thesis. Since simulation speed was crucial for applicability in cochlear implant stimulation, we incorporated some changes to the basic parameters of the active basilar membrane model of Baumgarte.

---

[15] Phase-locking is the ability of a neuron to fire action potentials that are time-locked to a periodic stimulus event.

As already stated in section 4.3.2, the BM is simulated in the time domain using Wave Digital Filters, so one critical parameter having great effect on speed is sampling rate (F_s). Because of inherent nonlinearities, Baumgarte recommends choosing a sampling rate about quadruple the highest signal frequency to avoid aliasing artifacts [Bau00:157]. The reference implementation employs 100000 sps for applications in music coding, but since cochlear implants usually do not deal with frequencies much higher than 8 to 10 kHz, we have chosen a sampling rate of 44100 sps 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 $\Delta z = 0.25$ Bark, which makes a mean distance of 0.34 mm between adjacent sections along the modeled cochlea. We have found spatial discretization in this case still to be fine enough to avoid reflections at section boundaries, which would lead to undesired artifacts by standing waves.

Reducing the number of BM sections also means to simulate less IHCs, which leads to a further gain in speed. In addition, as already stated in section 4.3.2.3, we use the original Meddis transmitter recycling model [Med86] instead of the stochastic version described by Sumner [SLOM02], which also reduces computational costs.

Last, but not least, several iterations of profiling and optimizing the source code combined with the use of advanced compiler options also contributed to better performance of the final executables.

4.3.3.2 Nonlinearity of the magnitude response

First tests with the Baumgarte model indicated a large deviation of the magnitude response nonlinearities from the expected characteristics. It turned out that the default parameter set specified in [Bau00] puts the modeled basilar membrane into the wrong operating range. To overcome this problem, we have introduced pre- and post-amplification factors (g_{pre} and g_{post}, respectively, see Table A.1 on page 131), with which the BM's working range can be tuned.

Once the working range was fixed, the saturating characteristic of the amplifier stage was calibrated. For that, the parameters of the symmetrical saturating function (within the feedback loop of the nonlinear mechanical model, labeled SAT in Figure 4.8) were changed. Baumgarte defines the SAT function between input voltage $u_{1,n} = g_1 u_{c,n}$ (cf. Figure 4.8) and output voltage $u_{sat,n}$ as shown in Equation 4.1.

$$u_{sat,n} = \frac{u_{1,n}}{\sqrt{1 + g_{s1} \cdot |u_{1,n}| + g_{s2} \cdot u_{1,n}^2}}.$$ (4.1)

Originally, the saturation factors g_{s1} and g_{s2} equaled to 20.0 and 10.0, respectively. To best fit model characteristic to experimental data published in [JPY86], these factors have been changed to 5.0 and 200.0, respectively. A comparison of uncalibrated vs. calibrated model characteristics as well as a plot of experimental data is presented in Figure 4.10.
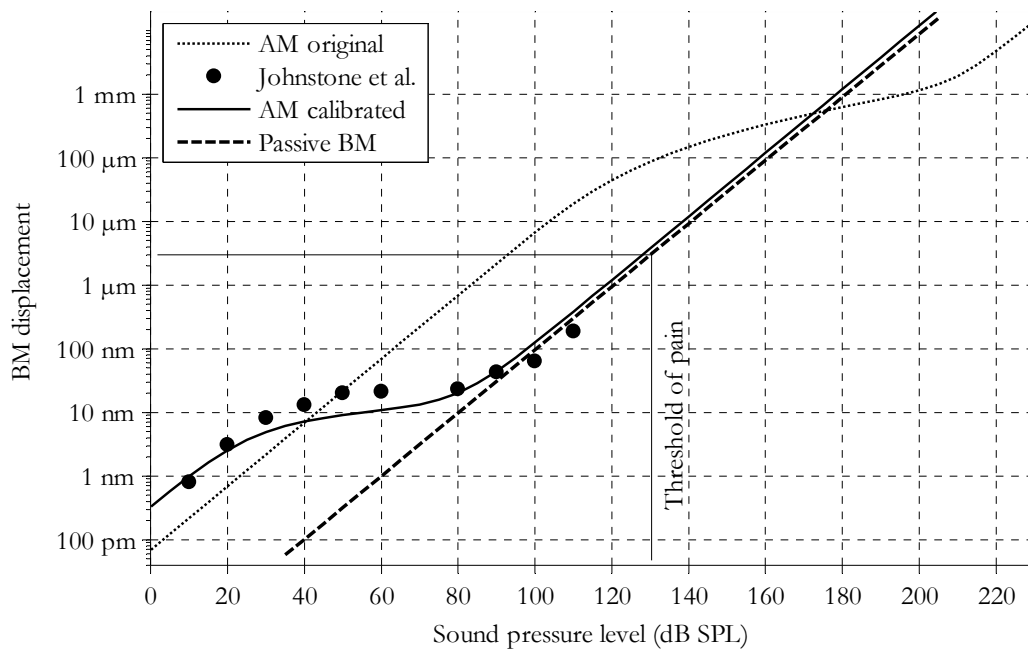
Figure 4.10: Input/output functions of the mechanical filtering stage of the auditory model. Data points are measured at sites of maximum displacement of the basilar membrane for 1 kHz pure tones at given levels.

Figure 4.10 illustrates the effect of calibration well. With the original parameters, the nonlinear effects first show up beyond the threshold of pain, which is not plausible. By fixing the working range and the saturation characteristic, there is a good correlation with experimental measurement data from [JPY86]. For reference, the linear characteristic of a passive (i. e., dead) basilar membrane is also plotted in the figure.

The magnitude response of the calibrated model has also been tested at BM positions not correlating to the frequency of the pure tone stimulus, see Figure 4.11. Results indicate good match with outcomes of e. g. [NDA91] and [OP97.]
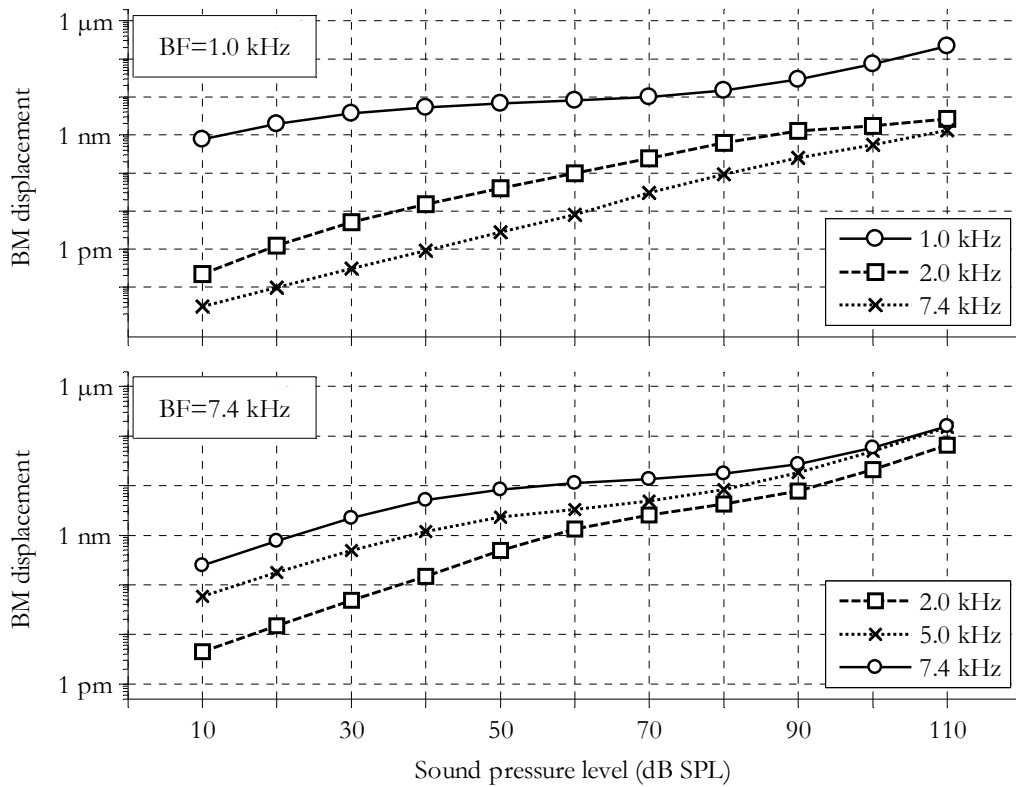
Figure 4.11: Input/output functions of the mechanical filtering stage of the auditory model. Data points are measured at BM-sites corresponding to the best frequencies (BF) shown in the top-left corners. Legends show frequencies of the pure tone input sounds.

4.3.3.3 Group delay

Sounds reaching the ear induce vibrations on the basilar membrane. Any wave on the BM proceeds from the base towards the apex, reaching maximum amplitude at the place corresponding to its characteristic frequency. Beyond that point, the wave damps out quickly. The velocity of the traveling wave decreases while getting closer to the apex. The *group delay* describes the characteristic travel time from base to the place of maximum displacement as a function of the frequency of the pure tone input.

Elberling et al. published a summary on group delay estimates in 2007 [EDCS07]. They compared the group delays given by Eggermont, De Boer, Don et al. and Neely et al., which had been determined by various measurement methods in humans (for references, please see the paper of Elberling et al.). For a visual comparison, see Figure 4.12.
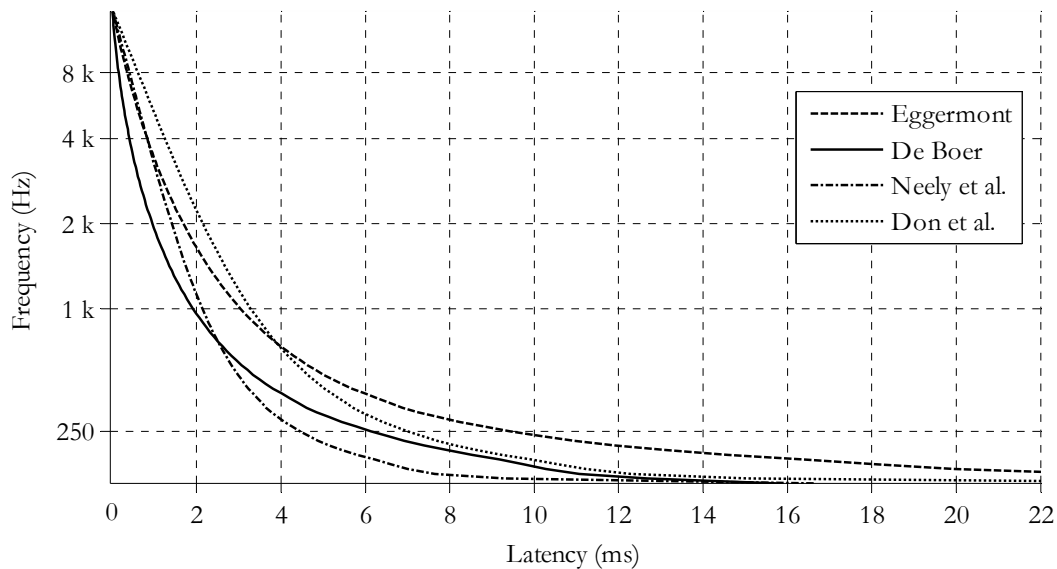
Figure 4.12: Human cochlear traveling wave delay approximations based on measure-
ments by Eggermont, De Boer, Don et al. and Neely et al. [EDCS07].

Although the four approximation functions exhibit great differences, our tests have shown that
the basilar membrane model of Baumgarte [Bau00] operates with longer latencies than any
measurement listed in [EDCS07] would suggest. Because the group delays have been captured
via different methods, it is not known which of them describes the traveling wave latency in the
human cochlea with the greatest precision. As suggested by Chilian [Chi10a], we use the mean
of the measured data as a reference for tuning the delay characteristic of the Baumgarte model.

As already stated on page 25 in section 2.3.1.1, the Baumgarte basilar membrane model is
based on that of Peisl [Pei90] and provides a Wave Digital Filter implementation to allow for
simulation in the time domain. The BM is modeled by multiply connected adjacent sections,
which mimic hydromechanics and the amplifying effects of the outer hair cells of the corre-
sponding cochlear segment. In this model, the mass of the longitudinal moving lymph liquid
and the combined mass of the basilar membrane and the axially moving lymph liquid, in each
case for a given section, are represented by the capacitance C_q and C_n, respectively. Their values
are calculated as given in Equations 4.2 and 4.3.

$$C_q = 12.5\,\text{nF} \cdot \frac{\Delta z}{0.1\,\text{Bark}} \tag{4.2}$$

$$C_n = \left(255\,\text{nF} + \frac{335\,\text{nF}}{24\,\text{Bark}} \cdot z_n \right) \cdot \frac{0.1\,\text{Bark}}{\Delta z} \tag{4.3}$$

One possibility to change the delay characteristic of the BM model (as documented by Chilian
[Chi10a]) is to change C_q and C_n. To yield a shorter group delay, C_q must be decreased while C_n

must be increased. For that, two scaling factors have been introduced, as shown in Equations 4.4 and 4.5.

$$C_q = g_{Cq} \cdot 12.5 \text{ nF} \cdot \frac{\Delta z}{0.1 \text{ Bark}} \tag{4.4}$$

$$C_n = g_{Cn} \cdot \left(255 \text{ nF} + \frac{335 \text{ nF}}{24 \text{ Bark}} \cdot z_n \right) \cdot \frac{0.1 \text{ Bark}}{\Delta z} \tag{4.5}$$

It should be noticed that if the scaling factors deviate too much from 1.0, instabilities on the simulated BM might occur. These can manifest in forms of standing waves or ringing (post-oscillation) at certain positions on the basilar membrane.
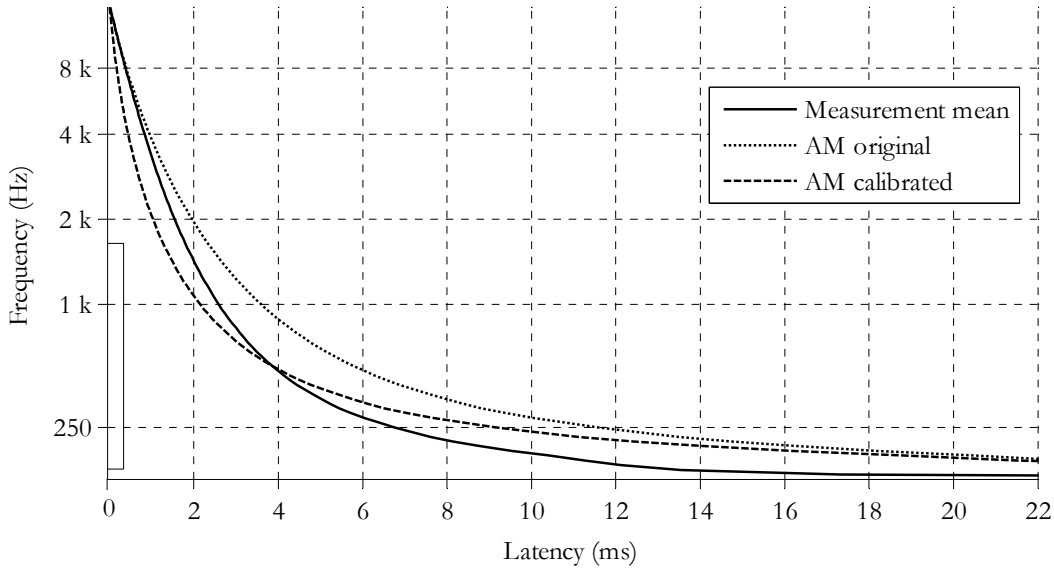
Figure 4.13: Traveling wave delay characteristics of both the original and the tuned Baumgarte model versus the mean of the delays based on measurements by Eggermont, De Boer, Don et al. and Neely et al. [EDCS07]. The frequency range marked on the y-axis indicates where the tuned BM model's delay characteristic is closer to the reference than that of the original model.

Chilian has found that the combination of g_{Cq}=0.6 and g_{Cn}=3.0 does not lead to such side effects while bringing the traveling wave delay characteristic nearer to the reference [Chi10a]. Results are shown in Figure 4.13.

The frequency range in which the tuned BM model's delay characteristic is closer to the reference than that of the original model correlates well with the range of phase-locking. Since timing is most important in this frequency range, outcomes of the delay tuning can be considered reasonable and useful.
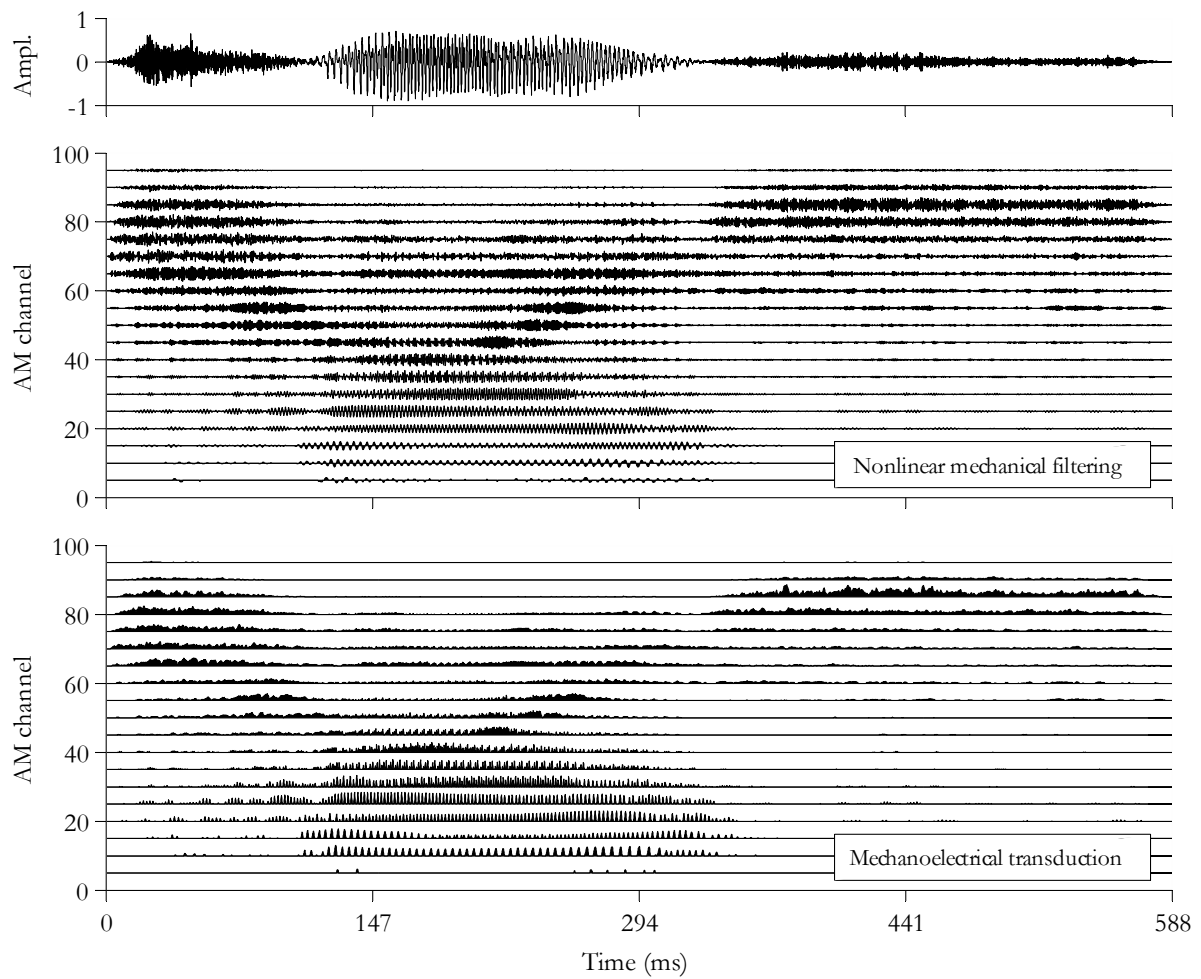
Figure 4.14: Response of the two main auditory model stages for the utterance "choice", presented at 65 dB SPL. Top: waveform of the utterance. Middle: output of the nonlinear mechanical filtering. Bottom: output of the mechanoelectrical transduction. For a clearer view, all plots are normalized and only every fifth auditory model channel is plotted.

Originally, the C_q and C_n values were determined by Oetinger and Hauser [OH61] for a highly idealized cochlea. These values were partially based on measurements by Békésy from the early 1940s, but they have already been changed by Lumer [Lum87a] empirically. In this context, using mass-scaling factors as high as 3.0 should not introduce remarkable model inaccuracies.

An example for the simulated response of the customized and tuned model of the nonlinear mechanical filtering and that of the mechanoelectrical transduction is given in Figure 4.14.

4.4   Stimulus coder

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 is calculated in SAM. An overview of the processing steps is shown on the right-hand side of Figure 4.1 on page 43. Coder parameters are summarized in Table A.2 on page 133.

4.4.1   Interfacing to the auditory filter bank

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 4.6 explains the decimation for a one-second piece of data. $\tilde{C}_n$ denotes the nth channel of the auditory model output, $\hat{C}_n$ is the decimated data and TPR is the total pulse rate.

$$
\begin{aligned}
\hat{C}_n[i] &= \max_{j = j_1, j_1 + 1, \ldots, j_2}\left(\tilde{C}_n[j]\right), i = 1,2,\ldots,TPR, \\
j_1 &= \left\lfloor f_s \cdot (i-1) / TPR + 1 \right\rfloor, \\
j_2 &= \left\lfloor f_s \cdot i / TPR \right\rfloor.
\end{aligned}
\tag{4.6}
$$

4.4.2   Frequency and amplitude 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 Figure 4.15, 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. [Cla03: 612-618, 623]), 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 A_m, $m=1,2,\ldots,M_{CI}$, where M_{CI} is the number of CI electrodes. A_m has the same data rate as $\hat{C}_n$. (Center frequencies of a typical electrode configuration are shown in the bottom part of Figure A.1 on page 139.)

To fit the huge dynamic range of the modeled ear into the operational range of the cochlear implant, a loudness mapping has also 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 [PBG06]. 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 $\vec{A}_m$. Frequency responses of the ACE and SAM filter banks are presented for comparison in Figure 4.15.
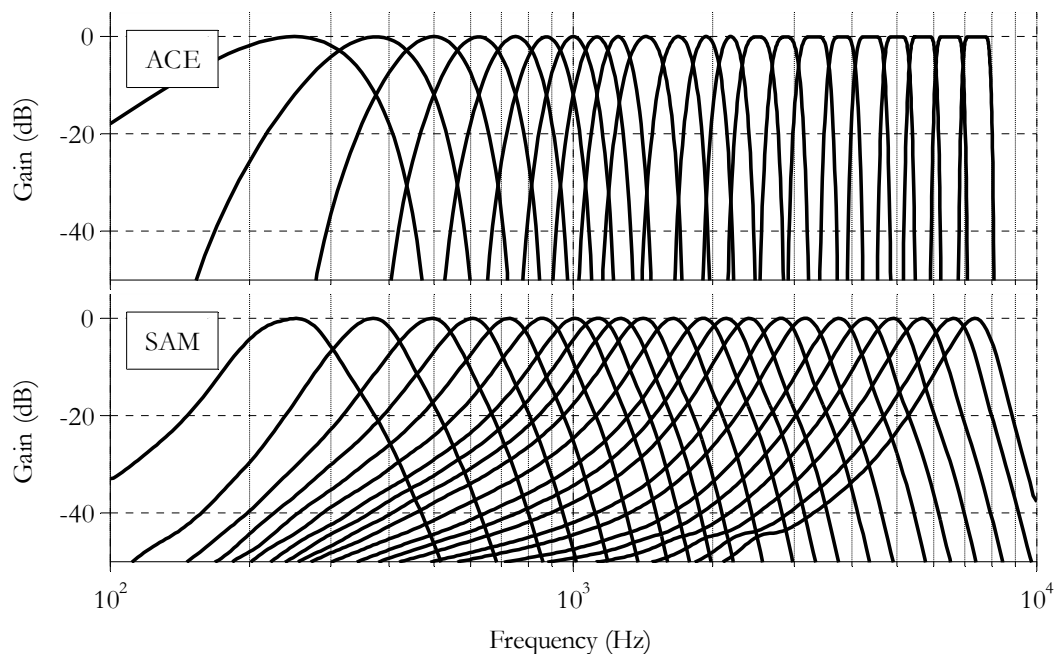
Figure 4.15: Normalized filter bank frequency responses showing the 22 electrode channels for ACE (top) and SAM with the same electrode frequency configuration. SAM response is measured at 55 dB SPL at the calculation stage of $\vec{A}_m$.

SAM does not use fixed channel stimulation rates by design (see also section 6.1.3 in the discussion on page 110), but it provides a mechanism to control the extent of 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"; 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 $R_m\to\infty$ 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 is applied. The LGF was introduced in cochlear implants with linear filter banks to better mimic the nonlinearities of loudness perception (see page 17 of section 2.2.4.1). Since the SAM strategy mimics the ear closer, the use of an LGF is not obligatory and is only required if the loudness growth should resemble other characteristic (like that of another CI strategy). LGF is applied as shown in Equation 4.7.

$$L_m[i] = \frac{\log\left(1 + c_m \cdot \hat{A}_m[i]\right)}{\log\left(1 + c_m\right)}. \tag{4.7}$$

In Equation 4.7, c_m is the curve-shaping factor, with which the steepness of the loudness growth can be varied. For $c_m\to 0$ the mapping is linear. For $c_m>0$ or $-1<c_m<0$ the effect is compression or expansion, respectively.

Since it was intended to test SAM on CI users with ACE experience, the default SAM configuration was set to use a curve-shaping factor of 1.95, with which the loudness growth was very similar to that of ACE (see also section 5.1.3.3).

4.4.3 Electrical 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 indices 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 Equation 4.8.

$$\begin{aligned}
P(M_1) &= 1 - \xi_p, \\
P(M_2) &= \xi_p \cdot (1 - \xi_p), \\
P(M_3) &= \xi_p^2.
\end{aligned} \tag{4.8}$$

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 repetition penalty is enabled, i. e., if $R_m>0$.

Next, the magnitude of the selected channel is translated into current level Y, as described in Equation 4.9. Y has no physical unit; it is used as an index to look up implant-typical currents in μA.

$$Y[i] = \left\lfloor 0.5 + THL_m + L_m[i] \cdot V_g \cdot \left(MCL_m - THL_m \right) \right\rfloor,$$
$$m \in \{M_1, M_2, M_3\}. \tag{4.9}$$

In Equation 4.9, THL_m, MCL_m and V_g denote the threshold level, most comfortable level and global volume ($0 \leq V_g \leq 1$), respectively. THL_m and MCL_m are determined for each channel for each CI user during the fitting sessions. The fitting procedure for the SAM strategy will be explained in section 4.6.

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 Equation 4.10.

$$I[i] = 17.5 \cdot 100^{Y[i]/255} \; \mu A. \tag{4.10}$$

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 [CR01] and sound source localization abilities [LM08], we have introduced ξ_j, with which the possibility is given to vary the phase gap duration of a biphasic stimulus in a stochastic manner, as shown in Equation 4.11. This random variation adds some irregularity to the stimulation signal, which reduces periodic characteristic, while preserving fine temporal structures.

$$t_{gap} \in \mathcal{U}\left(\left(1 - \xi_j \right) \cdot T_{pg}, \left(1 + \xi_j \right) \cdot T_{pg} \right). \tag{4.11}$$

In Equation 4.11, $\mathcal{U}$ denotes uniform distribution and T_{pg} is the standard phase gap value (typically around 8 μs). The phase gap duration used in the current stimulation cycle is denoted by t_{gap}. This value may also be limited by the CI hardware used. If ξ_j is zero, then $t_{gap} = T_{pg}$ holds for the whole duration of the stimulation.

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 $\underline{P}$.
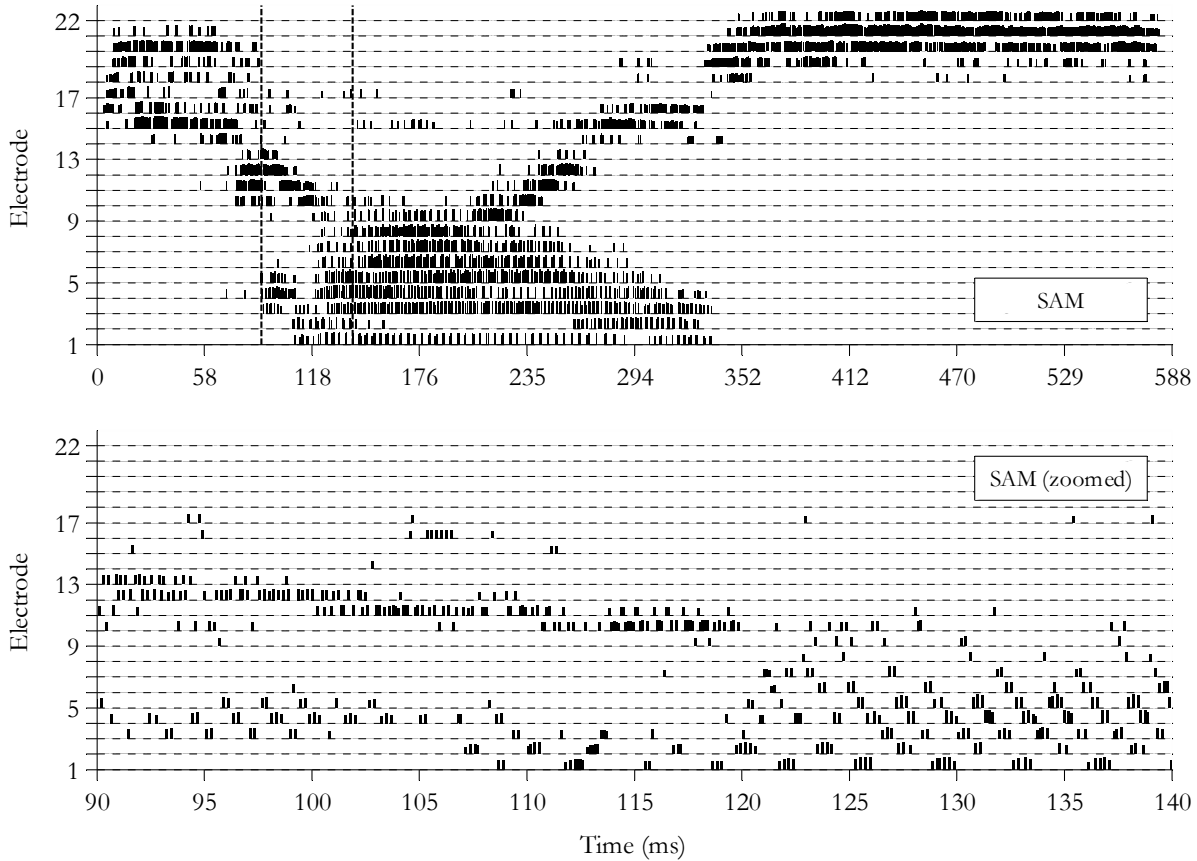
Figure 4.16: SAM stimulation pattern for the utterance "choice" (presented at 65 dB SPL) generated using the default parameters (as shown in Table A.1 and Table A.2, starting on page 131). The electrodogram on the top shows electrical discharges for the whole duration of the utterance, while the plot on the bottom visualizes data for only 50 ms (vertical dashed lines in the top plot define the time range of the bottom plot).

The presented stimulus rendering has similarities with both the CIS and N-of-M strategies (see section 2.2.4.1): 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). But in contrast to CIS (and 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. An example for the SAM coder output is presented in Figure 4.16.

4.5 Implementation details

Previous sections of this chapter have provided information about the SAM strategy from a scientific point of view. Since implementing the required software has been a significant part of

my work, this section is meant to reveal more details on that. Section 4.5.1 presents basic concepts of the software architecture and data flow, while section 4.5.2 elaborates on the safety measures that have been taken to minimize the chance of injury or discomfort of implantees caused by the developed software.

4.5.1 Software framework

The SAM strategy has been implemented in C/C++. Current version (v2.21) of the core software consists of approximately 13000 lines of source code (auditory model: 3300 lines, coder: 2700 lines, interfacing with OS and CI-hardware: 7000 lines). Since we intended to use this software in tests with CI users, a significant part of the code was about to serve the safety of the test subjects. SAM has been embedded into a proprietary software framework called CIX, which is an acronym for *Cochlear Implant eXperimental software.*
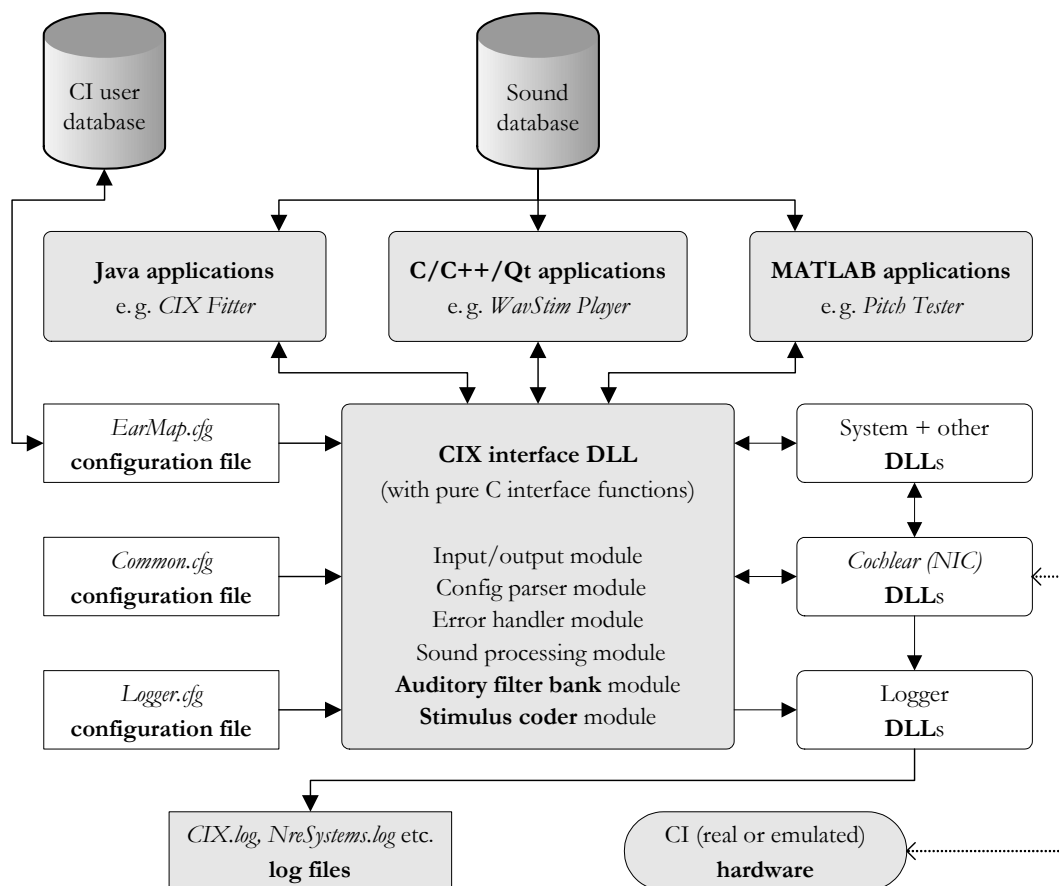
Figure 4.17: Schematic representation of the architecture of the CIX-based system. Boxes with edgy and round corners represent human-readable text files and software entities, respectively. The dotted line connecting "Cochlear DLLs" to the "CI hardware" illustrates galvanic isolation between the PC and the stimulation hardware.

The main concept is to keep all the functionality connected to the SAM strategy (including the auditory filter bank, the coder, and all related tools like sound file I/O, configuration management and error management) in a single DLL (dynamic-link library) file, which we call the "CIX interface". Through this interface, all relevant SAM functions can easily be used by many applications from various environments. The encapsulation of the functionality has several advantages. Most importantly, in the design phase, we were compelled to consider very carefully what SAM-related functionality we would make available through the DLL. We ended up having only a couple of functions (having pure C syntax) provided by the interface, which allowed us to test the system exhaustively and in a controlled manner from a MATLAB environment. See Appendix A.2 for an excerpt of the CIX interface DLL functions.

The CIX interface DLL solely depends on system DLLs (Windows and C++ development tool libraries), the Cochlear NIC (Nucleus Implant Communicator) libraries, and on the Log4CPlus DLL. Applications employing the CIX interface DLL, however, may have further dependencies. An overview of the CIX-based system architecture is shown in Figure 4.17.

As shown in Figure 4.17, the CIX interface DLL relies on three configuration files: the *EarMap.cfg* (short name for "ear model, CI user's map, and session configuration"), the *Common.cfg* and the *Logger.cfg*. The first two must be present and valid to initiate any calculation or stimulation.

The *EarMap.cfg* is always specific to a CI user. It contains strategy fitting parameters (like threshold and comfort levels, pulse width, pulse rate, etc.) and may include personal information about the user it belongs to. There is also a possibility to individualize any auditory model parameter via this configuration file.

The *Common.cfg* includes information about the software environment, the CI hardware to be used, and the default sound pre-processing parameters. In short, common configuration is specific to the test setup and can be left unchanged as long as the test environment remains the same. Sample configuration files can be found in Appendix A.4 and A.5 starting on page 136.

The *Logger.cfg* describes the requested thoroughness and format of the log. More information on logging is provided in section 4.5.2.2.

4.5.2   Safety measures

During the implementation work, we needed to keep in mind the high expectations on safety, as placed in requirement *R1* listed in section 4.1. Fortunately, one derivative of the centralized DLL concept is that if the functionality provided by the CIX interface library is safe, then all applications using this DLL are inherently a lot safer than otherwise. The CIX interface includes a set of measures to increase safety, and methods to determine and log the exact cause

of the error, if an error happens. These safety measures are listed and explained through the following sections.

4.5.2.1 Emergency stop

In spite of all precautions, testees may still encounter inconvenient auditory percepts for certain electrical stimuli. To ensure that the stimulation can be stopped at any time, the CIX interface DLL provides a function called *emergencyStopAndDisconnect()*. Even though the DLL is supposed to be used from a single program thread, it is compiled as a multi-threaded library to ensure that this function can always be called. The function stops the on-going stimulation (if any) at the next time instance provided by the operating system (i. e., practically immediately) and disconnects the CI hardware. All CIX applications provide a large stop button in their graphical user interfaces to facilitate the emergency stop feature.

4.5.2.2 Logger

The CIX interface DLL and all its modules contain a sophisticated logger [Smi13]. The logger gets initialized with the DLL and records all important events up to the point where the DLL gets unloaded. We defined six distinct log levels at the beginning of the project:

- **Trace**: is placed in the first line of each important function call (and in the last line of each destructor), so that the whole program run can be traced, if required.
- **Debug**: is used to monitor the behavior of critical variables, which may be crucial for the error-free run of the program.
- **Info**: is used to inform the user or developer about entering/leaving important sections of the program, and providing verbose information about processes.
- **Warning**: is used in "almost error" cases, from which the system can be recovered safely. For example, a less important configuration entry that is missing can be substituted safely by a default value. In this case, a warning message will be logged, but the execution will not be aborted. (However, we tried to avoid the use of this debug level as far as possible due to the critical nature of the project.)
- **Error**: is used in any software error case, which prevents the program from running correctly or safely. After encountering (and logging) an error, the program must be terminated in a safe way.
- **Fatal**: is used if a hardware error (NIC, soundcard, I/O, etc.) occurs. After encountering (and logging) a fatal error, the program must be terminated safely.

Even though the logger does not increase safety directly, it provides information on the program flow, which, in case of a failure, can turn to precious information to quickly resolve the situation. A sample logger output file (*CIX.log*) is listed in Appendix A.6 on page 138.

4.5.2.3 Exception handling and error codes

Even though the CIX interface DLL uses strict C syntax towards the "outer world", C++ is used inside it. One big advantage of C++ is its built-in capability for exception handling. Accordingly, critical sections of the CIX software (especially within the interface DLL) had been encapsulated into try-catch blocks. In this way, even errors that are very unlikely (memory allocation error, arithmetic errors, etc.) are handled properly without endangering the stability of the whole system. As a result, should a critical error occur, the system can be shut down safely and the appropriate lines can be added to the log file.

4.5.2.4 Configuration and data range checks

The only way to make the CIX interface calculate stimuli and communicate with a CI hardware is to set up its modules beforehand. This can be done solely by assigning the configuration files. Each configuration file that the CIX interface opens has to pass a sequence of checks before it will actually be used. The following general checks are done upon opening a configuration file:

- Does the specified file exist? Does it contain any characters at all?
- Can the whole file be read into a buffer?
- Do all the lines in the file have the required syntax?

Once these checks have passed, the configuration file is parsed and the data is supplied to all CIX modules that are affected by the given type of configuration file. Then, the modules do their own checks:

- Are all key-value pairs present that have no default definition in the given module?
- Are the values plausible (within a pre-defined range)?
- Are the values consistent with each other (e. g. pulse rate and phase width)?

Only if all modules report that the passed configuration data is valid will the CIX interface allow the use of the given configuration file.

Furthermore, since the software contains a large number of mathematical operations, we have spent much time ensuring that the implemented calculations would not yield domain er-

rors (e.g. logarithm of non-positive numbers) or arithmetic errors (e.g. division by zero). Critical program sections are guarded by exception handling to allow for smooth and safe termination of the program if an error should happen despite all precautions. In such a case, the logging feature would help identify the origin of the error.

4.5.2.5 Watchdog timer

We have witnessed that certain system configurations exist (like when parts of the software reside on a network drive), where the link between the host software and the NIC libraries is disturbed. This manifests itself in stuck function calls (e.g. never returning calls to the very important *nscStopStream()* and *nscGetStreamStatus()* functions) when targeting an emulated L34 processor. Even though conditions have been changed to avoid problematic system configurations, a safety measure has been taken to prevent this kind of failure making the whole software framework unresponsive: A watchdog timer has been built into the CIX function responsible for stimulation via NIC. This timer allows the stimulation loop to take only a certain time span to finish. The allowed duration is calculated from the length of the stimulus sequence by adding a given amount of tolerance (by default 200 ms + 50 % of the duration of the sequence, the latter can be set from the *Common.cfg* file, see STIM_DUR_TOL key in Appendix A.5). If the stimulation is not finished after the given amount of time, a fatal error is logged and the CIX interface DLL disconnects the CI hardware and requests the application to shut down.

4.5.3 Applications

We have also developed a series of CIX applications to support testing and evaluation of the SAM strategy. An incomplete list of these applications along with their short description is provided below:

- **CIX Fitter**: is a JAVA application with which the SAM strategy can easily be fitted. Its look and feel is similar to those of the CI fitting software of well-known manufacturers. This application is the only one that can not only read but also edit and save an *EarMap.cfg* file.

- **WavStim Player**: is a program written in C++. It provides a simple graphical user interface (GUI), where sound files can be chosen from a previously loaded database and streamed to the CI hardware. A special feature of this application is the precalculation feature: all sound files loaded in the GUI can be converted to SAM stimuli prior to the testing. In this way, the chosen stimulus can be streamed to the CI almost instantaneously, without having to wait for calculations.

- **OLSA Player**: is a C++ program providing a simple GUI to play back and stream any of the OLSA[16] [WKK99, WBK99a, WBK99b] sentences. A further feature of this application is the "three word only" mode, in which only the last three of the five words of each OLSA sentence will be used for the stimuli.

- **OLSA Tester**: is a JAVA application implementing the measurement method of the Oldenburg sentence test. It provides a simple GUI to control the test procedure and to view the results once the test has finished.

- **Real-time Audio Slicer**: is a C++ program providing a GUI and acting as a communication bridge between the audiologist and the CI user being connected to the clinical test processor (not transmitting sounds from the environment). The aim of this application is to monitor the input of the PC sound card continuously and to crop sentences (i.e., remove preceding and trailing silence) on the fly. Once a speech snippet is isolated, it gets pre-processed (high-pass filter, fade-in and fade-out, amplitude normalization) and streamed to the cochlear implant. In that way, the audiologist (or test leader) is able to communicate with the CI user in "almost real-time" without having to replace the clinical sound processor with the everyday processor of the implantee.

- **Sine Sequencer**: is a C++ program providing a GUI to specify sequences of pure tones. Duration, frequency and loudness of each tone as well as the pause between the tones can be specified. The parameters can also be randomized. The generated sequences can be played back and/or streamed to the cochlear implant.

- **Auralizer**: is a C++ program providing a simple GUI. With its help, wave files specified by the user can be turned to CI stimuli (using either the SAM or the ACE strategy) and then back to audio. For the latter, an acoustic simulation described in [CBH12] is used (see also section 5.4 for more details). This application helps get an impression how a CI user with a given configuration may perceive sound.

- **Real-time Visualizer**: is a C++ program providing real-time SAM calculation and visualization of the electrical stimuli. Its GUI allows for changing various auditory model and stimulus coder parameters on the fly. This application helps investigate many properties of SAM and is a good tool to study the effects of altering the parameters of the strategy.

---

[16] Oldenburg sentence test (German: Oldenburger Satztest), see also page 95.

- **Pitch Tester**: is a MATLAB script to test pitch perception abilities of CI users. The script employs an adaptive three-alternative forced choice method to vary frequency and is able to use various test signals like pure tones or sung vowels.

- **Delay Tester**: is a MATLAB script to find the just noticeable difference in traveling wave delay by adaptively changing SAM coder parameters. The script uses a three-alternative forced choice method to find the smallest difference that the CI user perceives as different (as compared to the physiological delay that SAM incorporates by default).

All the applications listed above depend on the CIX interface, so all of them benefit from the safety measures implemented in the DLL. Furthermore, all the applications do logging and append their log entries to the log file of the CIX interface DLL. Screenshots are presented in Appendix A.7 starting on page 139.

4.6 Fitting

The procedure of finding the threshold and most comfortable stimulation levels for each electrode is called fitting. This needs to be done with each CI recipient on a regular basis to ensure maximum benefit from the CI. The result of the fitting (i.e., the sum of all THL and MCL values plus other strategy parameters) is called a map. Allowedly, threshold levels of a map should be as high as possible but low enough to just not create auditory sensation, whereas stimuli at the most comfortable levels should be perceived as loud but still comfortable sounds. Depending on the subject and on the available objective measurement methods (ECAP[17], EABR[18], stapedius reflex, etc.) there may be several ways of fitting. The classical and (to date) most precise way is the behavioral one, where the clinician aims to find optimal settings based on auditory sensations reported by the CI recipient.

A behavioral fitting procedure for the SAM strategy contains the following steps:

1. Set all THL and MCL levels to zero and only disable faulty electrodes (if any).
2. Pick about six electrodes (preferably evenly distributed).

---

[17] Electrically Evoked Compound Action Potential

[18] Electrically Evoked Auditory Brainstem Response

3. Find threshold levels for the chosen electrodes and interpolate THL for the others. Then, for the previously selected electrodes, find MCL values that correspond to "medium loud" percept and interpolate MCL for the other ones.

4. Verify THL for all electrodes. Make adjustments if necessary.

5. Verify MCL for all electrodes. Make adjustments to achieve "loud but comfortable" percept, if necessary.

6. Test with real-life stimuli. Make changes if necessary.

7. A balancing procedure may be performed to make sure that MCLs generate approximately the same perceived loudness on all electrodes. (Balancing is usually carried out by stimulating four to five consecutive electrodes in turn at MCL and asking the CI user if any stimulus appears softer or louder than the others do.)

The repeated adjustments to the threshold and most comfortable levels are necessary, since SAM may induce stronger summation effects than conventional CI strategies. This can be attributed to two specific properties of SAM. First, there is no fixed channel stimulation rate, which means that the same electrode may be activated in quick succession. Second, SAM incorporates traveling wave delay, which makes the successive activation of neighboring electrodes probable, which may lead to higher charge densities.

The fitting method described above can be carried out efficiently with the CIX Fitter software (see Figure A.1 on page 139, please note that in the software THR and MCL are denoted by TL and CL, respectively). The software also supports many keyboard shortcuts and hotkeys to speed up the fitting procedure.

4.7 Summary

This chapter elaborated on the design, structure, implementation, and parameterization of the SAM strategy. Beyond posing basic requirements meant to steer own development work, a focus was placed on the two main building blocks of SAM: the auditory filter bank along with its multidimensional parameterization and customization, and the stimulus coder. The second focus was on the implementation of the strategy and related applications, as well as on safety measures that were required due to the nature of the project. Finally, yet importantly, the SAM fitting procedure was discussed, which is essential to actually put the strategy to work.

Chapter 5

Analysis of the SAM signal processing strategy

A lot of research, development, and implementation work were needed to move from a naive idea to the safe prototype implementation of SAM. The details of this transition were summarized in the previous chapter. However, to see if the new signal processing strategy makes any difference, it also needed to be evaluated.

This chapter first elaborates on the design requirements including safety, stability, psychoacoustic properties, and speed of the implementation, and checks if they have been met. Then, section 5.2 investigates the effects of changing key parameters of the processing strategy.

Next, section 5.3 focuses on the simulation studies that have shed light on some of the strengths and potentials of SAM.

Subsequently, section 5.4 reviews the most important test scenario: the evaluation with cochlear implant users. This section first describes materials and methods including detailed information about the measurements done. Then it elaborates on the results and draws comparisons between the measured performance with ACE and SAM.

Section 5.5 presents the evaluation with normal-hearing listeners along with a newly developed acoustic simulation algorithm, which has been used in these tests.

Finally, the chapter provides a recapitulation of the results and a general conclusion for the analysis of the new signal processing strategy.

5.1   Validation against requirements

5.1.1   Safety

We dedicated a great amount of time to make the prototype implementation of SAM safe. We implemented all properties proposed in section 4.1, including interruptibility and responsiveness of the user interface. As for the core component of the software, we developed our own safety measures as listed in section 4.5.1, including range checks, logging and exception handling. Furthermore, we have also followed the "NIC application hazard control recommendations" listed in [Coc10b]. We did not encounter any safety problem during the tests described later in this chapter.

5.1.2   Stability

During the development and evaluation of SAM we used a great number of various input signals and many different configurations of the system. As already addressed on page 42, some doubts emerged about stable operation over a large input dynamic range, considering that the filter bank consisted of cascaded nonlinear models of the auditory system. However, we did not find any irregularities regarding the stability of the auditory model embedded in SAM.
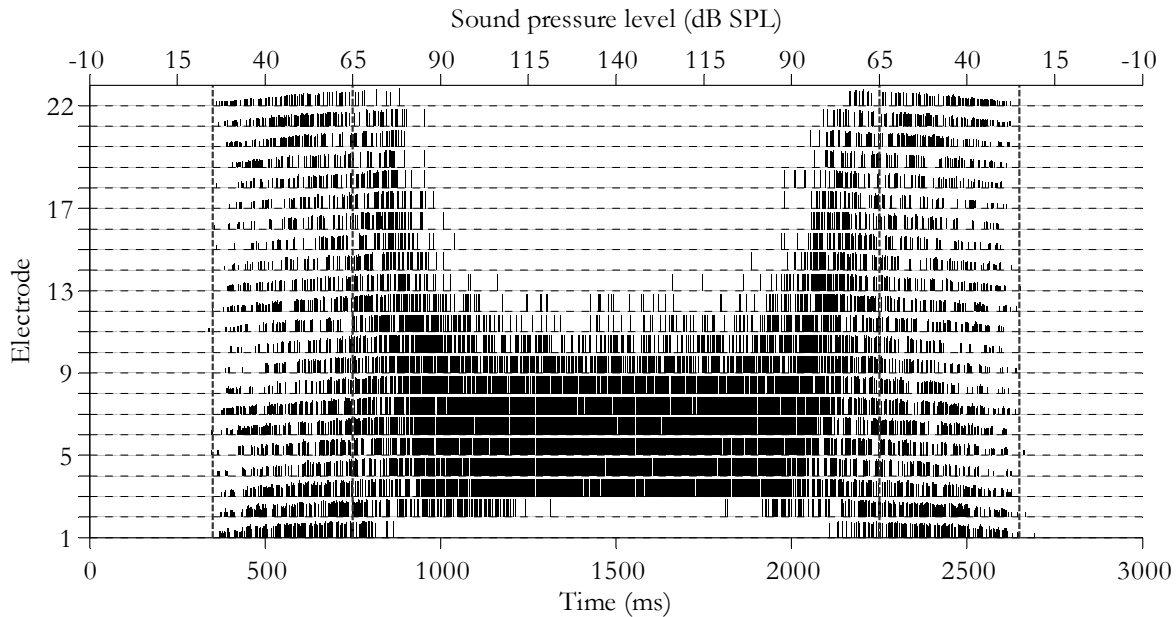
Figure 5.1:   Calculated SAM stimulation pattern for a 3 seconds long Gaussian white noise
              signal with logarithmically increasing (0 to 1.5 s) and then decreasing (1.5 to 3 s)
              amplitude. The corresponding model loudness ranges from -10 to 140 dB SPL.
              The gray vertical lines mark the 25 to 65 dB SPL ranges, within which the CI
              signal processing strategy (excluding preprocessing and preconditioning algorithms) is typically required to operate.

Still, to make sure that requirement *R2* (stability over the normal hearing range of 0 to 120 dB SPL, see page 42) is met, we have created a special test scenario, for which the SAM stimulation pattern is shown in Figure 5.1. During this test, a Gaussian white noise signal with logarithmically increasing and then decreasing amplitude is processed by SAM. The model loudness corresponding to the input amplitudes ranges from -10 to 140 dB SPL. Since the simulation is executed in one run, this test captures instabilities (if any) introduced by exceptionally high model loudness levels. As the resulting stimulation pattern shown in Figure 5.1 exhibits symmetry about the $t=1.5$ s axis, it can be stated that the system remains stable over the required loudness range.

Note that while the 25 to 65 dB SPL range in Figure 5.1 looks as expected, stimuli on electrodes corresponding to frequencies around 1 kHz tend to dominate for input levels over 80 dB SPL. This is not an error, but rather a characteristic feature of the auditory model emphasized by the coder. However, this feature would probably not be retained when inserting the SAM strategy into a complete CI speech processing chain. That is, because in real CI systems neither the microphone pre-amplifier nor the preprocessing algorithms (like automatic gain and sensitivity control) would allow for such large dynamic range. Systems from Cochlear typically omit signals softer than 25 dB SPL and limit loud signal parts to 65 dB SPL.

5.1.3 Features (psychoacoustic properties)

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. Most of the contents presented in the remaining parts of section 5.1 and section 5.2 have already been published in [HCH12]. Their reuse was permitted by the coauthors.
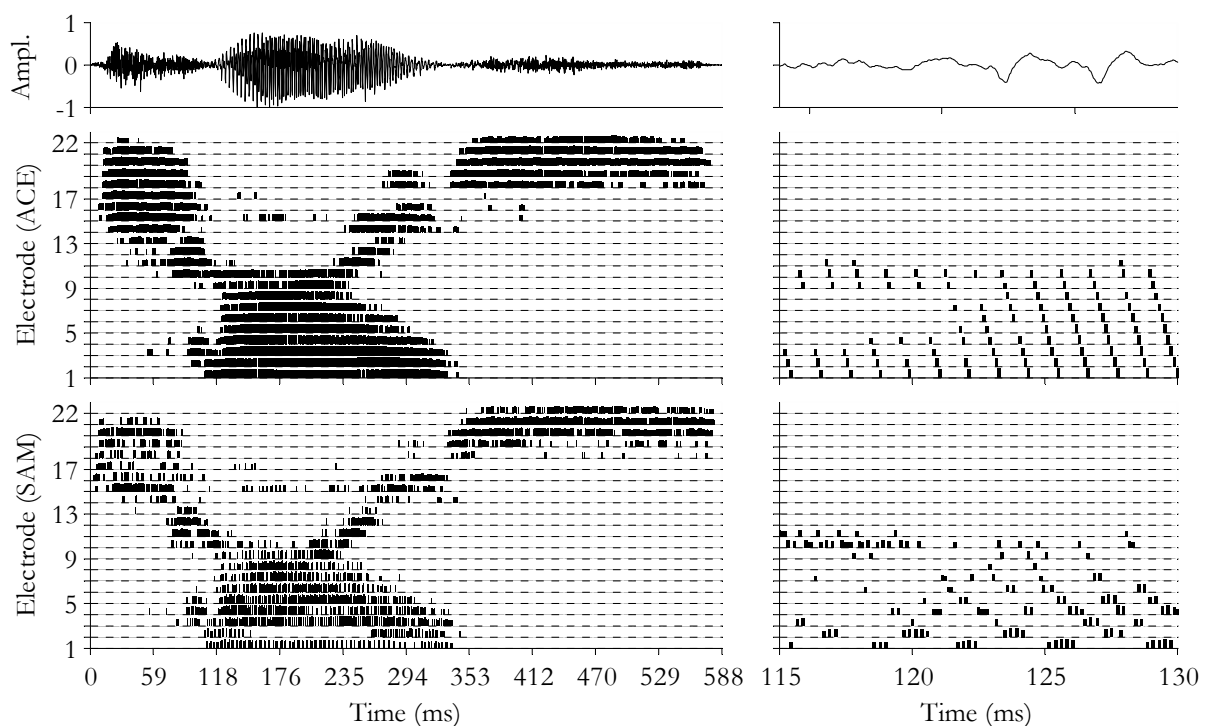
Figure 5.2: Example showing the waveform (first row), the ACE coder output (second row), and the SAM coder output (third row) for the word "choice". Left column shows data for the whole duration of the utterance, while the right column visualizes data for 15 ms only. While both SAM and ACE feature place coding, the temporal code is only present with SAM.

5.1.3.1 Cochlear delays

SAM stimulation patterns show realistic hyperbolic delay characteristic, which is caused by frequency specific group delays along the simulated cochlear partition. Delays range from about 0.2 ms (most basal electrode) to about 9.1 ms (most apical electrode) using the default parameters (as shown in Table A.1 and Table A.2, starting on page 131) in the presented 22-electrode setup. A positive 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.
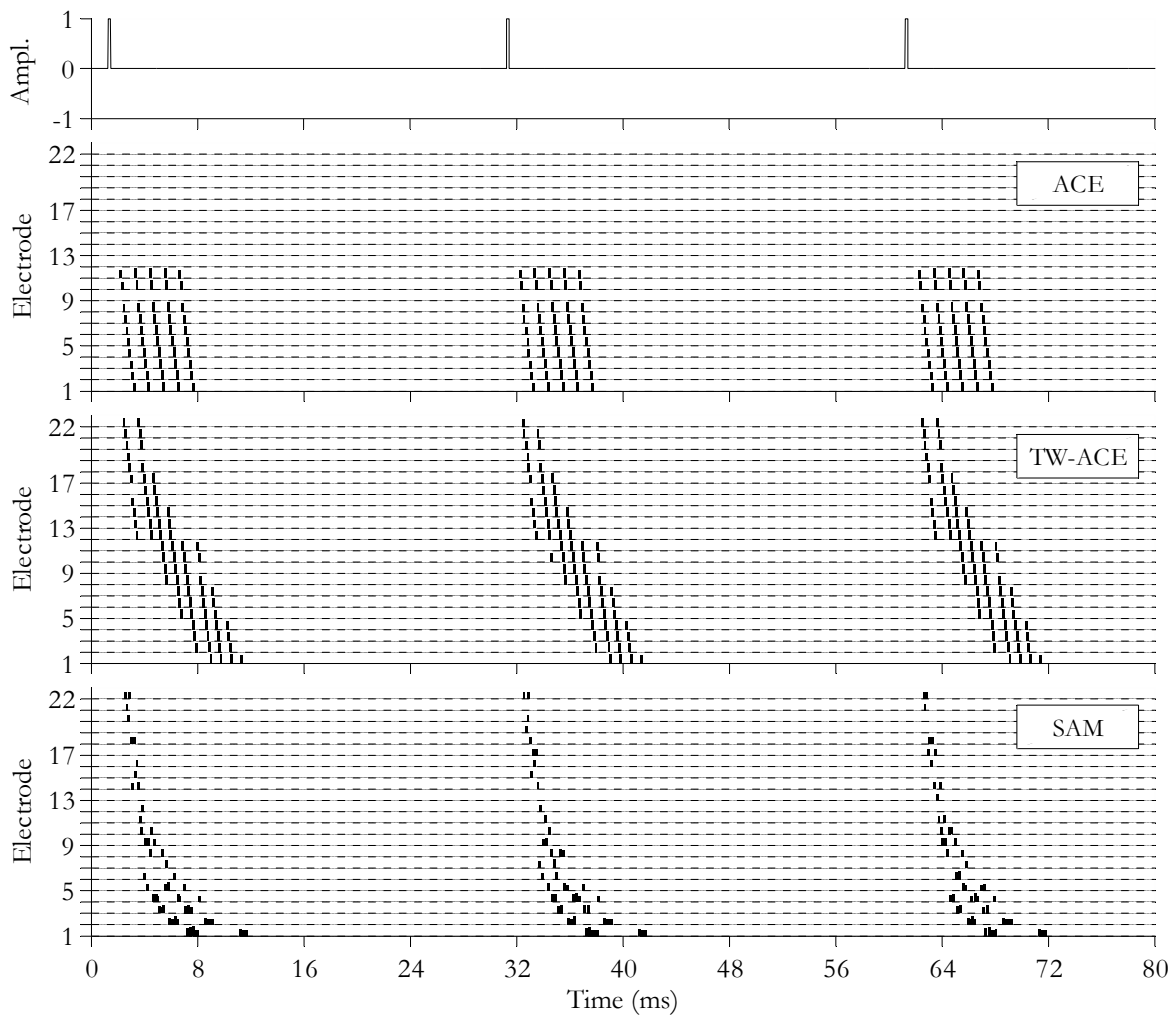
Figure 5.3: Comparison of delay characteristics of ACE, TW-ACE and SAM. Using the same basic parameterization (pulse rate, channel frequencies, etc.) for all three strategies, stimulation patterns show responses to a pulse (200 µs, 65 dB SPL). While ACE produces linear across-frequency delays, TW-ACE spreads the trajectory of pulses over time better. SAM produces well-defined hyperbolic delay trajectories corresponding well to the cochlear delays of the calibrated auditory model.

See Figure 5.3 for a comparison of cochlear delays in stimulation patterns of various CI signal processing strategies including SAM. As shown in Figure 5.3, SAM reflects the delay characteristics of the tuned hydromechanical model well (even though the stochastic coder adds a large amount of variation) and replicates physiological cochlear delays most appropriately (as given in [EDCS07]). For a direct comparison of the coder output of ACE and SAM for a speech signal, see Figure 5.2 on page 75. Cochlear delays with the SAM strategy can also be observed in e.g. Figure 3.1 and Figure 4.16.

5.1.3.2 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 abrupt changes would occur during the CI signal processing. Activation of the corresponding electrode was captured (see Figure 5.4).
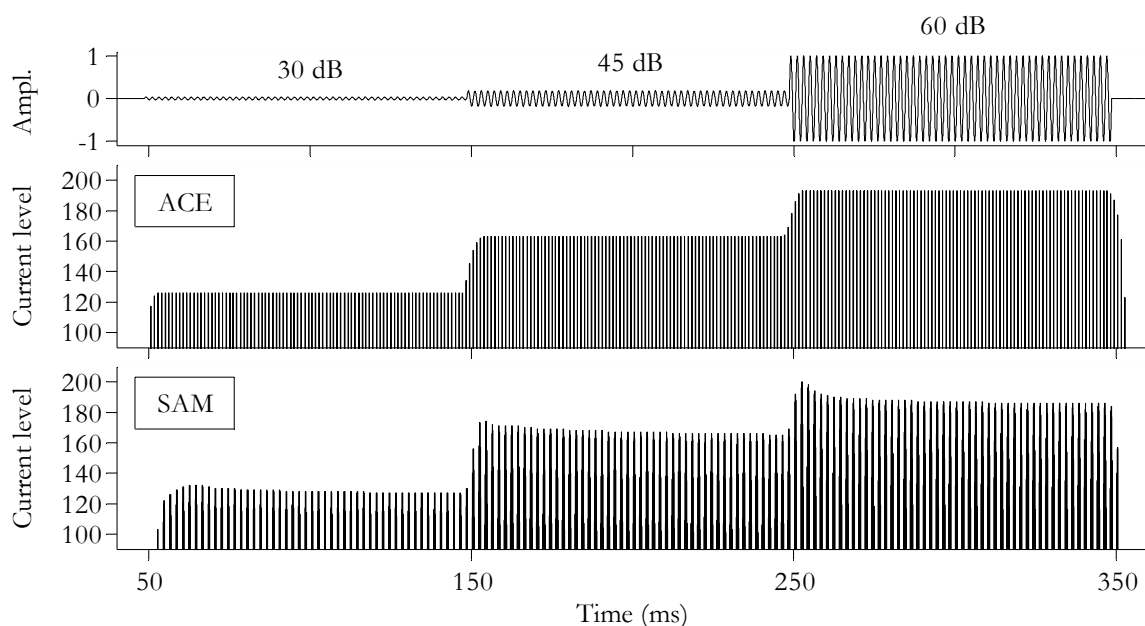
Figure 5.4: Comparison of adaptation characteristics of ACE and SAM. The intensity of a 500 Hz pure tone is abruptly varied from 0 to 30 to 45 to 60 and back to 0 dB SPL. The current level is captured on electrode 3, which has a characteristic frequency of approx. 500 Hz. With SAM, overshoot and subsequent adaptation can be observed upon each input intensity change. (For the sake of better visualization, peak selection randomness and repetition penalty have been turned off.) ACE does not show adaptation dynamics.

As expected, different current levels show up in the output for different input intensities. With SAM, however, every time the intensity suddenly increases there is also an overshoot, which is followed by subsequent adaptation to lower current levels until reaching the steady-state value. This phenomenon supports the detection of onsets, which can be very useful in consonant discrimination. Results also show that adaptation characteristic depends on both input intensity and stimulation history. ACE, in contrast to SAM, does not feature adaptation effects on its own. In CI systems, however, it is often combined with pre-processing methods incorporating simple types of adaptation to improve user experience mainly in noisy environments [PBG06].

5.1.3.3 Compression

In cochlear implants, compression of acoustic amplitudes is necessary to fit within the electrical dynamic range. Equation 4.7 on page 62 describes the loudness growth function, which is typically used for this purpose. SAM inherently includes compression due to the nonlinearities of the cochlear mechanics, and uses the function described in Equation 4.7 only to (almost perfectly) match the default loudness growth function of ACE. The achieved similarity varies slightly across frequencies and is presented for 1000 Hz in Figure 5.5. The plotted data was determined by finding the maximum stimulation level on the corresponding electrode for pure tones (with 300 ms fade-in) presented at various sound pressure levels.
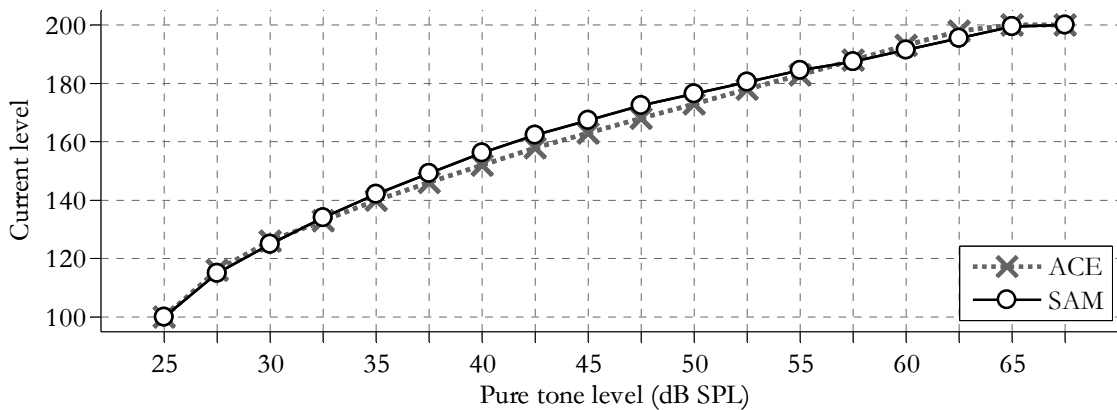
Figure 5.5: Loudness growth with SAM and ACE for a 1000 Hz pure tone at increasing sound pressure levels. The maximum stimulation level measured on electrode 7 (center frequencies: approx. 1016 Hz at 55 dB SPL with SAM, and 1000 Hz with ACE) is plotted. For the ACE processing an LGF setting of $Q=20$ (see section 3.1.4 of [Nog08]) was used, which translates to a curve shaping factor of approximately $\varrho=416$ (see section 2.2.4.1 on page 15).

5.1.3.4 Phase-locking

The tendency of the auditory nerve fibers 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 signal processing strategies (including ACE) use constant rate pulse carriers that contain no temporal information (cf. Figure 3.1 on page 33). By contrast, SAM is not restricted to a specific channel stimulation rate; hence, its coder is able to transmit temporal fine structure well: 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 DFT as described in [MM00]. 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.
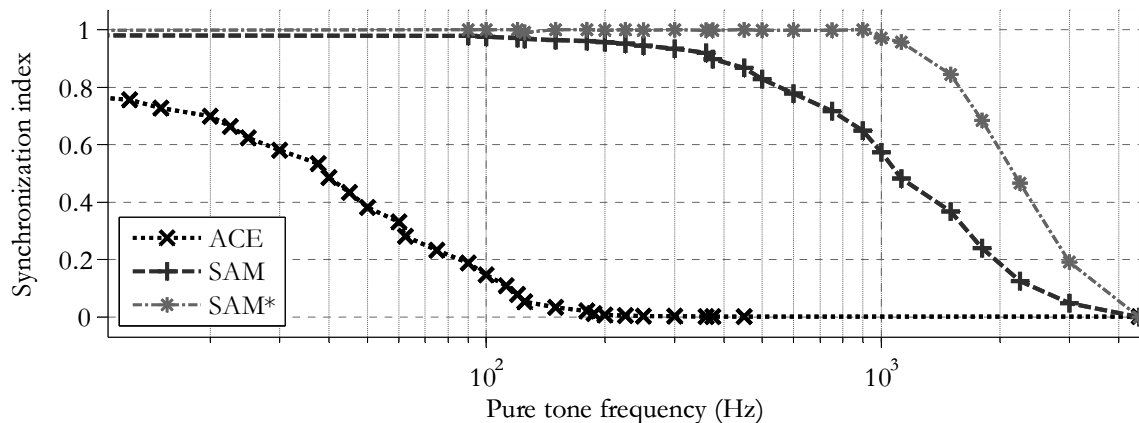
Figure 5.6: Synchronization index trends for SAM and ACE. Higher values indicate better synchronization to the frequency of the pure tone input. While ACE cannot temporally code frequencies above 200 Hz, SAM features noticeable temporal coding up to frequencies well above 1 kHz depending on configuration. 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.

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 DFT. 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 stimulation pattern. To avoid inaccuracies in the SI calculation due to spectral leakage, two conditions regarding the pure tone frequency have to be fulfilled. First, it has to be a divisor of the stimulation rate. Second, it must be divisible by the frequency resolution of the DFT spectrum. Considering these conditions, SI values were calculated using SAM and ACE for pure tones with 65 dB SPL intensity, about 1 s duration, and frequencies ranging from 10 Hz to 4.5 kHz. Results are presented above. As shown in Figure 5.6, synchronization index always falls consistently with increasing input frequency. SAM locks to stimulus phase for frequencies up to ca. 2 kHz, which complies with knowledge about phase-locking [SLOM02]. 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.

5.1.4   Speed

As already stated in section 4.5.1, SAM has been implemented in C/C++ and embedded into the Fraunhofer IDMT proprietary software framework called CIX. The C/C++ based implementation appeared to be the best choice to support real-time and continuous operation. These latter features were meant to support mobile operation of the strategy (e.g. running it on a netbook or a tablet PC, which can be used for at least several hours uninterruptedly). Unfortunately, because of limitations imposed by the Cochlear hardware/software, this was not possible at the time of the study[19]. Nonetheless, we still put emphasis on meeting requirement *R4* (see page 42) to ensure that SAM can be probed in other (possibly real-time) systems with ease.

Using the real-time implementations of ACE and SAM, computations required to process a given piece of audio data take about 30 times longer with SAM than with ACE. However, the time required to process an input sound of linearly increasing length is also increasing linearly for both strategies. A comparison is presented in Figure 5.7. Calculations were done on the same mid-range computer (Dell Opti-Plex 760, Intel® Core™ 2 Duo processor E8500, 3.16 GHz) under the same conditions. The two implementations (SAM and ACE) were optimized at about the same level for the target PC.

---

[19] In the summer of 2013, Cochlear announced a new release of the Nucleus Implant Communicator that supports continuous streaming.
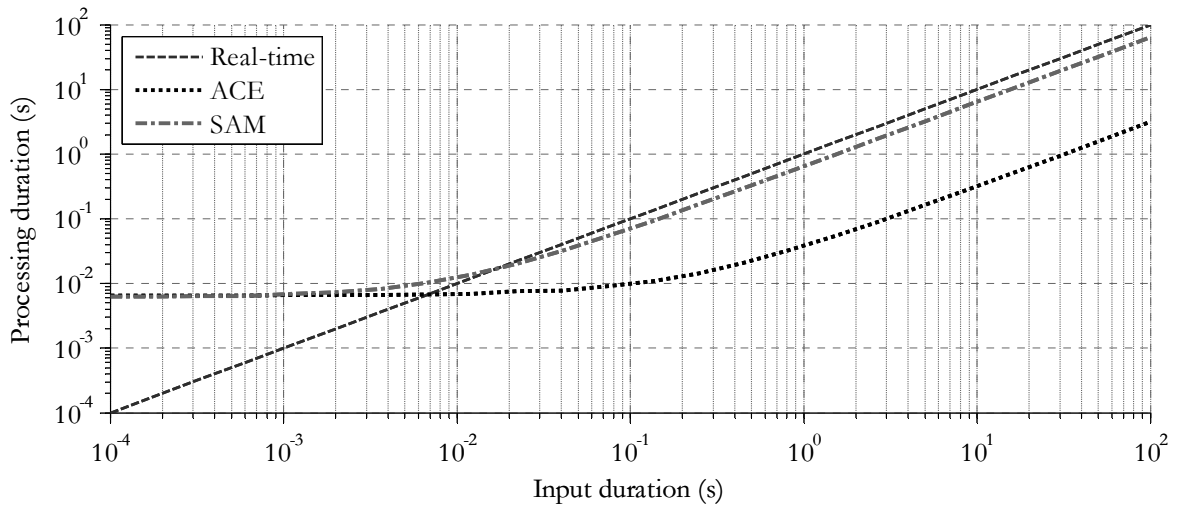
Figure 5.7: Run time (including initialization of variables and memory space of the given strategy, but excluding other overheads like process creation and file I/O times) as a function of input length.

As shown in Figure 5.7, strategy initialization overheads (around 50 to 70 ms on the test PC) dominate for short input signals, but this would not be an issue in a system that operates in a real-time and continuous manner.

We have also looked into the distribution of relative computational times within the SAM strategy. The result of the profiling is presented in Figure 5.8, which immediately reveals the most important targets for optimization: the building blocks of the auditory model. See sections 6.1.2, 6.1.4, and 7.1.2 for further information on this issue.
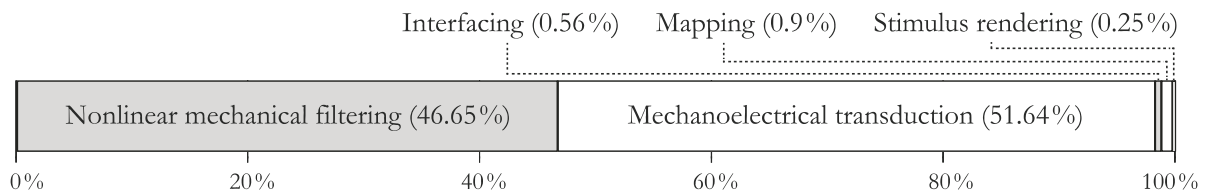
Figure 5.8: Relative computational times (excluding framework and other overheads) of the main building blocks of SAM, with default parameters.

5.2   Effects of changing coder parameters

In addition to the individually fitted physiologically induced parameters like THL_m and MCL_m there are several parameters in SAM that significantly affect coder behavior and may, thereby, affect perception. The next subsections illustrate how the stimulation pattern changes as an effect of varying three of those parameters, namely TPR, ξ_p and R_m.

5.2.1 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. In contrast, higher rates enable the coder to map temporally high-resolution parts of the auditory model output to the CI electrodes better. CI users seem to have their preferred stimulation rate for any given strategy, which may have various physiological reasons. A typical total pulse rate with the ACE strategy is 9000 pps, which we have adopted as default *TPR* for SAM. In Figure 5.9, direct comparisons of SAM and ACE stimulation patterns for lower (7200 pps) and higher (14400 pps) rates are given.
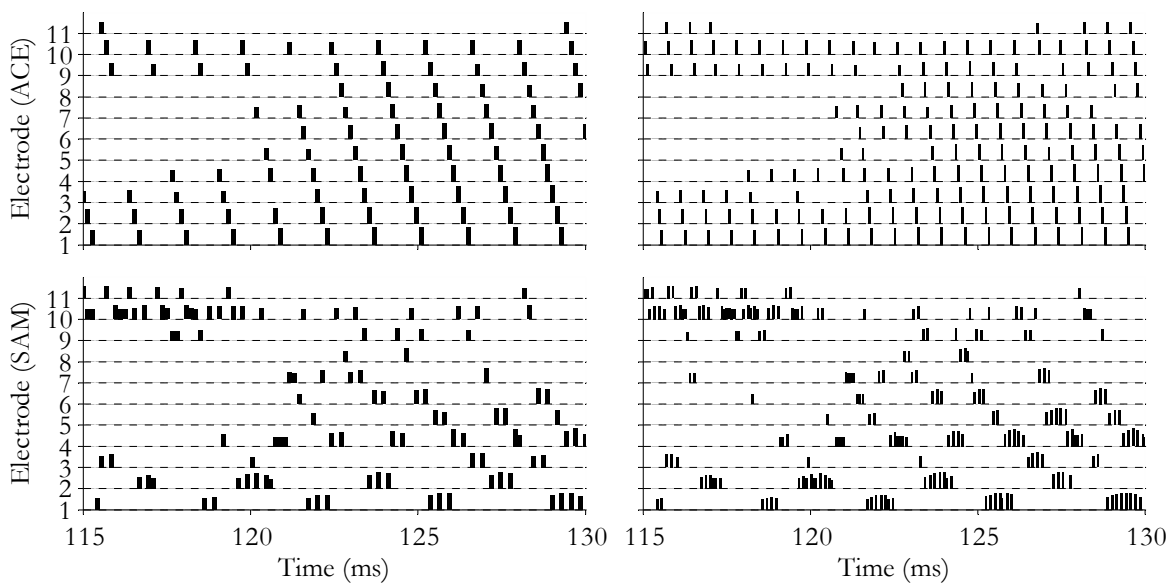
Figure 5.9: Comparison of coder outputs (first row: ACE, second row: SAM) for a short snippet of the word "choice" for total pulse rates of 7200 and 14400 pps (left and right columns, respectively). For a clearer view, only the discharges of the 11 most apical electrodes are shown. With SAM, cochlear delays and phase-locking effects are preserved over various total pulse rates. The amount of temporal information provided by ACE does not remarkably increase with higher rates.

During the ACE calculations with various total pulse rates, the number of peaks selected in each stimulation cycle was kept at 10. Note that the panels of the right column of Figure 5.2 can be used as reference with default parameter values. Figure 5.9 confirms that the fundamental structure of the stimulation pattern is stable across various stimulation rates for both strategies. With SAM, cochlear delays and phase-locking, though less precise, remain present at lower rates like *TPR*=7200 pps.

5.2.2 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.

Figure 5.10 shows a comparison for two distinct settings. $\xi_p=0$ and $R_m=0$ (Figure 5.10, left), and $\xi_p=0.2$ and $R_m=40\,$dB (Figure 5.10, right) represent deterministic and stochastic operational modes of SAM, respectively. Panels of the right column of Figure 5.2 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.

Since every CI electrode addresses hundreds or even thousands of neurons (depending on the individual configuration and circumstances), the reality (and hence the best configuration of SAM) may lie between the two depicted cases: A single neuron will never be able to fire continuously as a result of electrical stimulation, but a plurality of neurons –in total– might be.
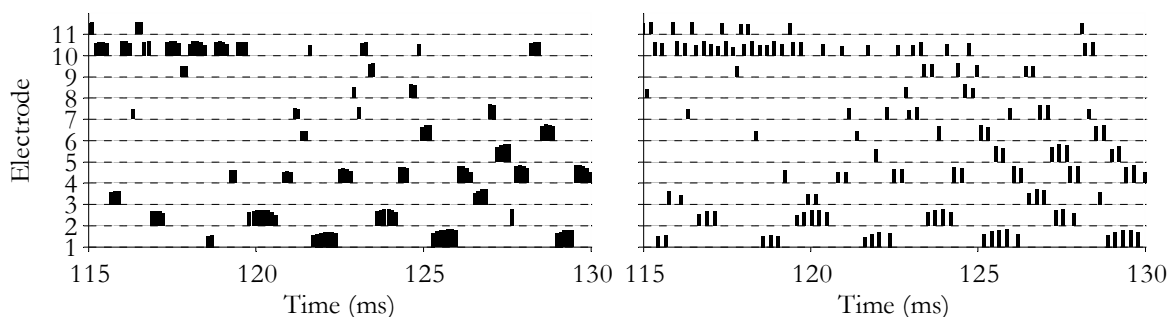
Figure 5.10: Comparison of SAM stimulation patterns for a short snippet of the word "choice" with different peak selection randomness and repetition penalty settings. Left: $\xi_p=0$ and $R_m=0$, right: $\xi_p=0.2$ and $R_m=40\,$dB. For a clearer view, only the discharges of the 11 most apical electrodes are shown. While in the first case (left) the same electrodes are often selected in consecutive stimulation cycles, in the other case case (right) no electrode is active in two successive cycles.

Conclusion

It seems that 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. This holds for a broad range of key parameter settings.

5.3   Computational evaluation

Beyond the evaluation work presented in sections 0 and 5.2, additional aspects of SAM have already been explored and studied, of which two will be outlined here: Model-based estimation of speech reception threshold and simulations to explore horizontal-plane sound source localization ability with SAM. Please note that most of the contents presented in the subsections of 5.3.1 and 5.3.2 have already been published in [HFHK11] and [HCK11], respectively. Their re-use was permitted by the coauthors.

5.3.1   Evaluation with a model of the human auditory speech processing

Traditional cochlear implant coding strategies (like ACE) present some information about the waveform or spectral features of the speech signal to the electrodes. However, neither of these approaches takes the cochlear traveling wave or the auditory nerve cell response into account, though these are given in acoustic hearing. To predict if the use of SAM would make a difference, we have compared the simulated electrical output of the SAM and ACE strategies by employing a modified version of the Hamacher model [Ham03] of the electrically stimulated auditory system. The model consisted of an auditory nerve cell population and included effects of the current spread and spatial and temporal integration (forward masking). The nerve cells generated pulses as action potentials in dependence on the spatial and temporal properties of the electric field produced by the electric stimuli. This model in combination with a dynamic time warping (DTW) speech recognizer was used to predict CI user performance in terms of speech intelligibility (with an approach similar to that proposed by Jürgens et al. [JFM+10]) for both coding strategies. The model is described in detail by Fredelake [Fre12].

5.3.1.1   Materials and methods

A model of the electrically stimulated auditory nerve by Hamacher [Ham03] has been adapted. Electric pulses, encoding a speech signal with a simulated CI, are multiplied with a spatial current function modeling the current spread inside the cochlea. A population of auditory nerve cells, which are based on the leaky-integrate-and-fire model, generates spikes, which are further processed to internal representations by modeling convergence and adaptation. An overview is presented in Figure 5.11. The model includes parts of both the peripheral and central auditory processing, which are abbreviated as PAP and CAP, respectively, throughout section 5.3.1. The joint model of auditory processing (PAP+CAP) will be referred to as MAP.
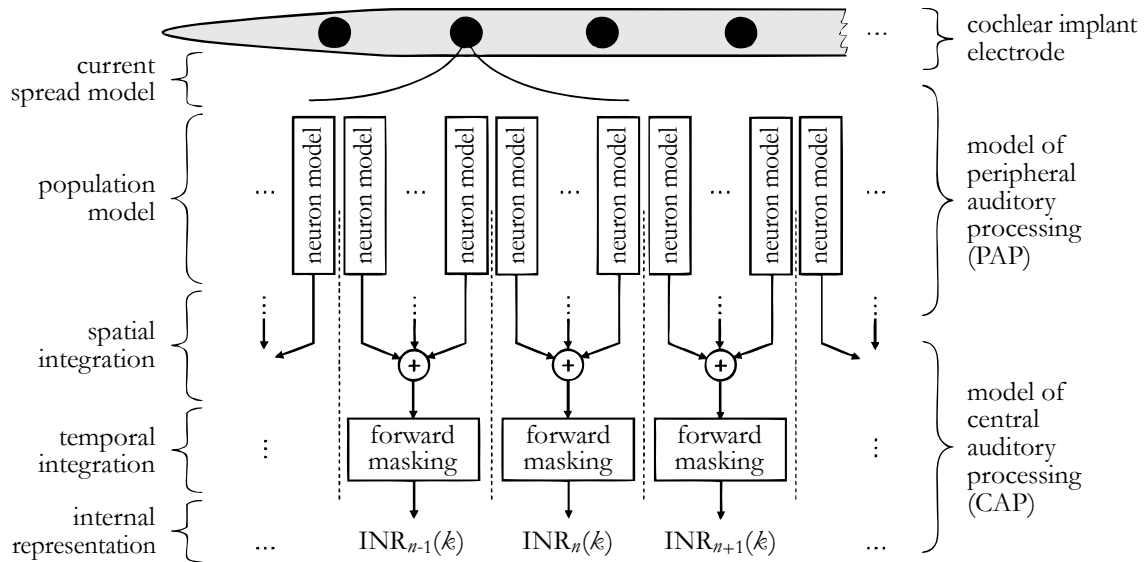
Figure 5.11: Overview of the model of auditory processing. Adapted from [Fre12].

In order to simulate different neural degeneration, the number of auditory nerve cells (N_C) is decreased, while the variable (λ_C) of the spatial current function (i. e., current spread, see also Equation 5.1 on page 89) is increased in a way that the total number of action potentials (APs) is kept constant for a given input current amplitude. For the threshold current level, the total number of APs was arbitrarily set to approximately 30 and for most comfortable level to 300, respectively. The PAP model configurations used in this study are listed in Table 5.1.

Table 5.1: Configurations of the model of auditory processing.

Configuration	v1	v2	v3	v4	v5	v6
N_C	10000	5000	2500	2000	1000	500
λ_C	0.25 mm	0.5 mm	1 mm	1.33 mm	2.5 mm	5 mm

For the speech reception threshold (SRT) estimation, the OLSA corpus [WKK99, WBK99a, WBK99b] in combination with stationary noise ("OLnoise", included in OLSA) is used with a limited vocabulary (50 words). The background noise level is fixed at 65 dB SPL and single words are mixed with OLnoise with SNRs ranging from -15 to 25 dB in 5 dB steps. For each word and each SNR the internal representations (INRs) are calculated via MAP. Afterwards, the INRs are classified with a dynamic time warping algorithm and therefrom the speech intelligibility function with the parameters SRT and slopes is calculated. The DTW speech recognizer has a speech memory consisting of pre-calculated INRs, called response alternatives. For every input word, it calculates the "perceptive distance" between the unknown INR and each

response alternative. A multiplicative internal noise (σ_{int}) simulates limited resolution ("cognitive factor"). An overview is presented in Figure 5.12.
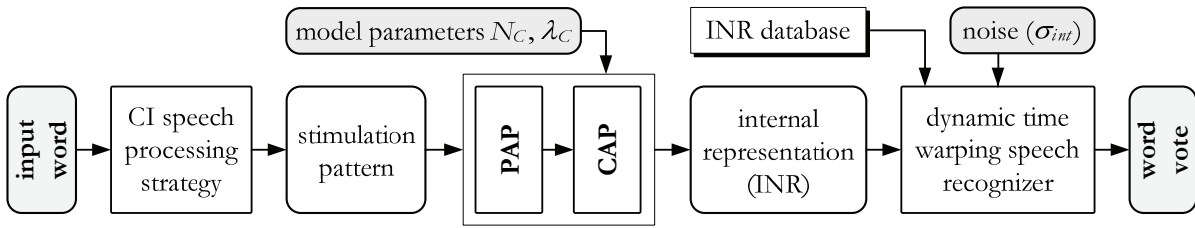
Figure 5.12: Overview of the system used for the evaluation with a model of the human auditory speech processing.

5.3.1.2 Results

The outcomes of the simulation study in terms of the modeled SRT and of speech intelligibility versus SNR is summarized in Table 5.2 and in Figure 5.13, respectively.

Table 5.2: Modeled SRT values (in dB SNR, lower is better) with OLSA for various model configurations, strategies and standard deviations σ_{int} of the internal noise. SRT values with gray background color indicate cases where the ACE strategy performed better.

Model configuration	$\sigma_{int}=0$		$\sigma_{int}=0.15$		$\sigma_{int}=0.25$		$\sigma_{int}=0.35$	
	ACE	SAM	ACE	SAM	ACE	SAM	ACE	SAM
v1	-1.6	-1.9	-1.9	-2.1	-2.0	-2.2	-1.7	-1.6
v2	-1.7	-1.9	-2.1	-2.2	-2.0	-2.4	-1.3	-1.8
v3	-2.9	-2.5	-2.9	-2.9	-1.8	-2.7	0.6	-1.2
v4	-3.3	-2.9	-3.0	-3.0	-1.7	-2.5	2.1	0.1
v5	-3.9	-3.3	-2.8	-3.1	1.0	-1.2	10.1	7.0
v6	-4.4	-3.6	-1.1	-2.5	10.2	8.6	25.2	28.2

Table 5.2 reveals for both strategies increasing SRTs with increased σ_{int}. The SRT increases with decreased cognitive performance especially in model configurations with few auditory nerve cells. Values with gray background color indicate lower SRTs for ACE in comparison to SAM, which means that ACE outperforms SAM in those cases.
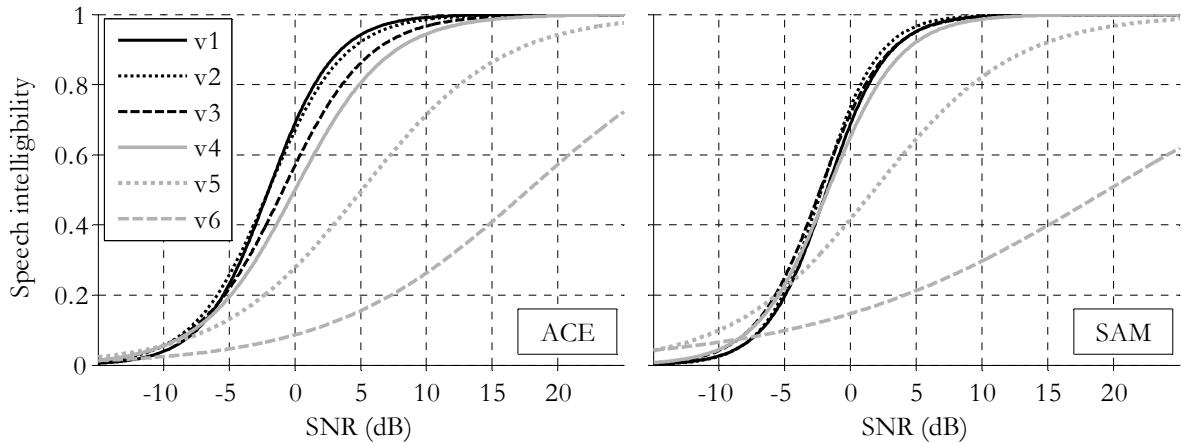
Figure 5.13: Modeled speech intelligibility with OLSA plotted against the SNR for various model configurations and σ_{int}=0.3.

However, differences are only little in most cases. Nevertheless, σ_{int}=0.35 led to lower SRTs for SAM than for the ACE strategy in model configurations v2–v5. Model configuration v6, however, could not be well fitted, because the SRT exceeded the highest SNR of 25 dB.

Conclusion

SAM is expected to mimic normal hearing processes in a more realistic way than the ACE strategy does. Still, the modeled SRTs show only little differences between the performance of SAM and ACE. With increasing internal noise (worse simulated cognitive condition), however, SAM outperforms ACE, especially in model configurations with fewer auditory nerve cells. Further investigations (not presented here, but included in [HFHK11]) show that while the two strategies deliver about the same amount of place pitch cues, SAM provides more temporal pitch cues, which may also contribute to better speech and pitch perception.

5.3.2   Simulation of sound source localization performance in bilateral use

Sound source localization capability of cochlear implant (CI) users has been a popular research topic over the past few years, because it has both social and safety implications. While it is widely accepted that unilateral implantation (witout contralateral hearing) does not provide enough information for localizing sound sources, conditions, algorithms and their parameterization for the best performance in the binaural case are still a focus of current research.

   While bilateral cochlear implantation is offered to a growing number of individuals, not all bilaterally implanted CI-users are fully satisfied. One possible cause for the dissatisfaction is the missing ability to localize sound sources reliably. The trend is to use <1 kpps channel stimulation rate and ≤9 kpps total pulse rate with n-of-m strategies like ACE, which, in fact, allows for only very limited localization performance based on temporal cues, as shown e.g. in [HWSB10]. Furthermore, the most common (and generally only) aim of CI fitting is to yield

better speech perception rates. Still, most bilateral recipients can localize sound sources to some extent, but only few can localize well [GAR+07, SBF04].

The primary goal of this study is to evaluate horizontal-plane localization with the SAM strategy, taking only interaural time differences (ITDs) into account. These have long been deemed unusable by CI recipients, but recent studies (see e.g. [DM10]) suggest otherwise. Furthermore, factors responsible for fair or poor localization ability with SAM are explored, and a performance-comparison between SAM and ACE is given. Full details can be found in [HCK11].

5.3.2.1 Materials and methods

This section describes the design of the experiments done for this study. The framework for the tests was built in MATLAB; calculations were done under Linux on a Dell Precision Work-Station 490. An overview of the processing steps is presented in Figure 5.14.
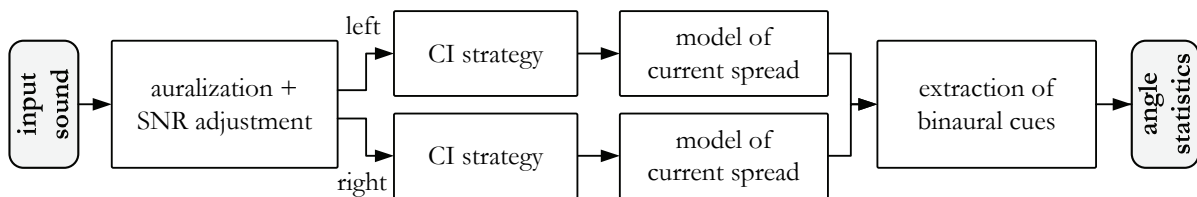
Figure 5.14: Overview of the system used for the evaluation of sound source localization performance in a bilateral simulation.

Test sounds

Two distinct test sounds are used: a speech recording and a noise signal. Both test sounds are sampled at 16 ksps with 16-bit resolution and are normalized to 0 dB FS peak value. The speech recording is a 1.34 s long part of a randomly picked wave file from the DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus [LKS86, ZSG90], where a male speaker utters "to open the store by eight" (TIMIT file *train/dr4/mbma0/sx232*, 0.77-2.11 s). This test signal will be referred to as the *Timit* signal. The noise signal is very similar to that used by Seeber and Fastl [SBF04] in their localization study. It contains 15 white noise pulses (pulse duration is 25 ms with 2 ms linear fade-in and fade-out) divided into 3 groups (pause between groups is 65 ms, pause between pulses within a group is 35 ms). This signal will be referred to as the *Pulses* signal.

Auralization and SNR adjustment

Auralization is done via the fast image method described by McGovern [McG09]. The distance between the sound source and the microphones (i.e., simulated ear positions) is fixed at

3.0 m. Reverberation conditions are listed in Table 5.3. RT_{60} and STI mean reverberation time and speech transmission index [SH80], respectively. The former is the time required for reflections of the direct sound to decay by 60 dB, whereas the latter is a well-established objective measurement predictor of how well a listener may understand speech using the given transmission channel. STI may vary between 0 (bad) and 1 (excellent). STI values presented in Table 5.3 were calculated for the *Timit* signal at 65 dB SPL presentation level after applying the given reverberation effect.

The added noise of the SNR adjustment step can be white noise (*WN*) or babble noise recorded at a train station (*TS*). Possible SNR settings are 5 dB, 20 dB and *clean* (no noise at all).

Table 5.3: Summary of reverberation conditions and related parameters.

Referred to as	Anechoic	Office	Corridor
Simulated environment	original room of recording	furnished office	narrow corridor
RT_{60}	70 ms	220 ms	530 ms
STI	0.986	0.96	0.467

CI strategy

The CI strategy used can be ACE or SAM. Tested channel stimulation rates of ACE are 720, 900, 1200, 1800, 2400, 3200 and 3500 pps, and the number of selected spectral peaks (*N*) is varied between 1 and 8. During comparisons, SAM's total stimulation rate is always picked to be equal to that of ACE.

Current spread model

A simple model of current spread is used to simulate the electrode-to-nerve interface by approximating the excitation along the cochlea (as shown in Nog08). An exponential decay function models the current density in the proximity of a stimulating pulse:

$$J_n(x) = e^{-|X(n)-x|/\lambda_c}, \tag{5.1}$$

where n is the number and $X(n)$ is the position of the stimulating electrode, x is the position on the cochlea where current density is to be measured, and λ_C represents the extent of current spread. Parameters x, $X(n)$, λ_C are measured in millimeter. Settings for current spread extent in this study are 0 (no current spread), 0.5 mm and 2.0 mm.

Extraction of binaural cues and statistics

The localization itself happens through generalized cross-correlation (GCC), using 30 ms window size, without overlapping. Mean and standard deviation of the localized degrees per window are calculated for each test file and used for further statistics. For each sound source position, the difference between the measured mean direction and the real direction is evaluated as the mean error. Similarly, the magnitude of the standard deviation over a test file reveals whether the localization has high certainty or high ambiguity.

Electrodograms

In this study, the output of the cochlear implant speech processing algorithms is always stored in the same matrix format, where the y dimension represents the CI electrodes, and the x dimension provides the time slots for possible pulses at the given total stimulation rate. (The matrix storage format ignores pulse-specific information like pulse shape, pulse width, and phase gap, but they are assumed to be identical among the strategies to be compared.) An ACE vs. SAM comparison is presented in Figure 5.15. It can be seen that due to the strict stimulation regime, most electrode channels of ACE contribute only little to the ITD-based binaural cues.
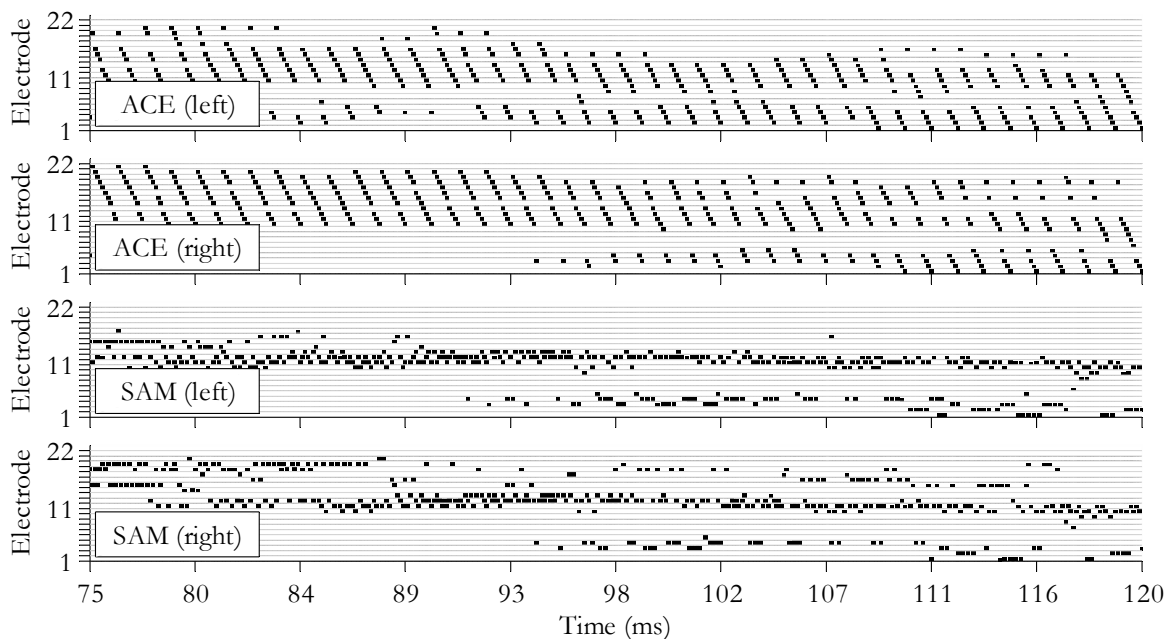
Figure 5.15: Electrodograms showing 45 ms of the CI stimulation patterns for the utterance "choice". The source was lateralized 65°. (Other conditions: *Anechoic* room, no added noise, *TPR*=9 kpps, λ=0, specific to ACE: *N*=10, *CSR*=900 pps.)

5.3.2.2 Results

This section presents the most important outcomes of the study. These are all based on the assumption that left and right CI processors are perfectly synchronized (which today is typically not the case). Estimations of the effects of missing synchronization to the localization performance are presented at the end of this section.

Localization performance can be best characterized by the error between the real and the localized directions of the sound source. Measured discrepancies for a predefined set of angles within a quadrant of the localization plane are used to yield a performance measure related to a specific setup. Figure 5.16 shows such comparisons between ACE and SAM.

In Figure 5.16, the central marks of the boxes indicate the medians. The 25th and 75th percentiles are represented by the bottom and the top edges of the boxes, respectively. The whiskers extend to the extreme data points, which are not yet considered outliers. Outliers are at least 1.5 interquartile ranges away from either end of the box. These are the default conventions regarding box-whisker plots all through this thesis.

The Shapiro-Wilk [SW65] test was run on groups of the results to be compared. It indicated for several groups that the data varied significantly ($\alpha = 0.05$, $p < 0.001$) from the pattern expected if it was drawn from a population with a normal distribution. Therefore, the non-parametric Wilcoxon two-sided rank sum test [Wil45, MW47] was employed to check subsets of the results for significant differences.
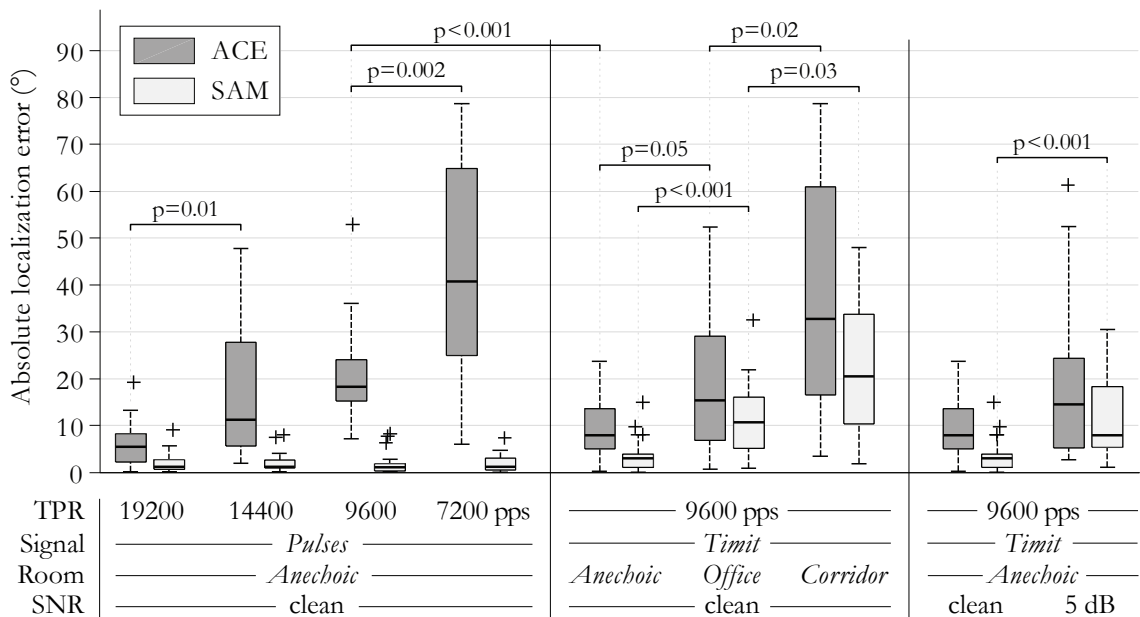
Figure 5.16: Box-whisker plots of absolute localization errors for signals played back from a distance of 3 meters at angles [5°, 10°, …, 90°]. Current spread λ_C was set to 0.5 mm. For ACE, $N = 8$ was fixed.

The null hypothesis was that compared data sets were independent samples from identical continuous distributions with equal medians, against the alternative that they did not have equal medians. Figure 5.16 presents the most influential effects on localization performance: pulse rate, amount of reverberation, noise level, and the type of the sound to be localized.

As expected, higher pulse rates allow for better resolution of ITDs, thus leading to smaller localization errors. Sound source localization using the ACE strategy at a channel stimulation rate of 900 pps (TPR=7200 pps) is practically impossible. There is a statistically significant difference (see p-values in Figure 5.16)[20] between the results obtained at TPR=7200 pps and at any other listed pulse rate. By contrast, the SAM strategy is barely affected by the pulse rate (at least within the tested boundaries). At each of the presented total pulse rates there is a significant difference (p=0.004 at TPR=19200 pps and p<0.001 in all other cases)[20] between the results obtained with ACE and SAM.

Reverberation and noise have similar effects on SAM as on ACE, but the magnitude of degradation is generally smaller with SAM. Between the reverberation conditions *Anechoic* and *Office* there is only 0.026 difference in STI (and 150 ms difference in RT_{60}), but, there is a significant difference (see p-values in Figure 5.16)[20] in the medians of the absolute localization errors measured in the two virtual environments. Raising the reverberation to the *Corridor* level results in further increase of errors. Though the medians of the localization errors in the presence of reverberation or noise seem smaller with SAM than with ACE, the Wilcoxon test rates the difference between the groups not great enough (α=0.05) to exclude the possibility that the difference is due to random sampling variability. That is, from the statistical point of view, ACE and SAM perform similarly bad under heavy noise or reverberation.

As for the effect of the choice of signal to be localized, the SAM strategy works similarly well with both the speech signal (*Timit*) and the pulse bursts (*Pulses*), whereas ACE performs significantly (p<0.001)[20] worse with *Pulses*.

The amount of current spread (at least the three settings tested) shows little to no effect on the localization performance (neither with ACE nor with SAM).

Asynchronous processors

Because of the block-by-block processing in ACE, the lag between the (unsynchronized) left and right processors may exceed one millisecond. This leads to huge ITD-based localization errors. The filter bank in SAM is based on the time domain simulation of the auditory system and processes input sound on a sample-by-sample basis. This method yields a maximum lag-range

---

[20] The test rejects the null hypothesis of equal medians at the 5% significance level with the given p-value.

of $\pm\frac{1}{2}\cdot TPR^{-1}$ seconds (about ten times shorter than in ACE), which leads to less localization errors even in unsynchronized systems.

However, a general problem with unsynchronized CI processors is as follows. Given that the internal clock speed of the two CI devices differ only a little, the direction of a fixed sound source may be perceived as continuously changing in a slow and periodical manner. This ever-changing percept of the sound direction may eventually lead to ignoring ITDs at neural level.

Conclusion

This study has shown that horizontal-plane localization of sound sources works well with SAM over a wide range of SAM's possible parameterization. The amount of cues for ITD-based localization preserved by SAM clearly exceeds that preserved by ACE, when compared using the same total pulse rate. Noise and reverberation seem to have less negative impact on the localization performance with SAM, but the difference is not significant. Furthermore, if the CI processors are not synchronized, time-domain filtering and sample-based processing in SAM makes the lag between the left and right devices less critical.

Nevertheless, these results are based on computer models and simulations thus completely ignoring real-world details as well as intraindividual differences in CI subjects.

5.4 Evaluation with cochlear implant users

In cooperation with the Cochlear-Implant Rehabilitationszentrum Thüringen (Erfurt, Germany) and with Cochlear Ltd., SAM was also tested in a pilot study with a group of five CI users. We investigated speech perception in quiet and in the presence of noise or reverberation, pitch discrimination abilities (for pure tones and sung vowels), and consonant discrimination. We also asked for subjective quality rating of speech and music snippets. Tests were repeated with the everyday strategy of the implantees and results were compared. Additionally, cognitive and linguistic skills were assessed, irrespective of hearing capabilities, by means of the visual text reception threshold test. Please note that most of the contents presented in this section have already been published in [HCK+14]. Their reuse was permitted by the coauthors.

5.4.1 Materials and methods

5.4.1.1 Subjects

Five post-lingually deafened adult CI users participated in the study. They were all native speakers of German and had at least two years of CI experience at the commencement of the study. Every subject had a Nucleus Freedom implant with a Contour Advance electrode together with

a Freedom sound processor from Cochlear. ACE was the everyday strategy of all study participants. More detailed demographic information is presented in Table 5.4. Bilateral (CI in both ears) users were tested only unilaterally by ignoring the later implanted side. Bimodal (CI in one ear, hearing aid in the other) users were tested exactly as unilateral implantees, since stimuli were streamed directly without the signal actually being played back aloud (see also section 5.4.1.5 on page 98).

Table 5.4: Subject demographics. *Deaf* denotes the duration in years the subject has spent unaided with practically zero bilateral speech intelligibility. *CI* shows amount of years since the first CI fitting. *CSR* and *N* denote the ACE parameters channel stimulation rate and number of selected spectral peaks, respectively.

Subject	Age (yr)	Deaf (yr)	CI (yr)	Most probable cause of hearing impairment	Lateralization	*CSR* (pps)	*N*
S1	37	3	4	Circulatory disorder	Bimodal	900	11
S2	70	1	5	Genetic	Unilateral	900	9
S3	69	15	2	Diphtheria	Bimodal	900	8
S4	50	1	5	Genetic / Traumatic	Bilateral	1200	10
S5	27	3	13	Meningitis	Bilateral	900	8

5.4.1.2 Fitting

All subjects were fitted with the SAM strategy as described in section 4.6 on page 71 at three different total pulse rates. After some exercise with each of these maps, every participant decided individually on which one to keep. Interestingly, all participants chose a map with a total pulse rate higher[21] than that of their everyday ACE map. For all presented ACE-SAM comparisons the chosen SAM map and the latest individual clinical ACE map was used.

5.4.1.3 Tests

Since as much as 10 hours per subject were allocated for fitting and testing (divided into five sessions), the performance of the implantees could be measured in various ways with a number of tests as follows. All comparative tests (ACE vs. SAM) of a given category were conducted on the same day, but separated by a 15-minute break.

Speech intelligibility in quiet

Speech intelligibility in quiet was planned to be assessed by means of the *Freiburg monosyllabic test* (German: Freiburger Einsilbertest [Hah53]). However, the first few tests with two sub-

---

[21] In average about 60% higher.

jects indicated saturation of results (using both ACE and SAM), so that a meaningful comparison could not have been done. Therefore, this test was aborted.

Speech intelligibility in noise

Speech intelligibility in speech-shaped noise was assessed with the OLSA test [WKK99, WBK99a, WBK99b]. The Oldenburg sentence test is a well-established sentence test in German language using the 50 % speech reception threshold scheme. The sentences are composed in a random fashion but always with the same structure (name, verb, numeral, adjective, object). The testee hears a five-word sentence with added noise and then repeats as much of the words as possible. Depending on the number of correctly recognized words, the level of the noise may be changed. The noise eventually reaches a level at which the subject is able to understand and quote two and a half words in average.

After only 20 sentences (about 4 minutes), the 90 % confidence region of the measured SRT value is shown to be ±0.82 dB [KLW+11]. The words are taken from a database (10 words per category), the modest size of which is responsible for some proven learning effects up to 2 dB after 120 sentences [WBK99b]. Despite that, the test-retest stability of SRTs has been shown to be high (0.7 dB) and learning effects to be low when using two training lists before testing [HHM+12].

We used two lists (one without and one with noise) for training purposes first and then presented 22 sentences in average to measure the SRTs.

Speech intelligibility in simulated reverberant environments

We also tested speech intelligibility in simulated reverberant environments using clean but reverberated OLSA sentences with four distinct magnitudes of reverberation. Twenty sentences were played back for each reverberation setting and the subject was asked to repeat them. The percentage of correctly repeated words was computed, whereas the first two words (of five in each sentence) were ignored, because these were not fully covered by reverberation, i. e., masked by reflected sounds.

Table 5.5: Summary of reverberation conditions and related parameters.

Configuration	Reverb-1	Reverb-2	Reverb-3	Reverb-4
Simulated environment	large living room	empty office	train station	wide stairwell (concrete walls)
RT_{60}	935 ms	1440 ms	1380 ms	2700 ms
STI	0.988	0.897	0.745	0.467

Reverberation conditions are listed in Table 5.5. STI values were calculated for the same (randomly selected) OLSA sentence at 65 dB SPL presentation level after applying the given reverberation effect. (For more information on STI see page 89 of section 5.3.2.1.)

Pitch discrimination thresholds with pure tones and sung vowels

We also measured pitch discrimination thresholds for pure tones and two sung vowels (an open central unrounded vowel with the IPA[22] notation "Λ", and a close front unrounded vowel wit the IPA notation "iː"). We measured in an adaptive three-interval three-alternative forced-choice ("3I-3AFC") procedure using the 1-up 2-down paradigm (see [Lev71]). Each presented sequence consisted of three tones (two identical reference and an odd one, each 600 ms long) separated by 400 ms pause. The F0-distance between the differing and the reference tones was varied adaptively with a quantization of one semitone. Frequencies (or fundamental frequencies in the case of sung vowels) of the tones, as expressed in notes, were determined to be symmetrical around the centre of the valid range for the given test variant (see Table 5.6). The initial distance was six semitones. The intensity of each tone was randomized by ±3 dB to reduce any unwanted effects of loudness variations on the subjects' ranking of pitches. Subjects were instructed to ignore loudness variations, if they perceived any. The task was to identify the tone that was different in pitch ("odd-one-out"), without having to tell which one was lower or higher in frequency. The discrimination threshold was then computed for each test tone type.

Table 5.6: Summary of conditions for the pitch discrimination test.

Tone class	Pure tones (C5)	Pure tones (C6)	Pure tones (C7)	Female sung "Λ" and "iː"	Male sung "Λ" and "iː"
Range	C4 (262 Hz) – C6 (1046 Hz)	C5 (523 Hz) – C7 (2093 Hz)	C6 (1046 Hz) – C8 (4186 Hz)	C4 (262 Hz) – F5 (698 Hz)	G2 (98 Hz) – A#3 (233 Hz)
Center of range	C5 (523 Hz)	C6 (1046 Hz)	C7 (2093 Hz)	G#4 (415 Hz)	D#3 (156 Hz)

Vandali et al. used sung vowels for pitch discrimination tests in an earlier study [VST+05] because of their relation to tonal languages. Won et al. states that for a pitch test very similar to ours the test-retest reliability was demonstrated to be good (intraclass correlation coefficient 0.85) [WDKR10].

---

[22] International Phonetic Alphabet

Consonant discrimination

We also assessed the discrimination ability of four consonant pairs (b/p, m/n, n/l, and k/t) using German minimal pair words. In each trial, either two similarly sounding (example in English: bark/park) or two identical words were played back in a sequence. There were 15 trials per consonant pair. The subject was asked to tell if the two words were the same or not. Percentage of correct answers was computed for each consonant pair.

Subjective quality rating

We also asked participants to rate the quality of speech and music in a double-blind experiment. A small inventory of five high-quality recorded sentences (three spoken by a female and two by a male colleague of the Cochlear-Implant Rehabilitationszentrum Thüringen, Erfurt, Germany) and five music snippets (each between 10 and 30 seconds long, from The Beatles, Brahms, Dire Straits, Spieltrieb [ST14], and Suzanne Vega) was prepared.

Our first approach was to play back the items in a random order (processed by either ACE or SAM) and ask the subjects to judge and score 18 pre-defined attributes (like warm, bright, strong, rough, firm, etc.) thus leading to a mean opinion score. However, the first tests made it clear that this kind of questioning was too complicated for the subjects. This was probably because subjects were not familiar with most of the attributes listed.

In the alternative approach a PC program shuffled an item, which was then played back twice (once with each CI strategy, the order of which was randomized). The subject's task was to express his or her preference for either the first or the second one. A neutral response was also possible. Subjects were instructed to pay attention to clarity, naturality and in case of speech to intelligibility. The test ended as soon as all items were played back once. The preferred strategy for speech and music was then determined for each subject by counting preferences and comparing the number of them.

Text reception threshold

The text reception threshold (TRT) test, developed by Zekveld et al. [ZGK+07], is a visual analogue of the widely used speech reception threshold test [PM79]. TRT "enables the quantification of modality-aspecific variance in speech-in-noise comprehension to obtain more insight into interindividual differences in this ability" [ZGK+07]. Meaningful use of TRT as a cognitive test has been demonstrated in several studies (see e.g. [BZK+12, HHM+12]) and Zekveld et al. found its test-retest reliability to be acceptable with an intraclass correlation coefficient of 0.74 [ZGK+07].

In our testing, the idea was to check whether TRT results (as a measure of cognitive and linguistic skills irrespective of hearing capabilities) would be in line with SRT results. Subjects had no known visual problem (by self-report) that would have deteriorated their reading capabilities. Participants started with two training lists in which the sentences were masked with dots or periodic bars, and the degree of masking was fixed at 15 %. The results of these were discarded. The actual measurement of the TRT took about 10 to 15 minutes and consisted of three different lists (20 sentences each) combined with three different maskers [HHM+12] as shown in Figure 5.17. For better visual comparability, TRT results are listed together with the measured SRT values in Table 5.7 on page 99.

Figure 5.17: Three types of masking used in the TRT test. (From top to bottom: random dots, periodic bars, and random bars.)

5.4.1.4 Assessment procedure

Within five sessions (each 90 minutes plus breaks), all tests were conducted with each participant using both the ACE and the SAM strategies. Subjects were provided a SAM-processed 10-minute excerpt of an audio book (Paulo Coelho's The Alchemist, in German language) prior to the measurement blocks (but at most once per session) to get accustomed to the new strategy. The text of the novel was available in printed form so that CI users could read along. Except for the duration of fitting and initial practice with SAM, subjects were blinded to the choice of processing strategy used in any test. As already mentioned on page 97, the subjective quality assessment procedure was even double blinded.

5.4.1.5 System setup

The Nucleus Implant Communicator (NIC) version 2 from Cochlear [Coc06, SM06] was employed to stream the stimuli directly from the PC to the CI. All computer programs developed and used during this study were able to process sounds by both the ACE and the SAM strategy (cf. section 4.5.1). This way, switching between strategies was easy for the operator and without attracting subjects' attention.

5.4.2 Results

This section summarizes the results of the tests listed in section 5.4.1.3. Nevertheless, before looking into the specific categories, one particular outcome needs to be emphasized: even though the stimulation patterns of ACE and SAM differ radically (cf. Figure 5.2 on page 75), every subject demonstrated good speech perception with SAM after only two to three minutes of use. Three of five subjects expressed acceptance of the new strategy after about 30 minutes of listening with it. This outcome is particularly important considering that CI users often require several weeks to get accustomed to a new strategy.

5.4.2.1 Tests based on the OLSA corpus

The standard OLSA test revealed that implant users S4 and S5 (already high-performer with the ACE strategy, i.e., OLSA SRT < 0 dB) could not benefit from switching to the SAM strategy in terms of speech intelligibility in speech-shaped noise [WKK99]. For the other three subjects (having about 10-15 dB worse speech reception thresholds using ACE than S4 and S5) the switch to SAM contributed to better SRTs. Results based on the reverberant OLSA corpus show similar trends: S4 and S5 did not gain performance with SAM, while the other three subjects achieved slightly better scores in at least two tests. For detailed results, see Table 5.7.

Table 5.7: Results of tests with the OLSA corpus. The first row shows speech reception thresholds (in dB SNR, lower is better) measured with the standard OLSA test procedure with speech-shaped noise. Rows two to five show percentage of correctly identified words of reverberant OLSA sentences at four fixed reverberation magnitudes (higher is better). Cells with gray background denote cases where the ACE strategy performed better. The last three rows show text reception thresholds (in degree of masking in percent, higher is better) measured with three kinds of maskers (random dots, periodic bars, and random bars, respectively).

Test configuration	S1		S2		S3		S4		S5	
	ACE	SAM	ACE	SAM	ACE	SAM	ACE	SAM	ACE	SAM
SRT (Standard)	4.9 dB	1.9 dB	6.3 dB	5.2 dB	10.9 dB	7.6 dB	-5.9 dB	-3.6 dB	-4.1 dB	-2.3 dB
Reverb-1	85%	91.7%	86.7%	86.7%	55%	80%	100%	96.7%	100%	100%
Reverb-2	83.3%	80%	75%	81.7%	51.7%	48.3%	100%	100%	100%	100%
Reverb-3	70%	73.3%	55%	70%	36.7%	36.7%	100%	93.3%	100%	95%
Reverb-4	20%	33.3%	15%	8.3%	18.3%	3.3%	71.7%	60%	61.7%	50%
TRT (Rnd. dots)	64.8%		53.5%		44.1%		61.9%		62.9%	
TRT (Per. bars)	61.9%		57.3%		53.3%		62%		55.5%	
TRT (Rnd. bars)	64.6%		54.2%		52%		60.9%		62.7%	

The given sample size (from five subjects only) does not allow for reliable analysis of covariance between SRT, reverberation, and TRT results. Still, it can be stated that the worst and second worst SRT results coincide with the worst and second worst TRT results, respectively. This may indicate cognitive factors being responsible for worse than average speech perception of subjects S2 and S3. By contrast, S1 reached best TRTs, whereas his SRTs are 5 to 10 dB worse than those of S4 and S5. This may indicate suboptimal parameters of the CI system and/or worse than average state of the auditory pathway.

As suggested by the results in section 5.3.1.2 on page 86, it is also possible that SAM's closer to natural processing helps the most in subjects with suboptimal cognitive states. There is, however, still insufficient evidence for or against this hypothesis.

5.4.2.2 Pitch discrimination

The SAM strategy was designed to provide a considerably amount more temporal pitch information than ACE does. Therefore, cochlear implant users were expected to discriminate smaller differences in pitch with SAM than with ACE. Test results showed that this expectation was reasonable: except for isolated cases, tests delivered better discrimination with the proposed new signal processing strategy. Table 5.8 shows 70.7 % discrimination thresholds (in semitones) as measured with the adaptive 3I-3AFC procedure for various signal types.

Table 5.8: Pitch discrimination thresholds in semitones (lower is better) measured using the adaptive 3I-3AFC procedure (with 1-up 2-down rule) that targeted 70.7 % correct discrimination level. Cells with gray background denote cases where ACE performed better.

Signal type	S1		S2		S3		S4		S5	
	ACE	SAM	ACE	SAM	ACE	SAM	ACE	SAM	ACE	SAM
Pure tones (C5)	8.5	2.3	2	4.6	3.9	2.2	2.3	1.4	2.5	1.5
Pure tones (C6)	8.7	3.3	2.5	3.3	2.5	1.8	1.7	1.5	1	1.3
Pure tones (C7)	4.1	2.8	2.7	1.5	3	1.5	3.5	3	2.3	1.8
Female sung "Λ"	6	10.3	5	6.4	7.4	6.2	6.6	4.3	7.5	5.9
Female sung "iː"	10.7	7.8	3.3	2.5	6.5	3.4	3.8	2	4	1.8
Male sung "Λ"	6.5	6	12.5	6	6.4	3.5	7.4	6.1	6.3	6.2
Male sung "iː"	7.7	4.5	13.5	7.7	10.4	4.8	6.6	4.8	7	6

Because the small number of samples does not allow for robust estimation of the distribution, the Wilcoxon two-sided rank sum test is employed to check subsets of the results for significant differences. The null hypothesis is that compared data sets are independent samples from identical continuous distributions with equal medians, against the alternative that they do not have equal medians.

There is a statistically significant difference ($p = 0.013$)[23] between the results obtained with ACE and SAM: the difference of the medians indicates a benefit of 2.6 semitones (ST) with the SAM strategy. (In addition, Wilcoxon matched-pairs signed rank tests reject the null hypothesis of zero medians for the difference in discrimination thresholds between paired samples, i.e., ACE vs. SAM for a given subject, at the 5% significance level for subjects S3, S4, and S5 with p-values of 0.016, 0.016, and 0.047, respectively.) This clearly shows that CI users are able to utilize the additional temporal information provided by the new strategy.

There is also a significant difference ($p = 4.12\text{e-}8$)[23] between the discrimination thresholds of pure tones and sung vowels: the difference of the medians indicates that sung vowels need an extra of 3.8 ST pitch difference to be discriminated at the same level as pure tones.

5.4.2.3 Consonant discrimination

Results of the consonant discrimination tests did not deliver clear trends. As shown in Table 5.9, CI users' performance was at about the same level with both strategies. The only notable outcome is that most of the results are better than chance level.

Table 5.9: Percentages of correct answers in the consonant discrimination test. Cells with gray background denote cases where the ACE strategy performed better.

Consonant pair	S1		S2		S3		S4		S5	
	ACE	SAM	ACE	SAM	ACE	SAM	ACE	SAM	ACE	SAM
b / p	86.7%	100%	100%	100%	86.7%	93.3%	66.7%	73.3%	100%	100%
m / n	33.3%	26.7%	66.7%	40%	40%	73.3%	80%	86.7%	67.7%	73.3%
n / l	40%	66.7%	60%	66.7%	66.7%	53.3%	80%	80%	60%	86.7%
k / t	93.3%	73.3%	80%	80%	86.7%	86.7%	80%	80%	100%	80%

5.4.2.4 Subjective quality rating

The results of the quality rating yielded some interesting findings. Subjects S2 and S4 preferred ACE to SAM for listening to speech. When listening to music, they did not favor either strategy (S4 even stated several times that "they all sounded awful" to him and expressed no preference among differently processed music snippets). S5 liked to hear speech coded with SAM more than coded with ACE, but had no preference with music, either. However, the answers of the two bimodal users (S1 and S3) were remarkable. They both perceived SAM stimuli to be more pleasant and natural, irrespective of the signal type (S3 actually "wanted to keep" this strategy).

---

[23] The Wilcoxon two-sided rank sum test rejects the null hypothesis of equal medians at the 5% significance level with the given p-value.

Since these subjects still had a more or less natural contralateral auditory perception to compare with (hearing aid in the contralateral ear), results suggest that stimulation via SAM elicits a closer to natural sensation.

Outcomes also indicated that the classical piece from Brahms was too complex for the participants. On the other hand, the music snippets from Spieltrieb and Susanne Vega were the most likeable items, probably because in these pieces a singing voice dominated while only a few instruments played.

Table 5.10: Number of times a subject preferred either of the processing strategies when listening to the specified signal. No preference appears as zero score for both strategies. Cells with gray background denote cases where ACE was the preferred strategy.

Signal	S1		S2		S3		S4		S5	
	ACE	SAM	ACE	SAM	ACE	SAM	ACE	SAM	ACE	SAM
Female speech (3 sentences)	0	2	2	1	0	3	1	0	0	3
Male speech (2 sentences)	0	1	2	0	0	2	1	0	0	0
Music (The Beatles)	0	1	1	0	0	1	0	0	1	0
Music (Brahms)	0	0	1	0	0	1	0	0	0	0
Music (Dire Straits)	0	1	0	0	1	0	0	0	1	0
Music (Spieltrieb)	0	1	0	1	0	1	0	0	0	1
Music (Susanne Vega)	0	1	0	1	0	1	0	0	0	1
Speech (total)	0	3	4	1	0	5	2	0	0	3
Music (total)	0	4	2	2	1	4	0	0	2	2

5.5   Evaluation with normal-hearing listeners

Clinical evaluation of new cochlear implant signal processing strategies can be very time-consuming and expensive. In addition, intra- and interindividual variability has to be taken into account. A simpler way to evaluate and compare CI strategies can be the acoustic simulation (also referred to as *auralization*) of cochlear implant hearing. These simulations generate an acoustic signal, which is similarly degraded as the signal presented through a CI. Hence, normal-hearing listeners can be used to gain insight into the sound perception of implantees. However, CI users report a dramatical change in perceived quality over time. A phenomenon made available by the plasticity of the brain. Therefore, an auralization is always just a rough approximation of what a CI user may hear.

Typical approaches of acoustic simulations of cochlear implant hearing were described by Shannon et al. [SZK+95] and Dorman et al. [DLR97]. In their algorithms, the input audio signal is band-pass filtered into m frequency bands. The temporal envelope is extracted from each

band and is subsequently used to modulate *m* carrier signals. The modulated carrier signals are finally recombined to generate an acoustic waveform. In numerous studies, the vocoder approach has been used with normal-hearing listeners to compare the performance of new signal processing strategies. However, this is often impractical, because current acoustic simulations with one specific parameterization generally mimic one specific CI strategy.

In our approach, we use a more general algorithm of acoustic simulation. Instead of extracting features from an audio signal and using them to synthesize a CI-like sound, we directly use the pulses generated by any CI signal processing strategy to modulate carrier signals. Since the algorithm uses a CI stimulation pattern as input, the algorithm is independent of the CI strategy used for generating the stimulation pattern. Furthermore, different technical and physiological aspects are modeled including current spread, loudness perception and frequency perception. (A more detailed description of the algorithm follows in section 5.5.1.2.) The auralization algorithm is used to evaluate and compare the ACE and SAM signal processing strategies in terms of speech intelligibility and pitch discrimination.

Please note that most of the contents presented in this section and its subsections originate from independent research of Anja Chilian and Elisabeth Braun, two former students supervised by me, see [Chi10b, Bra11]. We have also released a joint publication [CBH12] in 2012. The presented contents are reused with their permissions.

5.5.1 Materials and methods

5.5.1.1 Subjects

Twelve women and eighteen men were invited to participate in the study. Their ages ranged from 19 to 30 years. They all reported themselves normal hearing. Prior to the listening tests, they underwent a pure tone audiometry, regardless. Two subjects exhibited over 25 dB HL hearing threshold within the 250 Hz to 4 kHz range, bilaterally. A third subject had thresholds as high as 60 dB HL at specific frequencies. The answers of these three subjects were excluded from the results of the study.

The pure tone audiometry was performed by air conduction using an Oscilla® USB-300 device from Inmedico A/S running a semi-automatic test based on the Hughson-Westlake method [TW80]. The latter determines hearing threshold levels at 11 frequencies with 5 dB accuracy in about 5 to 10 minutes per ear. The results of the 27 participants (excluding the three subjects with worse than normal hearing) are shown in Figure 5.18. Please note that a small amount of uniformly distributed random jitter (in the range of ±1 dB) has been added to the plotted values for better visibility of the medians.
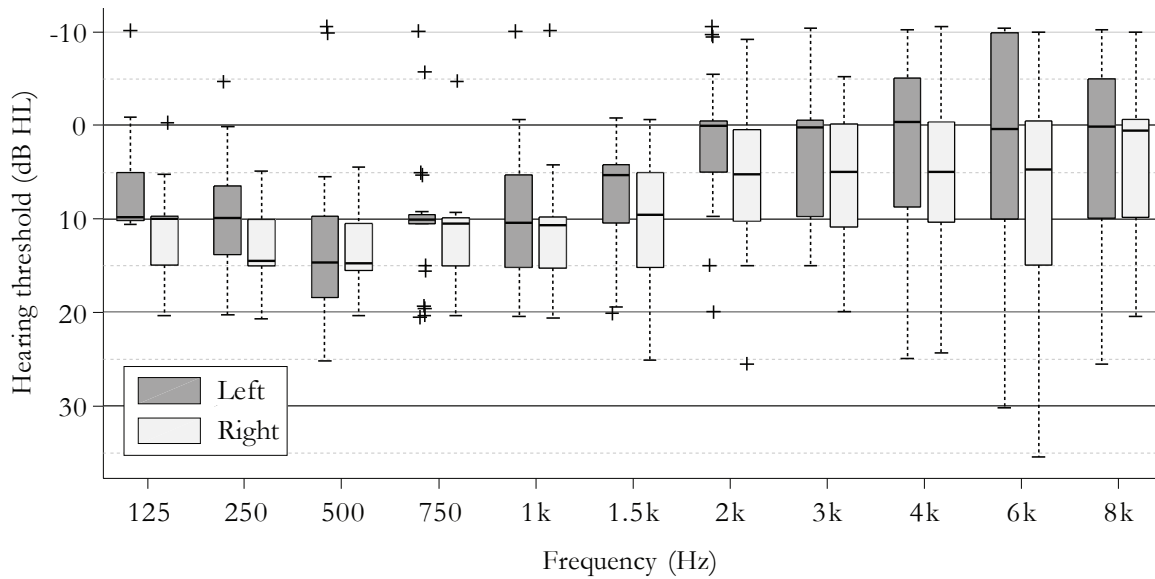
Figure 5.18: Box-whisker plots of the measured hearing thresholds of the participants.

5.5.1.2 Acoustic simulation

The acoustic simulation algorithm transforms a CI stimulation pattern directly into an audio signal. Different steps of signal processing mimic technical and physiological phenomena influencing speech perception in CI users. A block diagram of the main components of the algorithm is shown in Figure 5.19.

A stimulation pattern, which can be generated using any signal processing strategy, serves as input to the algorithm. First, based on the stimulation pattern, effects at the interface between the electrodes and the cochlear tissue are modeled. Stimulation current of the active electrode is spread widely along the cochlea due to the good conductivity of cochlear fluids. This limits the frequency resolution and therefore can degrade speech recognition abilities of CI users. The current spread λ_C is modeled as an exponential decay function as detailed on page 89.

Afterwards, simplified models of loudness and frequency perception are applied. First, current amplitudes are converted to simulate loudness perception in CI users. Second, transformed values are used for amplitude modulation of band-pass filtered carrier signals to simulate frequency perception and synthesize an audio signal. In addition, several simulation parameters can be adjusted to simulate different situations regarding speech perception of a CI listener.
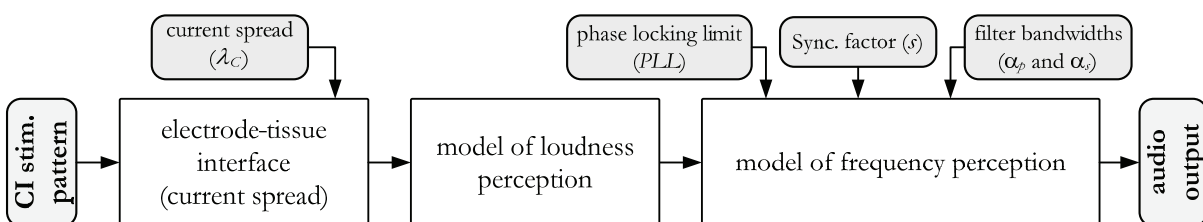
Figure 5.19: Overview of the acoustic simulation algorithm.

Model of loudness perception

Loudness perception in electrical stimulation depends on the stimulation current and the pulse width of the stimuli. We assume a constant pulse width of all electrical stimuli so that loudness perception is modeled to depend only on the current level. Calculated values of current spread for each cochlear section are therefore transformed using a function based on physiological data of loudness perception.

Fu and Shannon related the loudness function in acoustic hearing to the loudness function in electrical hearing [FS98]. The combination of their equations (1) and (2) yields the following relation between sound pressure level P_S and current level I:

$$P_S = k_p \cdot I^{4.5\dot{3}},$$ (5.2)

where k_p is an individual proportionality factor. In our model, the power function in Equation 5.2 is used to transform the calculated values of current spread into a value proportional to sound pressure.

Model of frequency perception

After transformation of the given stimulation current according to the current spread and to the loudness perception, an audio signal is synthesized simulating the frequency perception. Frequency information in electrical stimulation can be coded by both place and rate of stimulation. To simulate these two mechanisms, we extend the signal synthesis of the general vocoder approach by combining two different carrier signals. This approach is analogous to that by de la Torre Vega et al. discussed in [TMTQ04].

In common with the vocoder approach, modulation and band-pass filtering of carrier signals is applied. However, the present algorithm consists of two parallel synthesis paths, the results of which are combined. The first path simulates place pitch perception by using white noise as carrier. It is first band-pass filtered and then amplitude modulated using calculated values from previous stages of the simulation. In the second synthesis path, pulse trains are used as carrier to simulate temporal pitch perception. For this carrier, amplitude modulation is applied before band-pass filtering. The output values of the first and second paths are multiplied with the factors (1-s) and s, respectively, and then added together.

The filter bank consists of Butterworth band-pass filters. The center frequencies of the filters are consistent with the center frequencies of the CI electrodes. To determine edge frequencies of the filters, the center frequency f_c is first transformed into a position x_c along the cochlea using the following equation:

$$x_c = \frac{1}{a} \cdot \log\left(\frac{f_c}{A} + k\right),$$

(5.3)

with a=0.06, A=165.4, and k=0.88 [Gre90]. Then, two simulation parameters α_p and α_s are implemented to determine filter bandwidths. Parameter α_p determines the distance between positions along the cochlea associated with upper and lower passband frequencies around the position of the center frequency x_c, whereas α_s defines the distance between positions along the cochlea associated with upper and lower stopband frequencies. The calculated positions of the edge frequencies are subsequently transformed into frequency values using the inversion of Equation 5.3 (see Equation 2.1 on page 6). For sufficient overlapping of the filter bands, the parameter values should be α_p=0.75 mm and α_s=4.5 mm, which results in fourth-order Butterworth filters.

Temporal pitch perception with cochlear implants is additionally limited. Most CI users cannot discriminate changes above the so-called phase-locking limit. Therefore, we use an additional simulation parameter, which determines the phase locking limit (*PLL*). Up to this frequency, temporal pitch perception, i. e., synchronization with the rate of stimulation, is possible. For electrodes with center frequencies above *PLL*, the sound signal is generated using only white noise as carrier signal (i. e., s=0). Below *PLL*, both synthesis paths are used, the output values of which are combined according to the synchronization factor s.

Please refer to [CBH12] for more details and illustrations on the model of the frequency perception.

5.5.1.3 Tests

The presented algorithm of acoustic simulation was used to compare the CI signal processing strategies ACE and SAM. The strategies produced the stimulation patterns, which were processed by the acoustic simulation algorithm to synthesized sounds. These latter were presented to the normal-hearing listeners described in section 5.5.1.1.

Two of the tests described in section 5.4.1.3 were selected to be conducted with the normal-hearing listeners: the Freiburg monosyllabic test to measure speech intelligibility and the pitch discrimination test using the sung vowel "Λ" by both a female and a male singer. The former was conducted two times: once without added noise (*clean*) and once with speech-shaped noise [WKK99] from OLSA at 5 dB SNR. Each subject was presented with 15 monosyllables per noise condition and the percentage of correctly repeated words was determined. The pitch discrimination test was a 3I-3AFC test using the 1-up 2-down paradigm with one difference to

that described on page 96: subjects were not only asked to spot the tone different in pitch, but they also needed to specify if its pitch was higher or lower than the others'.

5.5.1.4 Assessment procedure

A test session took about 35 to 45 minutes per participant. At the beginning, each subject was briefly informed about cochlear implants, the aim of the measurement, and the tests to be conducted. Participants also had the opportunity to ask questions and to report their music listening habits. Then, hearing thresholds were determined by pure tone audiometry. Subsequently, two short excerpts of an audio book were played back (one minute acoustic simulation of ACE and one minute of SAM). Participants were allowed to adjust the loudness of the headphone to a comfortable level. This level was then maintained during the tests. Afterwards, the measurements began.

5.5.1.5 System setup

We used a beyerdynamic DT 770 headphone for playing back the signals. In the DT 770, the range of the variance in the frequency response is approximately 15 dB between 5 Hz and 35 kHz. Total harmonic distortion is below 0.2 %. The headphone was driven by an SB0490 external sound card from Creative Labs. It had linear frequency response between 8 Hz and 17.5 kHz (discrepancies smaller than 1 dB), which, considering the center frequencies of the simulated CI, was satisfactory for the task at hand.

Acoustic simulation parameters were $\lambda_C = 0.5$ mm, $PLL = 3000$ Hz, $s = 0.9$, $\alpha_p = 0.75$ mm, and $\alpha_s = 4.5$ mm. Strategy parameters were as follows: number of simulated CI electrodes $M = 22$, $TPR = 9000$ pps. $THL_i = 90$ and $MCL_i = 170$ for $i = 1, 2, \ldots, M$. For the ACE calculations, the number of selected spectral peaks $N = 10$ was set.

5.5.2 Results

Speech intelligibility tests showed a general increase of word recognition scores when comparing SAM to ACE. Further analysis (see next paragraph) indicated that this increase was statistically significant in the case of no added noise. Pitch discrimination thresholds were in all cases smaller with SAM than with ACE. Results show significant difference with both vowels. See Figure 5.20 for details. Please note that a small amount of uniformly distributed random jitter (in the range of ±1 % and ±0.1 semitone, respectively) has been added to the plotted values for better visibility of the medians.

The Shapiro-Wilk test was run on groups of the results to be compared. Except for the groups associated with the 1st, 3rd, and 4th (from left to right) box-whisker plots in Figure 5.20, the test indicated that the data varied significantly ($\alpha = 0.05$, $p < 0.007$) from the pattern ex-

pected if it was drawn from a population with a normal distribution. Therefore, the non-parametric Wilcoxon rank sum test was employed to check subsets of the results for significant differences. The null hypothesis was that compared data sets were independent samples from identical continuous distributions with equal medians, against the alternative that they did not have equal medians. In the indicated cases, the test rejected the null hypothesis of equal medians at the 5 % significance level with the given p-values as shown in Figure 5.20.
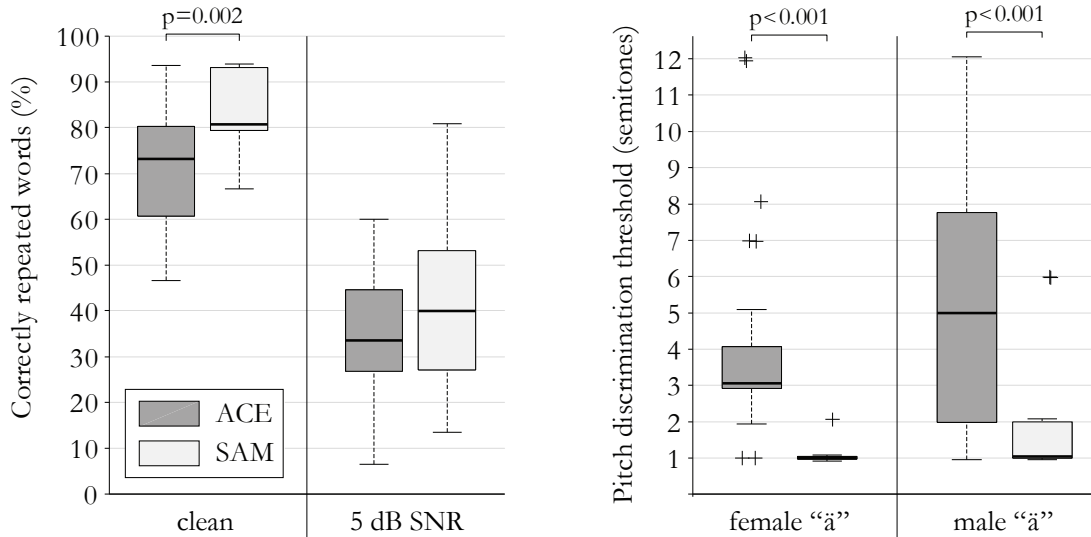
Figure 5.20: Box-whisker plots of the results of the speech intelligibility tests (on the left, higher is better) and of the pitch discrimination tests (on the right, lower is better).

5.6 Conclusions

In this chapter, a comprehensive analysis of the SAM strategy was presented. In addition to a system analytical approach, miscellaneous methods of performance evaluation via simulations and via listening tests with CI users and normal-hearing listeners were presented.

Independently of the choice of evaluation method, results are more than encouraging for SAM. Comparisons between results obtained using ACE and SAM reveal major advantages of the latter strategy. Much better pitch discrimination and sound source localization performance are two examples. Simulations also indicated that SAM may help implantees with impaired cognitive performance currently using ACE understand better.

Notwithstanding the above, the most important measure of success is the actual test with cochlear implant users. This test showed good quantitative results and good perceived quality, even though there was no possibility to give implantees enough time to get accustomed to the new processing algorithm. Further particularly important outcomes were the very quickly developing speech comprehension upon the first activation of SAM and the fast acceptance of the new strategy.

Chapter 6

Discussion

At the beginning of the work on SAM, there was one well-defined goal: make an auditory model based signal processing strategy and show it is better than others. During the development and evaluation process, I often had to make assumptions and simplifications, and also decisions on methods and procedures without knowing exactly what the opposite decision would have changed. As time passed, I noticed that all these uncertainties possibly could have made differences. Because of time and financial constraints, it is up to future research to follow up all these issues and to make use of the connection points left over for further investigation.

This chapter recapitulates key assumptions and their possible consequences as well as open questions related to the development (section 6.1) and analysis (section 6.2) of the SAM strategy, whereas Chapter 7 discloses plans of and shows directions for future research.

6.1   Issues related to the development of SAM

6.1.1   Requirements

As presented in section 4.1 starting on page 41, we declared main concepts and imposed high requirements on the SAM strategy and on its PC implementation. For the strategy, these were *stability* of the nonlinear auditory model over a large range of input levels and the presence of *psychoacoustic features* including cochlear delays, adaptation, compression and phase-locking. For the PC implementation, special emphasis was placed on *safety*, *speed*, *adaptability*, and *modularity*.

Section 5.1 provided evidence that all the requirements were fulfilled. Nevertheless, one could have put other requirements like real-time executability on a cochlear implant processor, limited power consumption, high-level individualization of the processing parameters, or the generation of as sparse as possible stimulation patterns minimizing the amount of "wasted" electrical pulses applied during refractory periods of the nerve fibers. None of these is currently featured in the implemented strategy, but all these possible future requirements are currently being worked towards, as discussed in Chapter 7.

6.1.2 Auditory model

As described in section 4.3.2, the auditory model used in SAM consists of models of the peripheral ear, the nonlinear mechanical filtering, and of the mechanoelectrical transduction. For the former two we employ models developed by Baumgarte [Bau00], whereas for the latter we use models developed by Sumner, Lopez-Poveda, O'Mard and Meddis [SLOM02].

Basically, it would be possible to use any combination of auditory submodels, as long as the compound model meets the requirements posed in section 4.1. We have chosen Baumgarte's model of nonlinear mechanical filtering, because it is based on cochlear hydromechanics, has had 40 years to mature (see section 2.3.1.1 on page 24), and promises great richness of detail. This choice has also implicated to use Baumgarte's outer and middle ear filter, which is fine-tuned to the subsequent basilar membrane model. The choice of the mechanoelectrical transduction model of Sumner et al. was based on the fact that it is widely accepted (as well as its predecessor by Meddis [Med86]) and almost ubiquitous in the field of auditory modeling.

Of course, there may be better and/or simpler (and computationally faster) models, but, unfortunately, no comprehensive comparison of auditory models exists (to my knowledge), which would declare one or the other model as best suited for use in cochlear implants. However, there are published studies comparing some specific aspects of auditory models (like masking [HC99], or compression [HE05]), or auditory models for other specific purposes (like for automatic speech recognition [Ste12]). A now outdated but extensive comparison of inner hair cell models was published by Hewitt and Meddis [HM91], which confirmed the advantages of the Meddis model. Some examples of more recent but not necessarily well-established inner hair cell models are those from Nam et al. [Nam05, NCG07] and by Lopez-Poveda and Eustaquio-Martín [LE06].

Eventually, key components and parameters of the SAM strategy responsible for good or bad performance of CI users need to be found (see also page 115 of section 6.2.3) to be able to decide if a better or a computationally faster auditory model would be more appropriate. For the meantime, the DRNL filter bank [MOL01] could be a candidate to substitute the nonlinear mechanical filtering part of the auditory model of SAM, because it offers a range of supported psychoacoustic features while it also promises a considerable optimization potential. For more information on this, see section 7.1.2.1.

6.1.3 Channel stimulation rate

Unlike ACE and other common strategies, SAM is not restricted to a pre-defined channel stimulation rate and it may activate electrodes in a stochastic manner. This mechanism allows the coder to distribute electrical stimuli among electrode channels as required, which we as-

sumed would better represent salient temporal features of the auditory model output in the stimulation pattern. As a side effect, however, channel stimulation rate may vary over time and might yield several thousand pulses per second (using the default parameters as shown in Table A.1 and Table A.2, starting on page 131). The refractory property of auditory nerve fibers would not justify the use of such high channel stimulation rate (see e.g. [Cla03:212-216]). This has at least been shown several times in combination with common strategies, see e.g. [FSC05]. However, the situation may 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 for pitch discrimination and contribute to speech perception. Future work should investigate this assumption. [HCH12]

The lack of a fixed channel stimulation rate in combination with a peak picking algorithm in the stimulus coder may also lead to a problem: A dominating static background signal (like a pure tone or a mixture of them) can make the transmission of speech impossible, because dominant spectral peaks are more likely to be selected by the stimulus renderer. While ACE, CIS, or FSP ensure the transmission of a given minimum number of spectral channels, SAM does not. Although the SAM strategy employs the "repetition penalty" mechanism (see page 61 of section 4.4.2) as a workaround for this issue, the use of a preprocessing algorithm to remove static background noise may be beneficial for speech understanding in noise.

Some studies (e.g. [BDG04, HK07]) have identified high channel stimulation rates as a possible source of fluctuations in perceived loudness. However, it was unclear if and how this artifact depended on the CI processing strategy under investigation. During our tests with cochlear implant users, we found no evidence of loudness fluctuation.

A positive side effect of high channel stimulation rates is the increasingly stochastic behavior of the peripheral auditory processing, since only a subset of the applied electrical currents will elicit time-correlated excitation of nearby nerve fibers. Such stochastic behavior in the electrical stimulation of the cochlea has been shown to have advantages [RWFA99, ZFM00, Wes04, LM08].

6.1.4   Hardware requirements

This thesis has illustrated that the SAM strategy is more complex than ACE. However, section 5.1.4 demonstrated that both strategies are real-time capable, and that for an input signal of linearly increasing length the required computational time is also increasing linearly, which gives a hint about the underlying time complexity. Further analysis also reveals that the amounts of

required working memory (related to space complexity) of the two strategies are in a comparable range.

Nevertheless, the most efficient implementations of ACE and SAM currently available require radically different amounts of computational power to run in real-time. Even though ACE computes a 128-point DFT 900 times a second (with *CSR*=900) using Fast Fourier Transform (FFT), it is able to run continuously on a Cochlear CI processor for about 24 hours powered by two button cells only (consuming about 1.7 Wh energy). Making SAM run on an energy-efficient PC for the same duration would consume about a thousand times more energy.

These facts underline the importance of properly matched hardware and software. ACE could not operate as energy efficiently if the algorithm was not executed on a customized piece of hardware. Conversely, SAM could operate approximately as energy efficiently as ACE does if it had its own customized hardware to run on.

Seen from this perspective, if SAM can prevail against other strategies in future studies, then it will be necessary to implement SAM on its own customized hardware to make it as efficient as possible. See section 7.1.2 on page 123 for an outlook related to this idea.

6.2   Issues related to the validation and evaluation of SAM

6.2.1   SAM vs. ACE: a fair comparison?

Contemporary auditory prostheses use various sound processing strategies. The most widespread are CIS, ACE, HiRes, and FSP as well as their variants (see section 2.2.4.2 on page 19). Even though it would have been much fairer to compare SAM with other experimental CI strategies (making use of auditory models or coding temporal fine structures), we have used ACE in comparison to SAM throughout this study. There are two reasons. First, ACE is currently the most prevalent CI signal processing strategy worldwide. Second, information on experimental strategies is generally scarcely available. Comparison against such a strategy would make it more difficult for the reader to understand differences between algorithms and to compare and reproduce results. Nevertheless, it is up to future studies to compare SAM to other strategies. Two interesting "opponents" would be FS4 and FAST (Fundamental Asynchronous Stimulus Timing) [SKJ+14], because both of them utilize phase information of the input signal and are likely to yield better results in pitch discrimination tests than ACE does.

As mentioned on page 16 of section 2.2.4.1, the standard ACE implementation employs a pre-emphasis filter as well as an automatic gain control circuitry at its input. We considered these stages as optional preprocessing and bypassed them consistently during testing, because the test sounds we used were recorded under ideal conditions and at a pre-defined constant

level. Ideally, all testing should have been repeated with the mentioned preprocessing methods turned on to see if they have an effect on the results. This is now up to future trials.

6.2.2   Computational evaluation

We believe that computational evaluation was an important step in the analysis of the SAM strategy. Although it could not fully predict CI user performance with the new strategy, it could contribute to more realistic expectations. Furthermore, it revealed properties of the strategies that would need more investigation.

The analysis of the features of the CI processing strategies (see sections 5.1.3 and 5.2) has shown that SAM clearly outperforms ACE in terms of psychoacoustic properties. To assess if these differences also make a difference in speech perception, we have chosen to use a modified version of the Hamacher model [Ham03] of the electrically stimulated auditory system to obtain a prediction of differences in speech reception threshold. I have witnessed some criticism of this model for its sometimes inexplicable results (like the tendency of values in the σ_{int}=0.15 column of Table 5.2 on page 86), but the general trends of the outcomes have been shown to be in line with expectations [Fre12]. The model results, in our case, showed only little differences between the performance of SAM and ACE in terms of speech reception threshold. However, with increased internal noise (a simulated condition for bad cognitive state) SAM gains the advantage. It would be inappropriate to state that the pilot study (with only five subjects) has proven this, but its outcomes did not refute the hypothesis, either.

Another outcome of the computational evaluation was the finding that SAM holds great potential for bilateral use. However, this conclusion is solely based on simulations and should definitely be verified with cochlear implant users as discussed in section 6.2.5.

Of course, neither the computational analysis of strategy features nor the use of methods of computational evaluation could substitute tests with CI users. On the downside, testing of CI signal processing strategies with implantees is expensive (patients' insurance and ethics approval), because of the inherent danger related to the direct in-vivo electrical stimulation of the human cochlea using non-certified software (and/or hardware). That is why developers of other new CI strategies also prefer to first use computational methods of evaluation.

Taft, for example, compares the number and level of pulse and noise stimuli between CI stimulation patterns generated either by the standard ACE strategy or by his ACE-based strategy using across-frequency delays [Taf09, TGB10]. (For details, see section 3.1.2 on page 32.) He also states that "Predictions of speech perception cannot be easily drawn from observations of stimulation patterns alone; such conclusions are best drawn from tests in patients. The met-

rics here only aim to quantify how different two stimuli are. This will guide the choice of parameters to test with patients in future chapters." [Taf09: 85]

Nogueira employed a vocoder interface for the evaluation with normal-hearing listeners as well as an acoustic model interface to be used with an automatic speech recognition system to evaluate his proposed CI strategies [Nog08]. The ASR system he used had been developed jointly and was used in several studies [HNSK07, NHE+07, SHKW09a, SHKW09b]. Nogueira justifies the use of computational evaluation methods as follows: "Results from CI subjects are generally subject to inter- and intra-subject variability. Results obtained from an objective method to measure speech intelligibility with signal processing strategies for cochlear implants would give more robustness to these studies." [Nog08: 87]

Kim et al. only performed acoustic simulation based evaluation with normal-hearing listeners [KCKK09]. This is definitely better than not involving listeners (whether CI users or normal hearing subjects) at all, but the lack of purely computational evaluation means that their sole results are prone to inter- and intra-subject variability.

Altogether, I think that computational evaluation methods are a meaningful addition to the experimental investigations. Although simulations are always based on models, which may miss important properties of the modeled entity, they are objective in a sense that subject variability is minimized and that they can be repeated any time yielding the same result.

6.2.3   Shortcomings of the pilot study

The most obvious shortcoming of the pilot study with CI listeners is the low number of participants. This is, however, not a flaw of the study design, but rather due to the lack of financial support. (As discussed on page 113, patients' insurance and ethics approval do represent significant costs.) It would definitely be desirable to expand on the original study by recruiting more participants, which could substantially improve the quality of the resulting statistics.

Should the pilot study be repeated or extended, a few things need to be revised. First, consonant discrimination tests need much more items to have a chance to yield significant results. Second, reverberation magnitudes listed in Table 5.5 on page 95 were poorly chosen (as we learnt during the tests): the difference between the configurations "Reverb-3" and "Reverb-4" should have been smaller. We have also learnt that trends in speech transmission index correlate well with trends in speech intelligibility of CI users under reverberant conditions, but there seems to hold a substantially different categorization of STI values for CI users from that valid for normal-hearing listeners.

A major factor definitely having influenced the results was the lack of time for the subjects to get accustomed to SAM. At the time of the study, all subjects had had at least two years of

experience with ACE. According to Rouger et al. [RLF+07], CI users generally need several months to develop good speech perception, and about two years to get fully accustomed to the signal processing strategy and reach their peak potential with the CI system (see also Figure 2.8 on page 22). By contrast, participants of the pilot study were only given 10 minutes per session to get accustomed to the new strategy, which, by all means, was too little. As already mentioned in section 5.1.4, the developed SAM software was meant to support mobile operation of the strategy. We planned to run it on a netbook (or a tablet PC), which could have been used for at least several hours uninterruptedly. (This would also have allowed for a more sophisticated A-B-A study design [HH76, HB78].) Even though that would not have been a "take home system", CI users could have used it inside the rehabilitation center hosting the experiments. Such an extended training with SAM could have had a positive effect on the results obtained with SAM. Unfortunately, because of limitations imposed by NIC v2 [Coc06], this was not possible at the time of the study. Now, it is up to future research to find the extent and course of the training effect with SAM.

Furthermore, it would also be interesting to conduct the same tests with users of cochlear implants of another manufacturer. Previous experience with a CI strategy that provides more temporal information than ACE could make a major difference in how well CI users can utilize the fine temporal structures provided in the SAM stimulation patterns. At the same time, it is also possible that implantees with no experience at all would benefit most from or improve at a faster rate after a treatment with SAM. To test these hypotheses, one should conduct a cohort study later on, once the advantages of SAM via acute tests are proven.

Critics say we have changed too much at once, so that causes for benefits with SAM cannot be identified unambiguously. As Zachary Smith (Coding Research Director at Cochlear Americas) once noted: "It seems that the biggest limitations of SAM are the unknowns. […] We do not know […] which parts are critical for its success." While this argumentation is true, this study has deliberately aimed towards a paradigm shift disregarding many traditions and limitations of previous signal processing strategies. In fact, it should be one of the next steps to identify key components and parameters of the SAM algorithm. This could most efficiently be done by combining computational and experimental studies with an in-depth analysis of the SAM strategy.

6.2.4   Psychoacoustic features

The compound auditory model embedded in SAM exhibits a number of properties known from the human auditory system (including psychoacoustic properties), which had been ex-

plored by the authors of the corresponding submodels (see e.g. [Bau99, Bau00, Med86, SLOM02, SLOM03]).

The first question is which of these features are still present in the output of the SAM strategy? We have managed to assess at least some of them during our investigations (as shown in section 5.1.3) and have found that tonotopy, group delays, adaptation, compression, and phase-locking are definitely present. The properties that are incorporated (at least in a simplified form) in all commercial CI strategies are place pitch cues (tonotopy), compression (loudness growth or map law function), and adaptation (multi-loop AGC). Features that are specific to some of the commercial strategies are temporal pitch cues (or phase-locking, in FSP/FS4/FS4p) and simultaneous masking (in MP3000).

As for masking, we recently started an investigation to see the extent to which masking effects of the auditory model are still present in the stimulation patterns generated by SAM. This work is still in progress, and, until now, we have found that simultaneous masking features (pure tones in white or filtered noise) are well preserved by SAM. Temporal masking (pre- and postmasking) has not yet been evaluated in-depth.

The second question related to psychoacoustic features of CI strategies, however, is, if CI users can make use of these. Even though some recent studies (like [WDKR10]) addressed this question, there does not seem to be a definite answer yet. Unfortunately, this thesis does not deliver the answer, either. However, it can be concluded that the question is relevant and that the clarification should be one of the next steps in future research.

6.2.5 Bilateral and bimodal use of SAM

Bilateral implantation is offered to a growing number of individuals in order to provide benefits of binaural hearing. One of these is the ability to localize sound sources. It has been shown, however, that localization performance with CIs is limited [GAR+07]. Depending on the strategy and its settings, localization can also be completely impossible [HWSB10]. The most common causes for difficulties are missing phase information and independent (asynchronous) operation of the CI processors (further worsened by block-by-block audio processing). As shown in section 5.3.2, SAM could also be employed in bilateral configurations, providing all known benefits of bilateral CI use, but showing superiority in terms of delivering cues for sound source localization. Because SAM is based on the time domain simulation of the auditory model, input data can be (and is) processed on a sample-by-sample basis. A positive side effect of this is that the largest possible lag between the two unsynchronized processors is TPR^{-1} second, which is only a fragment of that of ACE [HWSB10]. In addition, SAM provides a lot more phase information, which is crucial for localization tasks. [HCH12]

Given all these advantages and the promising simulation results presented in section 5.3.2.2, it would be useful to conduct binaural experiments with SAM.

Furthermore, it should be noted once more that both bimodal (unilateral CI and contralateral hearing aid) participants of the pilot study reported a more natural percept through their implant when using the SAM strategy. That makes SAM a perfect candidate for further bimodal tests. Considering that SAM also models cochlear delay trajectories, it would not be unexpected if bimodal use in the same ear (electric-acoustic stimulation, see page 10 of section 2.1.2.2) would also elicit percepts better than ACE could produce in combination with the hearing aid.

Chapter 7

Future research and conclusions

Previous chapters showed the development of an idea motivated by normal hearing, auditory modeling and state of the art of auditory prostheses towards the working and safe prototype implementation of the SAM signal processing strategy. Even though the development and evaluation work might seem finished, it is not. As already mentioned on page 109, time constraints forced us to ignore some particularly interesting ideas, questions, and scenarios, which arose throughout the course of the research. These are reviewed in the following section.

Subsequently, the chapter sums up the main research contributions of this dissertation and then ends with a few personal closing remarks.

7.1 Plans and directions for future research

While the proposed SAM strategy is a progression compared to the state of the art technology, there is still potential for improvement. Section 6.2 has already suggested further steps of evaluation, which may provide indications of strengths and weaknesses of SAM. This section also presents a number of worthwhile starting points for future research. These are divided into three categories as follows. Section 7.1.1 summarizes possibilities to make the strategy better in terms of energy efficiency and quality-related features. Section 7.1.2 is dedicated to ideas on how to make the calculation of SAM faster, without removing components or important properties. Finally, section 7.1.3 presents two new concepts possibly useful in future approaches of computational auditory processing, also applicable to SAM.

Some of the above directions represent ongoing research in our lab and they are identified as such in their respective sections. Regardless, I would be honored to be informed about and to provide help with research related to any aspect of the SAM strategy.

7.1.1 Enhancements

7.1.1.1 Sparse coding

One field of ongoing research in our laboratory is making the SAM stimulation patterns sparser. Sparser stimulation can conserve energy by eliminating electrical pulses that would

normally be applied during refractory periods of surrounding nerve fibers. For that, we constructed a function (see Figure 7.1) that is used to control the probability that the coder chooses an electrode in subsequent stimulation cycles that has just fired. We call it the electrode selection probability control (ESPC) function.
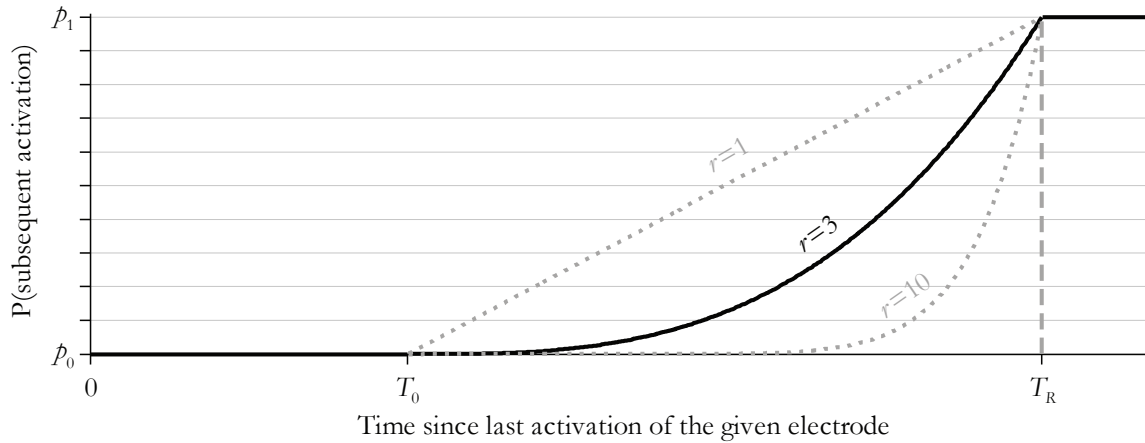
Figure 7.1: Illustration of the ESPC function currently used to control the probability of subsequent activation of an electrode in the sparse SAM strategy. T_0, T_R, r, p_0, and p_1 denote dead time, reduced activation probability time, the exponent of the power function, and the minimum and maximum probabilities, respectively.

This enhancement is very easy to insert into the current SAM implementation. A performance decrease as compared to the standard SAM strategy can also be prevented by using lookup tables. A current implementation (under testing) uses T_R=1 ms, r=10, p_0=0, p_1=1, whereas repetition penalty (section 4.4.2) and peak selection randomness (section 4.4.3) are turned off. The values of the function shown in Figure 7.1 are calculated and updated for each electrode in every stimulation cycle and used to weight filter bank values $L_m[i]$ (see Equation 4.7 on page 62) just before peak selection takes place. The results for various T_0 values are compared in Figure 7.2. However, both the approach and the results are preliminary.

The members of our lab are currently in the process of finding the most suitable spot of the stimulus coder to apply the weighting. Additionally, it is in discussion how the approach should include effects of neighboring electrodes and electrical pulses with amplitudes near to threshold level. To find answers to the above questions, we plan to create a more sophisticated model of the electrode-tissue interface along with the connected nerve fibers. Its (obviously computationally intensive) simulation will be used to find the effects of various conditions regarding electrical stimulation with SAM and to deduce a simple, yet improved, model and its optimal parameters to make stimulation patterns sparser, while keeping useful features of SAM.

Furthermore, it would be possible to have differently tuned ESPC functions per electrode. For that, however, it would be necessary to know the individual cochlear anatomy and status of

the auditory pathway as well as the exact intracochlear position of the CI electrode array of a given implantee. See also section 7.1.1.3 for ideas related to models of the three dimensional cochlea.
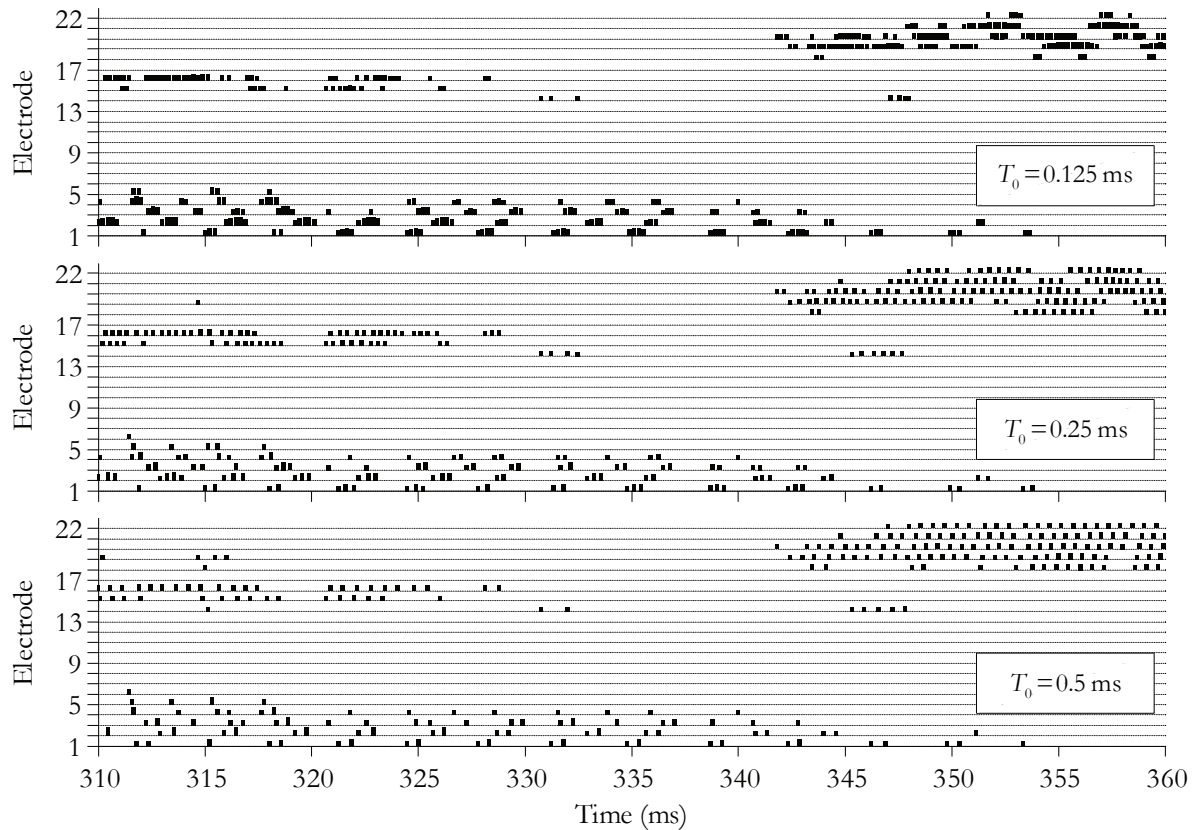
Figure 7.2: Electrodograms of sparse SAM stimulation patterns generated using T_R=1 ms, r=10, p_0=0, p_1=1, and 0.125 ms, 0.25 ms, and 0.5 ms dead time, respectively. Repetition penalty and peak selection randomness were turned off.

7.1.1.2 Parallel stimulation

The comparison of Figures 4.14 and 4.16 (on pages 59 and 64, respectively) makes it clear that only a highly simplified version of the auditory model output reaches the cochlear implant electrodes. The information mapping would be more precise if CI electrodes could be activated simultaneously. This would not only help represent spectrally complex scenes better, but could potentially counteract spectral smearing caused by current spread in the cochlea. The idea is not new and its mathematical background has also been established lately [Zie01, BDG04, HK07, ZRH+08]. Of relevance are the hardware and the insufficiently researched consequences of this kind of stimulation.

Current CIs from Cochlear do not feature multiple current sources [ZRH+08], which means that simultaneous activation of multiple electrodes is not possible. Although MED-EL implants do have multiple current sources, the system is restricted to eliciting electrical stimuli with com-

mon polarity [Zie01]. "With this constraint a flanker electrode cannot deliver an inverted current to cancel the skirt from an adjacent center electrode. Therefore, no stimulus voltage profile can ever be made narrower than the original monopolar spread function, and no focusing can be achieved" [HK07: 3713].

Van den Honert and Kelsall experimented with a modified system from Cochlear [HK07]. They replaced the transmitting coils (a bottleneck regarding implant power and available bandwidth for data to be transmitted) with a transcutaneous connector in two CI users to gain direct and unlimited access to the electrodes. They showed that simultaneous stimulation could indeed be used to defeat current spread and achieve focusing. However, they also found that perceived loudness decreased when stimulus focusing increased. Consequently, impractically high currents would have been required to maintain comfortably loud perception. Furthermore, it can be speculated that damaged cochlear regions become more important when applying focused stimulation.

A different but related approach is the incremental encoder by Blamey and colleagues [BDG04]. They conclude that there is "considerable evidence that the neural excitation evoked by a single pulse is spread spatially along the cochlea, and that specific loudness is summed across auditory channels within a time window of the order of milliseconds" [BDG04: 50]. Consequently, a truly parallel activation of electrodes is not necessary to give rise to "virtual electrodes". At the same time, focusing is not possible without simultaneous stimulation.

Work is ongoing in our lab to check which of the above-mentioned methods would best map the auditory model output to the CI electrodes while preserving more details than the current system does.

7.1.1.3 Effects of the 3D cochlea

The intracochlear current spread is generally approximated by an exponential decay function as shown on page 89. A 3D model of the implanted cochlea could help estimate the error of such a simple approximation.

Recently, a detailed 3D model of the implanted human cochlea was developed by Chilian [Chi13] at the Fraunhofer IDMT. The development of the geometrical model of the cochlea and of the implanted electrode array was based on histological cross sections and on publicly available specifications of the Nucleus straight electrode array by Cochlear, respectively. The 3D model was then transformed into a volume conduction model so that electrical potential distribution could be estimated by the finite element method. Beyond that, Chilian also managed to couple this model with a model of the electrically stimulated auditory nerve fibers,

through which the neural excitation patterns elicited by the cochlear implant stimulation can be simulated.

Currently, the model by Chilian is based on generalized parameter values. Outcomes of the simulations (like the extent of the current spread and its effects on refractoriness of nerve fibers) are incorporated in the early version of the sparse SAM strategy discussed in section 7.1.1.1. There are, however, approaches to parameterize 3D models of the cochlea based on measurements of single CI users (see e.g. [Mal09]). As a result of this individualization, the model would reproduce cochlear features of the given CI user much more faithfully and could provide valuable information for the CI fitting on a completely different (and higher) level. Should post-operative imaging become routine in cochlear implant treatment one day [CCA+14], so could CI strategies and CI fitting based on individualized cochlea models quickly become reality.

7.1.2   Ways of boosting the calculation speed

7.1.2.1   Complexity reduction

From Figure 5.8 (page 81) it is obvious that the auditory model simulation takes more than 50 times as long as the calculations comprising the stimulus coder. This also means that the AM is the most complex part of the SAM strategy and hence a perfect candidate for optimization. As discussed in section 6.1.2, the widespread DRNL model could be a promising alternative to the Baumgarte model.

The Baumgarte model evolved from the early hydromechanical laws laid down by Peterson and Bogert [PB50], Ranke [Ran50], and Oetinger and Hauser [OH61]. Its maturation required about 40 years and involved contributions from numerous scientists (see page 24 of section 2.3.1.1), which makes it reasonably trustworthy. However, since it is a transmission line model, the whole length of the cochlea needs to be simulated with a given minimum resolution even if the velocities only at some specific BM sections need to be processed further. For 22 (or less) such places, a phenomenological point model (like DRNL) might be more efficient. A point model does not have connections between the individual filter units tuned to different characteristic frequencies. This is different from the ear (and from the Baumgarte model) where different points on the basilar membrane are coupled. This also means that with a point model one can calculate the velocity of specific BM sections without having to calculate more than that. Kim et al. suggest that a simplified version of the DRNL filter bank could also be implemented in current CI processors [KCKK09: 835].

Work is ongoing in our lab to see if DRNL can effectively replace the current model of nonlinear mechanical filtering in SAM while being computationally more efficient.

As for the model of mechanoelectrical transduction, it is quite difficult to find a suitable alternative to the model of Sumner et al. Although alternatives do exist (see page 29 of section 2.3.2), most of them seem to be more complex than the currently employed model. The most promising (and probably the computationally least intensive) alternative may be the model of Lopez-Poveda and Eustaquio-Martín [LE06], which they also offer to provide as MATLAB source code. However, one drawback of this model is that its published parameters are based on guinea pig experimental data.

An alternative to reducing complexity is to "let physics do the computing"[24] and implement parts of the SAM strategy (that are computationally expensive on a microprocessor) on a more suitable piece of hardware. For more information on hardware implementation please read on to the next section.

7.1.2.2 Hardware implementation

As discussed in the previous section, hardware implementation of (parts of) the auditory model could significantly reduce the time needed to compute SAM stimulation patterns. That is because the AM represents sections of the cochlear structure, and, ideally, the simulation of the processes within each of these sections could be computed in parallel.

The above also implies that a "simple" DSP port of the auditory model code will not execute substantially faster on a single (or few) core digital signal processor. We thought it could be different with a general multiprocessor architecture. Our first target platform was an NVIDIA GeForce GTX 480 (1.4 GHz engine clock, 480 cores, 1536 MB GDDR5 memory with a bandwidth of 177.4 GB/s) using CUDA[25]. We started by executing the simulations related to the (uncoupled) inner hair cells in parallel, but found that managing the threads consumes almost as much time as the time benefit from the parallelization. As a consequence, a reasonable performance gain could only have been achieved if the entire auditory model (including nonlinear mechanical filtering) was parallelized. Unfortunately, we have found that the Baumgarte model is not well suited for parallelization on the GPU[26] architecture, because it incorporates lateral coupling between BM sections, which implies massive communication between threads, which is very time expensive. (By contrast, the DRNL filter bank can in fact be parallelized efficiently [San12], because its bands are not coupled to each other.)

A feasible solution could be to create a VLSI implementation of the auditory model, for example on a field-programmable gate array (FPGA). There are already numerous examples of

---

[24] A favorite saying of my former professor and supervisor, Tamás Roska, who I admired very much.

[25] Compute Unified Device Architecture

[26] Graphics Processing Unit

VLSI cochlea models [Wat93, HJST08, WB09], which could serve as a useful base for the implementation work.

7.1.3 Two new concepts to be used in computational auditory modeling

7.1.3.1 CNN universal machine

Cellular neural networks (CNN) are a parallel computing paradigm allowing direct communication between neighboring units only, as shown in Figure 7.3. The paradigm and its use in signal processing scenarios (initially image processing) are well described by Chua et al. [CYK91], who have also proposed an efficient VLSI implementation. All the units (or cells) of a CNN consist of the same circuit containing linear and nonlinear elements (capacitors, resistors, controlled and independent sources). The interconnection is realized by controlled sources, the weights of which can be split into two groups: feedback weights (called *template A*) and control weights (*template B*). The global behavior (or the "program") of a CNN is characterized by templates A and B and a bias I. Cells not directly connected may still affect each other's state indirectly through the propagation dynamics of the network.
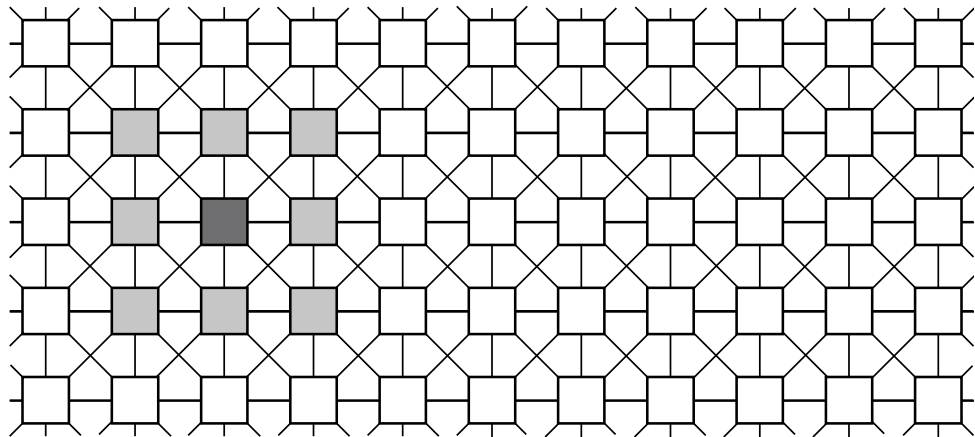
Figure 7.3: A 5x11 segment of a two-dimensional cellular neural network. As an illustrative example, one cell (dark gray) and its "3*3-neighborhood" (the cell itself and its light gray colored neighbors) have been emphasized.

In 1993, Roska and Chua introduced the CNN universal machine (CNN-UM) [RC93], in which analog (CNN) and digital circuitry (handling program logic) coexist, hence the alias *analogic array computer*. CNN-UMs have been shown to be Turing equivalent [CC96], which means that they can simulate any other real-world general-purpose computer or computer language.

Although CNNs were initially introduced as real-time image processors, they have been successfully used in subsequent studies related to complex spatio-temporal dynamics, self-

organization, and simulations of biological neural networks. The first publication (and the only one I am aware of) on CNNs in the field of cochlear implants originates from Zadák and Unbehauen [ZU93]. They managed to use CNNs to extract LPC-features from speech, which were meant to be used in a "hypothetical 22-channel cochlear implant system". Five years later, CNNs were also shown to efficiently compute the Discrete Fourier Transform [PFTP98].

I first heard about CNNs during my master studies at the University of Veszprém (Hungary). Later on, at the Pázmány Péter Catholic University in Budapest, I designed my first templates and CNN-UM *analogic algorithms* for simple image processing tasks, but remained quite inexperienced in the field. Still, considering the advantages of CNNs as outlined in the previous paragraphs and taking into account the properties (like *wave propagation*, *nonlinearity*, and *interconnectedness*) of the components of the auditory system, it seems apparent that CNNs could be used for tasks related to auditory simulations.

In 2007, Pazienza and Karacs published a method to automatically generate CNN-UM programs for given input-output pairs [PK07]. Encouraged by this possibility, I contacted the authors to clarify if CNN-UM-based simulation of parts of our auditory model would be feasible [PK08]. The outcomes of first considerations looked promising, but, unfortunately, our communications faded. The idea, however, to use cellular neural networks to efficiently simulate auditory processing is still there waiting to be explored.

7.1.3.2 Hough-transform

The Hough-transform is a feature extraction method generally used as part of a larger procedure to find imperfect instances of predefined shapes or objects within digital images. The algorithm uses a voting technique in a parameter space. Object candidates are determined as local maxima within the so-called accumulator space. The Hough-transform was generalized by Ballard [Bal81] after the related patent and paper of Paul Hough [Hou59].

Hubel and Wiesel demonstrated the existence of orientation columns in the visual cortex of the macaque brain [HWS78]. These structures are believed to perform a kind of parallel Hough-transform to extract shapes in real-time to serve the orientation of the monkey [Bla92].

We hypothesize that through the functional conservation property of the brain a similar function to that of the orientation columns exists in the auditory cortex to classify various cochlear delay trajectory shapes (and patterns) [HSK08]. The existence of structural conservation is already supported by the fact that columnar organization has also been found in the auditory cortex of cat and monkey [GC02].

We have experimented with various cochlear modeling stages to generate auditory images (AI) of individual phones [HSKK06, HKK06, HWS08] and, in later studies, complete utter-

ances [HSK08, SHKW09a, SHKW09b]. After an (optional) image stabilization step, which removed group delays (see section 4.3.3.3 on 56), a proprietary implementation of the generalized (parallel) Hough-transform was used to extract curvature information from the auditory images. Subsequently, we employed different classification methods to assess vowel or speech recognition rates.

Figure 7.4 shows the waveform, the stabilized auditory image of the basilar membrane displacement, and the accumulator space calculated via Hough-transforming the AI. Local maxima (darkest points within shaded areas) in the latter correspond to the most dominant curvatures occurring in the auditory image. In the depicted configuration, a local maximum at a normalized curvature value of 0.5 means a straight vertical line segment at the corresponding time instance in the AI. Curvatures of zero and one correspond to the steepest hyperbolic trajectories to the left and to the right, respectively, which the Hough-transform has been configured for.
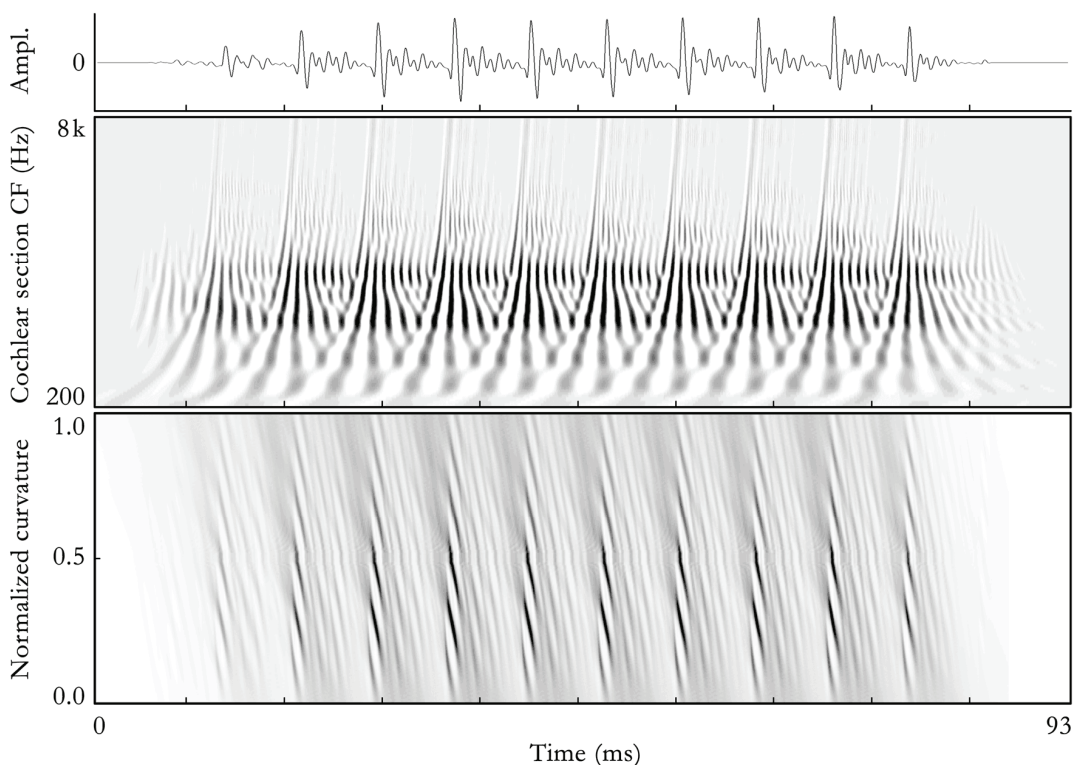
Figure 7.4: Different representations of the phone "*aː*" spoken in the word "harm": waveform (top), stabilized image of the basilar membrane model output (middle), and a time-aligned representation of the Hough-transform's accumulator space (bottom).

Our experiments confirm that cochlear delay trajectories contain features that can be extracted by the Hough-transform and be used in speech recognition tasks. Even though automatic speech recognition based on MFCCs (Mel-frequency cepstral coefficients) yields gener-

ally higher scores, ASR results using the extracted curvature information are more noise robust [HSK08]. Szepannek et al. showed that features acquired from the Hough-transformed auditory images could be used complementary to MFCCs to achieve higher ASR scores than with MFCCs alone [SHKW09a, SHKW09b].

Unfortunately, we have not been able to follow up with research on Hough-transform based speech recognition since 2009, but the approach still seems to be promising and possibly also useful in other fields of computational neuroscience.

7.2   Summary

This dissertation set out to investigate how a model of the human auditory system can play an important role in the signal processing of cochlear implants. Beyond the physiological and technological basis, the development and evaluation of the SAM strategy have been discussed in detail. The final chapters also showed room for improvement as well as possible starting points for future research. The main outcomes of this work can be summarized as follows:

- SAM came into being as a real-time capable PC implementation along with a variety of software tools for evaluation of CI signal processing strategies;
- SAM proved to have the potential to get accustomed to in a very short period of time (see section 5.4.2) and to provide good results even in acute tests;
- SAM was shown to provide significantly more pitch cues than the reference strategy, leading to much better pitch discrimination (see section 5.4.2.2);
- SAM was shown to provide significantly more interaural time difference cues than the reference strategy (see section 5.3.2.2), which may lead to substantial binaural benefit in CI hearing;
- Model results hinted that SAM might help subjects with impaired cognitive state understand speech better (see section 5.3.1.2).

Although the development and evaluation of SAM are complete within the scope of this thesis, SAM still offers a huge potential for further research.

7.3 Closing remarks

First, a remark I often miss from other scientific publications, which is related to resources not publicly available and utilized during the work: Upon request I can provide signal databases (e.g. the impulse responses used in the experiment explained on page 95 or the sung vowels mentioned on page 96 of section 5.4.1.3), SAM-processed signals, and –in specific cases– also executable SAM code. We are also currently working on a MATLAB version of the auditory model employed in SAM, which we plan to release as soon as practicable. I would be happy to support any research using these resources.

Finally, some personal remarks: At the beginning of my PhD studies, I wanted to learn as much as possible about cochlear implants. I would never have thought that working in an international and multidisciplinary scientific team at Fraunhofer IDMT would broaden my horizons in such a diverse way. Beyond the expert knowledge of cochlear implant related signal processing I gathered, I learned how to design, perform, and evaluate listening experiments. I even discovered statistics (and related software packages) to assess correlations and significance of the results. I substantially improved my programming skills (including coding style and version controlling) and acquired extensive experience with commercial software tools (examples include Adobe Creative Suite, Mathworks MATLAB, Microsoft Office, National Instruments LabVIEW, and Sony Sound Forge, just to name a few). I also recognized the importance of reference and knowledge management and imported about 400 of my favorite papers into Citavi, a reference management program by Swiss Academic Software. Furthermore, I rediscovered the Library of the Ilmenau University of Technology as the perfect place to write my thesis.

To sum up, I think I grew a lot during the SAM project, for which I am deeply grateful.

Appendices

A.1 SAM parameters

Table A.1. Parameters of the auditory model used in SAM. Parameters marked with a hash symbol ([#]) are not explicitly referred to in this thesis. Parameters marked with a circle (°) have been tuned. "Model of the nonlinear mechanical filtering", "inner hair cell", and "neurotransmitter" are abbreviated as NLM, IHC, and NT, respectively.

	Parameter description	Symbol	Value(s)	Ref.
Outer and middle ear	Outer ear model resistor	R_O	6.35 Ω	[Bau00]
	Outer ear model capacitor	C_O	1.575 μF	
	Outer ear model inductor	L_O	1.7875 mH	
	Middle ear model resistor 1	R_{M1}	12.7 Ω	
	Middle ear model resistor 2	R_{M2}	31.75 Ω	
	Middle ear model inductor	L_M	3.175 mH	
Nonlinear mechanical filtering	Number of NLM sections °	N_{BM}	101	[Bau00]
	Resolution of the NLM °	Δz	0.25 Bark	[Bau00]
	NLM pre-amplification °	g_{pre}	-42 dB	-
	NLM post-amplification °	g_{post}	-86 dB	-
	NLM friction loss resistor	R_q	400 kΩ	[Bau00]
	NLM endolymph mass scaling factor °	g_{Cq}	0.6	[Chi10a]
	NLM endolymph mass capacitor °	C_q	18.75 nF	[Bau00]
	NLM section n scaling factor °	g_{Cn}	3.0	[Chi10a]
	NLM section n mass capacitor °	C_n	(4.2·n+310) nF	[Bau00]
	NLM helicotrema resistor	R_H	281.1 kΩ	
	NLM helicotrema capacitor	C_H	56.6 nF	
	NLM amplification factor 1	g_1	100	
	NLM amplification factor 2	g_2	20	
	NLM saturation factor 1 °	g_{s1}	5	
	NLM saturation factor 2 °	g_{s2}	200	
	NLM to IHC cilia coupling gain [#]°	C_{cilia}	-65 dB	-
	NLM to IHC cilia time constant [#]	τ_c	2.13 ms	[Med06a]

	Parameter description	Symbol	Value(s)	Ref.
Biophysical model	Max. mechanical conductance [#]	G_{cilia}^{max}	8 nS	[Med06a]
	Apical resting conductance	G_a	1.974 nS	
	Displacement sensitivity 1 [#]	s_0	85e-9 m^{-1}	
	Displacement offset 1 [#]	u_0	7 nm	
	Displacement sensitivity 2 [#]	s_1	5e-9 m^{-1}	
	Displacement offset 2 [#]	u_1	7 nm	
	Cell capacitance	C_M $(=C_A+C_B)$	15 pF	[WM04]
	Potassium conductance	G_k	18 nS	[Med06a]
	Endocochlear potential	E_t	100 mV	
	Potassium reversal potential	E_k	-70.45 mV	
	Resistance of the supporting cell	$R_p = R_t$	40 mΩ	
Calcium kinetics	Calcium reversal potential [#]	E_{Ca}	66 mV	[Med06a]
	Maximum calcium conductance [#]	G_{Ca}^{max}	7.2 nS	
	Calcium current constant 1 [#]	β_{Ca}	400	
	Calcium current constant 2 [#]	γ_{Ca}	130	
	Calcium current time constant [#]	$\tau_{I_{Ca}}$	100 μs	
	Calcium diffusion time constant [#]	$\tau_{[Ca]}$	100 μs	
	Calcium concentration threshold [#]	$[Ca^{2+}]_{thr}$	0 mol	
	Release rate conversion scalar [#]	z	2e32	[SLOM02]
Transmitter dynamics	NT loss rate	l	2580 s^{-1}	[Med06a]
	NT recovery rate	r	6580 s^{-1}	
	Maximum free NT quanta [#]	M	10	
	NT reprocessing rate	x	66.3 s^{-1}	[SLOM02]
	NT replenishment rate	y	10 s^{-1}	[SLOM02]
	Time step of the simulation [°]	t_s $(=1/f_s)$	$1/44100 \approx 22.68$ μs	-

Table A.2. Default parameters of the stimulus coder. Parameters marked with asterisk (*) share the same default value for all electrode channels even though they can be configured for each channel independently.

Parameter description	Symbol	Value(s)
Number of CI electrodes	M_{CI}	22
Total pulse rate	TPR	9000 pps
Threshold level *	THL_M	100
Most comfortable level *	MCL_M	200
Curve-shaping factor *	c_m	1.95
Phase width	T_{pw}	25 µs
Phase gap	T_{pg}	8 µs
Phase gap variance	ξ_j	0
Peak selection randomness	ξ_p	0.1
Repetition penalty *	R_m	10 dB
Global volume	V_g	1

A.2 CIX interface documentation (excerpt)

CIX_Interface.h File Reference:

#include "CIX_ErrorCodes.h"

Defines

#define **CIX_Interface_Base_Version** "2.21"
Version information.

Functions

Method for DLL initialization:

DLL_INTERFACE void **init** ()
Initialize DLL internal members. This function must be called once upon start!

Methods for hardware connection:

DLL_INTERFACE bool **connectToCiHardware** ()
Connect to nucleus implant communicator (NIC) at the given address (as specified in Common.cfg).

DLL_INTERFACE bool **getCiHardwareStatus** (int &nicStatus)
Get current state of the NIC (as enumerated by Cochlear).

DLL_INTERFACE bool **emergencyStopAndDisconnect** (void)
Immediately stop any operation on the NIC and disconnect.

DLL_INTERFACE bool **disconnectFromCiHardware** (void)
Disconnect from the NIC hardware.

Methods for loading configuration files:

DLL_INTERFACE bool	**assignCommonConfigFile** (const char *const configFileName) Define which file to use as 'common configuration' and check its contents.
DLL_INTERFACE bool	**assignEarMapConfigFile** (const char *const configFileName) Define which file to use as 'EarAnalyzer, Patient Data, MAP and session configuration', short 'EarMap config' and check its contents.

Methods for file-based calculation and stimulation:

DLL_INTERFACE bool	**calculateAndWriteEarOut** (const char *const waveFileName, const double loudnessDbSpl, const char *const earOutFileName) Calculate auditory model output and write data to file.
DLL_INTERFACE bool	**calculateAndWriteStimOut** (const char *const waveFileName, const double loudnessDbSpl, const char *const stimOutFileName, const double volume) Calculate CI stimulation pattern and write data to file.
DLL_INTERFACE bool	**calculateAndStimulate_Calculate** (const char *const waveFileName, const double loudnessDbSpl, const double volume) Calculate CI stimulation pattern based on the ear model and store output in RAM.
DLL_INTERFACE bool	**calculateAndStimulate_Stimulate** (void) Start streaming the CI stimulation pattern, which was previously calculated and stored in RAM.
DLL_INTERFACE bool	**calculateAndStimulate_Clear** (void) Clear CI stimulation pattern, which was previously calculated and stored in RAM (free up memory).

Methods for error handling:

DLL_INTERFACE errorCode	**getLastErrorCode** (void) Get last error code as 'errorCode' enum.
DLL_INTERFACE const char *const	**getLastErrorString** (void) Get last error as C-string.
DLL_INTERFACE void	**clearLastError** (void) Clear last error (last error becomes 'NO_Error').

Detailed Description

(C) Copyright Fraunhofer IDMT (2009) All Rights Reserved

Cochlear Project, CI_Experimental

Author:
Tamas Harczos, Andras Katai and Anja Chilian

Date:
2009-12-17

Version:
See CIX_Interface_Version definition.

Generated for CIX by **doxygen** 1.6.1

A.3 CIX error codes documentation (excerpt)

CIX_ErrorCodes.h File Reference:

Detailed Description

(C) Copyright Fraunhofer IDMT (2009) All Rights Reserved

Cochlear Project, CI_Experimental

Author:
Tamas Harczos and Andras Katai

Date:
2009-12-17

Version:
0.17

Enumeration Type Documentation

enum errorCode

Enumerator:

ERROR_LibUninitialized

The DLL has not yet been initialized.

ERROR_NotImplemented

The requested functionality is not yet implemented.

ERROR_ParameterError

The function parameter passed is invalid.

ERROR_MemAllocError

New operator has thrown an exception.

ERROR_FileCannotOpen

File could not be opened for reading.

ERROR_FileCannotCreate

File could not be created.

ERROR_FileReadError

Unspecified read error occured.

ERROR_FileWriteError

Unspecified write error occured.

ERROR_InitCommonCfgError

Error parsing the Common configuration file.

ERROR_InitEarMapCfgError

Error parsing the EarMap configuration file.

ERROR_MissingCfgEntry

An important config entry was not found.

ERROR_PreprocClipping

Heavy clipping (>20dB) occured during preprocessing. Requested loudness might be too high.

ERROR_CalcNoCommonCfg

A Common configuration file has not been assigned yet.

ERROR_CalcNoEarMapCfg

An EarMap configuration file has not been assigned yet.

ERROR_DataTooLong

The passed input file is longer than the defined maximum. This could cause memory allocation errors.

ERROR_StimAlreadyRunning

Stimulation is already ongoing through the NIC. Wait to finish or stop stimulation first.

ERROR_StimNoData

No stimulation data has been calculated until now. Call a calculation function first.

ERROR_HardwareNotConnected

Cannot communicate (or disconnect), since no connection is built up to the CI hardware.

ERROR_HardwareAlreadyConnected

Cannot initialize, since a connection is already built up to the CI hardware. Disconnect first.

ERROR_HardwareInitError

CI hardware cannot be initialized.

ERROR_HardwareTimeout

CI hardware did not respond within a specified duration of time.

ERROR_HardwareFailure

CI hardware is not functioning correctly any more.

A.4   Sample EarMap configuration file (excerpt)

```
########################################################################
##                                                                    ##
## EarMap config file version by T.Harczos (3rd January 2012)         ##
##                                                                    ##
########################################################################


# Patient data
PATIENT_NAME= "User Name";
PATIENT_BIRTHDATE= "1980";



########################################################################
## Coder parameters                                                   ##
########################################################################


# Active channels (1: active, 0: inactive)
MAP_CHANNEL_AC= 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1;

# Threshold Level (first TL corresponds to the lowest frequency channel)
MAP_CHANNEL_TL= 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100,
100, 100, 100;

# Comfort level (first TL corresponds to the lowest frequency channel
MAP_CHANNEL_CL= 200, 200, 200, 200, 200, 200, 200, 200, 200, 200, 200,200, 200, 200, 200, 200, 200, 200, 200,
200, 200, 200;

# Gain correction per channel in dB
MAP_CHANNEL_GAIN= 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0;

# Total pulse rate (1/s)
MAP_TOT_PULSRATE= 9000;

# Pulse width (us) along all channels
MAP_PHASE_WIDTH= 25.0;

# Inter-phase gap (us) along all channels
MAP_PHASE_GAP= 8.0;

# Inter-phase gap jitter (1.0: maximum jitter [max], 0.0: no jitter [min])
MAP_GAP_JITTER= 0.125;

# Level of randomness in maxima selection (0.0: deterministic [min], 0.5: stochastic [max])
MAP_MAX_RANDOMNESS= 0.1;

# Level [dB FS] with which a FB band should be attenuated if already stimulated in the last cycle
MAP_REP_PENALTY= 10.0;



########################################################################
## Auditory model parameters                                          ##
########################################################################

# BM parameters
SFREQ= 44100.0;
BM_NO_SECTIONS= 101;
BM_MAX_LAT_COUPL= 8;
BM_DELTA_Z= 0.25;

# BM scalers (dB FS)
BM_BM_PREAMP= -42.0;
BM_FACTOR= -86.0;

# Factors to change the trajectory shapes
BM_HWM_CQNU_FACTOR= 0.6;
BM_HWM_C_FACTOR= 3.0;
```

```
# Factors to change non-linearity
BM_USAT_FACTOR1= 5.0;
BM_USAT_FACTOR2= 200.0;


# Number of BM channels to process further
POST_BM_CHANNELS= 22;


# Number AM channels to be mapped to CI electrodes (enumeration begins with zero!)
BM_CHANNEL_MAP= 11, 16, 21, 25, 29, 33, 37, 40, 43, 47, 51, 55, 58, 61, 65, 68, 72, 75, 78, 81, 84, 86;


# Type of frequency mapping ("PICK" / "MAX" / "AVG")
BM_CHANNEL_MAP_TYPE= "PICK";



########################################################################
## Stereocilia coupling, IHC, and synaptic cleft parameters        ##
########################################################################


# ...
```

A.5 Sample Common configuration file (excerpt)

```
########################################################################
## Base parameters                                                 ##
########################################################################


# Full path (without trailing slash) to a fast and spacious temp directory
TEMP_PATH= "R:\\Temp";


# Maximum duration (in seconds) of audio to be processed and used for stimulation
MAX_DURATION= 10.0;


# Hardware string (Cochlear style) of the device to be used
CI_DEVICE_NAME= "L34-CIC4-1";


# Stimulation duration tolerance factor (>=1.0)
STIM_DUR_TOL= 1.5;



########################################################################
## Sound preprocessor parameters                                   ##
########################################################################


# Noise floor of the recording system (dB FS RMS)
PREPROC_NFL_RMS= -65.0;


# EarModel input (reference) loudness (dB FS RMS)
PREPROC_REF_RMS= -18.0;


# Size of the overlapping windows during preprocessing (seconds)
PREPROC_WINSIZE= 0.6;
```

A.6 Sample CIX.log file (excerpt)

```
11:50:12 INFO  CIX          <> ###########################################################################################
11:50:12 INFO  CIX          <> CIX Interface (v2.21, build 18-04-2013 10:22:01) initialized on 2013-05-31
11:50:12 INFO  CIX          <> ###########################################################################################
11:50:12 INFO  CIX          <> [INITIALIZATION] Logger created.
11:50:13 INFO  CIX          <> [INITIALIZATION] DLL initialized.
11:50:13 INFO  CIX          <> Common configuration file "configs\Common_Emu.cfg" assigned.
11:50:13 INFO  CIX          <> [CONNECTION] Connected to CI-hardware "L34-CIC4-0".
11:50:13 DEBUG CIX          <> connectToCiHardware(): dll_nicClient= 00D207C0.
11:50:25 INFO  CIX          <> EarMap configuration file "configs/EarMap.cfg" assigned.
11:50:53 INFO  CIX          <> [CALCULATION] Calculating "datasets/Dataset50_Music_Snippets/B_Jude.wav" (65 dB SPL, volume=0.95).
11:50:58 DEBUG CIX.Strategy <> init("configs/EarMap.cfg"): CIC4 implant mode 5 selected.
11:50:58 INFO  CIX.Strategy <> ------------------------------------- Statistics -----------------------------------------
11:50:58 INFO  CIX.Strategy <> Channel   c00 c01 c02 c03 c04 c05 c06 c07 c08 c09 c10 c11 c12 c13 c14 c15 c16 c17 c18 c19 c20 c21
11:50:58 INFO  CIX.Strategy <> Electrode E22 E21 E20 E19 E18 E17 E16 E15 E14 E13 E12 E11 E10 E09 E08 E07 E06 E05 E04 E03 E02 E01
11:50:58 INFO  CIX.Strategy <>
11:50:58 INFO  CIX.Strategy <> CL        200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200
11:50:58 INFO  CIX.Strategy <> MaxCurLev 191 195 195 195 195 195 195 195 195 195 189 189 186 178 177 174 181 175 164 179 178 176
11:50:58 INFO  CIX.Strategy <> MinCurLev 100 101 101 101 104 100 101 103 101 101 104 104 101 100 100 101 101 105 102 104 113 102
11:50:58 INFO  CIX.Strategy <> TL        100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100
11:50:58 INFO  CIX.Strategy <>
11:50:58 INFO  CIX.Strategy <> AvgCurLev 157 159 159 161 165 162 156 157 152 148 149 146 148 145 141 137 142 144 142 146 142 140
11:50:58 INFO  CIX.Strategy <> AvgCLPerc 57% 59% 58% 60% 64% 61% 56% 56% 51% 47% 48% 45% 47% 45% 40% 37% 41% 43% 42% 45% 42% 40%
11:50:58 INFO  CIX.Strategy <> StimPerc   8%  7%  7% 10% 11%  7%  8%  8%  6%  4%  3%  3%  3%  1%  2% <1%  1% <1% <1% <1% <1% <1%
11:50:58 INFO  CIX.Strategy <> ------------------------------------------------------------------------------------------
11:50:58 DEBUG CIX          <> getCiHardwareStatus(...): calling nscGetStreamStatus(00D207C0,...) ...
11:50:58 INFO  CIX          <> [SENDING] Sending sequence to processor ...
11:51:00 INFO  CIX          <> [STIMULATION] Duration=10101.1 ms (15351.7 ms with tolerance), polling every 100 ms.
11:51:01 DEBUG CIX          <> stimRunningSinceMs=287 ms (2.84127 %).
11:51:02 DEBUG CIX          <> stimRunningSinceMs=1214 ms (12.0185 %).
11:51:03 DEBUG CIX          <> stimRunningSinceMs=2219 ms (21.9679 %).
11:51:04 DEBUG CIX          <> stimRunningSinceMs=3234 ms (32.0163 %).
11:51:05 DEBUG CIX          <> stimRunningSinceMs=4239 ms (41.9657 %).
11:51:06 DEBUG CIX          <> stimRunningSinceMs=5245 ms (51.925 %).
11:51:07 DEBUG CIX          <> stimRunningSinceMs=6251 ms (61.8843 %).
11:51:08 DEBUG CIX          <> stimRunningSinceMs=7659 ms (75.8233 %).
11:51:09 DEBUG CIX          <> stimRunningSinceMs=8765 ms (86.7726 %).
11:51:10 DEBUG CIX          <> stimRunningSinceMs=9771 ms (96.7319 %).
11:51:11 DEBUG CIX          <> stimRunningSinceMs=10174 ms (100 %).
11:51:11 DEBUG CIX          <> getCiHardwareStatus(...): calling nscGetStreamStatus(00D207C0,...) ...
11:51:11 DEBUG CIX          <> calling nscStopStream() ...
11:51:14 DEBUG CIX          <> calling nscDelete(00D207C0) ...
...
12:23:50 ERROR CIX          <> assignCommonConfigFile("Common_Bad.cfg"): TEMP_PATH value ("X:\Temp") is invalid! Cannot write or
                                delete in the given directory.
...
12:31:22 ERROR CIX          <> assignEarMapConfigFile("EarMap_Bad.cfg"): MAP_CHANNEL_AC key may only have 0/1 elements!
...
12:35:30 FATAL CIX          <> connectToCiHardware(): ERROR_HardwareTimeout
```

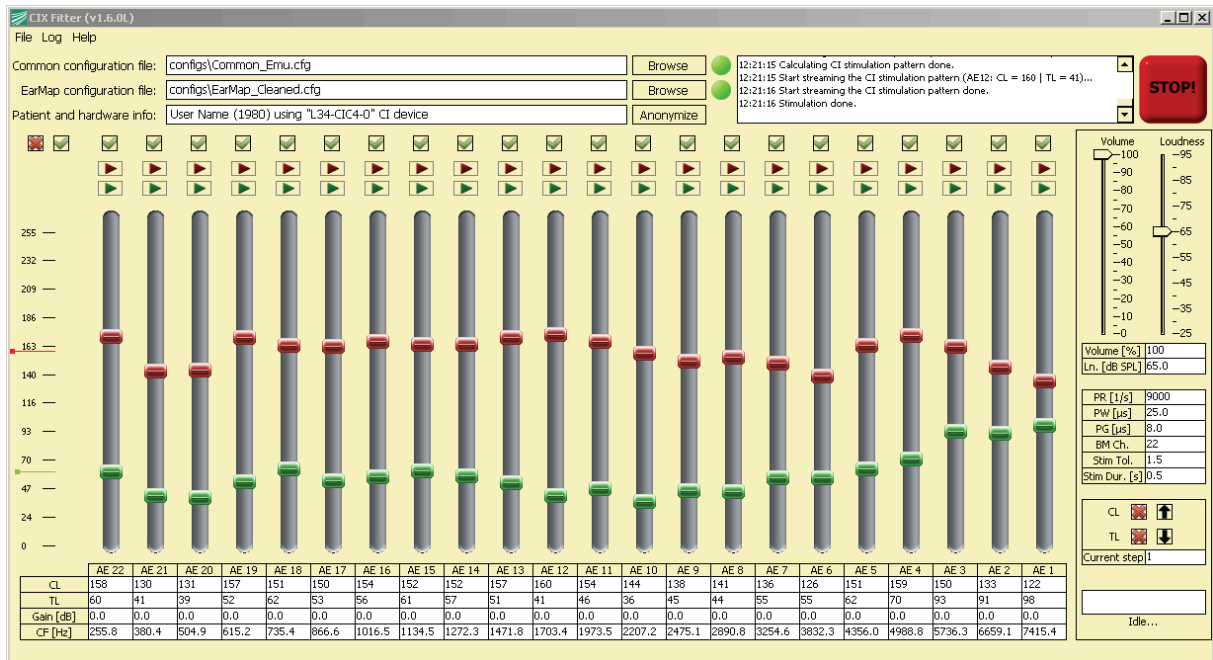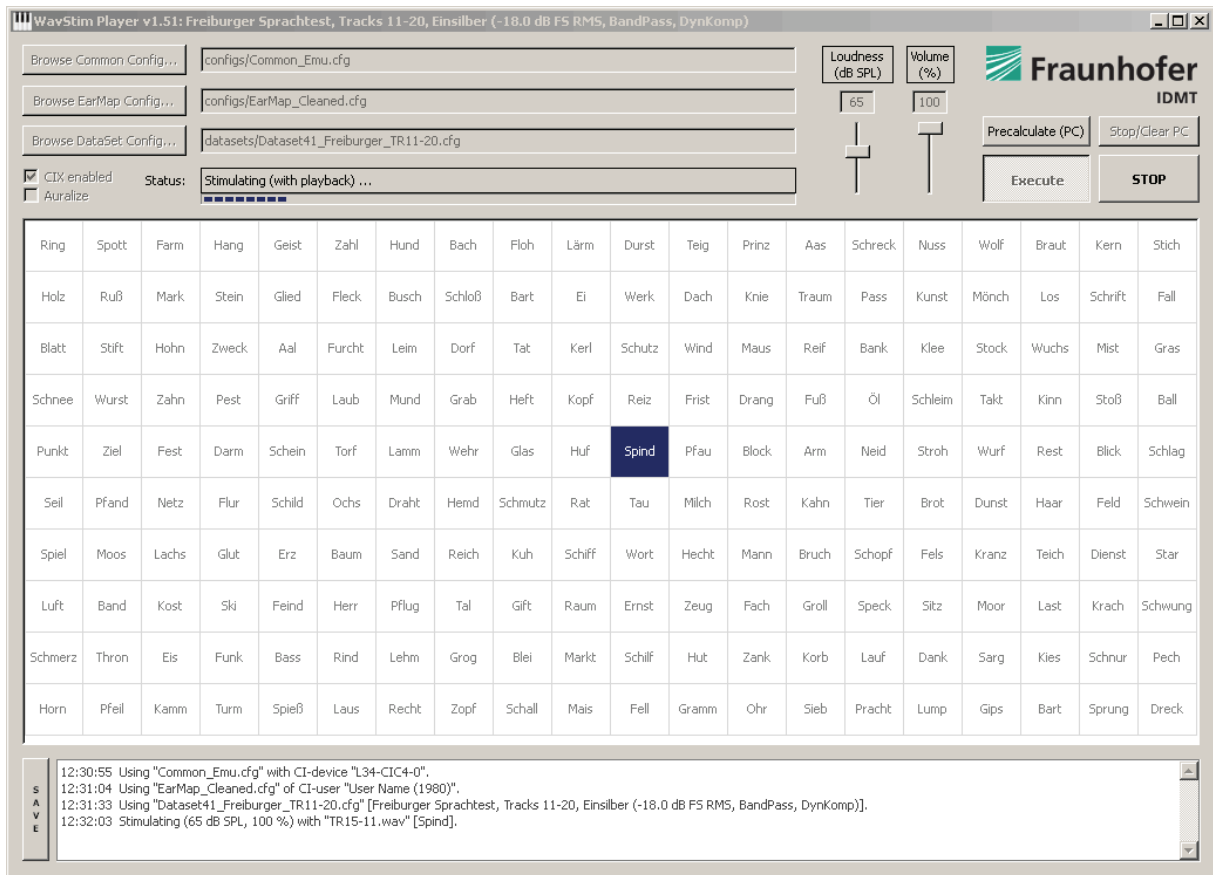## A.7   Screenshots of CIX applications
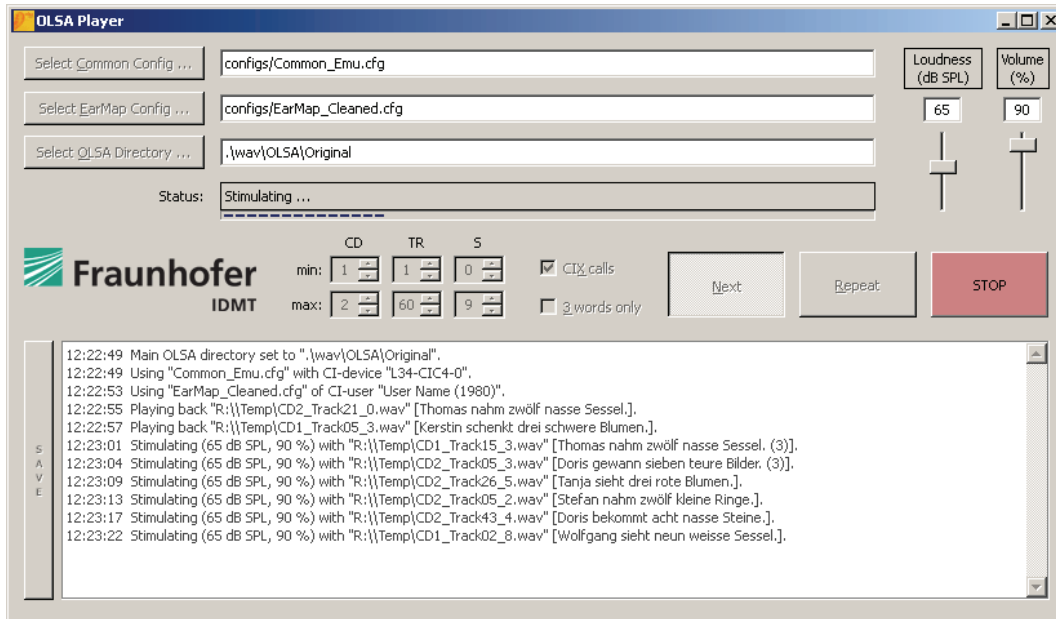

Figure A.1: CIX Fitter.

Figure A.2: WavStim Player.
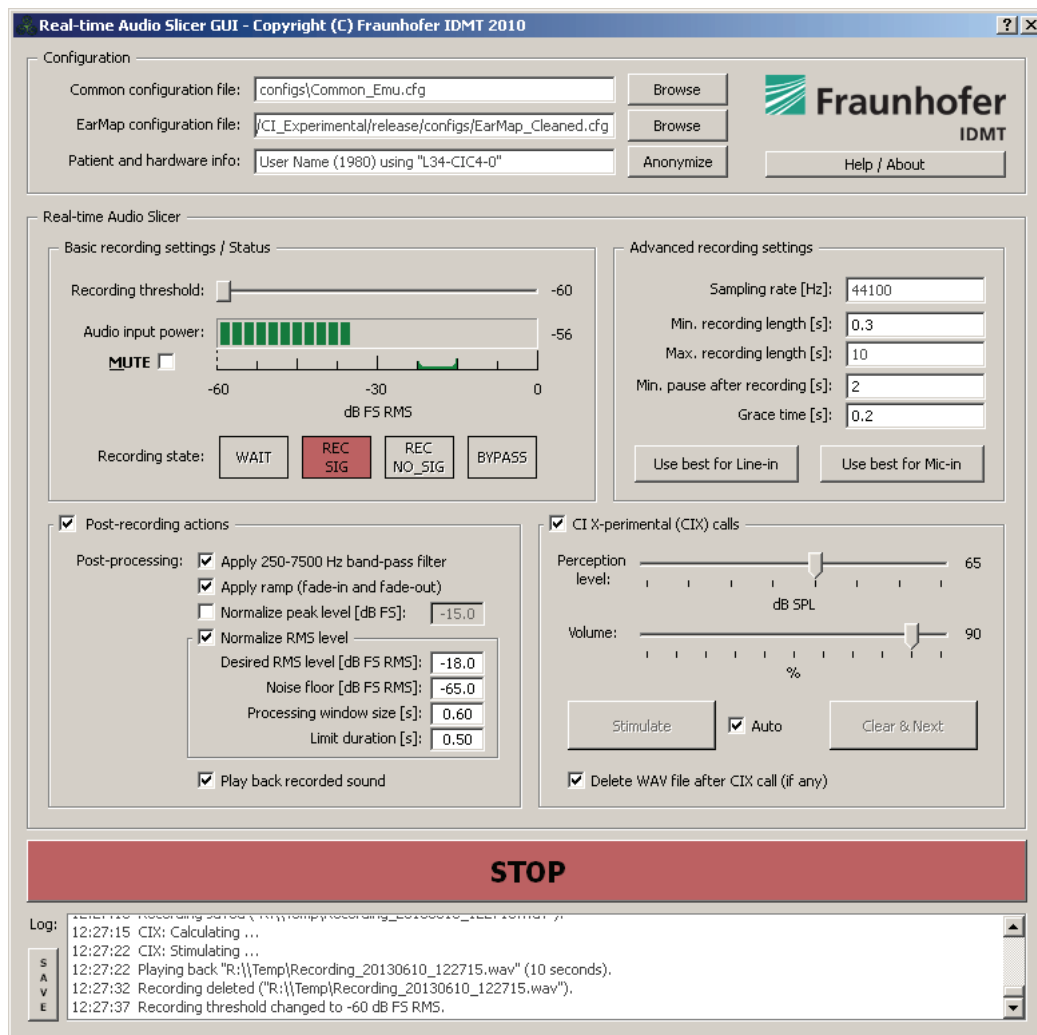
Figure A.3: OLSA Player.
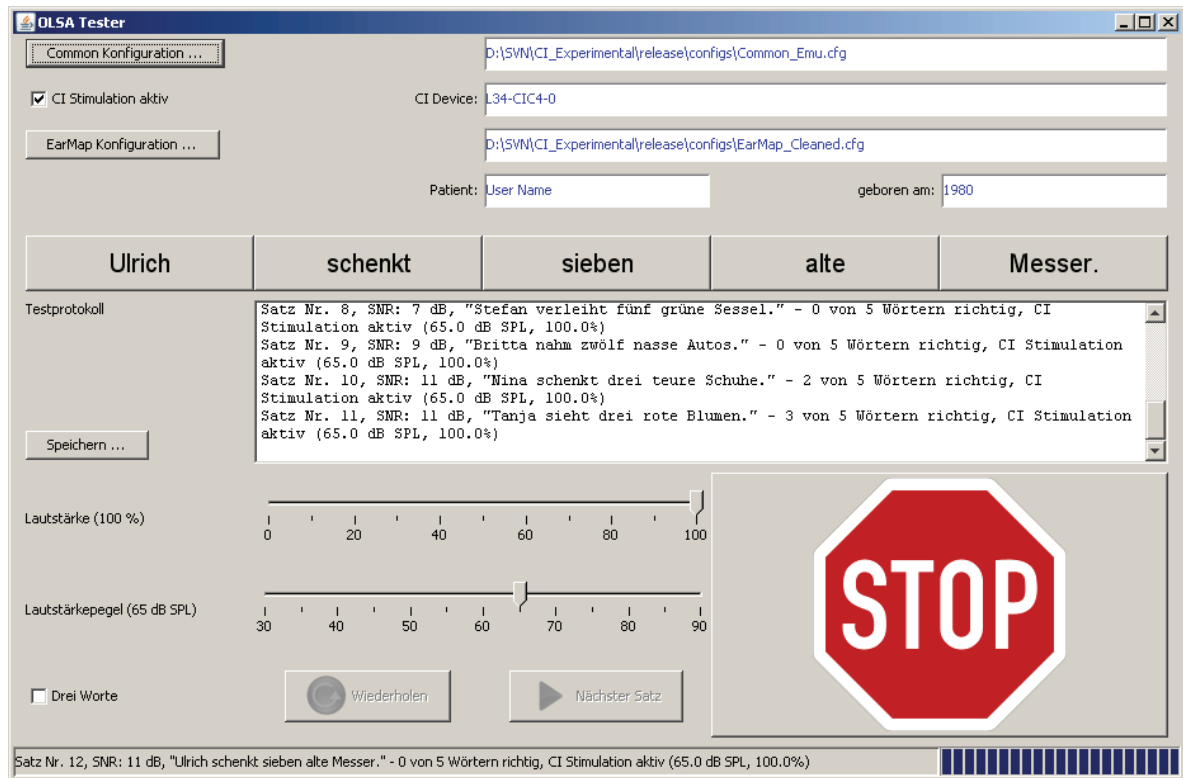

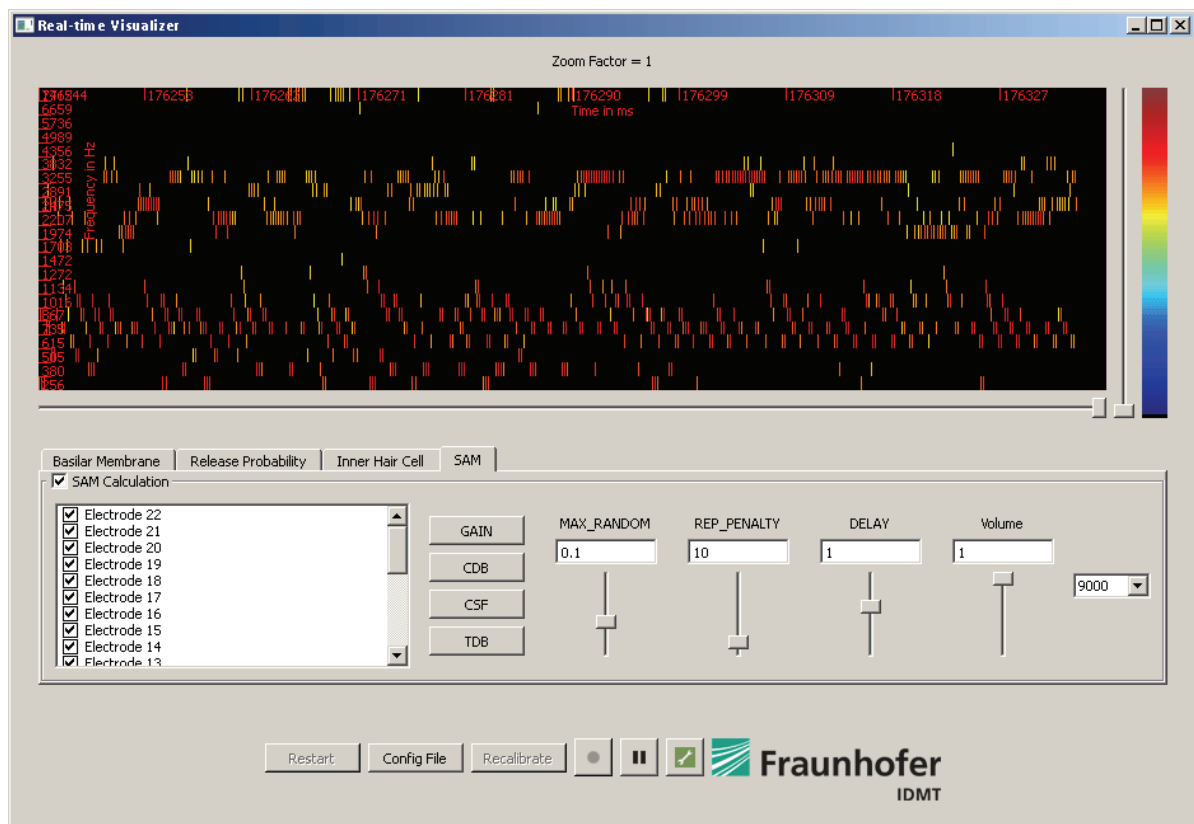Figure A.4: Real-time Audio Slicer.

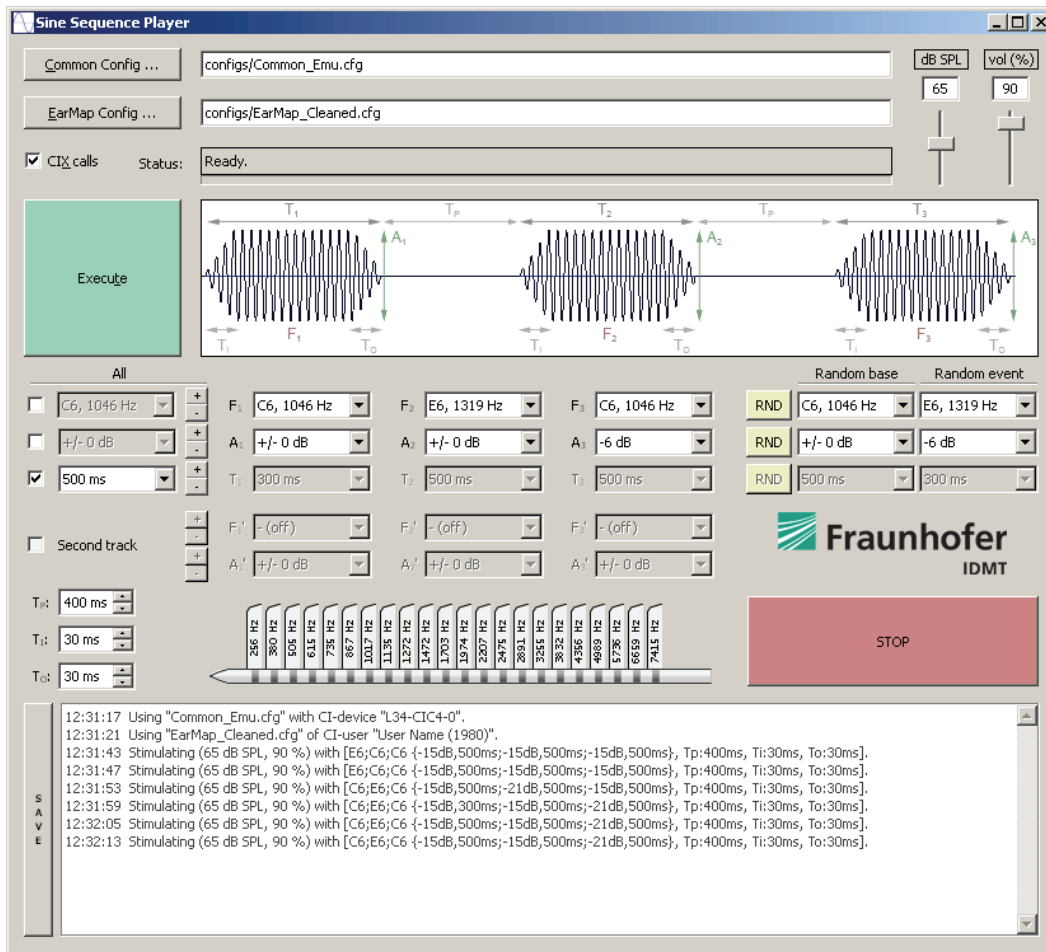Figure A.5: OLSA Tester.

Figure A.6: Real-time Visualizer.

Figure A.7: Sine Sequenceer.

References

[AJ80]   A. M. H. J. Aertsen and P. I. M. Johannesma, "Spectro-temporal receptive fields of au-
            ditory neurons in the grassfrog: I. Characterization of Tonal and Natural Stimuli," *Biol.*
            Cybern, vol. 38 (4), pp. 223–234, 1980.

[APJ+07]   S. K. An, S.-I. Park, S. B. Jun, C. J. Lee, K. M. Byun, J. H. Sung, B. S. Wilson, S. J. Reb-
            scher, S. H. Oh, and S. J. Kim, "Design for a Simplified Cochlear Implant System,"
            IEEE Trans. Biomed. Eng, vol. 54 (6), pp. 973–982, 2007.

[Aris30]   Aristotle, *De Anima.* (English translation *On the soul* by J. A. Smith, Stilwell, Kansas,
            USA: Digireads.com, 2006), c. 330 BC.

[Bal81]   D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern*
            Recognition, vol. 13 (2), pp. 111–122, 1981.

[Bau99]   F. Baumgarte, "A Physiological Ear Model for the Emulation of Masking," *ORL*, vol. 61,
            pp. 294–304, 1999.

[Bau00]   F. Baumgarte, "Ein Psychophysisches Gehörmodell zur Nachbildung von Wahr-
            nehmungsschwellen für die Audiocodierung," PhD dissertation (in German), Faculty of
            Electrical Engineering and Computer Science, University of Hannover, Germany, 2000.

[BBS+11]   A. Buechner, A. Beynon, W. Szyfter, K. Niemczyk, U. Hoppe, M. Hey, J. Brokx, J. Eyles,
            P. Van de Heyning, G. Paludetti, A. Zarowski, N. Quaranta, T. Wesarg, J. Festen, H.
            Olze, I. Dhooge, J. Müller-Deile, A. Ramos, S. Roman, J.-P. Piron, D. Cuda, S. Burdo, W.
            Grolman, S. R. Vaillard, A. Huarte, B. Frachet, C. Morera, L. Garcia-Ibáñez, D. Abels, M.
            Walger, J. Müller-Mazotta, C. A. Leone, B. Meyer, N. Dillier, T. Steffens, A. Gentine, M.
            Mazzoli, G. Rypkema, M. Killian, and G. Smoorenburg, "Clinical evaluation of cochlear
            implant sound coding taking into account conjectural masking functions, MP3000™,"
            Cochlear Implants Int, vol. 12 (4), pp. 194–204, 2011.

[BDG04]   P. J. Blamey, B. Dickson, and L. M. Grant, "An incremental excitation scale for cochlear
            implants," *ARLO*, vol. 5 (2), pp. 50–55, 2004.

[Bék28]   Gy. Békésy, "Zur Theorie des Hörens: die Schwingungsform der Basilarmembran," *Phy-*
            sik. Z, vol. 29, pp. 793–810, 1928.

[Bék47]   Gy. Békésy, "The Variation of Phase Along the Basilar Membrane with Sinusoidal Vi-
            brations," *J. Acoust. Soc. Am*, vol. 19 (3), pp. 452–460, 1947.

[Bék49]   Gy. Békésy, "The Vibration of the Cochlear Partition in Anatomical Preparations and in
            Models of the Inner Ear," *J. Acoust. Soc. Am*, vol. 21 (3), pp. 233–245, 1949.

[Bék52]   Gy. Békésy, "Direct Observation of the Vibrations of the Cochlear Partition Under a Mi-
            croscope," *Acta Otolaryngol*, vol. 42 (3), pp. 197–201, 1952.

[Bék60]   Gy. Békésy, *Experiments in Hearing.* Translated and edited by E. G. Wever. New York,
            USA: McGraw-Hill Book Co, 1960.

[Bla92]   G. G. Blasdel, "Orientation selectivity, preference, and continuity in monkey striate cor-
            tex," *J. Neurosci*, vol. 12 (8), pp. 3139–3161, 1992.

[Boe75]   E. de Boer, "Synthetic whole-nerve action potentials for the cat," *J. Acoust. Soc. Am*, vol. 58 (5), pp. 1030–1045, 1975.

[Bos03]   S. Boswell, *The Mind Hears: Tuning In With a Cochlear Implant*, 2003. Available: http://www.asha.org/public/hearing/treatment/mind_hears.htm (January 13, 2012).

[Bra11]   E. Braun, "Evaluierungstools für Signalverarbeitungsalgorithmen von Cochlea-Implantaten," B.Sc. thesis (in German), Institute of Biomedical Engineering and Informatics, Ilmenau University of Technology, Germany, 2011.

[BZ73]   E. Borg and J.-E. Zakrisson, "Stapedius reflex and speech features," *J. Acoust. Soc. Am*, vol. 54 (2), pp. 525–527, 1973.

[BZK+12]   J. Besser, A. A. Zekveld, S. E. Kramer, J. Rönnberg, and J. M. Festen, "New Measures of Masked Text Recognition in Relation to Speech-in-Noise Perception and Their Associations With Age and Cognitive Abilities," *J. Speech Lang. Hear. Res*, vol. 55 (1), pp. 194–209, 2012.

[Car93]   L. H. Carney, "A model for the responses of low-frequency auditory-nerve fibers in cat," *J. Acoust. Soc. Am*, vol. 93 (1), pp. 401–417, 1993.

[CBH12]   A. Chilian, E. Braun, and T. Harczos, "Acoustic simulation of cochlear implant hearing," in *Proc. 3rd Int. Symp. Audit. Audiol. Res. (ISAAR 2011): Speech perception and auditory disorders*, Nyborg, Denmark: The Danavox Jubilee Foundation, pp. 425–432, 2012.

[CC96]   K. R. Crounse and L. O. Chua, "The CNN Universal Machine is as universal as a Turing Machine," *IEEE Trans. Circ. Syst. I*, vol. 43 (4), pp. 353–355, 1996.

[CCA+14]   A. Coombs, P. J. Clamp, S. Armstrong, P. J. Robinson, and D. Hajioff, "The role of post-operative imaging in cochlear implant surgery: A review of 220 adult cases," *Cochlear Implants Int*, DOI: 10.1179/1754762814Y.0000000071, 8 pages, 2014.

[Cha02]   M. Chasin, "Bone anchored and middle ear implant hearing aids," *Trends in Amplif*, vol. 6 (2), pp. 33–38, 2002.

[Chi10a]   A. Chilian, "Heranführen eines auditorischen Modells an den Einsatz als Sprachprozessor für Cochlea-Implantate," Internship report (in German): Fraunhofer Institute for Digital Media Technology IDMT, Institute of Biomedical Engineering and Informatics, Ilmenau University of Technology, Germany, 2010.

[Chi10b]   A. Chilian, "Entwicklung von Tools und Methoden zur Evaluierung von Signalverarbeitungsstrategien für Cochlea-Implantate," B.Sc. thesis (in German), Institute of Biomedical Engineering and Informatics, Ilmenau University of Technology, Germany, 2010.

[Chi13]   A. Chilian, "Entwicklung eines Modells der elektrisch stimulierten Cochlea," M.Sc. thesis (in German), Institute of Biomedical Engineering and Informatics, Ilmenau University of Technology, Germany, 2013.

[Chr01]   T. U. Christiansen, "The Meddis inner hair-cell model," in *Proc. 1st Seminar on Auditory Models*, Ørsted, Denmark: Acoustic Technology, Technical University of Denmark, 2001.

[Cla03]   G. Clark, *Cochlear implants: Fundamentals and applications*. Berlin, Germany: Springer, 2003.

[Coc06]   Cochlear Ltd, "E11318RD: NIC v2 Software Interface Specification," Cochlear Ltd, Lane Cove, New South Wales, Australia, 2006.

[Coc10a]   Cochlear Ltd., *Types of hearing loss*, 2010. Available: http://www.cochlear.com/au/ hearing-and-hearing-loss/types-hearing-loss-adults (January 23, 2012).

[Coc10b]   Cochlear Ltd, "E13985RA: NIC v2.1 Hazards Analysis and Conformance Guide," Cochlear Ltd, Lane Cove, New South Wales, Australia, 2010.

[Com91]   D. van Compernolle, "Development of a Computational Auditory Model," Technical Report, Institute for Perception Research (IPO), Eindhoven, The Netherlands, 1991.

[CR01]   M. Chatterjee and M. E. Robert, "Noise Enhances Modulation Sensitivity in Cochlear Implant Listeners: Stochastic Resonance in a Prosthetic Sensory System?," *JARO*, vol. 2 (2), pp. 159–171, 2001.

[CRSC03]   L. T. Cohen, L. M. Richardson, E. Saunders, and R. S. Cowan, "Spatial spread of neural excitation in cochlear implant recipients: comparison of improved ECAP method and psychophysical forward masking," *Hearing Research*, vol. 179 (1-2), pp. 72–87, 2003.

[CYK91]   L. O. Chua, L. Yang, and K. R. Krieg, "Signal processing using cellular neural networks," *J. VLSI Sig. Proc. Syst*, vol. 3 (1-2), pp. 25–51, 1991.

[Dav62]   H. Davis, "Advances in the Neurophysiology and Neuroanatomy of the Cochlea," *J. Acoust. Soc. Am*, vol. 34 (9B), pp. 1377–1385, 1962.

[Dav83]   H. Davis, "An active process in cochlear mechanics," *Hear. Res*, vol. 9 (1), pp. 79–90, 1983.

[DBP+10]   F. Di Lella, A. Bacciu, E. Pasanisi, V. Vincenti, M. Guida, and S. Bacciu, "Main peak interleaved sampling (MPIS) strategy: effect of stimulation rate variations on speech perception in adult cochlear implant recipients using the Digisonic SP cochlear implant," *Acta Otolaryngol*, vol. 130 (1), pp. 102–107, 2010.

[DFW+11]   C. A. J. Dun, H. T. Faber, M. J. F. de Wolf, C. W. R. J. Cremers, and M. K. S. Hol, "An Overview of Different Systems: The Bone- Anchored Hearing Aid," in *Implantable Bone Conduction Hearing Aids*, M. Kompis, M.-D. Caversaccio, and M. Kompis, Eds, Basel, Switzerland: S. Karger AG, pp. 22–31, 2011.

[DGG+07]   N. Deggouj, M. Gersdorff, P. Garin, S. Castelein, and J.-M. Gérard, "Today's indications for cochlear implantation," *B-ENT*, vol. 3, pp. 9–14, 2007.

[DLR97]   M. F. Dorman, P. C. Loizou, and D. Rainey, "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," *J. Acoust. Soc. Am*, vol. 102 (4), pp. 2403–2411, 1997.

[DM10]   M. Drapal and P. Marsalek, "Stochastic model shows how cochlear implants process azimuth in real auditory space," *Chin. J. Physiol*, vol. 53 (6), pp. 439–446, 2010.

[DR93]   G. S. Donaldson and R. A. Ruth, "Derived band auditory brain-stem response estimates of traveling wave velocity in humans. I: Normal-hearing subjects," *J. Acoust. Soc. Am*, vol. 93 (2), pp. 940–951, 1993.

[Dud39]   H. Dudley, "Remaking Speech," *J. Acoust. Soc. Am*, vol. 11 (2), pp. 169–177, 1939.

[Dui76]   H. Duifhuis, "Cochlear nonlinearity and second filter: Possible mechanism and implications," *J. Acoust. Soc. Am*, vol. 59 (2), pp. 408–423, 1976.

[Dui04]   H. Duifhuis, "Comment on "An approximate transfer function for the dual-resonance nonlinear filter model of auditory frequency selectivity" [J. Acoust. Soc. Am. 114, 2112–2117] (L),"" *J. Acoust. Soc. Am*, vol. 115 (5), pp. 1889–1890, 2004.

[Eag45]   C. C. Eaglesfield, "Carrier Frequency Amplifiers: The Unit Step Response of Amplifiers with Single and Double Circuits," *Wireless Engr*, vol. 22, pp. 523–532, 1945.

[EDCS07]   C. Elberling, M. Don, M. Cebulla, and E. Stürzebecher, "Auditory steady-state responses to chirp stimuli based on cochlear traveling wave delay," *J. Acoust. Soc. Am*, vol. 122 (5), pp. 2772–2785, 2007.

[EE68]   A. M. Engebretson and D. H. Eldredge, "Model for the Nonlinear Characteristics of Cochlear Potentials," *J. Acoust. Soc. Am*, vol. 44 (2), pp. 548–554, 1968.

[Egg73]   J. J. Eggermont, "Analog Modelling of Cochlear Adaptation," *Biol. Cybern*, vol. 14 (2), pp. 117–126, 1973.

[Egg75]   J. J. Eggermont, "Cochlear adaptation: A theoretical description," *Biol. Cybern*, vol. 19 (4), pp. 181–189, 1975.

[FDA12]   U. S. Food and Drug Administration, *Cochlear Implant Devices at FDA*. Available: http://www.accessdata.fda.gov/scripts/cdrh/devicesatfda/index.cfm?search_term=cochlear%20implant (February 03, 2012).

[Fet86]   A. Fettweis, "Wave digital filters: Theory and practice," *Proceedings of the IEEE*, vol. 74 (2), pp. 270–327, 1986.

[Fla60]   J. L. Flanagan, "Models for approximating basilar membrane displacement," *J. Acoust. Soc. Am*, vol. 32 (7), p. 937, 1960.

[Fla65]   J. L. Flanagan, *Speech Analysis Synthesis and Perception*. Berlin, Germany: Springer, 1965.

[Fre12]   S. Fredelake, "Model-based prediction of the benefit with rehabilitative hearing devices," PhD dissertation, School of Mathematics and Science, Carl von Ossietzky University of Oldenburg, Germany, 2012.

[Fri90]   D. H. Friedman, "Implementation of a nonlinear wave-digital-filter cochlear model," in *Proc. Int. Conf. Acoust, Speech, Signal Process. (IEEE ICASSP)*, pp. 397–400, 1990.

[FS98]   Q.-J. Fu and R. V. Shannon, "Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am*, vol. 104 (5), pp. 2570–2577, 1998.

[FSC05]   L. M. Friesen, R. V. Shannon, and R. J. Cruz, "Effects of Stimulation Rate on Speech Recognition with Cochlear Implants," *Audiol Neurotol*, vol. 10 (3), pp. 169–184, 2005.

[GAR+07]   D. W. Grantham, D. H. Ashmead, T. A. Ricketts, R. F. Labadie, and D. S. Haynes, "Horizontal-Plane Localization of Noise and Speech Signals by Postlingually Deafened Adults Fitted With Bilateral Cochlear Implants," *Ear Hear*, vol. 28 (4), pp. 524–541, 2007.

[GC02]   G. J. Goodhill and M. Á. Carreira-Perpiñán, "Cortical Columns," in *Encyclopedia of Cognitive Science*, L. Nadel, Ed, London, UK: Macmillan Publishers Ltd, pp. 845–851, 2002.

[Gei68]   C. D. Geisler, "A Model of the Peripheral Auditory System Responding to Low-Frequency Tones," *Biophys. J*, vol. 8 (1), pp. 1–15, 1968.

[GFRM05] T. Green, A. Faulkner, S. Rosen, and O. Macherey, "Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification," *J. Acoust. Soc. Am*, vol. 118 (1), pp. 375–385, 2005.

[GLF+93] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "DARPA TIMIT acoustic-phonetic continuous speech corpus," Technical Report NISTIR 4930, National Institute of Standards, Gaithersburgh, MD, USA, 1993.

[Gna12] D. Gnansia, Neurelec Scientific & Clinical Research, "The MPIS strategy", personal communication (e-mail), Feb. 2012.

[Gol90] J. L. Goldstein, "Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering," *Hear. Res*, vol. 49 (1-3), pp. 39–60, 1990.

[Gol95] J. L. Goldstein, "Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane: data and MBPNL model," *Hear. Res*, vol. 89 (1-2), pp. 52–68, 1995.

[Gre90] D. D. Greenwood, "A cochlear frequency-position function for several species: 29 years later," *J. Acoust. Soc. Am*, vol. 87 (6), pp. 2592–2605, 1990.

[GW94a] C. Giguère and P. C. Woodland, "A computational model of the auditory periphery for speech and hearing research. I. Ascending path," *J. Acoust. Soc. Am*, vol. 95 (1), pp. 331–342, 1994.

[GW94b] C. Giguère and P. C. Woodland, "A computational model of the auditory periphery for speech and hearing research. II. Descending paths," *J. Acoust. Soc. Am*, vol. 95 (1), pp. 343–349, 1994.

[GW99] L. Geurts and J. Wouters, "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants," *J. Acoust. Soc. Am*, vol. 105 (4), pp. 2476–2484, 1999.

[Hah53] K.-H. Hahlbrock, "Über Sprachaudiometrie und neue Wörterteste," *Arch Ohren Nasen Kehlkopfheilkd*, vol. 162 (5), pp. 394–431, 1953.

[Ham03] V. Hamacher, "Signalverarbeitungsmodelle des elektrisch stimulierten Gehörs," PhD dissertation (in German), Faculty of Electrical Engineering and Information Technology, RWTH Aachen University, Germany, 2003.

[HB78] R. D. Horner and D. M. Baer, "Multiple Probe Technique: A Variation of the Multiple Baseline," *J. Appl. Behav. Anal*, vol. 11 (1), pp. 189–196, 1978.

[HC33] W. Hughson and S. J. Crowe, "Experimental Investigation of the Physiology of the Ear," *Acta Otolaryngol*, vol. 18 (3), pp. 291–339, 1933.

[HC99] L. G. Huettel and L. M. Collins, "Using computational auditory models to predict simultaneous masking data: model comparison," *IEEE Trans. Biomed. Eng*, vol. 46 (12), pp. 1432–1440, 1999.

[HCH12] T. Harczos, A. Chilian, and P. Husar, "Making use of auditory models for better mimicking of normal hearing processes with cochlear implants: the SAM coding strategy," *IEEE Trans. Biomed. Circuits Syst*, vol. 7 (4), pp. 414–425, 2012.

[HCK11]   T. Harczos, A. Chilian, and A. Kátai, "Horizontal-plane localization with bilateral co-chlear implants using the SAM strategy," in *Proc. 3rd Int. Symp. Audit. Audiol. Res. (ISAAR 2011): Speech perception and auditory disorders*, Nyborg, Denmark: The Danavox Jubilee Foundation, pp. 339–345, 2011.

[HCK+14] T. Harczos, A. Chilian, A. Kátai, F. Klefenz, I. Baljić, P. Voigt, and P. Husar, "Making use of auditory models for better mimicking of normal hearing processes with cochlear implants: first results with the SAM coding strategy," in *Proc. 4th Int. Symp. Audit. Audiol. Res. (ISAAR 2013): Auditory plasticity - Listening with the brain*, Nyborg, Denmark: The Danavox Jubilee Foundation, pp. 317–324, 2014.

[HE05]   J. M. Harte and S. J. Elliott, "A comparison of various nonlinear models of cochlear compression," *J. Acoust. Soc. Am*, vol. 117 (6), pp. 3777–3786, 2005.

[Hel63]   H. L. F. von Helmholtz, *Die Lehre von den Tonempfindungen als Physiologische Grundlage für die Theorie der Musik*. (Corrections and translation *On the Sensations of Tone as a Physiological Basis for the Theory of Music* by A. J. Ellis, New York, USA: Longmans, Green, 1912), 1863.

[HFHK11] T. Harczos, S. Fredelake, V. Hohmann, and B. Kollmeier, "Comparative evaluation of cochlear implant coding strategies via a model of the human auditory speech processing," in *Proc. 3rd Int. Symp. Audit. Audiol. Res. (ISAAR 2011): Speech perception and auditory disorders*, Nyborg, Denmark: The Danavox Jubilee Foundation, pp. 331–338, 2011.

[HGH07]   M. Holmberg, D. Gelbart, and W. Hemmert, "Speech encoding in a model of peripheral auditory processing: Quantitative assessment by means of automatic speech recognition," *Speech Commun*, vol. 49 (12), pp. 917–932, 2007.

[HH76]   D. P. Hartmann and R. V. Hall, "The Changing Criterion Design," *J. Appl. Behav. Anal*, vol. 9 (4), pp. 527–532, 1976.

[HHM+12] S. Haumann, V. Hohmann, M. Meis, T. Herzke, T. Lenarz, and A. Büchner, "Indication criteria for cochlear implants and hearing aids: impact of audiological and non-audiological findings," *Audiol. Res*, vol. 2 (1), pp. 55–64, 2012.

[HJST08] T. J. Hamilton, C. Jin, A. van Schaik, and J. Tapson, "An active 2-D silicon cochlea," *IEEE Trans. Biomed. Circuits Syst*, vol. 2 (1), pp. 30–43, 2008.

[HK07]   C. van den Honert and D. C. Kelsall, "Focused intracochlear electric stimulation with phased array channels," *J. Acoust. Soc. Am*, vol. 121 (6), pp. 3703–3716, 2007.

[HKK06]   T. Harczos, F. Klefenz, and A. Kátai, "A neurobiologically inspired vowel recognizer using Hough-transform: A novel approach to auditory image processing," in *Proc. 1st Int. Conf. on Computer Vision Theory and Applications (VISAPP)*, pp. 251–256, 2006.

[HL88]   A. J. Hudspeth and R. S. Lewis, "Kinetic analysis of voltage- and ion-dependent conductances in saccular hair cells of the bull-frog, Rana Catesbeiana," *J. Physiol*, vol. 400, pp. 237–274, 1988.

[HM91]   M. J. Hewitt and R. Meddis, "An evaluation of eight computer models of mammalian inner hair-cell function," *J. Acoust. Soc. Am*, vol. 90 (2), pp. 904–917, 1991.

[HNJ+06]   I. Hochmair, P. Nopp, C. Jolly, M. Schmidt, H. Schößer, C. Garnham, and I. Anderson, "MED-EL Cochlear Implants: State of the Art and a Glimpse into the Future," *Trends in Amplif*, vol. 10 (4), pp. 201–220, 2006.

[HNSK07]   T. Harczos, W. Nogueira, G. Szepannek, and F. Klefenz, "Comparative Evaluation of Successive Cochlear Modeling Stages as Possible Front-Ends for Automatic Speech Recognition," in *Proc. 19th Int. Congr. Acoustics (ICA)*, Madrid, Spain, 2007.

[Hou59]   P. V. C. Hough, "Machine analysis of bubble chamber pictures," in *2nd Int. Conf. on High-Energy Accelerators (HEACC'59)*, Geneva, Switzerland, pp. 554–558, 1959.

[HSK08]   T. Harczos, G. Szepannek, and F. Klefenz, "Towards automatic speech recognition based on cochlear traveling wave delay trajectories," in *Proc. 1st Int. Symp. Audit. Audiol. Res. (ISAAR 2007): Auditory Signal Processing in Hearing-Impaired Listeners*, Helsingør, Denmark: The Danavox Jubilee Foundation, pp. 83–91, 2008.

[HSKK06]   T. Harczos, G. Szepannek, A. Kátai, and F. Klefenz, "An auditory model based vowel classification," in *Proc. IEEE Biomedical Circuits and Systems Conf. (BioCAS)*, pp. 69–72, 2006.

[HSMS97]   I. Hochmair-Desoyer, E. Schulz, L. Moser, and M. Schmidt, "The HSM sentence test as a tool for evaluating the speech understanding in noise of cochlear implant users," *Am. J. Otol*, vol. 18 (6, suppl.), p. S83, 1997.

[HSOM04]   S. D. Holmes, C. J. Sumner, L. P. O'Mard, and R. Meddis, "The temporal representation of speech in a nonlinear model of the guinea pig cochlea," *J. Acoust. Soc. Am*, vol. 116 (6), pp. 3534–3545, 2004.

[HWS08]   T. Harczos, S. Werner, and G. Szepannek, "Formant map counterpart in auditory processing based on cochlear pressure wave trajectories," in *Proc. IEEE Biomedical Circuits and Systems Conf. (BioCAS)*, pp. 45–48, 2008.

[HWS78]   D. H. Hubel, T. N. Wiesel, and M. P. Stryker, "Anatomical demonstration of orientation columns in macaque monkey," *J. Comp. Neurol*, vol. 177 (3), pp. 361–379, 1978.

[HWSB10]   T. Harczos, S. Werner, G. Szepannek, and K. Brandenburg, "Evaluation of cues for horizontal-plane localization with bilateral cochlear implants," in *Proc. 2nd Int. Symp. Audit. Audiol. Res. (ISAAR 2009): Binaural Processing and Spatial Hearing*, Helsingør, Denmark: The Danavox Jubilee Foundation, pp. 37–46, 2010.

[IP97]   T. Irino and R. D. Patterson, "A time-domain, level-dependent auditory filter: The gammachirp," *J. Acoust. Soc. Am*, vol. 101 (1), pp. 412–419, 1997.

[IP01]   T. Irino and R. D. Patterson, "A compressive gammachirp auditory filter for both physiological and psychophysical data," *J. Acoust. Soc. Am*, vol. 109 (5), pp. 2008–2022, 2001.

[IP06]   T. Irino and R. D. Patterson, "A Dynamic Compressive Gammachirp Auditory Filterbank," *IEEE Trans. Audio, Speech, Lang. Process*, vol. 14 (6), pp. 2222–2232, 2006.

[JFD+06]   C. J. James, B. Fraysse, O. Deguine, T. Lenarz, D. Mawman, Á. Ramos, R. Ramsden, and O. Sterkers, "Combined Electroacoustic Stimulation in Conventional Candidates for Cochlear Implantation," *Audiol Neurotol*, vol. 11 (1, suppl.), pp. 57–62, 2006.

[JFM+10]   T. Jürgens, S. Fredelake, R. M. Meyer, B. Kollmeier, and T. Brand, "Challenging the Speech Intelligibility Index: Macroscopic vs. Microscopic Prediction of Sentence Recognition in Normal and Hearing-impaired Listeners," in *Proc. 11th Ann. Conf. Int. Speech Comm. Assoc.: INTERSPEECH*, pp. 2478–2481, 2010.

[Joh72]   P. I. M. Johannesma, "The pre-response stimulus ensemble of neurons in the cochlear nucleus," in *Proc. IPO Symp. Hear. Theory*, pp. 58–69, 1972.

[JPY86]   B. M. Johnstone, R. Patuzzi, and G. K. Yates, "Basilar membrane measurements and the travelling wave," *Hear. Res*, vol. 22, pp. 147–153, 1986.

[Jut11]   S. Jutzi, *Technik ersetzt Körperteile* (in German), 2011. FOCUS Magazin Online. Available: http://www.focus.de/gesundheit/ratgeber/zukunftsmedizin/therapie/tid-21244/ forschung-und-technik-technik-ersetzt-koerperteile_aid_590947.html (February 06, 2012).

[KCKK09]   K. H. Kim, S. J. Choi, J. H. Kim, and D. H. Kim, "An Improved Speech Processing Strategy for Cochlear Implants Based on an Active Nonlinear Filterbank Model of the Biological Cochlea," *IEEE Trans. Biomed. Eng*, vol. 56 (3), pp. 828–836, 2009.

[KDL06]   A. G. Katsiamis, E. M. Drakakis, and R. F. Lyon, "Introducing the Differentiated All-Pole and One-Zero Gammatone Filter Responses and their Analog VLSI Log-domain Implementation," in *Proc. IEEE Int. Midwest Symp. Circuits Syst*, pp. 561–565, 2006.

[KDL07]   A. G. Katsiamis, E. M. Drakakis, and R. F. Lyon, "Practical Gammatone-Like Filters for Auditory Processing," *EURASIP J. Audio Speech Music Process*, vol. 2007, Article ID 63685, 15 pages, 2007.

[KKK07]   K. H. Kim, J. H. Kim, and D. H. Kim, "An improved speech processor for cochlear implant based on active nonlinear model of biological cochlea," *in Proc. 29th Ann. Int. Conf. IEEE EMBS*, pp. 6352–6355, 2007.

[KLW+11]   B. Kollmeier, T. Lenarz, A. Winkler, M. A. Zokoll, H. Sukowski, T. Brand, and K. C. Wagener, "Hörgeräteindikation und -überprüfung nach modernen Verfahren der Sprachaudiometrie im Deutschen," *HNO*, vol. 59 (10), pp. 1012–1021, 2011.

[KMP73]   D. O. Kim, C. E. Molnar, and R. R. Pfeiffer, "A system of nonlinear differential equations modeling basilar-membrane motion," *J. Acoust. Soc. Am*, vol. 54 (6), pp. 1517–1529, 1973.

[Kri06]   B. S. Krishna, "Comment on "Auditory-nerve first-spike latency and auditory absolute threshold: A computer model" [J. Acoust. Soc. Am.119, 406–417 (2006)]," *J. Acoust. Soc. Am*, vol. 120 (2), pp. 591–593, 2006.

[KW90]   R. C. Kidd and T. F. Weiss, "Mechanisms that degrade timing information in the cochlea," *Hear. Res*, vol. 49, pp. 181–208, 1990.

[LBA07]   E. A. Lopez-Poveda, L. F. Barrios, and A. Alves-Pinto, "Psychophysical estimates of level-dependent best-frequency shifts in the apical region of the human basilar membrane," *J. Acoust. Soc. Am*, vol. 121 (6), pp. 3646–3654, 2007.

[LBL+10]   D. S. Lazard, P. Bordure, G. Lina-Granade, J. Magnan, R. Meller, B. Meyer, E. Radafy, P.-E. Roux, D. Gnansia, V. Péan, and E. Truy, "Speech perception performance for 100 post-lingually deaf adults fitted with Neurelec cochlear implants: comparison between Digisonic® Convex and Digisonic® SP devices after a 1-year follow-up," *Acta Otolaryngol*, vol. 130 (11), pp. 1267–1273, 2010.

[LE06]   E. A. Lopez-Poveda and A. Eustaquio-Martín, "A Biophysical Model of the Inner Hair Cell: The Contribution of Potassium Currents to Peripheral Auditory Compression," *JARO*, vol. 7, pp. 218–235, 2006.

[Lev71]   H. Levitt, "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am*, vol. 49 (2B), pp. 467–477, 1971.

[LKD10]   R. F. Lyon, A. G. Katsiamis, and E. M. Drakakis, "History and Future of Auditory Filter Models," in *Proc. IEEE Int. Symp. Circuits Syst.* (ISCAS), New York, USA: IEEE Press, pp. 3809–3812, 2010.

[LKS86]   L. F. Lamel, R. H. Kassel, and S. Seneff, "Speech database development: Design and analysis of the acoustic-phonetic corpus," in *Proc. DARPA Speech Recognition Workshop*, Report No. SAIC-86/1546, pp. 100–109, 1986.

[LLM07]   A. Lopez-Najera, E. A. Lopez-Poveda, and R. Meddis, "Further studies on the dual-resonance nonlinear filter model of cochlear frequency selectivity: Responses to tones," *J. Acoust. Soc. Am*, vol. 122 (4), pp. 2124–2134, 2007.

[LM01]   E. A. Lopez-Poveda and R. Meddis, "A human nonlinear cochlear filterbank," *J. Acoust. Soc. Am*, vol. 110 (6), pp. 3107–3118, 2001.

[LM08]   B. Laback and P. Majdak, "Binaural jitter improves interaural time-difference sensitivity of cochlear implantees at high pulse rates," *PNAS*, vol. 105 (2), pp. 814–817, 2008.

[Loi98]   P. C. Loizou, "Mimicking the human ear: An Overview of Signal-Processing Strategies for Converting Sound into Electrical Signals in Cochlear Implants," *IEEE Signal Process. Mag*, vol. 15 (5), pp. 101–130, 1998.

[Loi06]   P. C. Loizou, "Speech Processing in Vocoder-Centric Cochlear Implants," in *Advances in Oto-Rhino-Laryngology*, vol. 64, *Cochlear and Brainstem Implants*, A. R. Møller, Ed, Basel, Switzerland: Karger, pp. 109–143, 2006.

[LOM98]   E. A. Lopez-Poveda, L. P. O'Mard, and R. Meddis, "A revised computational inner hair cell model," in *Psychophysical and Physiological Advances in Hearing*, London, UK: Whurr Publishers, pp. 112–121, 1998.

[Lop03]   E. A. Lopez-Poveda, "An approximate transfer function for the dual-resonance nonlinear filter model of auditory frequency selectivity," *J. Acoust. Soc. Am*, vol. 114 (4), pp. 2112–2117, 2003.

[Lop04]   E. A. Lopez-Poveda, "Reply to "Comment on 'An approximate transfer function for the dual-resonance nonlinear filter model of auditory frequency selectivity'" (L)," *J. Acoust. Soc. Am*, vol. 115 (5), p. 1891, 2004.

[Lum87a]   G. Lumer, "Computer Model of Cochlear Preprocessing (Steady State Condition) I. Basics and Results for one Sinusoidal Input Signal," *Acta Acust. united Ac*, vol. 62 (4), pp. 282–290, 1987.

[Lum87b]   G. Lumer, "Computer Model of Cochlear Preprocessing (Steady State Condition) II. Two-tone Suppression," *Acta Acust. united Ac*, vol. 63 (1), pp. 17–25, 1987.

[LWM04]   J. Laneau, J. Wouters, and M. Moonen, "Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees," *J. Acoust. Soc. Am*, vol. 116 (6), pp. 3606–3619, 2004.

[Lyo84]   R. F. Lyon, "Computational models of neural auditory processing," in *Proc. Int. Conf. Acoust, Speech, Signal Process. (IEEE ICASSP)*, pp. 41–44, 1984.

[Lyo96]   R. F. Lyon, "The all-pole gammatone filter and auditory models," in *Proc. Forum Acusticum*, Stuttgart, Germany: S. Hirzel Verlag, 1996.

[Lyo97]   R. F. Lyon, "All-pole models of auditory filtering," in *Diversity in Auditory Mechanics*, E. R. Lewis, G. R. Long, R. F. Lyon, C. R. Steele, and E. Hecht-Poinar, Eds, River Edge, New Jersey, USA: World Scientific, pp. 205–211, 1997.

[Lyo11]   R. F. Lyon, "Cascades of two-pole–two-zero asymmetric resonators are good models of peripheral auditory function," *J. Acoust. Soc. Am*, vol. 130 (6), pp. 3893–3904, 2011.

[LZOS10]   A. Lorens, M. Zgoda, A. Obrycka, and H. Skarżynski, "Fine Structure Processing improves speech perception as well as objective and subjective benefits in pediatric MED-EL COMBI 40+ users," *Int. J. Pediatr. Otorhinolaryngol*, vol. 74 (12), pp. 1372–1378, 2010.

[Mal09]   T. K. Malherbe, "Development of a method to create subject specific cochlear models for electric hearing," M.Eng. thesis, Engineering, the Built Environment and Information Technology, University of Pretoria, South Africa, 2009.

[McG09]   S. G. McGovern, "Fast image method for impulse response calculations of box-shaped rooms," *Appl Acoust*, vol. 70 (1), pp. 182–189, 2009.

[Med86]   R. Meddis, "Simulation of mechanical to neural transduction in the auditory receptor," *J. Acoust. Soc. Am*, vol. 79 (3), pp. 702–711, 1986.

[Med88]   R. Meddis, "Simulation of auditory–neural transduction: Further studies," *J. Acoust. Soc. Am*, vol. 83 (3), pp. 1056–1063, 1988.

[Med06a]   R. Meddis, "Auditory-nerve first-spike latency and auditory absolute threshold: A computer model," *J. Acoust. Soc. Am*, vol. 119 (1), pp. 406–417, 2006.

[Med06b]   R. Meddis, "Reply to comment on "Auditory-nerve first-spike latency and auditory absolute threshold: A computer model" (L)," *J. Acoust. Soc. Am*, vol. 120 (3), pp. 1192–1193, 2006.

[Mei12]   D. Meister, MED-EL Research & Development, "The FSP, FS4, and FS4p strategies", personal communication (e-mail), Mar. 2012.

[MFLP10]   R. Meddis, R. R. Fay, E. A. Lopez-Poveda, and A. N. Popper, Eds, *Computational Models of the Auditory System*. Boston, Massachusetts, USA: Springer Science + Business Media LLC, 2010.

[MHS90]   R. Meddis, M. J. Hewitt, and T. M. Shackleton, "Implementation details of a computation model of the inner hair-cell auditory-nerve synapse," *J. Acoust. Soc. Am*, vol. 87 (4), pp. 1813–1816, 1990.

[ML10]   R. Meddis and E. A. Lopez-Poveda, "Auditory Periphery: From Pinna to Auditory Nerve," in *Springer Handbook of Auditory Research*, vol. 35, *Computational Models of the Auditory System*, R. Meddis, R. R. Fay, E. A. Lopez-Poveda, and A. N. Popper, Eds, Boston, Massachusetts, USA: Springer Science+Business Media LLC, pp. 7–38, 2010.

[MM00]   U. Moissl and U. Meyer-Base, "A comparison of different methods to assess phase-locking in auditory neurons," in *Proc. 22nd Ann. Int. Conf. IEEE EMBS*, pp. 840–843, 2000.

[MO05]   R. Meddis and L. P. O'Mard, "A computer model of the auditory-nerve response to forward-masking stimuli," *J. Acoust. Soc. Am*, vol. 117 (6), pp. 3787–3798, 2005.

[MOL01]   R. Meddis, L. P. O'Mard, and E. A. Lopez-Poveda, "A computational algorithm for computing nonlinear auditory frequency selectivity," *J. Acoust. Soc. Am*, vol. 109 (6), pp. 2852–2861, 2001.

[Møl65]   A. R. Møller, "An experimental study of the acoustic impedance of the middle ear and its transmission properties," *Acta Otolaryngol*, vol. 60 (1-6), pp. 129–149, 1965.

[Møl06a]   A. R. Møller, *Hearing: Anatomy, physiology, and disorders of the auditory system*, 2nd ed. San Diego, California, USA: Academic Press, 2006.

[Møl06b]   A. R. Møller, "History of Cochlear Implants and Auditory Brainstem Implants," in *Advances in Oto-Rhino-Laryngology*, vol. 64, *Cochlear and Brainstem Implants*, A. R. Møller, Ed, Basel, Switzerland: Karger, pp. 1–10, 2006.

[MS03]   A. McEwan and A. van Schaik, "An Analogue VLSI Implementation of the Meddis Inner Hair Cell Model," *EURASIP J. Appl. Signal Process*, vol. 7, pp. 639–648, 2003.

[MS04]   A. McEwan and A. van Schaik, "An Alternative Analog VLSI Implementation of the Meddis Inner Hair Cell Model," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, New York, USA: IEEE Press, vol. 4, pp. 928–931, 2004.

[MS07]   R. H. Margolis and G. L. Saly, "Toward a standard description of hearing loss," *Int. J. Audiol*, vol. 46 (12), pp. 746-458, 2007.

[MW47]   H. B. Mann and D. R. Whitney, "On a test of Whether one of two random variables is stochastically larger than the other," *Ann. Math. Statist*, vol. 18 (1), pp. 50–60, 1947.

[Nam05]   J.-H. Nam, "A Computational Study on the Structure, Dynamics and Mechanoelectric Transduction of Vestibular Hair cell," PhD dissertation, Department of Mechanical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA, 2005.

[NCG07]   J.-H. Nam, J. R. Cotton, and W. Grant, "A Virtual Hair Cell, I: Addition of Gating Spring Theory into a 3-D Bundle Mechanical Model," *Biophys. J*, vol. 92 (6), pp. 1918–1928, 2007.

[NDA91]   A. L. Nuttall, D. F. Dolan, and G. Avinash, "Laser Doppler velocimetry of basilar membrane vibration," *Hear. Res*, vol. 51 (2), pp. 203–213, 1991.

[Nee85]   S. T. Neely, "Mathematical modeling of cochlear mechanics," *J. Acoust. Soc. Am*, vol. 78 (1), pp. 345–352, 1985.

[NHE+07]   W. Nogueira, T. Harczos, B. Edler, J. Ostermann, and A. Büchner, "Automatic Speech Recognition with a Cochlear Implant Front-End," in *Proc. 8th Ann. Conf. Int. Speech Comm. Assoc.: INTERSPEECH*, pp. 2537–2540, 2007.

[NIH10]   U. S. National Institutes of Health, *Cochlear Implants: Facts Sheet*, 2010. Available: http://report.nih.gov/NIHfactsheets/Pdfs/CochlearImplants(NIDCD).pdf (February 06, 2012).

[NK83]   S. T. Neely and D. O. Kim, "An active cochlear model showing sharp tuning and high sensitivity," *Hear. Res*, vol. 9 (2), pp. 123–130, 1983.

[NKH+07]   W. Nogueira, A. Kátai, T. Harczos, F. Klefenz, A. Buechner, and B. Edler, "An Auditory Model Based Strategy for Cochlear Implants," *in Proc. 29th Ann. Int. Conf. IEEE EMBS*, pp. 4127–4130, 2007.

[NLE+09]   W. Nogueira, L. M. Litvak, B. Edler, J. Ostermann, and A. Büchner, "Signal Processing Strategies for Cochlear Implants Using Current Steering," *EURASIP J. Adv. Signal Process*, vol. 2009, Article ID 531213, 20 pages, 2009.

[Nog08]   W. Nogueira, "Design and Evaluation of Signal Processing Strategies for Cochlear Implants based on Psychoacoustic Masking and Current Steering," PhD dissertation, Faculty of Electrical Engineering and Computer Science, Leibniz Universität Hannover, Germany, 2008.

[NSZ05]   K. Nie, G. Stickney, and F.-G. Zeng, "Encoding frequency modulation to improve cochlear implant performance in noise," *IEEE Trans. Biomed. Eng*, vol. 52 (1), pp. 64–73, 2005.

[OH61]   R. Oetinger and H. Hauser, "Ein elektrischer Kettenleiter zur Untersuchung der mechanischen Schwingungsvorgänge im Innenohr," *Acustica*, vol. 11, pp. 161–177, 1961.

[OP97]   A. J. Oxenham and C. J. Plack, "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," *J. Acoust. Soc. Am*, vol. 101 (6), pp. 3666–3675, 1997.

[OSB99]   A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time signal processing*, 2nd ed., Upper Saddle River, New Jersey, USA: Prentice Hall, 1999.

[PA95]   R. D. Patterson and M. H. Allerhand, "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," *J. Acoust. Soc. Am*, vol. 98 (4), pp. 1890–1894, 1995.

[PB50]   L. C. Peterson and B. P. Bogert, "A Dynamical Theory of the Cochlea," *J. Acoust. Soc. Am*, vol. 22 (3), pp. 369–381, 1950.

[PBG06]   J. F. Patrick, P. A. Busby, and P. J. Gibson, "The Development of the Nucleus® Freedom™ Cochlear Implant System," *Trends in Amplif*, vol. 10 (4), pp. 175–200, 2006.

[Pei90]   W. Peisl, "Beschreibung aktiver nichtlinearer Effekte der peripheren Schallverarbeitung des Gehörs durch ein Rechnermodell," PhD dissertation (in German), Faculty of Electrical Engineering and Information Technology, Technical University of Munich, Germany, 1990.

[Pfe70]   R. R. Pfeiffer, "A Model for Two-Tone Inhibition of Single Cochlear-Nerve Fibers," *J. Acoust. Soc. Am*, vol. 48 (6B), pp. 1373–1378, 1970.

[PFTP98]   M. Perko, I. Fajfar, T. Tuma, and J. Puhan, "Fast Fourier transform computation using a digital CNN simulator," in *Proc. 5th IEEE Int. Workshop on Cellular Neural Networks and Their Applications (CNNA'98)*, New York, USA: IEEE Press, pp. 230–235, 1998.

[PK07]   G. E. Pazienza and K. Karacs, "An automatic tool to design CNN-UM programs," in *Proc. 2007 European Conference on Circuit Theory and Design (ECCTD'2007)*, New York, USA: IEEE Press, pp. 492–495, 2007.

[PK08]   G. E. Pazienza and K. Karacs, "Feasability of CNN-based simulation of auditory models", personal communication (e-mail), Aug. 2008 – Sep. 2010.

[PM79]   R. Plomp and A. M. Mimpen, "Improving the reliability of testing the speech reception threshold for sentences," *Int. J. Audiol*, vol. 18 (1), pp. 43–52, 1979.

[PNHR88]   R. D. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice, "An efficient auditory filterbank based on the gammatone function," APU report 2341, MRC, Applied Psychology Unit, Cambridge, U.K, 1988.

[POD02]   C. J. Plack, A. J. Oxenham, and V. Drga, "Linear and Nonlinear Processes in Temporal Masking," *Acta Acust. united Ac*, vol. 88 (3), pp. 348–358, 2002.

[PRH+92]   R. D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and M. H. Allerhand, "Complex sounds and auditory images," in *Proc. 9th Int. Symp. Hearing: Auditory physiology and perception*, Oxford, England, UK: Pergamon, pp. 429–446, 1992.

[PUI03]   R. D. Patterson, M. Unoki, and T. Irino, "Extending the domain of center frequencies for the compressive gammachirp auditory filter," *J. Acoust. Soc. Am*, vol. 114 (3), pp. 1529–1542, 2003.

[Ran50]   O. F. Ranke, "Theory of Operation of the Cochlea: A Contribution to the Hydrodynamics of the Cochlea," *J. Acoust. Soc. Am*, vol. 22 (6), pp. 772–777, 1950.

[RC93]   T. Roska and L. O. Chua, "The CNN universal machine: an analogic array computer," *IEEE Trans. Circ. Syst. II*, vol. 40 (3), pp. 163–173, 1993.

[RE99]   A. Robert and J. L. Eriksson, "A composite model of the auditory periphery for simulating responses to complex sounds," *J. Acoust. Soc. Am*, vol. 106 (4), pp. 1852–1864, 1999.

[RLF+07]   J. Rouger, S. Lagleyre, B. Fraysse, S. Deneve, O. Deguine, and P. Barone, "Evidence that cochlear-implanted deaf patients are better multisensory integrators," *PNAS*, vol. 104 (17), pp. 7295–7300, 2007.

[RPL+85]   J. J. Rosowski, W. T. Peake, T. J. Lynch, R. Leong, and T. F. Weiss, "A model for signal transmission in an ear having hair cells with free-standing stereocilia. II. Macromechanical stage," *Hear. Res*, vol. 20 (2), pp. 139–155, 1985.

[RWFA99]   J. T. Rubinstein, B. S. Wilson, C. C. Finley, and P. J. Abbas, "Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation," *Hear. Res*, vol. 127 (1-2), pp. 108–118, 1999.

[San12]   N. Sanghvi, "Parallel Computation of the Meddis MATLAB Auditory Periphery Model," M.Sc. thesis, Department of Electrical and Computer Engineering, The Ohio State University, Columbus, Ohio, USA, 2012.

[SAO11]   A. L. L. Sampaio, M. F. S. Araújo, and C. A. C. P. Oliveira, "New Criteria of Indication and Selection of Patients to Cochlear Implant," *Int. J. Otolaryngol*, vol. 2011, Article ID 573968, 13 pages, 2011.

[SB11]   J. Saunders and D. Barrs, "Cochlear Implantation in Developing Countries as Humanitarian Service: Physician Attitudes and Recommendations for Best Practice," *Otolaryngol Head Neck*, vol. 145 (1), pp. 74–79, 2011.

[SBF04]   B. U. Seeber, U. Baumann, and H. Fastl, "Localization ability with bimodal hearing aids and bilateral cochlear implants," *J. Acoust. Soc. Am*, vol. 116 (3), pp. 1698–1709, 2004.

[Sch85]   D. Schofield, "Visualisations of speech based on a model of the peripheral auditory system," NPL Report DITC 62/85, Division of Information Technology and Computing, National Physical Laboratory, Teddington, England, UK, 1985.

[Sch03]   A. van Schaik, "A small analog VLSI inner hair cell model," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, New York, USA: IEEE Press, vol. 1, pp. 17–20, 2003.

[SCW+86]   S. A. Shamma, R. S. Chadwick, W. J. Wilbur, K. A. Morrish, and J. Rinzel, "A biophysical model of cochlear processing: Intensity dependence of pure tone responses," *J. Acoust. Soc. Am*, vol. 80 (1), pp. 133–145, 1986.

[SH74]   M. R. Schroeder and J. L. Hall, "Model for mechanical to neural transduction in the auditory receptor," *J. Acoust. Soc. Am*, vol. 55 (5), pp. 1055–1060, 1974.

[SH80]   H. J. M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am*, vol. 67 (1), pp. 318–326, 1980.

[SHKW09a] G. Szepannek, T. Harczos, F. Klefenz, and C. Weihs, "Combining different auditory model based feature extraction principles for feature enrichment in automatic speech recognition," in *Proc. 13th Int. Conf. on Speech and Computer (SPECOM)*, pp. 205–210, 2009.

[SHKW09b] G. Szepannek, T. Harczos, F. Klefenz, and C. Weihs, "Extending Features for Automatic Speech Recognition by Means of Auditory Modelling," in *Proc. 17th European Signal Proc. Conf. (EUSIPCO)*, pp. 1235–1239, 2009.

[Sie65]   W. M. Siebert, "Some Implications of the Stochastic Behavior of Primary Auditory Neurons," *Biol. Cybern*, vol. 2 (5), pp. 206–215, 1965.

[SKJ+14]   Z. M. Smith, A. Kan, H. G. Jones, M. Buhr-Lawler, S. P. Godar, and R. Y. Litovsky, "Hearing better with interaural time differences and bilateral cochlear implants," *J. Acoust. Soc. Am*, vol. 135 (4), p. 2190, 2014.

[SL93]   M. Slaney and R. F. Lyon, "On the Importance of Time: A Temporal Representation of Sound," in *Visual representations of speech signals*, M. Cooke, S. Beet, and M. Crawford, Eds, New York, USA: J. Wiley & Sons, pp. 95–115, 1993.

[Sla93]   M. Slaney, "An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank," Technical Report #35, Perception Group - Advanced Technology Group, Apple Computer Inc, Cupertino, California, USA, 1993.

[SLOM02]   C. J. Sumner, E. A. Lopez-Poveda, L. P. O'Mard, and R. Meddis, "A revised model of the inner-hair cell and auditory-nerve complex," *J. Acoust. Soc. Am*, vol. 111 (5), pp. 2178–2188, 2002.

[SLOM03]   C. J. Sumner, E. A. Lopez-Poveda, L. P. O'Mard, and R. Meddis, "Adaptation in a revised inner-hair cell model," *J. Acoust. Soc. Am*, vol. 113 (2), pp. 893–901, 2003.

[SM06]   B. Swanson and H. Mauch, "E10511DD: Nucleus MATLAB Toolbox 4.20 Software User Manual," Cochlear Ltd, Lane Cove, New South Wales, Australia, 2006.

[Smi13]   T. E. Smith, *Log4CPLUS: a C++ port of the Log for Java (log4j) logging library*. Available: http://log4cplus.sourceforge.net (2013-05-31).

[SOLM03]   C. J. Sumner, L. P. O'Mard, E. A. Lopez-Poveda, and R. Meddis, "A nonlinear filter-bank model of the guinea-pig cochlear nerve: Rate responses," *J. Acoust. Soc. Am*, vol. 113 (6), pp. 3264–3274, 2003.

[ST14]   Homepage of Spieltrieb, a duo of musicians from Oldenburg, Germany. Available: http://www.spieltriebhome.de/.

[Ste12]   R. M. Stern, "Applying physiologically-motivated models of auditory processing to automatic speech recognition," in *Proc. 3rd Int. Symp. Audit. Audiol. Res. (ISAAR 2011): Speech perception and auditory disorders*, Nyborg, Denmark: The Danavox Jubilee Foundation, pp. 283–294, 2012.

[Str85]   H. W. Strube, "A Computationally Efficient Basilar-Membrane Model," *Acustica*, vol. 58, pp. 207–214, 1985.

[SW65]   S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52 (3-4), pp. 591–611, 1965.

[SWWL03]   R. Schatzer, B. S. Wilson, R. D. Wolford, and D. T. Lawson, "Speech processors for auditory prostheses: Signal processing strategies for a closer mimicking of normal auditory functions," Sixth Quarterly Progress Report, NIH Contract N01-DC-2-1002, Neural Prosthesis Program, National Institutes of Health, Bethesda, Maryland, USA, 2003.

[SZK+95]   R. V. Shannon, F.-G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech Recognition with Primarily Temporal Cues," *Science*, vol. 270 (5234), pp. 303–304, 1995.

[Taf09]   D. A. Taft, "Cochlear Implant Sound Coding with Across-frequency Delays," PhD dissertation, Department of Electrical and Electronic Engineering and Department of Otolaryngology, The University of Melbourne, 2009.

[TC03]   Q. Tan and L. H. Carney, "A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide*," J. Acoust. Soc. Am*, vol. 114 (4), pp. 2007–2020, 2003.

[TEF10]   The Ear Foundation, *Cochlear Implants: 2010*. Available: http://www.earfoundation.org.uk/downloads/ci2010.pdf (February 06, 2012).

[TGB09]   D. A. Taft, D. B. Grayden, and A. N. Burkitt, "Speech coding with traveling wave delays: Desynchronizing cochlear implant frequency bands with cochlea-like group delays," *Speech Commun*, vol. 51 (11), pp. 1114–1123, 2009.

[TGB10]   D. A. Taft, D. B. Grayden, and A. N. Burkitt, "Across-Frequency Delays Based on the Cochlear Traveling Wave: Enhanced Speech Presentation for Cochlear Implants," *IEEE Trans. Biomed. Eng*, vol. 57 (3), pp. 596–606, 2010.

[TMTQ04]   Á. de la Torre Vega, M. B. Martí, R. de la Torre Vega, and M. S. Quevedo, "Cochlear Implant Simulation version 2.0: Description and usage of the program," University of Granada, Spain, 2004.

[Tuc46]   D. G. Tucker, "Transient Response of Tuned-Circuit Cascades," *Wireless Engr*, vol. 23, pp. 250–258, 1946.

[TW80]   R. S. Tyler and E. J. Wood, "A Comparison of Manual Methods for Measuring Hearing Levels," *Audiology*, vol. 19 (4), pp. 316–329, 1980.

[VPV10]   K. Vermeire, A. K. Punte, and P. Van de Heyning, "Better Speech Recognition in Noise with the Fine Structure Processing Coding Strategy," *ORL*, vol. 72 (6), pp. 305–311, 2010.

[VST+05]   A. E. Vandali, C. Sucher, D. J. Tsang, C. M. McKay, J. W. D. Chew, and H. J. McDermott, "Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies," *J. Acoust. Soc. Am*, vol. 117 (5), pp. 3126–3138, 2005.

[VWPC00]   A. E. Vandali, L. A. Whitford, K. L. Plant, and G. M. Clark, "Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system," *Ear Hear*, vol. 21 (6), pp. 608–624, 2000.

[VZ10]   H. F. Voigt and X. Zheng, "The Cochlear Nucleus: The New Frontier," in *Springer Handbook of Auditory Research*, vol. 35, *Computational Models of the Auditory System*, R. Meddis, R. R. Fay, E. A. Lopez-Poveda, and A. N. Popper, Eds, Boston, Massachusetts, USA: Springer Science+Business Media LLC, pp. 39–63, 2010.

[Wat93]   L. Watts, "Cochlear Mechanics: Analysis and Analog VLSI," PhD dissertation, Computation and Neural System, California Institute of Technology, CA, USA, 1993.

[WB09]   B. Wen and K. Boahen, "A Silicon Cochlea with Active Coupling," *IEEE Trans. Biomed. Circuits Syst*, vol. 3 (6), pp. 444–455, 2009.

[WBK99a]   K. Wagener, T. Brand, and B. Kollmeier, "Entwicklung und Evaluation eines Satztests in deutscher Sprache II: Optimierung des Oldenburger Satztests (Development and evaluation of a sentence test in German language II: Optimization of the Oldenburg sentence test)," *Z. Audiol*, vol. 38 (2), pp. 44–56, 1999.

[WBK99b]   K. Wagener, T. Brand, and B. Kollmeier, "Entwicklung und Evaluation eines Satztests in deutscher Sprache III: Evaluation des Oldenburger Satztests (Development and evaluation of a sentence test in German language III: Evaluation of the Oldenburg sentence test)," *Z. Audiol*, vol. 38 (3), pp. 86–95, 1999.

[WD08a]   B. S. Wilson and M. F. Dorman, "Cochlear implants: A remarkable past and a brilliant future," *Hear. Res*, vol. 242 (1-2), pp. 3–21, 2008.

[WD08b]   B. S. Wilson and M. F. Dorman, "Cochlear implants: Current designs and future possibilities," *J. Rehabil. Res. Dev*, vol. 45 (5), pp. 695–730, 2008.

[WDKR10]   J. H. Won, W. R. Drennan, R. S. Kang, and J. T. Rubinstein, "Psychoacoustic abilities associated with music perception in cochlear implant users," *Ear Hear*, vol. 31 (6), pp. 796–805, 2010.

[Wei63]   T. F. Weiss, "A Model for Firing Patterns of Auditory Nerve Fibers," Quarterly Progress Report no. 69, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, pp. 217–223, 1963.

[Wei64]   T. F. Weiss, "A model for firing patterns of auditory nerve fibers," PhD dissertation, Research Lab of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, 1964.

[Wei66]   T. F. Weiss, "A model of the peripheral auditory system," *Biol. Cybern*, vol. 3 (4), pp. 153–175, 1966.

[Wes04]   T. Wesarg, "Entwicklung und klinische Untersuchung von Sprachkodierungsstrategien für mehrkanalige Cochleaimplantat-Systeme zur elektrischen Stimulation des Hörnervs," PhD dissertation (in German), Faculty of Electrical Engineering and Information Technology, Otto von Guericke University of Magdeburg, Germany, 2004.

[WFL+93]   B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, and M. Zerbi, "Design and evaluation of a continuous interleaved sampling (CIS) processing strategy for multichannel cochlear implants," *J. Rehabil. Res. Dev*, vol. 30 (1), pp. 110–116, 1993.

[WHB10]   S. Werner, T. Harczos, and K. Brandenburg, "Overview of numerical models of cell types in the cochlear nucleus," in *Proc. 2nd Int. Symp. Audit. Audiol. Res. (ISAAR 2009): Binaural Processing and Spatial Hearing*, Helsingør, Denmark: The Danavox Jubilee Foundation, pp. 61–70, 2010.

[Wil45]   F. Wilcoxon, "Individual comparisons by ranking methods," *Biomet. Bull*, vol. 1 (6), pp. 80–83, 1945.

[WKK99]   K. Wagener, V. Kühnel, and B. Kollmeier, "Entwicklung und Evaluation eines Satztests in deutscher Sprache I: Design des Oldenburger Satztests (Development and evaluation of a sentence test in German language I: Design of the Oldenburg sentence test)," *Z. Audiol*, vol. 38 (1), pp. 4–15, 1999.

[WL24]   R. L. Wegel and C. E. Lane, "The Auditory Masking of One Pure Tone by Another and its Probable Relation to the Dynamics of the Inner Ear," *Phys. Rev*, vol. 23 (2), pp. 266–285, 1924.

[WL85]   T. F. Weiss and R. Leong, "A model for signal transmission in an ear having hair cells with free-standing stereocilia. III. Micromechanical stage," *Hear. Res*, vol. 20 (2), pp. 157–174, 1985.

[WLS10]   B. S. Wilson, E. A. Lopez-Poveda, and R. Schatzer, "Use of auditory models in developing coding strategies for cochlear implants," in *Springer Handbook of Auditory Research*, vol. 35, *Computational Models of the Auditory System*, R. Meddis, R. R. Fay, E. A. Lopez-Poveda, and A. N. Popper, Eds, Boston, Massachusetts, USA: Springer Science+Business Media LLC, pp. 237–260, 2010.

[WLZF92]	B. S. Wilson, D. T. Lawson, M. Zerbi, and C. C. Finley, "Speech processors for auditory prostheses: Virtual channel interleaved sampling (VCIS) processors," First Quarterly Progress Report, NIH Contract N01-DC-2-2401, Neural Prosthesis Program, National Institutes of Health, Bethesda, Maryland, USA, 1992.

[WM04]	L. Wiegrebe and R. Meddis, "The representation of periodic sounds in simulated sustained chopper units of the ventral cochlear nucleus," *J. Acoust. Soc. Am*, vol. 115 (3), pp. 1207–1218, 2004.

[WPR85]	T. F. Weiss, W. T. Peake, and J. J. Rosowski, "A model for signal transmission in an ear having hair cells with free-standing stereocilia. I. Empirical basis for model structure," *Hear. Res*, vol. 20 (2), pp. 131–138, 1985.

[WRZ+98]	B. S. Wilson, S. Rebscher, F.-G. Zeng, R. V. Shannon, G. E. Loeb, D. T. Lawson, and M. Zerbi, "Design for an inexpensive but effective cochlear implant," *Otolaryngol Head Neck*, vol. 118 (2), pp. 235–241, 1998.

[WSL+05]	B. S. Wilson, R. Schatzer, E. A. Lopez-Poveda, X. Sun, D. T. Lawson, and R. D. Wolford, "Two New Directions in Speech Processor Design for Cochlear Implants," *Ear Hear*, vol. 26 (4, suppl.), pp. 73S-81S, 2005.

[WSL06]	B. S. Wilson, R. Schatzer, and E. A. Lopez-Poveda, "Possibilities for a closer mimicking of normal auditory functions with cochlear implants," in *Cochlear implants*, S. B. Waltzman and J. T. Roland, Eds. 2nd ed, New York, USA: Thieme Medical, pp. 48–56, 2006.

[WWS+03]	B. S. Wilson, R. D. Wolford, R. Schatzer, X. Sun, and D. T. Lawson, "Speech processors for auditory prostheses: Combined use of DRNL filters and virtual channels," Seventh Quarterly Progress Report, NIH Contract N01-DC-2-1002, Neural Prosthesis Program, National Institutes of Health, Bethesda, Maryland, USA, 2003.

[You94]	S. J. Young, "The HTK Hidden Markov Model Toolkit: Design and Philosophy," CUED/F-INFENG/TR.152, Engineering Department, Cambridge University, England, U.K, 1994.

[YWR90]	G. K. Yates, I. M. Winter, and D. Robertson, "Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range," *Hear. Res*, vol. 45 (3), pp. 203–219, 1990.

[Zbo97]	G. Zboray, *Az érzékszervek (összehasonlító anatómiai előadások XII.)*, Budapest, Hungary: Eötvös Kiadó, 1997.

[ZFM00]	F.-G. Zeng, Q.-J. Fu, and R. Morse, "Human hearing enhanced by noise," *Brain Research*, vol. 869 (1-2), pp. 251–255, 2000.

[ZGK+07]	A. A. Zekveld, E. L. J. George, S. E. Kramer, S. T. Goverts, and T. Houtgast, "The Development of the Text Reception Threshold Test: A Visual Analogue of the Speech Reception Threshold Test," *J. Speech Lang. Hear. Res*, vol. 50 (3), pp. 576–584, 2007.

[ZHBC01]	X. Zhang, M. G. Heinz, I. C. Bruce, and L. H. Carney, "A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression*," J. Acoust. Soc. Am*, vol. 109 (2), pp. 648–670, 2001.

[Zie01]   C. M. Zierhofer, "Electrical nerve stimulation based on channel specific sampling sequences", World Patent WO 01/13991 A1, 2001.

[ZP90]   E. Zwicker and W. Peisl, "Cochlear preprocessing in analog models, in digital models and in human inner ear," *Hear. Res*, vol. 44, pp. 209–216, 1990.

[ZRH+08]   F.-G. Zeng, S. Rebscher, W. Harrison, X. Sun, and F. Haihong, "Cochlear Implants: System Design, Integration, and Evaluation," *IEEE Rev. Biomed. Eng*, vol. 1, pp. 115–142, 2008.

[ZSG90]   V. Zue, S. Seneff, and J. Glass, "Speech database development at MIT: TIMIT and beyond," *Speech Commun*, vol. 9 (4), pp. 351–356, 1990.

[ZU93]   J. Zadák and R. Unbehauen, "An application of mapping neural networks and a digital signal processor for cochlear neuroprostheses," *Biol. Cybern*, vol. 68 (6), pp. 545–552, 1993.

[Zwi48]   J. J. Zwislocki, "Theorie der Schneckenmechanik: qualitative und quantitative Analyse," *Acta Otolaryngol Suppl. (Stockh)*, vol. 72, pp. 1–76, 1948.

[Zwi62]   J. J. Zwislocki, "Analysis of the Middle-Ear Function. Part I: Input Impedance," *J. Acoust. Soc. Am*, vol. 34 (9B), pp. 1514–1523, 1962.

[Zwi79]   E. Zwicker, "A Model Describing Nonlinearities in Hearing by Active Processes with Saturation at 40 dB," *Biol. Cybern*, vol. 35 (4), pp. 243–250, 1979.

[Zwi86]   E. Zwicker, "A hardware cochlear nonlinear preprocessing model with active feedback," *J. Acoust. Soc. Am*, vol. 80 (1), pp. 146–153, 1986.

This document was *not* created with $\text{\LaTeX}$.

That made me win a bet.