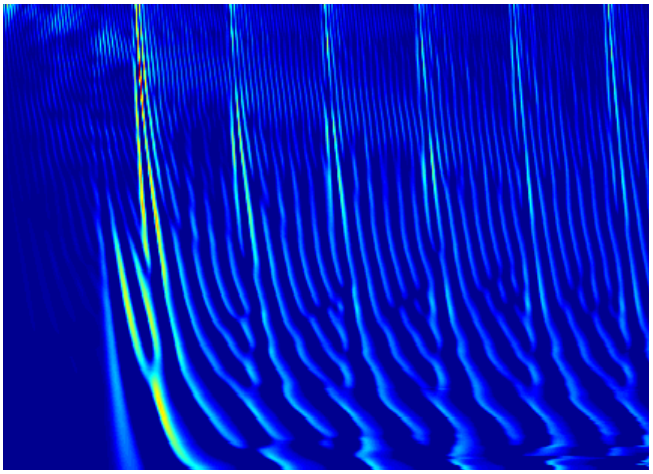
Technical University of Denmark

CONTRIBUTIONS TO
HEARING RESEARCH

Volume 12

*Filip Munch Rønne*

# Modeling auditory evoked potentials to complex stimuli



DTU Electrical Engineering
Department of Electrical Engineering

The defense was held on December 10th, 2012.

This PhD-dissertation is the result of a research project at the Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark (Kgs. Lyngby, Denmark).

# Supervisors

## Main supervisor

Prof. Torsten Dau
Centre for Applied Hearing Research
Department of Electrical Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

## Co-supervisor

Ass. Prof. James Harte
Institute of Digital Healthcare, WMG
University of Warwick, Coventry
CV4 7AL, United Kingdom

## Co-supervisor

Dr. Med. Claus Elberling
William Demant Holding A/S
Kongebakken 9, DK-2765 Smørum, Denmark

# Abstract

The auditory evoked potential (AEP) is an electrical signal that can be recorded from electrodes attached to the scalp of a human subject when a sound is presented. The signal is considered to reflect neural activity in response to the acoustic stimulation and is a well established clinical and research tool to objectively assess the function and integrity of the auditory nervous system. However, the physiological generation of AEPs represents a complicated interaction between linear and nonlinear cochlear and neural processes and is not well understood in humans. This thesis presents and evaluates a phenomenological model of AEP generation that can predict key experimental observations of recorded AEPs. The purpose of the study was to investigate the role of the different stages of auditory signal processing and their effects on AEP generation.

In recent years, there has been a push both clinically and in research towards using realistic and complex stimuli, such as speech, to electrophysiologically assess the human hearing. However, to interpret the AEP generation to complex sounds, the potential patterns in response to simple stimuli needs to be understood. Therefore, the model was used to simulate auditory brainstem responses (ABRs) evoked by classic stimuli like clicks, tone bursts and chirps. The ABRs to these simple stimuli were compared to literature data and the model was shown to predict the frequency dependence of tone-burst ABR wave-V latency and the level-dependence of ABR wave-V amplitude for clicks and chirps varying sweeping rates. The model was also evaluated based on ABR recordings evoked by speech syllables, and was shown to account for the differences in the responses observed between the stimuli. It was demonstrated that the generation of the syllable-evoked ABRs was highly influenced by cochlear and afferent neural processing, which supported the importance of cochlear processing for the generation of AEPs.

A second major contribution of this study was the investigation of whether auditory steady-state responses (ASSRs) can be used to assess human cochlear compression. Sensorineural hearing impairments is commonly associated with a loss of outer hair-cell functionality, and a measurable consequence is the decreased amount of cochlear

compression at frequencies corresponding to the damaged locations in the cochlea. In clinical diagnostics, a fast and objective measure of local cochlear compression would be of great benefit, as a more precise diagnose of the deficits underlying a potential hearing impairment in both infants and adults could be obtained. It was demonstrated in this thesis, via experimental recordings and supported by model simulations, that the growth of the ASSR amplitude with stimulus level can indeed be used as such an estimate of local cochlear compression.

# Resumé

Akustisk udløste potentialer (Auditory evoked potentials, AEP)[1] er elektriske signaler, der kan måles via elektroder fastgjort til hovedskallen af en person, når lyd bliver præsenteret for personen. Signalet, der forventes at reflektere den neurale aktivitet, der sker i respons til et akustisk stimulus, er et anerkendt værktøj til at evaluere funktionaliteten og integriteten af det auditive neurale system. De fysiologiske mekanismer, der genererer AEPer, repræsenterer dog en kompliceret interaktion mellem både lineære og ulineære processer, som ikke er særlig godt forstået. Denne afhandling præsenterer og evaluerer en fænomenologisk model af AEP generering, som kan simulere vigtige eksperimentelle AEP observationer.

I de senere år har der været en øget interesse i både den kliniske verden og i forskningsverdenen for at bruge realistiske og komplekse stimuli, som fx tale, til elektrofysiologisk at evaluere den menneskelige hørelse. For at kunne fortolke AEPer målt med komplekse stimuli er det dog nødvendigt at have forstået de potentielle mønstre, som mere simple stimuli genererer. I denne afhandling blev AEP modellen derfor brugt til at simulere akustiske hjernestammeresponser (Auditory brainstem response, ABR) til klassiske stimuli som klik, tonepulser og chirps. De simulerede ABRer blev sammenlignet med data fra litteraturen. Det blev vist, at modellen kunne simulere frekvensafhængigheden af ABR bølge-V-latenstiden, når stimuli var tonepulser, samt simulere niveauafhængigheden af ABR bølge-V-amplituden når stimuli var enten klik eller chirps med varierende stigningstid. Modellen blev også evalueret på ABR målt med stavelsesstimuli, og det blev vist at den kunne redegøre for målte responsforskelle mellem forskellige stavelser. Det blev demonstreret, at genereringen af AEPer målt med stavelsesstimuli var meget påvirket af cochleær processeringen og den afferente neurale processering. Dette understreger vigtigheden af cochleær processeringen i AEP-genereringen.

Et andet vigtigt bidrag i denne afhandling er studiet af, hvorvidt det akustiske

[1] I dette resume er begreber så vidt muligt oversat til dansk, hvorimod forkortelser, for at undgå forvirring, er bibeholdt i deres originale engelske version.

steady-state respons (Auditory steady-state response, ASSR) kan bruges til at evaluere cochleær kompression i mennesker. Sensorineurale høretab bliver ofte associeret med tab af funktionaliteten af de ydre hårceller. En målbar konsekvens af dette er en nedsat cochleær kompression ved de frekvenser, der svarer til de beskadigede steder i øresneglen. I klinisk diagnostik ville en hurtig og objektiv test af lokal cochleær kompression være meget værdsat, da mere præcise diagnoser af den underliggende fysiologiske årsag til et potentielt høretab ville kunne blive stillet for både spædbørn og voksne. I denne afhandling blev det demonstreret, via eksperimentelt arbejde såvel som simulationer, at væksten i ASSR-amplitude med stigende stimulusniveau kan bruges som et sådant værktøj til at estimere den lokale cochleær kompression.

# Preface

In March 2009 I started the journey towards this Ph.D. thesis. It has been a journey full of experiences, challenges and good times. I have thoroughly enjoyed it. In the three years I have come across quite a number of people to whom I owe my deepest thanks.

First, naturally, Torsten Dau who gave me this opportunity. Torsten has been very supportive and has engaged in hours of interesting discussions with me. Torsten, I have always enjoyed your great enthusiasm for this field of research and for the people conducting it. It has been a true privilege to have a boss and supervisor who I also call my friend.

James Harte, for being so generous with help and discussions. I thank you for all our enjoyable on- and off-topic conversations at DTU, over Skype and naturally in your home in England. Also thanks for tirelessly explaining, describing and discussing academic content with me.

Claus Elberling, for being an inspiration and role model. Also thank you, for sharing your unpublished work with me, and for pointing out numerous literature studies I otherwise would have missed.

My colleagues at CAHR, DTU, for making my time spent there so enjoyable. Our talks over a cup of coffee, a game of table soccer, or across the office. Thank you.

Thanks to Ian Bruce, for providing me with the humanization of the AN model, and for answering many questions.

Also a thanks to friends and family who made these three years some of the best in my life. A special thanks to my fiancée Kirstine. Thanks for support and thanks for

bearing with me when my Ph.D. study (occasionally) was hard and frustrating. I love you.

And finally to the reader. I hope you find this work interesting.

*Filip Munch Rønne, July 7, 2012.*

# Related publications

## Journal articles

- Rønne F.M., Harte J., Elberling C. and Dau T. (2012). *Modeling auditory evoked brainstem responses to transient stimuli.* J. Acoust. Soc. Am., 131 (5), pp 3903 - 3913.
- Rønne F.M., Harte J., Møllenbach S.K. and Dau T. (2012). *Investigating the potential of auditory steady-state responses to assess loss of cochlear compression.* J. Acoust. Soc. Am., submitted.

## Conference papers

- Rønne F.M., Harte J., Elberling C. and Dau T. (2011). *Modelling the level-dependent latency of the auditory brainstem response,* in Proceedings of Forum Acusticum 2011.
- Rønne F.M., Gøtsche-Ramussen K. (2011). *Low-frequency versus high-frequency synchronization in chirp-evoked auditory brainstem responses,* in Proceedings of International Symposium on Auditory and Audiological Research (ISAAR) 2011, pp 275 - 282.
- Harte J., Rønne F.M., and Dau T. (2010). *Modeling human auditory evoked brainstem responses based on nonlinear cochlear processing,* in Proceedings of the 20th International Congress on Acoustics, ICA 2010.

# Contents

# 1 Introduction

The auditory evoked potential (AEP) is a sub branch of electroencephalography (EEG) that has been in use since the 1930s. It is an electrical signal that can be recorded from electrodes attached to the scalp of a human subject, when a sound is presented. The signal is believed to reflect neural activity in response to the acoustic stimulation, and can as such be used as a tool to objectively assess the function and integrity of the auditory nervous system. Unlike psychoacoustic measures, it does not necessarily (depending on the particular AEP understudy) require the attention of the test subject, making it specifically interesting to use with small children. The AEP is well established as a clinical tool to screen the hearing of infants. Besides clinical usage, it represents a powerful tool for research purposes. The AEP is objective, fairly fast to record and reproducible. It can be recorded from all stages in the auditory pathway, from the auditory nerve (AN) over the brainstem to the cortex. The earliest responses, stemming from the AN or brainstem, provides an assessment of the integrity of the mechanoelectrical transduction of sound in the auditory periphery and initial neural encoding prior to higher order cognitive processing, and thus offers a more direct method to investigate the auditory system than traditional psychoacoustic methods allow. However, much interest is typically focused on the behavioral outcome measures and the link between the two is currently not well understood. Given the type of stimulus and the recording settings, the neural generation site of the AEP can be varied. Common for all types of AEPs, independent of generation site and stimulus type, is that they produce a one-dimensional AEP pattern, where the electrical potential varies as a function of time. This pattern reflects a complex signal in the brain, where individual nerves contribute in various degrees to the recorded AEP. Further, the acoustic stimulation evoking the response has been processed through the entire auditory periphery including the nonlinear cochlea. The AEP is thus the result

of an acoustic stimulation and the processing through the middle ear, the nonlinear cochlear, and various subsequent neural sites, all of which produce complex neural activity that is then recorded on the scalp of the human subject. It is difficult to deduce the contributions from the different underlying physiological mechanisms, based on recorded AEPs. There is thus a need for models such as the AEP model presented in this thesis.

The generation of AEPs depend on various linear and nonlinear processes along the auditory pathway. One way to test hypotheses about the generator mechanisms underlying AEPs is to develop a model. Such a model should be able to predict key experimental observations of AEPs to various stimuli, as a benchmark. The present study develops and evaluates such an AEP model. The model is phenomenological implying that it has been built to mimic experimentally measured phenomena instead of strictly modeling the physiology of the auditory pathway. The model is divided into stages similar to the auditory system, and a link between simulated phenomena and the model stage producing the key feature of the phenomena can be established. This means that, if a simulated AEP predicts key features of experimentally measured AEPs, it is likely that the underlying physiology behind the actual recording resembles the functionality that has been modeled. The AEP model built in this study is capable of testing current hypotheses regarding the functionality of the different stages of the auditory pathway, and open for investigations of where the current knowledge is limited. Furthermore, the model can be a valuable tool to understand the consequences of hearing loss on the formation of AEPs and can help to improve the use of AEPs as a diagnostic tool.

The present thesis is structured as a collection of papers, where each chapter is based on a peer-reviewed paper published in a journal or a conference proceeding. The only exceptions are chapter 7 which is based on a submitted journal paper and chapter 8 which presents recent work, not yet submitted for a journal publication.

1. Chapter 2 provides a background overview of auditory evoked potentials and reviews existing models of AEP generation and selected models of the auditory periphery. This provides the reader with a historic overview of the field and

presents the approaches attempted to auditory and AEP modeling found in the literature.

2. Chapter 3 is based on Rønne et al. (2012) and develops an auditory brainstem (ABR) model capable of simulating transiently evoked potentials. The modeling framework and the underlying assumptions, used throughout this thesis, are presented in this chapter. The developed model contributes with insights into the complex nature of ABR generation, and the importance of the auditory periphery. Further, the model has been made available online[1] and can be used to investigate the representation of other types of stimuli as well.

3. Chapter 4 is based on Rønne et al. (2011). It investigates a limitation of the ABR model found in chapter 3, that the level-dependent latency of click-evoked ABRs is under-estimated. A second model, based on a different simulation of the auditory periphery, is developed in this chapter. This is done to investigate whether the implementation of the peripheral model has a significant influence on this limitation. The chapter contributes with a discussion of the potential stages in the auditory periphery that are likely to affect the level-dependency of ABR latency. The chapter highlights that the ABR model fails to simulate a realistic ABR latency behavior even though two established models of peripheral processing, the auditory-nerve (AN) model and the dual-resonance non-linear (DRNL) filter model, are used.

4. Chapter 5 is based on Rønne and Gøtsche-Rasmussen (2011) and presents a study of the alignment of high- and low-frequency content when recording rising-chirp-evoked ABRs. This study is motivated based on the simulations using the ABR model, and evaluates the hypothesis found in literature that chirps evoke larger ABR amplitudes than clicks due to the time-alignment of low-frequencies.

5. When investigating AEPs evoked by longer-duration stimulus, a key feature of the auditory system becomes the adaptation of the inner-hair-cell (IHC) -

[1] The ABR model is included in the Auditory Modeling (AM) toolbox (Søndergaard et al., 2011) and can be downloaded from: http://amtoolbox.sourceforge.net/.

auditory-nerve (AN) synapse in the cochlea. Chapter 6, which is based on Harte et al. (2010), investigates this adaptation using experimental recordings and simulations of click trains. The chapter contributes by discussing the extent to which the modeling approach can be used to simulate responses of longer-duration stimuli.

6. Chapter 7 is based on Rønne et al. (2012a) and presents an investigation of whether auditory steady-state responses (ASSR) can be used to assess cochlear compression in humans. This study examines two potential experimental paradigms, level-growth and modulation-growth functions, using an analytical approach, ASSR recordings in humans, and an extended version of the ABR model (referred to as the ASSR model). The clear recommendation given in this chapter is to use the level-growth function. This is a potential clinical application that could be of interest in both infant hearing screening and in hearing aid fitting procedures for both children and adults.

7. Chapter 8 evaluates the ABR model capabilities to simulate speech evoked AEPs. It is demonstrated that, even with highly complex stimuli such as speech syllables, the model captures key features of the AEP responses, demonstrating the importance of peripheral processing for the generation of ABRs evoked by complex stimuli. Further, the chapter contributes with a discussion of the effects of cochlear tuning on the neural encoding of speech syllables.

8. Chapter 9 provides a general discussion of the modeling approach and its limitations. Further, the implications and perspectives of this study are presented.

# 2

# Background

## 2.1 Auditory evoked potentials

In 1875, Richard Caton recorded electrical activity from the brain of a rabbit. What he recorded became known as electroencephalography (EEG) and has since developed into a major diagnostics and research tool. Fifty years later, Berger (1929) became the first to record EEGs in human subjects. Wever and Bray (1930) recorded cochlear microphonics in animals and were the first to use EEG for audiological purposes. The first reported measurement of acoustically evoked responses in humans was undertaken by Fromm et al. (1935). Since 1935 the recordings of auditory evoked potentials (AEP) have developed fast and now represents a well known and used technique both for clinical and research purposes (see Collura (1993) and Hall (1992) for a historical review).

A common setup for recording AEPs includes a computer generating digital sounds, a D/A converter and an acoustic transducer presenting the sounds to the subject. On the recording side, the setup includes electrodes attached to the scalp, a recording amplifier including an A/D converter and a computer to store and post-process the recordings. The recordings are time-aligned with the stimulus and, by using multiple repetitions and averaging, the noise can be suppressed sufficiently to record a signal where the response to the stimulus is detectable. Noise remains though a major obstacle to AEP recordings, and post-processing like filtering and artifact rejection schemes are often applied. The AEP formation is highly dependent on the location of the electrodes on the scalp. An often used configuration is to record differentially between the vertex and the ipsi-lateral mastoid, with a ground electrode placed on the

forehead. This configuration is sensitive to sources of electrical activity originating from the brainstem, whereas other configurations are used depending on the AEP of interest. Throughout this thesis, the vertex / ipsi-lateral mastoid configuration, is used both for modeling work and experimental work.

AEPs represent the summed electric potential from many remotely located neurons firing in response to an acoustic stimulus. They are often classified in terms of time of occurrence after stimulus onset, specifically when transient stimuli are used. The AEPs are thus called auditory brainstem responses (ABRs) with latencies between 1 and 15 ms (first described by Jewett, 1970), middle-latency responses (MLRs) with latencies in the range of 15-50 ms (first described by Geisler et al., 1958) and auditory late responses (ALRs) with latencies in the range of about 75-200 ms (first described by Davis et al., 1939). The latencies can be associated with generation place, such that longer latencies corresponds to higher generation sites in the auditory pathway. The generation site of the AEP has also alternatively been used to classify recordings, such that; AEPs from the hair cells in the cochlea are called cochlear microphonics (CM) (e.g Withnell, 2001); AEPs from the distal end of the auditory nerve (AN) are called compound action potentials (CAPs) (e.g Chertoff et al., 2010); AEPs from the brainstem are called auditory brainstem responses (ABR); and AEPs from the cortex has been named cortical auditory evoked potentials (CAEPs) (e.g Sharma and Dorman, 1999). The term ABR is in the present study used to denote both an AEP evoked by a transient signal producing a response with a latency between 1 and 15 ms, and also as an AEP recorded at brainstem level to any arbitrary stimulus. A third potential classification is to classify recordings according to the stimulus that evokes them. An AEP can be evoked by any acoustic stimulation, however, in literature some stimuli have been studied intensively and have been established as de facto standards for investigating AEP generation and the underlying physiology. These stimuli include transients like clicks, chirps and tone-bursts (e.g. Jewett, 1970; Jewett and Williston, 1971; Dau et al., 2000, and chapter 3 of this thesis), steady-state signals such as amplitude modulated (AM) tones (e.g. John and Picton, 2000; Galambos et al., 1981; Kuwada et al., 1986; Picton et al., 1987; Rees et al., 1986, and chapter 7 of this thesis), but also more complex signals like speech syllables (e.g., Warrier et al., 2004; Agung et al., 2006; Aiken and Picton, 2008; Akhoun et al., 2008;

Lalor and Foxe, 2010; Chandrasekaran and Kraus, 2010, and chapter 7 of this thesis). Auditory steady-state responses (ASSRs) are often associated with the special case where a pure tone carrier is modulated by a lower-frequency tonal modulator. The response to complex stimuli like syllables have often been referred to as complex auditory brainstem responses (cABR) (e.g. Skoe et al., 2011) or frequency following responses (FFR) (e.g. Dau, 2003; Swaminathan et al., 2008). In this study, the syllable evoked cABR (studied in chapter 8) will be denoted ABR, as the division between a "complex" and "non-complex" stimulus is difficult to define.

To summarize, important parameters for the AEP generation are generation site, electrode montage, onset latency, amplitude range, subject attention and plasticity[1], as well as stimulus characteristics like duration, intensity, frequency content and variation over time. Table 2.1 summarizes the differences in the characteristics between the different types of responses. The responses have been grouped to aid clarity. The ASSR has not been included in the table as the modulation frequency alters both the generation site and the dependence on subject arousal. At high modulation rates, the ASSR would belong in the column alongside the ABR whereas at lower modulation rates, the ASSR would behave as the CAEP. The first chapters (3, 4, 5 and 6) of this study focus on the transiently-evoked ABR, because these are reproducible and largely unaffected by subject arousal. Chapter 7 investigates the low modulation rate 40-Hz ASSR (i.e. an ASSR belonging in the CAEP column). Besides being an interesting clinical tool, the 40-Hz ASSR challenges the developed AEP model of the present study, as it includes higher-stage neural processing and adds potential complications of subject arousal to the model. Chapter 8 investigates the syllable-evoked ABR (cABR). This challenges the model further, as it has been suggested in the literature that the syllable-evoked ABR is subject to plasticity.

[1] physiological changes of the nervous system due to e.g. learning

| | CM | CAP / ABR / MLR FFR / cABR | CAEP / ALR |
|---|---|---|---|
| Generation site | Cochlea | Auditory nerve (AN) and brainstem | Cortex |
| Typical electrode montage | Within ear canal | Vertex and mastoid | Multiple electrodes |
| Onset latency | < 1 ms | 1 to 50 ms | > 50 ms |
| Amplitude range | $\mu v$ | $nv$ | $\mu v$ |
| Subject arousal | Unaffected | Largely unaffected. Subjects can sleep | Eliminated in sleeping subjects |
| Plasticity | Unaffected | Experience slightly alters the AEP | Experience alters the AEP |
| Stimulus intensity | no latency shifts | latency shifts | latency shifts |

Table 2.1: Differences between groups of AEPs.

## 2.2 Auditory models

Several models of the (human and animal) auditory pathway have been proposed. Some of which aim at modeling cochlear mechanics and the underlying physiology as strictly as possible while others model the observed responses without having the intention of strictly modeling each stage of the physiological pathway. The latter is called a phenomenological model. This section describes two well-established phenomenological auditory models, the auditory nerve (AN) model and the dual-resonance non-linear (DRNL) model. The AN model is used as the basis for the AEP model developed in this study. The DRNL model, is considered as an alternative AEP model (Rønne et al., 2011, chapter 4).

### 2.2.1 The AN model

The AN model is a phenomenological model developed over many years (1993 - to present), designed to simulate AN responses of cats. The original AN model (Carney, 1993) simulates single-fiber responses which are linked to a specific place on the BM with a specific characteristic frequency (CF). Even though the model only simulates responses from one fiber at a time, the simulation of the response to broad-band stimuli is possible, as the BM filter stage of the model simulates the contributions from both on- and off-frequency stimulation to the single-fiber response. This ensures that the simulated single-fiber responses can be compared to experimental single-fiber AN recordings (in this model from cats). The first stage of the original AN model is a time-varying BM filter, implemented as a symmetric gamma-tone filter, with a feed-back control path simulating broadening tuning with increasing stimulus level. The output is delayed in time to simulate the traveling-wave delay on the BM. The signal path of the model does further contain an inner hair-cell (IHC) non-linearity that gives a physiologically-inspired half-wave rectification. Combined with a low-pass filter, this simulates the transition between responses following the fine-structure of the stimulus at low stimulus-frequencies and responses following the stimulus-envelope at higher stimulus-frequencies. The IHC-AN stage produces adaptation (similar to Westerman and Smith, 1988) resulting in an onset emphasis and a slight suppression of the late part of a long duration response. The last stage in the model simulates the refractoriness of the neural AN responses. The refractory stage is not included in the AEP model developed on basis of the AN model.

The AN model has been modified several times. Zhang et al. (2001) exchanged the feed-back control path of the BM filtering with a feed-forward control path. Further, the control path filter was made broader than the signal path filter, and the tip was shifted slightly towards a higher CF. These updates made the model capable of simulating two-tone suppression, asymmetrical growth of suppression and the offset of suppression tuning curves (compared to excitatory tuning curves). The Zhang et al. (2001) cat-version of the AN model was transformed into a human version by Heinz et al. (2001) which was later used by Dau (2003) to develop a human ABR model.

Tan and Carney (2003) implemented a middle-ear filter and exchanged the gammatone BM filters by chirping BM filters. The latter was done to simulate best frequency[2] shifts with stimulus level and frequency glides in the impulse responses independent of stimulus levels. Experimentally, the best frequency has been observed to shift upwards with increasing stimulus levels. The frequency glides, also accounted for by the Tan and Carney (2003) AN model, is based on the experimental observation that the early part of the impulse response of a BM filter is not dominated by the same frequency components as the later part of the impulse response (Carney et al., 1999). The frequency glide was found to be independent of stimulus level, such that the zero-crossings of the fine structure was independent of level whereas the envelope of the response changes with level.
Zilany and Bruce (2006) and Zilany and Bruce (2007) modified the model to be able to account for the effect of high stimulus levels. Tones presented at high stimulus levels have been shown to be subject to a sharp transition of up to 180° of the phase-level function (Kiang, 1990). This transition is called the component 1 (C1)/component 2 (C2) transition, where C1 is the response to low stimulus-levels and C2 the response to high stimulus levels. At the levels of the C1/C2 transition, approximately 90 to 105 dB SPL, "peak splitting" occurs (Kiang, 1990). Peak splitting describes the phenomena that the phase-locked response to a high-level tone can result in a doubling of the number of peaks in the recorded time histograms. In this case, peaks in-between the phase-locked peaks appear when the stimulus-level approaches 95 dB SPL and grows with stimulus-level until they completely dominates above 105 dB SPL, resulting in a 180° phase shift (C1/C2 transition). These two related effects, the C1/C2 transition and peak splitting, were implemented in the model as a parallel C2 filter path, complementing the regular (C1) signal path.

The latest version of the model (Zilany et al., 2009) exchanged the double logarithmic adaptation with a combined logarithmic and power-law adaptation which have been shown to provide a more realistic IHC-AN adaptation behavior. Among the achieved improvements were a more accurate prediction of forward-masking, an

[2] The best frequency was defined as the frequency at which the fiber response is strongest at a certain stimulus level, whereas the CF can be defined as the frequency where the threshold of the fiber is lowest (Tan and Carney, 2003)

improved recovery of the AN response after stimulus-onset and improved predictions of the response-synchrony to amplitude-modulated tones. In table 2.2, the differences between the AN model versions are shown with respect to the phenomena they can simulate.

The Zilany and Bruce (2007) AN model was used here as basis for the ABR model discussed in chapters 3, 4, 5 and 6. The Zilany et al. (2009) AN model was used as basis for the ASSR model and the ABR model presented in chapters 7 and 8.

### 2.2.2 The DRNL model

Another well established auditory model is the DRNL model which build upon the DRNL filter. The DRNL filter (Meddis et al., 2001) resembles the BM stage of the AN model such that it also is a phenomenological model that simulates the response of a single place on the BM. Furthermore, the original DRNL model (Meddis et al., 2001) was also evaluated on animal data (chinchilla and guinea pig). The input to the model is stapes velocity and the output BM motion. Although the model is implemented in a different way as the AN model (the DRNL filter is implemented as the sum of two parallel processes, one linear and one nonlinear, whereas the AN model uses a feed-forward control path to control the BM filter), it is capable of simulating many of the same BM related phenomena, such as compressive input/output functions, two tone suppression and frequency glides[3]. Lopez-Poveda and Meddis (2001) exchanged the animal-fitted parameters of the Meddis et al. (2001) DRNL model with human-data-fitted parameters, and added an outer- and middle-ear filter stage before the DRNL filter. In the model presented by Meddis (2006)[4] an advanced model of the IHC functionality was added to the DRNL model[5]. Furthermore, a spike generating AN stage which includes refractoriness was implemented. The output of the peripheral part of the DRNL model, including stages from the outer ear to the AN, were used as

[3] Tan and Carney (2003) argued that the frequency glides of the DRNL model are level-dependent, contrary to experimental data

[4] parts of the work was presented in Sumner et al. (2002) and Sumner et al. (2003)

[5] this far more complicated model is sometimes referred to as the “model of the auditory periphery”, however, to avoid confusion the term DRNL model are used here. The term DRNL model has to be distinguish from the DRNL filter described previously

| Auditory function / processing stage | Carney 1993 | Zhang 2001 | Tan 2003 | Zilany 2007 | Zilany 2009 |
|---|---|---|---|---|---|
| Middle ear | | | | | |
| Middle ear filtering | | | + | + | + |
| Basilar membrane filtering | | | | | |
| Broadening tuning with stim. level | + | + | + | + | + |
| Compressive input/output functions | + | + | + | + | + |
| Traveling wave delay | + | + | + | + | + |
| Two-tone suppression | | + | + | + | + |
| Assym. growth of suppression | | + | + | + | + |
| Frequency glides | | | + | + | + |
| Best frequency shifts | | | + | + | + |
| C1/C2 transition | | | | + | + |
| Peak splitting | | | | + | + |
| IHC transduction | | | | | |
| Physiological rectification | + | + | + | + | + |
| Upper limit of phase locking | + | + | + | + | + |
| IHC-AN synapse | | | | | |
| Double logarithmic adaptation | + | + | + | + | + |
| Power law adaptation | | | | | + |
| Spike generator | | | | | |
| Refractoriness | + | + | + | + | + |

Table 2.2: Overview of the AN model development, with respect to the phenomena simulated by the respective version of the model. The corresponding papers correctly referenced are; Carney (1993), Zhang et al. (2001), Tan and Carney (2003), Zilany and Bruce (2007) and Zilany et al. (2009). The Heinz et al. (2001) model is similar to the Zhang et al. (2001) only human- instead of cat-fitted parameters are used.

input to a neural model of a single cochlear nucleus chopper neuron, effectively being a decision making stage that compares inputs from several modeled AN responses (tuned to different CFs).

## 2.3 Modeling AEPs

In this section, a convolutive approach to simulating AEPs is described. This convolutive approach has been used in the present study to develop an ABR model, as convolution between single fiber responses produced by the AN model and a so-called unitary response (UR).

### 2.3.1 Convolutive approach

Elberling (1976) defined a "unit function" as the recorded electrical waveform that is synchronous to a single event (one spike) in one neuron. Given the assumption that the same waveform is generated by all types of neurons, deBoer (1975) developed a model of CAP generation. The CAP model was based on linear BM filtering, half-wave rectification and envelope extraction. The output of this peripheral part of the model was the firing rate function. To model CAPs, recorded in the ear canal, deBoer (1975) proposed a unit function describing the waveform recorded in the ear canal when a distal AN neuron discharges. Theoretically the summation of contributions from all AN neurons would lead to the CAP, assuming that there are no other interfering electrical potentials. However, as summation of all neurons were not computationally viable, deBoer (1975) suggested to use 64 representative neurons, each related to a different BM filter tuned to a specific CF. The summed activity pattern, from the 64 channels, convolved with a (calibrated) unit function provided the simulated CAP. Following this concept, Melcher and Kiang (1996) suggested, in a more general description, that the potential produced at surface mounted electrodes by any cell in the auditory pathway, including higher neural stages, can be described by the convolution of the instantaneous discharge rate with a unitary response (UR). This concept was adopted by Dau (2003) who developed an ABR model. Dau (2003)

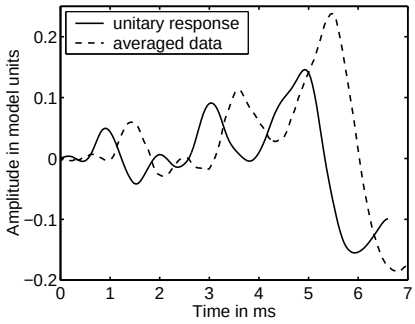



Figure 2.1: The UR derived by Dau (2003) and corresponding click evoked ABR.

used the AN model by Heinz et al. (2001) to produce instantaneous discharge rates and a summed activity pattern. As deBoer (1975), this was done using independent channels tuned to different CFs. In Dau (2003), 500 channels between 100 Hz and 10 kHz were considered. The summed activity pattern represented the activity at the distal end of the AN and was in Dau (2003) convolved with a UR representing not only contributions from wave-I (the CAP) of the ABR (as the unit function of deBoer, 1975), but also contributions from wave-II to wave-VII, i.e. components spanning the first 7 ms of the neural processing. The UR idea was thus that a single spike in IHC-AN traveling up the auditory pathway, will elicit potentials at several places, each delayed and scaled compared to the previous one. Dau (2003) assumed the UR to be a linear function independent of stimulus, thus arguing that convolving the instantaneous discharge rate functions with the UR and adding the contributions after wards, yields the same result as convolving the summed activity pattern with the UR. The UR was calculated as the deconvolution between the summed activity pattern evoked by a click stimulus, and an experimentally recorded ABR evoked by an identical click stimulus. Figure 2.1 shows the derived UR and the recorded click evoked ABR (reprinted with permission from Dau, 2003). The UR bears a large resemblance to the recorded click evoked ABR, and waves corresponding to wave-I, -III and -V can be detected.

In contrast to the deBoer (1975) model, not only the CAP component but also later waves of the ABR response was considered in Dau (2003). Furthermore, the effect of

nonlinear BM processing on the potential pattern was considered, while deBoer (1975) used a linear model. However, both models were based on the same assumptions that, 1) the complete set of AN fibers can be replaced by a limited set of simulated fibers (channels), each corresponding to a representative place on the BM tuned to a specific CF. 2) The individual channels creating the instantaneous discharge rates act independently of one another. 3) The UR is linear. Meaning that it is invariant to the type of stimulation, subject and the type of neurons involved.

The third assumption was evaluated by Chertoff (2004), who found that his unit function was slightly dependent on both stimulus-frequency and stimulus-level; however, the stimulus-dependencies were small and no general description was attempted.

Regarding the simulation of steady-state responses, Bohorquez and Oezdamar (2008) presented a convolution approach to predict the 40-Hz ASSR. This convolution approach has little resemblance to the UR method described above. Bohorquez and Oezdamar (2008) modeled the 40-Hz ASSR as a convolution between a click-train and the single-click evoked MLR, thus modeling the ASSR as a linear convolution between two linear functions. The click-evoked MLR consist of three main peaks, the ABR wave-V the $P_a$ and the $P_b$, each of which are typically separated by approximately 25 ms. When a click-train at a rate of 40 Hz is presented to the auditory system, the components of the MLR were argued to add up in phase, such that the $N_a$ peak of one click will add up in phase with the $N_b$ peak of the previous click. A convolutive approach thus seems to be modeling the 40 Hz ASSR well.

### 2.3.2 Dipole modeling and ABR physiology sources

Scherg and von Cramon (1985a) developed a spatio-temporal dipole model of AEP generation. The model was focusing on the electrical dipole components in the brain, and had therefore no model of the auditory periphery. The basic assumption was that scalp potentials result from the superposition of all charges within the brain. Further, it was argued that, as the net charge in the brain is zero, only pairs of positive and negative charges exist. Each pair is thus producing a dipole field. The primary idea

was that the scalp potentials result from the superposition of the far fields of many microscopic dipoles, i.e. the same assumption as deBoer (1975), Elberling (1976), Melcher and Kiang (1996) and Dau (2003) used to argue for the UR idea. The main difference between the approaches was the use of multi-channel recordings in both Scherg and von Cramon (1985a) and Scherg and von Cramon (1985), and the fact that the UR of Dau (2003) includes neural processing whereas Scherg and von Cramon (1985a) only considers the propagation from the dipole to the electrodes. The aim of Scherg and von Cramon (1985) and Scherg and von Cramon (1985a) were to provide a full description of the waveforms at all electrodes simultaneously. By searching for the minimal number of equivalent dipoles sufficient to explain the scalp potential, each dipole source (defined by stationary location and orientation) could be associated with a hypothesized anatomical source. This approach led to the conclusion that the spread of local potential to the electrodes was only dependent on the location and orientation of the recording electrodes and dipole sources. The UR of, e.g., Dau (2003) can thus be seen as the special case, where the electrode locations where at vertex and mastoid (giving the orientation as direction between them), and where only the dipole sources in the brain aligned with this orientation (or weighted according to their misalignment) were effective. Scherg and von Cramon (1985) found that the generation of wave-I of the classic ABR was located to the distal end of the auditory nerve. Wave-III of the ABR, was located to be in, or near to, the cochlear nucleus. Wave-IV and wave-V could not be located precisely; however, an origin in the early parts of the brainstem was suggested.

## 2.4 Background summary

This chapter reviewed the literature on some of the key aspects of this study. It was outlined how this study models the transiently evoked ABR, the 40 Hz ASSR and the syllable-evoked ABR. Furthermore, the present study develops an AEP model based on the convolutive approach, where the AN model produces a summed activity pattern, that is convolved with a linear UR, to produce the simulated AEP. The following

chapter is based on Rønne et al. (2012), which develops and evaluate the ABR model, designed to simulate transiently evoked ABRs.

# 3

# Modeling auditory evoked brainstem responses to transient stimuli

This chapter develops an ABR model. The theoretical modeling framework is presented, as is the main implementation details on how the AN model has been humanized. Both the theoretical framework and the humanization is used throughout this thesis. This chapter can thus be read both as an independent study of modeling transiently evoked ABRs, and as the method section for the rest of this thesis. The chapter is based on Rønne et al. (2012).

## 3.1 Abstract

A quantitative model is presented that describes the formation of auditory brainstem responses (ABR) to tone pulses, clicks and rising chirps as a function of stimulation level. The model computes the convolution of the instantaneous discharge rates using the "humanized" nonlinear auditory-nerve (AN) model of Zilany and Bruce (2007) and an empirically derived unitary response function which is assumed to reflect contributions from different cell populations within the auditory brainstem, recorded at a given pair of electrodes on the scalp. It is shown that the model accounts for the decrease of tone-pulse evoked wave-V latency with frequency but underestimates the level dependency of the tone-pulse as well as click-evoked latency values. Furthermore, the model correctly predicts the nonlinear wave-V amplitude behavior in response to the chirp stimulation both as a function of chirp sweeping rate and level. Overall, the results support the hypothesis that the pattern of ABR

generation is strongly affected by the nonlinear and dispersive processes in the cochlea.

## 3.2 Introduction

When sound is presented to the ear, it is possible to record auditory evoked potentials (AEPs) on the surface of the human scalp. AEPs represent the summed electric potential from many remotely located neurons firing in response to the stimulus applied. They are typically grouped in terms of time of occurrence after stimulus onset and are thus denoted as auditory brainstem responses (ABRs) with latencies between 1 and 7 ms, middle-latency responses (MLRs) with latencies in the range of 15-50 ms, and auditory late responses (ALRs) with latencies in the range of about 75-200 ms.

AEPs have been used to asses the neural encoding of sound both for clinical and research purposes. Various types of stimuli have been considered, such as transients like clicks, chirps and tone-bursts (e.g., Jewett and Williston, 1971; Dau et al., 2000), steady-state signals such as amplitude modulated (AM) tones (e.g. John and Picton, 2000; Galambos et al., 1981; Kuwada et al., 1986; Picton et al., 1987; Rees et al., 1986), but also more complex signals like speech (e.g., Warrier et al., 2004; Agung et al., 2006; Swaminathan et al., 2008; Aiken and Picton, 2008; Akhoun et al., 2008; Lalor and Foxe, 2010; Chandrasekaran and Kraus, 2010). Tone-burst evoked ABRs have been studied to objectively estimate frequency-specific hearing sensitivity, for example in newborn and young children (e.g. Ribeiro and Carvallo, 2008) or to estimate effects of cochlear group delay as a function of frequency and level of stimulation (e.g. Gorga et al., 1988; Harte et al., 2009; Neely et al., 1988; Murray et al., 1998). Broadband rising chirps have recently been developed for ABR recordings to maximize synchronous firing of nerve fibers across frequency, leading to an increase of ABR wave-V amplitude and a higher signal-to-noise ratio compared to traditional click stimulation (e.g. Dau et al., 2000; Elberling and Don, 2008; Fobel and Dau, 2004; Junius and Dau, 2005; Shore and Nuttall, 1985). It is argued (Dau et al., 2000), that these broadband chirp stimuli compensate for the frequency-

dependent group delay seen in the basilar membrane (BM) velocity/displacement traveling waves. In a recent study, Elberling et al. (2010) presented five chirps with different frequency-delay functions and investigated the resulting wave-V amplitude of their responses at stimulation levels of 20, 40 and 60 dB normal hearing level (nHL). Their results demonstrated that the dispersion function, or sweeping rate, of the chirp that evoked the largest wave-V amplitude was a function of stimulation level. With increasing level, the "optimal" chirp that created the largest wave-V response was found to become progressively shorter (Elberling et al., 2010), i.e. to have the fastest sweeping rate.

It is well known that the *frequency* dependency of wave-V latency is related to the tonotopical coding of frequency on the BM in the cochlea. High-frequency stimulation excites basal parts of the BM and thus produces a shorter delay than low-frequency stimulation that mainly excites apical parts of the BM (Gorga et al., 1988; Greenwood, 1990; Harte et al., 2009; Neely et al., 1988; Murray et al., 1998). The *level* dependency of wave-V latency is not so well understood. Cochlear tuning is known to be level dependent, where an increase of the stimulus level results in broader auditory filters and thus a broader excitation pattern on the BM (Glasberg and Moore, 1990; Recio and Rhode, 2000). This means that regions of the BM with characteristic frequencies further away from the stimulus frequency are also excited. Elberling (1976) and Folsom (1984) reasoned that the broadening of excitation with level might result in shorter latencies, as more basal regions of the BM are activated that are associated with shorter implicit delays. Another inherent feature of the filter tuning is the change in the envelope of the BM impulse response at a given location, as level is increased. The timing of the individual peaks of the physiological impulse response are level independent but the amplitude of the earlier peaks are more emphasized as the stimulus level increases (e.g., Kiang (1965), Recio and Rhode (2000)). This change in the envelope, as stimulus level is increased, results in an onset emphasis that could result in a decrease of the wave-V latency. Adaptation in the inner-hair cell (IHC)-AN synapse similarly enhances the onset of a signal while attenuating later parts (Westerman and Smith, 1988) in the stimulus. Thus, adaptation in the IHC-AN synapse might also contribute to the level-dependence of wave-V latency.

The wave V amplitude is both stimulus frequency and stimulus level dependent. The general shape of the frequency dependence is considered to be mainly controlled by the transfer functions of the outer and middle ear effectively acting as a band-pass filter (Pascal et al., 1998; Puria, 2003), with maximal transduction at 1-2 kHz. The level dependence of the wave-V amplitude results from the summation of the individual neural responses after the non-linear processing through the BM at the individual characteristic frequencies (CFs), where compressive behavior has been found for medium-level stimulation at the CF while linear behavior has been found for low-level stimulation (e.g. Ruggero et al., 1997). The chirp-evoked ABRs obtained in Elberling et al. (2010) demonstrated non-monotonic level-dependent behavior, assumed to result from the broadening of neural excitation with increasing level (Harte et al., 2010). At low levels, each frequency component of the chirp might excite a narrow region on the BM and, given the timing associated with each component, might add up in phase (e.g. Dau et al., 2000). At high stimulus levels, each frequency component excites a broader region on the BM, due to upwards spread of excitation (Rhode and Recio, 2000). Thus, a specific location on the BM is excited by a broader range of frequency components. These different components contribute with different timing which results in desynchronization and a reduction of ABR wave-V amplitude (Elberling et al., 2010).

However, while it appears obvious that cochlear processing affects ABR amplitudes and latencies, only very few studies have actually attempted to provide quantitative predictions of ABR data. In the present study, a computational model is presented that simulates evoked responses to tone pulses of various frequencies and levels, upward chirps with different sweep rates and levels as well as click stimuli. The key stages in the model are (i) the nonlinear processing in the cochlea, including key properties such as compressive basilar-membrane filtering, inner hair-cell (IHC) transduction, and IHC-AN synapse adaptation, and (ii) the (linear) transformation between the neural representation at the output of the AN and the recorded potential at the scalp. This approach was inspired by Goldstein and Kiang (1958), who described evoked responses as a linear convolution of an elementary unit waveform of a given neuron, called the unitary response, with the instantaneous auditory nerve (AN) discharge rate

in response to a given stimulus. This approach was applied to simulate cat compound action potentials (CAP) by deBoer (1975).

Based on the work of Goldstein and Kiang (1958), deBoer (1975) and Melcher and Kiang (1996), Dau (2003) proposed a model for the generation of ABRs and frequency following responses (FFR) to tones. In Dau (2003), the unitary response was estimated empirically based on measured ABR data, via deconvolution of average click-evoked responses and the simulated neural activity pattern at the output of an AN model. Dau (2003) demonstrated that the auditory periphery strongly affects the simulated ABR patterns and could account for some of the key features observed in the recordings of chirp- versus click-evoked responses. However, while that study provided a proof of concept, it did not consider a more detailed analysis of the responses as a function of stimulation frequency and level. Furthermore, significant discrepancies between the predicted and measured wave-V latencies were observed but not further evaluated. Here, the original modeling framework of Dau (2003) was extended to include current advances in AN modeling, such as linear BM filters at high stimulus levels, peak splitting (Kiang, 1990) and a shift of best frequency with level. The AN model developed by Zilany and Bruce (2007) was used here which is based on current knowledge derived from both behavioral and objective measures of cochlear processing. The model was originally developed for cat but also adopted by the same authors for humans including corresponding middle-ear filtering and BM filter tuning (Ibrahim and Bruce, 2010).

## 3.3 Model for ABR generation

### 3.3.1 Convolution model of ABR generation

Melcher and Kiang (1996) described the generation of ABR in cats as a summation of individual brainstem cell potentials, $v_i$, in response to a given stimulus, $s$;

$$\mathrm{ABR}(t,\bar{x_1},\bar{x_2},s) = \sum_i v_i(t,\bar{x_1},\bar{x_2},s) \tag{3.1}$$

where $\bar{x_1}$ and $\bar{x_2}$ are the locations of the electrodes on the scalp. The potential, $v_i$, in response to a given acoustic stimulus, can be determined by a convolution between the instantaneous firing rate of the $i^{th}$ cell, $r_i(t,s)$, and a unitary response function, $u(t,\bar{x_1},\bar{x_2})$. This latter function is defined as the potential produced between the electrode positions on the scalp, $\bar{x_1}$ and $\bar{x_2}$, each time the cell discharges;

$$v_i(t,\bar{x_1},\bar{x_2},s) = r_i(t,s) \star u_i(t,\bar{x_1},\bar{x_2}) \tag{3.2}$$

where $\star$ denotes the convolution operation. To obtain an ABR with this method, all cells need to be considered individually, which would be computationally prohibitive. To avoid this, Melcher and Kiang (1996) suggested the use of the cell population potential, $V$. Cells can be grouped by the physio-anatomical type of the cell, $p$, where $P$ is the number of different cell types:

$$\mathrm{ABR}(t,\bar{x_1},\bar{x_2},s) = \sum_{p=1}^{P} V_p(t,\bar{x_1},\bar{x_2},s) \tag{3.3}$$

It is reasonable to assume that all cells of the population described have the same unitary response (UR), $u(t,x1,x2)$, as they have the same morphological and electrical properties (Melcher and Kiang, 1996). The combination of eqn. (3.2) and (3.3) yields a general expression for ABR generation:

$$\mathrm{ABR}(t,\bar{x_1},\bar{x_2},s) = u(t,\bar{x_1},\bar{x_2}) \star \sum_{p=1}^{P} \sum_{i=1}^{N_p} r_{pi}(t,s) \tag{3.4}$$

where $N_p$ is the total number of cells of type, $p$. The three main peaks in the click-evoked ABR are waves I, III and V. Dau (2003) made the assumption that the instantaneous firing functions in the medial superior olive (MSO), anterior ventral cochlear nucleus (AVCN) are the same as in the AN, following the suggestion by Melcher and Kiang (1996). Thus, the instantaneous firing functions for the different cell populations are given by $r_{i,\mathrm{MSO}} = r_{i,\mathrm{AVCN}} = r_{i,\mathrm{AN}} = r_i$, simplifying Eqn. 3.4 to:

$$\mathrm{ABR} = u(t, \overline{x_1}, \overline{x_2}) \star \sum_{i=1}^{N} r_i(t, s) \tag{3.5}$$

The generation of an ABR is thus represented as the sum of the instantaneous firing from all cells, convolved with a UR that is dependent on the electrode location on the scalp but assumed to be independent of cell type, efferent influence and stimulus.

### 3.3.2 Model structure

The structure of the ABR model is shown in Fig. 1. The AN model calculates the instantaneous discharge rate for individual AN fibers, in response to a given acoustic stimulus defined in pascals. Each AN fiber is tuned to a specific characteristic frequency (CF). The CFs chosen were spaced according to the human cochlear map of Greenwood (1990). The number of fibers included was a trade-off between computational time and model accuracy. Throughout this study, 500 fibers ranging from 100 Hz to 16 kHz were used in all simulations. The fibers were chosen so they were spaced equally on the BM according to the human cochlear map (Greenwood, 1990). The output of the AN model, the instantaneous firing rate of all the AN fibers, were summed and convolved with the UR function.

The AN model of Zilany and Bruce (2006) is shown schematically in Fig. 2. The input to the AN model is the instantaneous pressure waveform of the stimulus in units of pascals. The output of the AN model is the spike rate in response to the stimulus pressure. The model includes a number of key functional stages: a middle-ear filter; a feed-forward control path; a primary signal-path filter (C1) representing the basilar membrane (BM) filtering adapted by the control path; a parallel-path filter (C2) for high-level stimuli; an inner-hair cell (IHC) section followed by a synapse model and a stochastic AN spike discharge generator. In Fig. 2, the following abbreviations are used: outer hair cell (OHC), low-pass (LP) filter, static nonlinearity (NL), characteristic frequency (CF) and inverting nonlinearity (INV). $C_{\mathrm{OHC}}$ and $C_{\mathrm{IHC}}$ are scaling constants that indicate the OHC and IHC status, respectively. The black and gray curves in the filter stages represent the tuning at low and high sound pressure

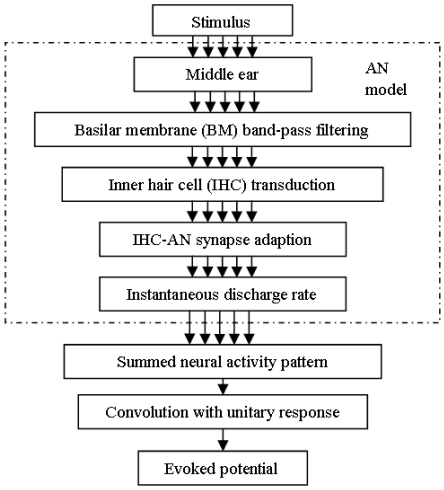



Figure 3.1: Schematic structure of the ABR model. 500 AN fibers tuned to different CFs are individually simulated by the AN model. The summed activity, integrated across frequency, is then convolved with a unitary response and represents the simulated ABR to a given stimulus.

levels, respectively. The wide band C2 filter shape is fixed and is the same as the broadest possible C1 filter. The black and gray functions in the stage following the C1 filter indicate the nonlinearity in the IHC input/output functions in normal and impaired (scaled down according to $C_{\mathrm{IHC}}$) hearing, respectively. Details about the model implementation can be found in Zilany and Bruce (2006). In the present study, the spikes/s output from the synapse model was used, rather than the stochastic output from the spike generator. The stochastic spike generator requires averaging over many repetitions before it becomes repeatable and thus usable to ABR modeling.

### 3.3.3 Features of the humanized AN model

The parameters of the AN model of Zilany and Bruce (2006) and Zilany and Bruce (2007) were originally fitted to cat AN data. Later, the model was modified to estimate human responses by the same authors. First, the original cat middle-ear transfer function was replaced by a human middle-ear transfer function, based on the linear

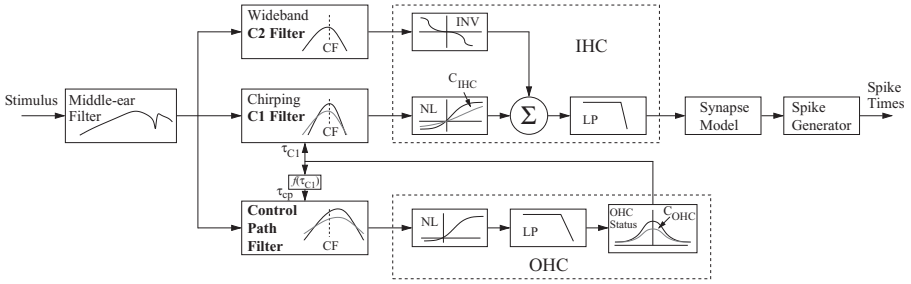



Figure 3.2: Diagram of the auditory-nerve model developed by Zilany and Bruce (2006). Reprinted from Zilany and Bruce (2006) with permission from the Acoustical Society of America (©2006). The input to the AN model is the instantaneous pressure waveform of the stimulus in units of pascals. This waveform is band pass filtered by a middle-ear filter. A feed-forward control path filter determines the characteristics of the main C1 filter path which is mainly active at levels below approximately 96 dB SPL. A parallel C2 filter path is mainly active at higher stimulus levels. The two filter paths are followed by a nonlinear inner hair-cell (IHC) stage and a nonlinear synapse model. The output of the AN model, used in this study, is the instantaneous discharge rate obtained at the output of the synapse model.

circuit model of Pascal et al. (1998). Second, the cat BM tuning was replaced by human BM tuning (see Ibrahim and Bruce, 2010, for details). Two prominent and different estimates of BM tuning exist in the literature, hence the relative broad tuning by Glasberg and Moore (1990) and the sharper tuning by Shera et al. (2002). In this study, the tuning from Shera et al. (2002) was used. It has been argued that humans have this significantly sharper BM mechanical tuning than experimental animals such as cats and guinea pigs (Shera et al., 2002, 2010; Bentsen et al., 2011). The sharper human tuning is also probable in light of the recent findings by Joris et al. (2011) who showed that macaque monkeys have sharper tuning than rodents and cats. Further, the simulations using the ABR model produced the best results with the Shera et al. (2002) tuning compared to the alternative broader tuning presented by Glasberg and Moore (1990). To incorporate the sharper tuning, the model equivalent rectangular bandwidth (ERB) quality factor, $Q_{\mathrm{ERB}}$, for cochlear tuning was modified to be:

$$Q_{\mathrm{ERB}} = 12.7 \left( \frac{CF}{1000} \right)^{0.3} \tag{3.6}$$

where CF is the center frequency of the BM filter. According to Shera et al. (2002),

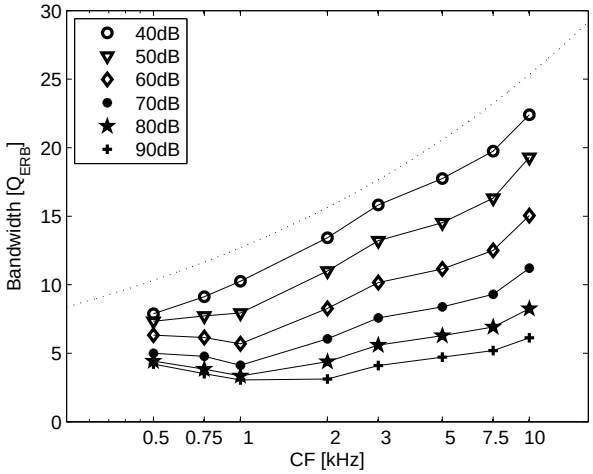



Figure 3.3: Filter bandwidths, $Q_{\mathrm{ERB}}$, derived from the output of the C1 filter path (from Fig. 2). The dashed curve shows $Q_{\mathrm{ERB}}$ estimates based on Shera et al. (2002)'s data obtained at a stimulation level of 40 dB pe SPL.

this function is applicable to humans at frequencies at and above 1 kHz. To map the $Q_{\mathrm{ERB}}$ to the $Q_{10}$ estimates used by the AN model the following mapping function was used (Ibrahim and Bruce, 2010):

$$Q_{10} = 0.2085 + 0.505 Q_{\mathrm{ERB}} \tag{3.7}$$

Fig. 3.3 shows the quality factor, Q, for the model's filters for different levels and CFs derived from simulated responses. The Q-values were derived from tuning curves by evaluating the magnitude response at CF to a number of pure tones with equal amplitude covering the frequency range around CF. The output from the C1 filter path was used for this calculation.

Third, cochlear suppression tuning curves have been found to have a peak at a higher frequency than the tip of an excitatory tuning curve (Delgutte, 1990), i.e., maximum suppression has been observed when stimulating at a higher frequency than CF. This was implemented in the original Zilany and Bruce (2006) model by

basally shifting the CF of the so-called control path filter by 1.2 mm on the BM. The 1.2 mm basal shift was retained in the humanized model, but Greenwood (1990)'s human frequency-place mapping was implemented to link the 1.2 mm shift to the corresponding characteristic frequency.

## 3.4 Method

### 3.4.1 Estimation of the unitary response

The unitary response (UR) was obtained by deconvolving a "template" click-evoked ABR with the summed neural activity pattern generated by the AN model in response to a click stimulus. Given the assumed superposition, any stimulus should in theory be usable. In this study, a click stimulus was chosen as it is most commonly used in clinics. The deconvolution is an ill-posed mathematical problem and has an infinite number of solutions. A stable and probable solution was, like in Dau (2003), found using Tikhonov regularization (Tikhonov, 1963) as implemented in the MATLAB Regularization Tools of Hansen (1998). The UR is subject dependent. In an attempt to employ a general UR, Elberling et al. (2010)'s grand average ABR data (left panel of Fig. 4) was used for the deconvolution. The resulting general UR was advantageous as the simulations presented in this study were compared to reference data, typically averaged across many subjects.

The grand average ABR (Elberling et al., 2010) was made by aligning wave-V peaks across recordings from 20 ears. The stimulus was a 100 $\mu$s standard click presented at 60 dBnHL ($\approx$ 95.2 dB pe SPL, see section III.B.3 for conversion factor). The alignment procedure created a standardized click-evoked ABR that had the disadvantage that the wave-V amplitude was smaller than in an individually measured ABR, due to inter-subject variability of the individual wave-forms. The UR was therefore scaled such that the simulated click-evoked ABR at 40 dBnHL had the same amplitude as the mean ABR amplitudes (rather than the amplitude of the grand averaged waveforms) from Elberling et al. (2010). The right panel of Figure 3.4 shows the UR, obtained with the grand averaged ABR (from the left panel) as the target. The

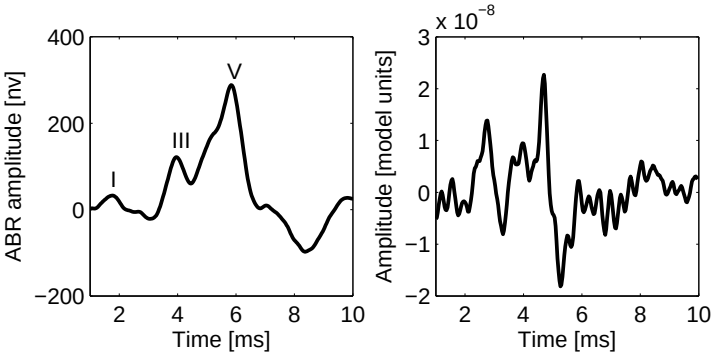



Figure 3.4: Left panel: Grand average ABR evoked by 60 dBnHL click (Elberling et al., 2010). Right panel: The derived unitary response function used throughout this study. This was calculated as the deconvolution of the grand average ABR and the summed neural activity pattern generated by the AN model in response to an identical click stimulus.

UR function is similar to the one obtained in Dau (2003). The ABR model using this UR is also capable of simulating the latency of wave-I. Given the linearity of the UR function the wave-I to wave-V interval will remain constant. Simulated wave-I amplitudes will however be smaller due to the way the UR was derived from the grand average ABR. If the model were to simulate wave-I amplitudes, the UR should either be scaled according to a representative wave-I amplitude, or be recalculated based on a click-response where the wave-I is more faithfully represented. In the present model, linear superposition was assumed above the level of the AN synapse; thus, the derived UR function was applied to any input stimulus at any level.

### 3.4.2 Stimuli

#### Tone bursts

Hanning-windowed tone bursts as in Harte et al. (2009) were used as stimuli. The tone bursts with center frequencies of 2 kHz and above included approximately 10 cycles and therefore ranged from 5 to 1.25 ms (see Table 3.1). The number of cycles during the rise time period was reduced to 7.5 at 1.5 kHz and 5 at 1.0 kHz. These durations represent a trade-off between having an equal number of cycles across frequencies and

| Frequency | Total Length | |
|---|---|---|
| kHz | ms | cycles |
| 0.5 | 10 | 5 |
| 0.75 | 7 | 5.25 |
| 1 | 5 | 5 |
| 1.5 | 5 | 7.5 |
| 2 | 5 | 10 |
| 3 | 3.4 | 10.2 |
| 4 | 2.5 | 10 |
| 6 | 1.7 | 10.2 |
| 8 | 1.25 | 10 |

Table 3.1: Tone burst stimuli used, with durations represented in ms and as number of cycles.

a relatively narrow spread in their spectrum. Levels of 40 to 100 dB peSPL were used, in steps of 10 dB.

**Broadband chirps and clicks**

Five chirps with different delay functions were used as defined in Elberling et al. (2010). The frequency-dependent delays of the chirps were defined as:

$$\tau = k \cdot \mathrm{CF}^{-d} \tag{3.8}$$

where $\tau$ represents the latency associated with frequency CF, and $k$ and $d$ are paired constants. Table 3.2 lists the parameters representing the individual chirps, following the choices of Elberling et al. (2010). The delay difference between 710 and 5700 Hz for the chirps 1 to 5 were thus 1.86, 2.56, 3.32, 4.12 and 5.04 ms, respectively. For comparison, a "standard" click stimulus of 100 $\mu$s duration was presented at 20, 40 and 60 dB nHL. The five chirps were calibrated such that they had the same spectrum level as the click.

| k | d | Chirp |
|---|---|---|
| 0.4501 | 0.6373 | 5 |
| 0.2207 | 0.5468 | 4 |
| 0.1083 | 0.4563 | 3 |
| 0.0531 | 0.3658 | 2 |
| 0.0260 | 0.2753 | 1 |

Table 3.2: Values of the paired parameter, k and d, which define the delay-frequency function (eq. 3.8)

**Calibration of the stimuli**

As the experimental data were described in dB pe SPL or dB nHL, it was necessary to acoustically calibrate the transient stimuli used in this study with an IEC 60711 coupler. The tone bursts and the click were measured acoustically with an Etymotic ER2 earphone connected to an IEC 60711 coupler (Brüel and Kjær 4157) through a Brüel and Kjær external ear simulator DB 2012. For each stimulus in the tone burst simulation (6 tone bursts and 1 click), the amplitude was adjusted until the acoustically measured peak-to-trough amplitude was similar to the peak-to-trough amplitude of a reference 1-kHz pure tone signal. A scaling factor was found to calibrate the numerical model.

As in Elberling et al. (2010), the chirps were adjusted to have the same spectrum level (rather than dB pe SPL) as the calibrated click. Elberling et al. (2010) provided the click and chirp levels in dB nHL, and the stimuli needed to be converted to dB peSPL at the eardrum before being presented to the model. The correct conversion factor was found to be 35.2 dB[1] (Richter and Fedtke, 2005), and hence

[1] The ISO 389-6:2007 standard specifies that the peak-to-peak reference equivalent threshold sound pressure level (peRETSPL) is 43.5 dB peRETSPL, for an ER2 earphone connected to an IEC 60711 coupler through the external ear simulator DB 0370. Unfortunately, the tube diameter for the standard ear tip for the ER2 earphone (ER1-14) is 1.37mm whereas it is 3mm for the DB 0370. This mismatch creates an acoustic horn effect which affects the spectrum (Richter and Fedtke, 2005; Elberling et al., 2012) and thus the level. Richter and Fedtke (2005) also measured the peak-to-peak reference equivalent threshold sound pressure level (peRETSPL) for an ER2 earphone connected to a head and torso simulator (HATS) and found it to be 35.2 dB. The change of the external ear simulator from the DB 0370 (ISO 389-6:2007) to the HATS (Table 7 Richter and Fedtke, 2005), results thus in a 8.3 dB change in the peRETSPL. As the acoustic horn effect is not present in human fittings, the ISO 389-6:2007 does not represent the pe SPL at the eardrum. For the modeling presented in the present study, the HATS measurements from Richter and Fedtke (2005) were therefore used as the reference.

the levels corresponding to 20, 40 and 60 dB nHL were found to be 55.2, 75.2 and 95.2 dB peSPL, respectively.

## 3.5 Results

### 3.5.1 Simulation of tone-burst evoked wave-V latencies

Figure 3.5 shows the simulated tone-burst evoked ABR wave-V latencies obtained with the ABR model (symbols connected with solid lines). For direct comparison, functions fitted to measured data from Neely et al. (1988) are indicated as dashed lines. Neely et al. (1988)'s fitted lines were described by:

$$\tau_b = a + bc^{-(i/100)}(\mathrm{CF}/1000)^{-g} \tag{3.9}$$

where i is the tone-burst intensity in SPL (divided by 100), CF is the tone burst center frequency in Hertz, and $a$ = 5 ms, $b$ = 12.9 ms, $c$ = 5.0 and $g$ = 0.413 were fitted constants. Additionally, measured data obtained in Harte et al. (2009) at a level of 66 dB peSPL are shown as a dotted line. The differences between Neely et al. (1988) and Harte et al. (2009)'s stimuli resulted in negligible differences in simulation results, therefore only Harte et al. (2009)'s stimuli are simulated here. The inter-subject variability (the standard deviation) on the Harte et al. (2009) data is 1.36 ms for 1 kHz, 0.93 ms for 2 kHz, and 0.71 ms for 8 kHz. Neely et al. (1988) does not explicitly state any inter-subject variability. The click data (Elberling et al., 2010) showed an inter-subject variability of 0.61 ms, 0.92 ms and 0.91 ms for hence 20 dB HL, 40 dB HL and 60 dB HL stimulus level.

The simulated and measured ABR wave-V latencies decrease exponentially as a function of frequency. At the highest stimulation levels, the simulated latencies are close to those observed in Neely et al. (1988). With decreasing level, the rate of change of latency with frequency increases both in the simulations and the measured data. However, the dynamic range of latencies across levels is smaller in the predictions than in the data. This effect is dominant towards higher tone-burst

frequencies where latencies of about 6-7 ms were predicted in contrast to 6-8 ms in the measured data. The squared correlation coefficient (the zero lag of the normalized covariance function) between tone-burst data and simulations is found to be $R^2 = 0.90$, showing a nice covariance between simulations and data. The simulated click-evoked latencies are indicated by the symbols next to the 8-kHz tone-pulse results. The filled circles on the right show the corresponding measured click data taken from Elberling et al. (2010). The stimulus levels used for the simulations were the same as those for the tone-burst simulations, whereas the levels of the click in the experimental study of Elberling et al. (2010) are stated next to the respective data points. As for the high-frequency tone pulses, the model predicts a reduced dynamic range of wave-V latencies across levels compared to the measured values.

### 3.5.2 Simulation of broadband chirp-evoked wave-V amplitudes and latencies

The black lines in Fig. 3.6 shows the simulated wave-V amplitudes obtained for the five chirps described in Elberling et al. (2010), at the three levels tested. In addition, click-evoked wave-V amplitudes for the same stimulation levels are shown on the left. The "change of delay" abscissa refers to the delay differences between the 5700-Hz component to the 710-Hz component of the stimulus. This reflects that a chirp with a faster sweeping rate has a shorter duration. The click is represented by a 0-ms change of delay as all the frequency components have the same delay. The gray lines of Fig. 3.6 shows the corresponding measured data from Elberling et al. (2010). The squared correlation coefficient between data and simulations is $R^2 = 0.90$, demonstrating good covariance between simulations and data. The measured data shows that, for the highest stimulation level of 60 dB nHL, the chirp with a relatively short duration (chirp 2) i.e. a small delay difference between the low- and high-frequency stimulus components, had the largest wave-V amplitude. Chirp 2 thus represents the stimulus that is most effective at synchronizing the neural output across frequency. In contrast, for the stimulation levels of 40 dB nHL and 20 dB nHL, the corresponding maxima were found with chirp 3 and chirp 5, respectively, suggesting that other sweeping rates provided maximal synchronization across frequency. These

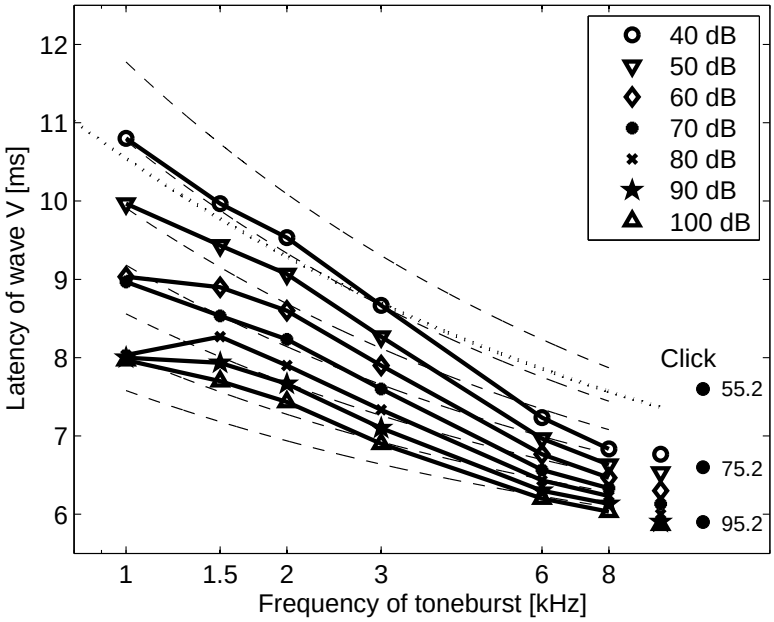



Figure 3.5: Simulated (solid curves) and modeled (dashed curves based on eq. 3.9, dotted curve, based on Harte et al., 2009) ABR wave-V latencies as a function of tone-burst center frequency and level. Each line fitted to Neely et al. (1988)'s empirical data corresponds to one simulated level. Open symbols to the right show simulated click-evoked ABR wave-V latencies, filled symbols show Elberling et al. (2010) measured click latencies. All levels are given in dB pe SPL.

key features observed in the measured data are also reflected in the simulations. The click-evoked responses show a smaller amplitude than those obtained with all chirps both in the data and the predictions. However, the maxima in the simulated functions are slightly shifted towards chirps with shorter durations. Overall, the correspondence between simulations and measured data is remarkable and the results support the hypothesis that the dynamic nonlinear processes in the cochlea strongly affect ABR formation.

Figure 3.7 shows wave V latencies simulated (black lines) by the ABR model and measured (gray lines) by Elberling et al. (2010) in response to the click and the five chirps. The squared correlation coefficient between data and simulations is

found to be $R^2 = 0.96$, indicating covariance of simulations and data. $R^2$ does not tell anything about the agreement between absolute latency values, it only shows that the data and simulation co-vary to a large degree. The measured latencies can probably be explained in terms of upwards spread of excitation (Elberling et al., 2010) and the fact that the frequency region dominating the ABR response is 2 to 4 kHz (Eggermont and Don, 1980) for the lower levels of 20 and 40 dB HL (for higher levels the region broadens towards higher frequencies). As stimulus level is increased, the BM filters broaden and lower frequency parts of the stimulus will excite the main frequency region. The longer the chirp is, the earlier is the low frequency part of the stimulus presented and an early excitation of the main frequency region is possible. Thus, at high levels (e.g. 60 dB HL) and long chirp delays (e.g. chirp 5), the latency will be very short due to the early presentation of low frequencies and the upward spread of excitation. The simulated results show the same trends, i.e. that the shortest duration is observed for high stimulus levels and long chirp delays. However, the level-dependence seems, as in the previous simulation of tone bursts and clicks, much compressed.

## 3.6 Discussion

This study evaluated the developed ABR model by comparing simulations with literature data, using clicks, tone-bursts and chirps as stimuli. The wave-V amplitudes simulated in response to a click presented at three stimulus-levels showed good correspondence to literature data, demonstrating that the overall calibration of the model was correct. Further, the correct level-dependence indicates that cochlear compression was well implemented. The latencies of the simulated tone-burst evoked ABRs showed good frequency-dependence, whereas the level-dependence was somewhat compressed. First, this shows that the traveling wave delay (the frequency-dependence) was modeled well. Second, the compressed level-dependence suggests that either the level-dependence of the BM tuning or the adaptation of the AN-IHC synapse was modeled imprecisely, or alternatively, that the assumptions underlying the UR were too extensive. This will be further discussed below. The chirp simulations

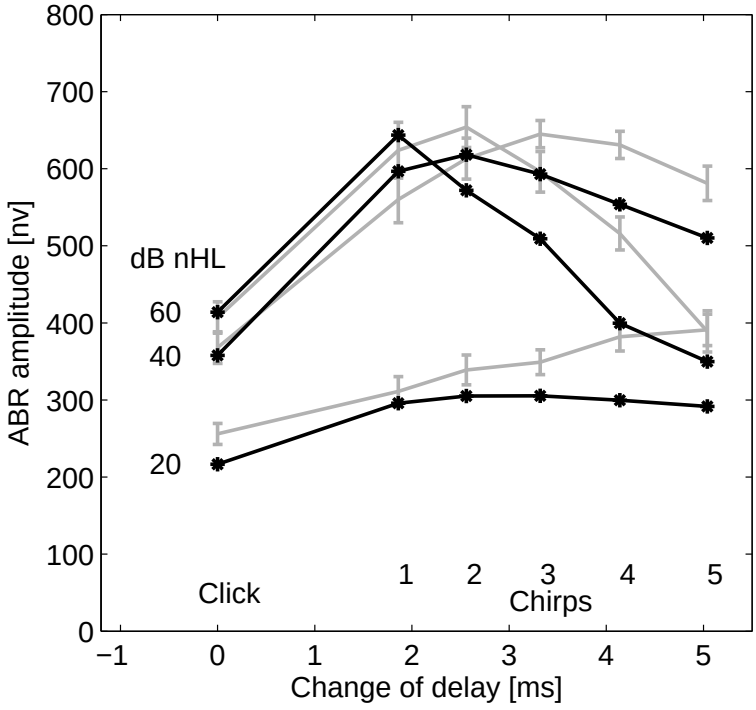



Figure 3.6: Black lines: Simulated ABR wave-V amplitudes evoked by click and 5 chirps with different frequency-delay functions at three different stimulus levels. gray lines: ABR wave-V amplitudes evoked by the click and five chirps (Recorded by Elberling et al., 2010). All simulations are well within one standard deviation of the measured value. Note that the error bars in the figure represents one standard error.

showed a good correlation with literature data. The simulations of the five chirps with different sweeping rates at three different levels demonstrated that the current model was capable of simulating responses to complex stimuli and that the interaction between the traveling wave delay and the level-dependent BM tuning seems to be working well.

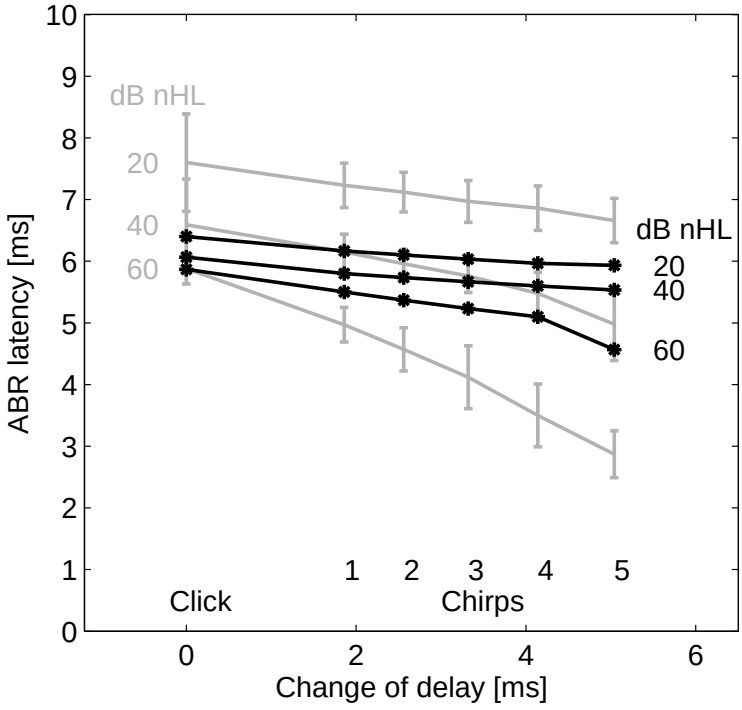



Figure 3.7: Black lines: Simulated ABR wave-V latencies evoked by click and 5 chirps with different frequency-delay functions at three different stimulus levels. gray lines: ABR wave-V latencies evoked by the click and five chirps (Recorded by Elberling et al., 2010). Note that the error bars represents one standard deviation.

### 3.6.1 Limitations of the conceptual approach

The assumption that all nonlinearity is restricted to the BM and AN and that the remaining processing is linear is an obvious over-simplification given the high complexity of neural processing within the brainstem. Specifically, the assumption that the rate functions in the MSO and AVCN within the brainstem are the same as in the AN is most likely erroneous (Dau, 2003). For example, it has been shown that neural synchronization in the AVCN can be enhanced compared with AN fibers, due to the convergence of inputs from two or more AN fibers on an AVCN cell and

postsynaptic cells that require coincident input spikes before firing (Joris et al., 1994). Furthermore, even though the human ABR may be largely generated by brainstem cells in the spherical cell pathway (Melcher and Kiang, 1996), there is probably also some contribution from other cell types such as globular and multipolar cells. There is still some controversy about the exact generating sites of the ABR peaks beyond wave I. The whole modeling approach should therefore be considered as a rough approximation of the real neural mechanisms involved in the generation of brainstem potentials. Nevertheless, it appears that the chosen approach represents an effective approximation since major characteristics of the measured data can be accounted for. These major characteristics include the wave-V amplitude, the frequency dependence of the wave-V latency and, to a lesser degree, the level-dependence of the wave-V latency.

### 3.6.2 Effects of the unitary response function

In the present study, the UR was empirically obtained by deconvolving a grand average click ABR with the discharge rate function at the output of the AN model. The UR was only obtained once, for this 95.2 dB SPL click evoked grand averaged ABR, and all other stimulus conditions made use of this UR. Only using one UR derived from a single waveform ensured that the generality of the modeling framework could be tested. Simple linear convolution of a UR might be an over-simplification for several reasons. First, the UR can be assumed to be subject dependent. In the present study, all simulations were rerun using individually estimated UR functions from three different subjects (not shown explicitly). However, this only resulted in a change to the overall simulated response amplitudes, and introduced an individual latency offset. The differences were minimal and reflected inter-subject differences, keeping the same broad dynamics as observed for the grand averaged UR. Second, Chertoff (2004) investigated the level and frequency dependency of a UR used to model compound action potentials (CAP) in Mongolian gerbils. He showed that the UR has both a slight level and frequency dependence in this species (the first peak of the CAP-UR shifts up to 0.1 ms). However, no general formulation of the dependency was stated and no formulation of a level-dependent UR for humans has yet been

attempted in the literature. Further, the interval between wave-I and wave-V peaks has been shown to be remarkably robust across stimulus level in ABR recordings (Don and Eggermont, 1978; Eggermont and Don, 1980), indicating that a level-dependent UR is not required. Contradictory to this, however, Chertoff et al. (2010) measured compound action potential (CAP) latency in humans, and demonstrated that CAPs could have a smaller latency change with level than what has been reported for ABR wave-V latency (Serpanos et al., 1997; Dau, 2003; Elberling et al., 2010). This would tend to suggest that the wave-I (which is believed to have the same origin as the CAP) to wave-V interval, and thus the UR, should be level-dependent. It is unclear from the literature whether a level-dependent UR is in fact needed.

### 3.6.3 Wave-V latency dependency on frequency and level

Taking the variability on the measured data into account, the simulated tone-burst evoked response latencies showed reasonable agreement with the measured data (Harte et al., 2009; Neely et al., 1988) for the frequency range 1 - 8 kHz and for a level range of 40 - 100 dB SPL. In particular, for a given stimulation level, the change of latency with frequency can be accounted for quite well by the model. However, the latency change with level was smaller in the simulations than in the data, particularly at high frequencies. Click-evoked ABRs were also simulated to test the model's performance when considering broadband excitation. The simulated click-evoked latencies of the present study decreased by only 0.6 ms for a 40 dB increase of stimulus level (from 55 to 95dB pe SPL), corresponding to -0.015 ms / dB, which is in contrast to the decrease of a little less than 2 ms observed in the Elberling et al. (2010) data, corresponding to -0.043 ms / dB. Other literature studies report latency decreases in the order of -0.043 ms / dB (Serpanos et al., 1997) and -0.046 ms / dB (Dau, 2003) for similar stimulus ranges. Even though the variability on the individual data set was high (a standard deviation of 0.81 ms on average for Elberling et al., 2010), the discrepancy between model and data is noticeable.

BM filter tuning and IHC-AN synapse adaptation determine the level dependency of ABR wave-V latency in the model. The ABR model latency change of -0.015 ms / dB

is a small improvement over the earlier modeling study by Dau (2003) who obtained latency changes of -0.005 ms / dB for a similar stimulus level range. Additional simulations, where the BM tuning was altered (and reported in Rønne et al., 2011), demonstrated that the improvement was the result of the use of the humanized version of Zilany and Bruce (2007)'s AN model instead of the model by Heinz et al. (2001). The humanized AN model uses the sharper tuning estimates from Shera et al. (2002) (see Ibrahim and Bruce, 2010) while Heinz et al. (2001) used the estimates of Glasberg and Moore (1990). The filters of Shera et al. (2002) (derived at only 40 dB SPL) are more sharply tuned than those described in Glasberg and Moore (1990) since they were estimated based on behavioral forward-masking data and otoacoustic emission data. In contrast, the estimates of Glasberg and Moore (1990) are based on behavioral simultaneous masking, which is affected by peripheral suppression (Shera et al., 2002; Bentsen et al., 2011). However, there is still a substantial discrepancy between the simulated and the measured latency-level range. As shown in Fig. 3.3, the model incorporates a level dependence in the C1 filter tuning factor. While the empirical evidence for the frequency dependence of the tuning factor (Shera et al., 2002, 2010; Bentsen et al., 2011) is well documented, there is little data existing for the level dependence in humans. This quality factor level dependence will strongly affect wave-V latency and could be one reason for the underestimation observed in the simulations. Additionally, neural adaptation in the IHC-AN synapse enhances the onset and leads to shorter delays. For analysis purposes (data not shown in this paper, see Rønne et al., 2011), click-evoked wave-V latencies were simulated using an altered version of the ABR model where the IHC output of the AN model was used, thus not including any adaptation process. However, while adaptation affected the absolute value of the wave-V latency in the framework of the present model, it did not have a major impact on the latency variation with level. A possible level-dependence of the UR, though not implemented in the model, could also affect the ABR wave-V latency. As discussed above, the literature is inconclusive on this matter. Further, Chertoff et al. (2010)'s CAP latencies decrease by -0.030 ms / dB over the level range of 75 to 105 dB SPL. So, even if a level-dependent UR was implemented to account for the difference in latency change between Chertoff et al. (2010) and Elberling et al. (2010), the AN model would still under predict the wave-V latency.

It thus remains unclear why the model fails to account more accurately for the level-dependent behavior of wave-V latency.

### 3.6.4 Across-frequency synchronization for broadband stimulation

When considering effects of level-dependent neural synchronization across frequency, the simulations illustrate the crucial role of nonlinear cochlear processing for the formation of brainstem responses to transient stimuli. The chirps presented in Elberling et al. (2010) were considered here as "critical" stimuli to challenge and evaluate the model. The results support the hypothesis that the dynamic behavior of ABR generation is mainly due to peripheral mechanisms as all processing at higher neural stages beyond the level of the AN was essentially considered as a linear filter. Further, the results reinforce the need to have level dependent chirp stimuli to get maximum wave-V amplitude clinically (Elberling and Don, 2010).

### 3.6.5 Perspectives

The model might be useful as a tool for studying consequences of different types of cochlear hearing impairment on the evoked potential waveform, provided that pathology can be adequately simulated in the model. Furthermore, brainstem responses to complex stimuli (cABR), such as consonant-vowel utterances, have been considered as an objective index of the neural transcription of features (e.g. temporal, spectral) that are important for speech understanding in quiet and noise (e.g. Anderson et al., 2011). The model could be used to analyze which spectro-temporal characteristics of the speech-evoked patterns can be accounted for by cochlear processes. Finally, an important step would be to consider "steady-state" responses (SSR) obtained with temporally fluctuating stimuli such as complex tones or amplitude modulated tones or noises. These responses are assumed to be generated by units in the auditory brainstem and in the primary auditory cortex (e.g. Kuwada et al., 1986). Therefore, the corresponding unitary response would have to be extended

by a middle-latency component. It is not clear, to what extent such a convolution approach can be successfully applied to middle-latency responses (MLR), to transients as well as amplitude modulation following responses. Regarding MLRs, at least, it has been shown that the "classical" SSR to click trains presented at a 40 clicks/sec repetition rate can be modeled reasonably well using a linear convolution approach (Bohorquez and Oezdamar, 2008; Junius and Dau, 2005).

## 3.7 Summary and conclusion

A computational model for the generation of ABRs to transient stimuli was presented. The model was based on the assumption that an ABR can be simulated as the convolution between an instantaneous discharge rate function and a unitary response. The instantaneous discharge rate function was obtained from a state-of-the-art nonlinear AN model (Zilany and Bruce, 2006). The UR was derived "empirically" as the deconvolution between the simulated instantaneous discharge rate AN function in response to a click stimulus and measured average click-evoked ABR.

The model was evaluated by comparing the predicted responses to measured ABR data from the literature. It was shown that a realistic simulation of the level-dependent signal processing in the cochlea is essential for the interpretation of ABR to tone pulses, clicks and chirps presented at various stimulation levels. In particular, the model could account reasonably well for the nonlinear wave-V amplitude behavior as a function of chirp stimulus level and sweeping rate which supports the strong role of cochlear nonlinearities, such as compression and level-dependent tuning, for the formation of ABR. However, the model clearly underestimated the level dependence of the response (wave-V) latency and it remained unresolved in the framework of the modeling work presented here what mechanisms are responsible for the relatively large latency changes with level observed in the data.

Overall, the developed model can provide insight into the complex nature of ABR generation. It can be used to investigate the representation of other types of stimuli (such as speech in noise) or to study effects of (different types of cochlear) hearing

impairment on the predicted potential patterns. Furthermore, the modeling approach might provide a basis for the investigation of longer-latency responses, such as steady-state responses to amplitude modulated tones and noises.

The ABR model including, grand average ABR, UR, and key simulations, is included in the Auditory Modeling (AM) toolbox (Søndergaard et al., 2011) and can be downloaded from: http://amtoolbox.sourceforge.net/ (date last viewed 02/14/12).

# 4

# Modeling the level-dependent latency of the auditory brainstem response

This chapter is based on Rønne et al. (2011). In the framework of the thesis, this is an expanded discussion on why the level-dependent latency of the click-evoked ABR is under predicted by the ABR model.

## 4.1 Abstract

Auditory brainstem responses (ABR) are used for both clinical and research purposes to objectively assess human hearing. A prominent feature of the transient evoked ABR is the level-dependent latency of the distinct peaks in its waveform. The latency of the most prominent peak, wave-V, is about 8 ms at a peak equivalent sound pressure level of 55 dB, and reduces for increasing level by approximately 1 ms / 20 dB. A classical explanation for this finding asserts that an increasing stimulus levels lead to a broadened excitation pattern on the basilar membrane. This results in further activation of the basal regions of the cochlea. Given the physical properties of the basilar membrane, increased basal activation is believed to cause a decreasing ABR latency. An Auditory Nerve (AN) model and the Dual Resonance Non-Linearity (DRNL) filter model are considered as separate front-end cochlear models to simulate ABRs. Even though both models incorporate level-dependent tuning and synapse adaptation, and thus theoretically should be capable of simulating level-dependent latencies, both models under-predict the latencies. The failure to produce accurate simulations suggests, that the level-depending tuning in the models is not accurately

modeled. The level dependency of the basilar membrane filter tuning in humans is not well described in the literature and could therefore cause the modeling difficulties.

## 4.2 Introduction

ABRs in response to transient sound stimuli represent the summed electric potential from many remotely located neurons, recorded via scalp electrodes. The ABR has 7 distinct waves, where wave-V is the most prominent. One key feature of the ABR wave-V is the peak latency which is dependent on both frequency (Neely et al., 1988) and level (Dau, 2003). This frequency dependence is due to the tonotopic mapping on the basilar membrane (BM) with high frequency at base and low frequency at apex. The result is that high frequency auditory nerve responses occur earlier than low frequency responses. The level-dependence is not as well understood, but is thought to be determined by the frequency specificity of the basilar membrane (BM), i.e. its tuning, and the inner hair cell (IHC) - auditory nerve (AN) synapse adaptation. This study has investigated the ability of two established auditory models, when used as a front-end in an ABR model, to simulate level-dependent wave-V latency in response to click stimuli. Both the Auditory Nerve (AN) model (Zilany and Bruce, 2006, 2007) and the Dual Resonance Non-Linearity (DRNL) filter model (Meddis, 2006) were assumed to contain the nonlinear processes required to account for level-dependent wave-V latency. Two front-end models are used to minimize the potential effect of implementation errors, and to evaluate whether the individual differences between the two models are important.

### 4.2.1 Level-dependent latency theory

Cochlear tuning is level-dependent, where an increase in stimulus level results in broader auditory filters. On the BM, the broader filters result in broader excitation patterns, i.e. regions of the BM with characteristic frequencies further from the center-frequency of a stimulus are recruited. Elberling (1976) and Folsom (1984) discussed how this broadening in excitation with level results in shorter latencies as more basal

regions of the BM are activated, i.e. regions with shorter implicit delays. Another inherent feature of the filter tuning is the change in the envelope of the local BM impulse response. An increase in level will result in an inherently shorter impulse response. The delay of the individual peaks will be constant but the amplitude of the earlier peaks will be emphasized, and given the associated delay will decrease with increasing stimulus level. Recio and Rhode (2000) demonstrated that this phenomena can be physiologically measured on the chinchilla BM, and Kiang (1965) showed that the effect is also measurable in the cat AN. Across many filters, the envelope change with increasing stimulus level acts as an onset emphasis that results in a decrease of wave-V latency. The IHC-AN synapse adaptation has similar properties, amplifying the onset of a signal and attenuating later parts (Westerman and Smith, 1988). This effect enhances the level-dependent effects on wave-V latency created by the filter tuning.

## 4.3 ABR Model structure and unitary response

The structure of the ABR model is shown in Fig. 4.1. The ABR model uses either the DRNL filter model (DRNL-ABR) (Meddis, 2006) or the AN model (AN-ABR) (Zilany and Bruce, 2006, 2007) as the front-end cochlear model. The AN model calculates the instantaneous discharge rate for individual AN fibers, in response to a given stimulus defined in Pascals. Equivalently, the DRNL filter model calculates the vesicle release probability also for single AN fibers. Each fiber (in both models) is tuned to a specific characteristic frequency (CF). The CFs chosen were spaced according to the human cochlear map of Greenwood (1990). The number of fibers included was a trade-off between computational time and model accuracy. Throughout this study, 500 fibers ranging from 100 Hz to 16 kHz were used in all simulations. The output of the front-end cochlear models was summed across all fibers and convolved with a unitary response (UR) function, derived separately for the two models. The UR is defined as the potential produced between the electrode positions on the scalp each time a cell discharges. The URs, one for each of the models, were obtained by deconvolving a template 95.2 dB peSPL click-evoked ABR Elberling et al. (2010),

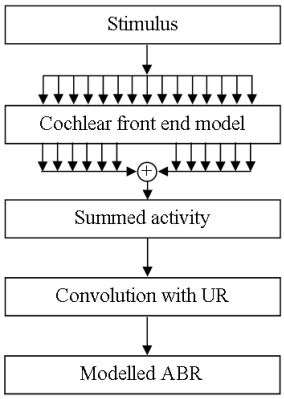



Figure 4.1: Schematic structure of the ABR model. 500 AN fibers tuned to different CFs are individually simulated by the AN model. The summed activity, integrated across frequency, is then convolved with a unitary response and represents the simulated ABR to a given stimulus.

shown in the left panel of Fig. 3.4, with the summed neural activity pattern generated by either front-end model in response to a similar click stimulus. The deconvolution is an ill-posed mathematical problem and has an infinite number of solutions. A stable and probable solution was, like in Dau (2003), found by using the Tikhonov regularization Tikhonov (1963), and the MATLAB toolbox from Hansen (1998). Figure 4.2 (right) shows the unitary responses, obtained with a grand averaged ABR at 95.2dB peSPL as the target. Linear superposition was assumed above the level of the AN synapse, and thus the calculated unitary response functions given in Fig. 4.2 was used for any input stimulus level. As expected, the two derived URs are almost identical (see Harte et al. (2010) for further information on the modeling framework).

### 4.3.1 Cochlear models

The input to the auditory nerve (AN) Zilany and Bruce (2006, 2007) model is the instantaneous pressure waveform of the stimulus in units of *Pa*. The output of the AN model is the spike rate in response to the stimulus pressure. The model includes a number of key functional stages: a middle-ear filter; a feed-forward control path

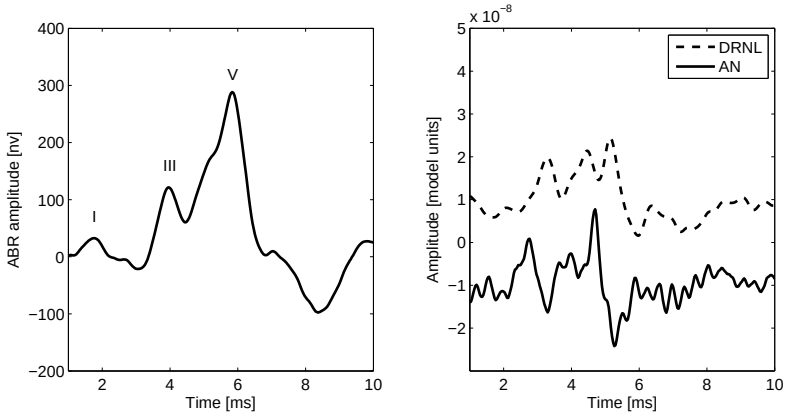



Figure 4.2: Left panel: Grand average template ABR evoked by a 95.2 dB peSPL click Elberling et al. (2010). Right panel: Derived unitary response functions for hence the AN-ABR and the DRNL-ABR model. Both are calculated as the deconvolution of the grand average ABR and the summed neural activity pattern generated by the front-end cochlear model in response to an identical click stimulus. The two URs has for display been shifted in amplitude.

representing the active mechanism; a primary signal-path filter (C1) representing the basilar membrane (BM) filtering adapted by the control path; a parallel-path filter (C2) for high-level stimuli; an inner-hair cell (IHC) section followed by a synapse model and a stochastic AN spike discharge generator. In the present study, the spikes/s output from the synapse model was used, rather than the stochastic output from the spike generator. The input to the dual-resonance nonlinear (DRNL) filter model Meddis (2006) is also the instantaneous pressure waveform in *Pa*. The output from the model is the vesicle release probability. The model Meddis (2006) used in this work consists of an outer and middle ear-filter, the DRNL filter (BM filter stage), an inner hair cell (IHC) transduction stage and a IHC-AN synapse. The DRNL (Lopez-Poveda and Meddis, 2001; Meddis et al., 2001) filter is a computational algorithm which aims at simulating a number of features characteristic of the basilar membrane. One of many features is a compressive input-output function, and consequently level-dependent tuning. The output from both models were deterministic and the effects of refractoriness were thus not considered in this work.

### 4.3.2 Stimuli and calibration

As the literature data are described in dB peSPL it was necessary to acoustically calibrate the transient stimuli used. The click were measured acoustically in an IEC 60711 coupler. The numerical stimulus peak-to-trough amplitude of a reference 1-kHz pure tone signal was adjusted until the acoustically measured peak-to-trough amplitude was similar to that of the click. A scaling factor, defined as the ratio between the stimulus peak-to-trough amplitude of the pure tone and the stimulus peak-to-trough amplitude of the transient signals, was derived as;

$$S = \frac{L_{Signal}}{L_{Reference}} \tag{4.1}$$

where $S$ is the scaling factor, $L_{Signal}$ is the stimulus peak-to-trough amplitude of the transient signal, and $L_{Reference}$ is the stimulus peak-to-trough amplitude of the reference pure tone. The AN model was calibrated such that the root-mean-square value of a reference pure tone signal was 1, whereas the DRNL model is calibrated such that the peak value of a reference pure tone signal was 1. The amplitude of the numerical click in Elberling et al. Elberling et al. (2010), used as stimuli to the models, was thus scaled by the derived factor $S$ for the DRNL model, and by $\frac{S}{\sqrt{(2)}}$ for the AN model.

## 4.4 Results

The left panel of Fig. 4.3 shows ABRs simulated by the AN-ABR model in response to clicks at 50, 70 and 90 dB peSPL. A shift in the wave-V peak to shorter latencies with increasing stimulus level is clearly observed. The right panel of Fig. 4.3 shows simulated click-evoked ABR wave-V latencies as a function of stimulus level. Also shown are recorded click ABR latencies Dau (2003). Simulations were done with both the DRNL-ABR and the AN-ABR model. The two models produce similar results for stimuli levels between 70 and 100 dB peSPL. For lower levels, the DRNL-ABR model no longer produces a distinct wave-V, thus deriving a latency associated to those

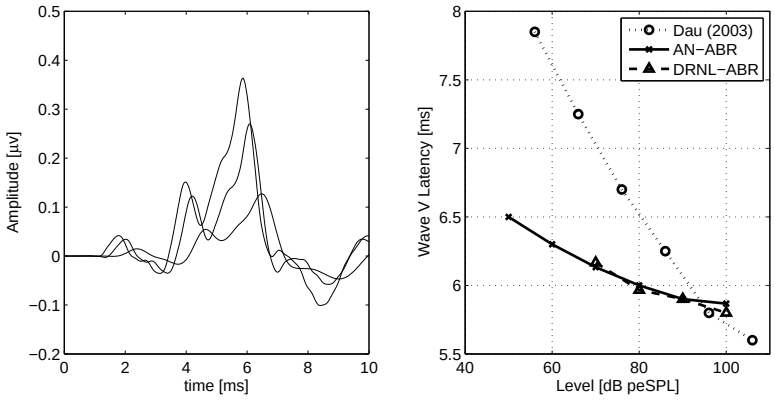



Figure 4.3: Left panel: AN-ABR model simulations to click stimulus at 50, 70 and 90 DB peSPL. Note the latency change of the wave-V peak. Right panel: Simulations of click-evoked ABR wave-V latencies across stimuli levels, using both the AN-ABR and the DRNL-ABR model. Both models show compressed level-dependent latencies compared to Dau Dau (2003) experimental data.

levels was not possible. As expected, it is seen that both models simulated reduced wave latency for increasing stimulus level. However, a clear disparity between both sets of simulations and the recorded reference data is observed. The recorded data shows a decrease in wave-V latency of approximately 2 ms for a 40 dB stimulus level increase, whereas the models simulates approximately 0.6ms decrease for 40dB increase in stimulus level.

## 4.5 Discussion

Fig. 4.3 (right) showed that both models under-predicted the ABR latency. The classical theoretical explanations of the ABR latency change with stimulus level says that the IHC-AN synapse adaptation and the cochlear tuning should be the key features. To quantify whether these features were captured, the impact of the tuning and the adaptation in the AN-ABR model was investigated. The focus was on the AN-ABR model as it produces the most reliable results over the widest range of input stimulus levels. To be able to interpret the model correctly, URs for each new version

of the model were derived. The URs were derived from the same 95.2 dB peSPL click-evoked template ABR, thus results shown in this section have by default correct latency estimation at 95.2 dB peSPL.

The key feature producing the level dependency of wave-V latency was the filter tuning. Fig. 4.4 (left) shows the effect of exchanging the Shera et al. (2002) filter tuning, originally implemented in the AN model, with the less sharply tuned Glasberg and Moore (1990) filters, on wave-V latency. It is observed that the latency change with stimulus level is approximately halved. Both Shera et al. (2002) and Glasberg and Moore (1990) describe the frequency dependence of the filter tuning. Thus, exchanging Shera et al. (2002) tuning with the Glasberg and Moore (1990) tuning makes all the filters broader, independent of level. The reason for the larger latency change with stimulus level found when using Shera et al. (2002) tuning is that sharper filters increase the frequency specificity and thus limit the upward spread of excitation at low levels. At higher levels, there is thus room for a significant increase in upward spread of excitation, thus creating larger latency changes with level. Shera et al. (2002) measured the filter tuning using a forward masking paradigm. The tonal target stimuli was presented at 40dB SPL. Literature data obtained at higher levels and high frequencies, measured with this paradigm, are however not available. For the high levels, the lack of data is likely due to the practical limitations of presenting an off-target masker that does not get uncomfortably loud when measuring the skirts of the filters. As the sharpness of the tuning was shown to be important for the level dependency of wave-V latency, the lack of trustworthy data is however a large uncertainty. Getting the level-dependency of the tuning correctly could prove to be key when modeling wave-V latencies. Fig. 4.5 shows filter bandwidths, $Q_{ERB}$, at different center frequencies and levels, derived from the current AN model. Data to which these simulated $Q$-values could be compared with, would be beneficial.

The right panel of Fig. 4.4 shows simulated click-evoked ABR wave-V latencies, generated by the AN-ABR model where the adaptation of the IHC-AN synapse has been left out. The removal of the adaptation clearly shows a reduction of latency change with stimulus level. Note that the UR was calculated based on a 95.2 dB peSPL click, and that the latency of the simulations around this level by default therefore is

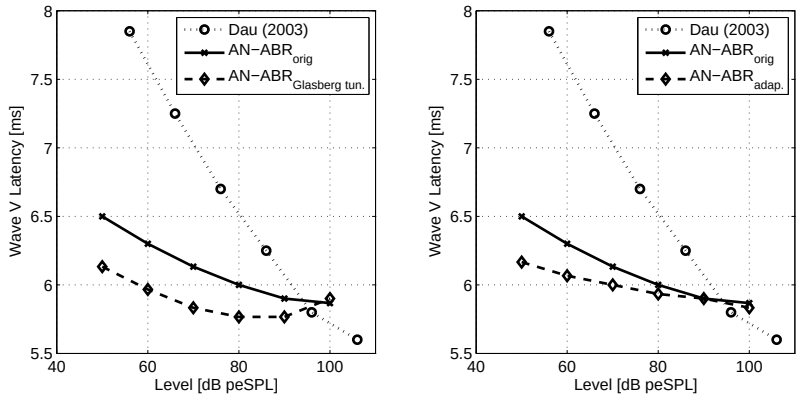



Figure 4.4: Simulations of click-evoked ABR wave-V latencies across stimuli levels. In both figures are the data recorded by Dau (2003) and the simulations from Fig. 4.3 shown as reference. The figure to the left additionally shows wave-V latencies simulated by the AN-ABR model where the filter tuning was based on Glasberg and Moore (1990). The figure to the right shows wave-V latencies simulated by an AN-ABR model that excluded the IHC-AN synapse adaptation.

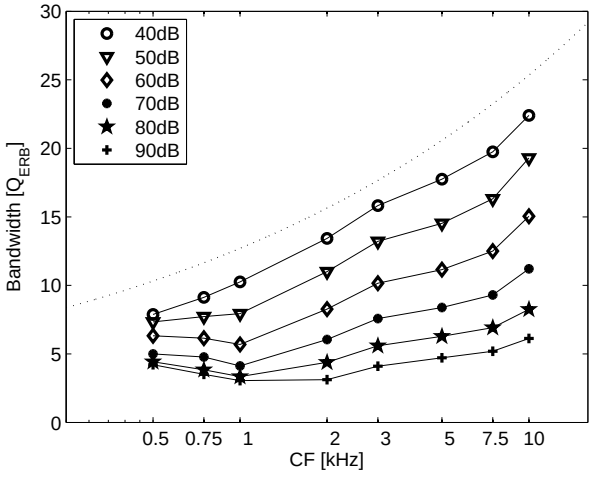



Figure 4.5: Filter bandwidths, $Q_{ERB}$, derived from the output of the C1 filter path. The dashed curve shows $Q_{ERB}$ based on Shera et al. (2002)'s experimentally derived function for a stimulation level of 40 dBpeSPL.

correct. The "correct" picture when removing the adaptation should therefore have been a curve shifted upwards, as the inclusion of adaptation sharpens the onset and thus leads to shorter delays. However, the simulated results show that removing the adaptation approximately halves the latency change with level. This was supported by an additional simulation (not shown) where the adaptation was removed from the AN-ABR model based on the Glasberg and Moore (1990) tuning. The IHC-AN synapse adaptation used in the AN model was revised by Zilany et al. (2009). Additional simulations were performed using this synapse model; however, no effect on the level-dependent latency was found. The adaptation is thus argued to be important for wave-V latency but not the reason for the under-estimated latency change.

Two other modeling features could be thought to affect the ABR latency. The first is the unitary response (UR). Chertoff (2004) investigated the level dependency of a UR used to model compound action potentials (CAP) in Mongolian gerbils. Chertoff (2004) showed that the UR was level-dependent in this species. However, no general formulation of the dependency was stated, and no formulation of a level-dependent UR for humans has been found in the literature. It cannot be excluded that a level-dependent UR would affect the latencies. The interval between wave-I and wave-V, is however, remarkably robust across stimulus level. The UR models the auditory pathway from the wave-I generation site, argued to be the IHC-AN synapse, to the wave-V generation site. Thus, it is not likely that a level-dependent UR would have a major impact on the latencies. The second alternative feature that could affect the ABR latency is the auditory nerve refractory period which was not included in the AN-ABR model of the present study. This choice was made to make the model computationally faster. Additionally simulations were carried out where the refractory period was included. However, no improvement on the wave-V latency change with level was observed.

## 4.6 Conclusion

Two ABR models were build, both using a principle where a cochlear front-end model was convolved with a unitary response (UR). Both ABR models were shown

to significantly under-estimate the click-evoked ABR wave-V latency change with stimulus level. The two models should, given classical explanations, be able to model click-evoked ABR latencies. The fact that they fail leads to the suggestion that the cochlear tuning is likely to be imprecise at high levels and high frequencies.

# 5

# Low-frequency versus high-frequency synchronization in chirp-evoked auditory brainstem responses

In chapter 3 the ABR model was developed. It was quantified that the model was capable of simulating ABR wave V latencies and amplitudes to click, tone bursts and chirps. First, this chapter develops two tools to illustrate details of the ABR model simulations. These illustration tools, the AN-spectrogram and the AN-UR-spectrogram, has proven a valuable tool aiding stimulus creation for experiments, as well as the understanding of simulations. Here they are used to motivate the "Low-frequency versus high-frequency synchronization in chirp-evoked auditory brainstem responses" study [1].

## 5.1 The ABR model used as an illustration tool

### 5.1.1 Stimuli

The two stimuli used, a click and a chirp, were both taken from Elberling et al. (2010) and were thus identical to the click stimulus and the "chirp-3" stimulus in chapter 3. Both stimuli were band-limited from 100Hz to 10 kHz. All simulations were carried out at 75.2 dB peSPL, corresponding to 40dB HL for the click (see section 3.4.2)

[1] This study is based on Rønne and Gøtsche-Rasmussen (2011)

### 5.1.2 Spectrograms

Fig. 5.1 and 5.2 show hence a simulated click and chirp evoked ABR. Wave I, III and V are clearly visible. The latency and amplitude of the wave V's were naturally similar to the ones presented in Fig. 3.7 and 3.6. Each simulated ABR was the summation of 500 channels, each tuned to a different CF. Fig. 5.3 shows click evoked AN responses in a AN-spectrogram representation. The Y-axis shows the 500 AN fibers characterized by their CF. Each horizontal line in the figure are thus the click evoked response of the humanized Zilany and Bruce (2007) AN model tuned to a CF. The color represents the instantaneous discharge rate at a specific time in a specific fiber. Fig. 5.4 shows AN-UR-spectrogram representation, created by convolving each horizontal line in Fig. 5.3 with the unitary response (UR, see section 3.4.1). As the convolution was a linear process, the summation over channels of this figure give the ABR shown in Fig. 5.1. The color represents each channels contribution to the summed ABR potential (unit of $\mu$v).

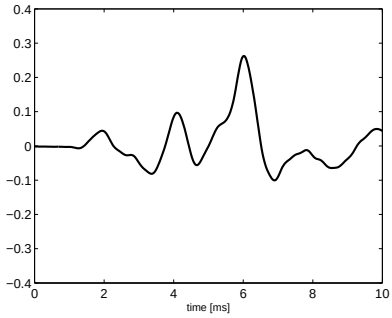



Figure 5.1: Simulated ABR evoked by Elberling et al. (2010) click.

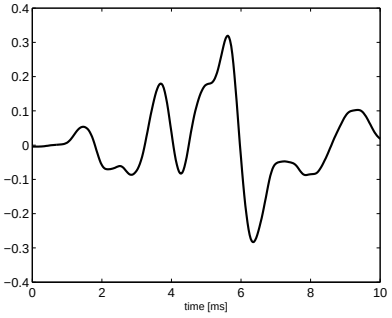



Figure 5.2: Simulated ABR evoked by Elberling et al. (2010) chirp-3.

In the spectrograms, details of the underlying processing can be observed. In the AN-spectrogram it can be observed that the fine-structure information is available at low frequencies whereas only the envelope seems to be tracked at higher frequencies. This is seen as the impulse responses at low frequencies (a single horizontal line) has multiple peaks, with a periodicity corresponding to the fiber CF. In the AN-UR-

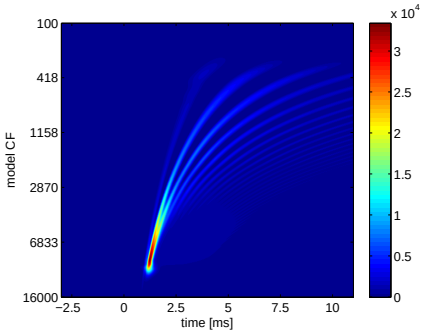



Figure 5.3: AN-spectrogram showing the simulated neural activity at the AN in response to click stimulus.

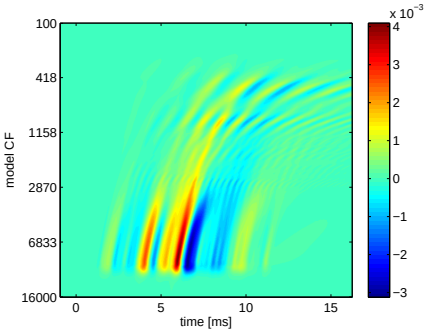



Figure 5.4: AN-UR-spectrogram visualizing the components that sums up to form the simulated ABR. This figure is created by convolving Fig. 5.3 line by line with the UR.

spectrogram the ABR wave III and V are visible as the two red lines occurring around 4 and 6 ms. A clear latency shift from the AN-spectrogram is observed due to the UR.

Fig. 5.5 and 5.6 shows spectrograms evoked by the Elberling et al. (2010) chirp-3. It is clearly observed that much of the activity in the AN-UR-spectrogram is time-aligned at the discrete values of 4,5 and 6 ms. It is further observed that the impulse responses have a long duration at low frequencies. This has the consequence that it is impossible to time-align all of the activity stemming from low frequencies. It is however observed, that the peaks of the low frequency impulse responses are aligned with the peaks of the high frequency contributions. A larger wave-V amplitude is thus observed using a chirp stimulus than a click stimulus.

### 5.1.3 Motivation for the following study

It was shown that the simulated the low frequency contributions to the click-evoked ABR was not time-aligned with the high-frequency contributions and were thus not adding up in phase (Fig. 5.4). The chirp evoked ABR showed a much more time-aligned response at low frequencies (Fig. 5.6); however, it was also indicated that the alignment of high-frequencies was significantly better. This contradicts the

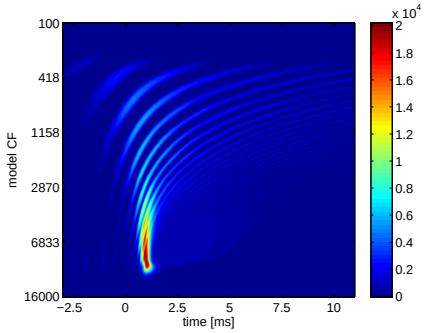



Figure 5.5: AN-spectrogram showing the simulated neural activity at the AN in response to an Elberling et al. (2010) chirp-3 stimulus.

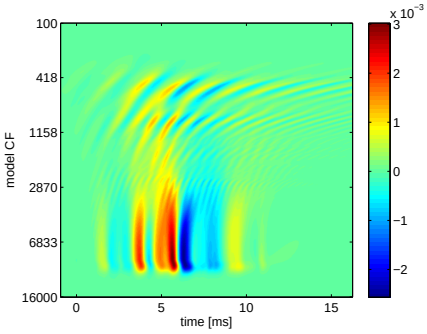



Figure 5.6: AN-UR-spectrogram visualizing the components that sums up to form the simulated ABR. This figure is created by convoliving Fig. 5.5 line by line with the UR.

common belief in literature (e.g. Shore and Nuttall, 1985; Dau et al., 2000) where it has been argued that the alignment of the low frequencies was the only reason for the larger wave-V amplitude evoked by a chirp rather than a click. This deviation between literature explanations and simulations led to the following study, where it was investigated whether the better alignment of the high-frequencies contribute significantly to the larger chirp evoked ABR wave-V amplitude.

## 5.2 Abstract

This study investigates the frequency specific contribution to the auditory brainstem response (ABR) of chirp stimuli. Frequency rising chirps were designed to compensate for the cochlear traveling wave delay, and lead to larger wave-V amplitudes than for click stimuli as more auditory nerve fibers fire synchronously. Traditional click stimuli were believed to only excite high-frequency fibers synchronously. It is still currently unclear whether the broad-band chirp stimulus leads to increased synchronization of both low- and high-frequency fibers. It is also unclear if both these groups of fibers contribute significantly to the overall wave-V amplitude. In the present study, ABRs were recorded from 10 normal-hearing listeners using low-

and high-frequency band-limited chirps and clicks (0.1 - 1.5 kHz and 1.5 - 10 kHz) presented at a level of 40 dB HL. The results showed significantly larger wave-V amplitudes for both low and high-frequency band-limited chirps than for the filtered clicks. This demonstrates that the synchronization of nerve fibers occurs across the entire frequency range at this presentation level, and this leads to significant increases in wave-V amplitudes. The increase for the low-frequency chirp was found to be clearly larger than that obtained at the higher frequencies.

## 5.3 Introduction

ABRs in response to transient sound stimuli represent the summed electric potential from many remotely located neurons, recorded via scalp electrodes. The click evoked ABR has 7 distinct waves, where wave-V is the most prominent. One key feature of the ABR wave-V is the peak latency which is dependent on both stimulus frequency (Neely et al., 1988) and level (Dau, 2003). The frequency dependence is due to the tonotopic mapping on the basilar membrane (BM) with high-frequency at base and low-frequency at apex (Greenwood, 1990). Each frequency component of a stimulus is associated with a certain delay, and a click stimulus will thus elicit responses over a relatively large time span. This limits the synchrony of the response, and thereby reduces the ABR amplitude evoked by such a stimulus (Elberling et al., 2007). Frequency rising chirps have been designed to compensate for the cochlear traveling wave delay. The use of chirp stimulus lead to larger wave-V amplitudes than for click stimuli as more auditory nerve fibers fire synchronously (see Elberling et al., 2007, for review). The increase in synchrony has traditionally been argued to occur mainly at low frequencies, where the peaks of the individual nerve responses are most delayed. E.g. Shore and Nuttall (1985) and Dau et al. (2000) argue that the low frequencies are the key to the improved wave-V amplitudes, as low frequencies are least synchronous with the more aligned high frequencies and the room for improvement thus is largest. However, the impulse responses of the nerve fiber responses at low frequencies are much longer in time citepKiang1965, and it is thus not possible to align all the excitation at low frequencies. A chirp is though designed to align all frequencies

(Elberling and Don, 2008), and the better alignment of high frequencies, with short impulse responses, could thus be an alternative hypothesis. It is still currently unclear whether the broad-band chirp stimulus leads to increased synchronization of both low- and high-frequency fibers. It is also unclear if both of these groups of fibers contribute significantly to the overall wave-V amplitude. The research questions addressed in this paper are: 1) Is the increased wave-V amplitude (increased nervous synchrony) observed for both high and low frequencies when stimulating with chirps instead of clicks? 2) Are high or low frequencies key to the increased wave-V amplitude observed when stimulating with broad-band chirps?

## 5.4 Test design

Six stimuli were created. A broad-band click and a broad-band chirp, containing the frequencies from 100 Hz to 10 kHz, were used as reference. The click was a 100 ţs standard click, and the chirp was identical to "chirp 3" in (Elberling et al., 2010). Further were low-frequency and high-frequency versions of hence click and chirp created. The method described by (Elberling et al., 2007) was used. The phase delays for hence chirps and clicks were the same as used to create the broad-band stimuli. Both the high-frequency and low-frequency cut-off frequency was 1500 Hz. Fig. 5.7 and Fig. 5.8 shows the time series representation of the three hence click and chirp stimuli. The power spectra of the two broad-band stimuli were identical. The summed versions of hence the low-frequency and high-frequency click, and the low-frequency and high-frequency chirp has also identical power spectra as the broad-band versions. The power of hence the low-frequency (-3.1 dB relative to broad-band condition) high-frequency (-0.6 dB relative to broad-band condition) stimulus are thus smaller than the power of the broad-band versions. Fig. 5.9 shows the power spectra of the stimuli, note that hence the two broad-band stimuli, the two low-frequency stimuli and the two high-frequency stimuli have identical spectra. The six stimuli were linked to each other in terms of the power spectra as described above. Therefore only the broad-band click was calibrated, and the rest adjusted correspondingly. By inserting ER1-14 ear plug in a B&K Ear Simulator Type 4157 (IEC 60711) using adapter B&K DB

2012 the click was calibrated to a level of 75.2 dB peSPL. The reference equivalent threshold sound pressure level (RETSPL) for the click calibrated this way is 35.2 dB RETSPL (taken from the corresponding head and torso simulator measurement of Richter and Fedtke (2005), and the measurements are thus carried out at 40 dB HL.

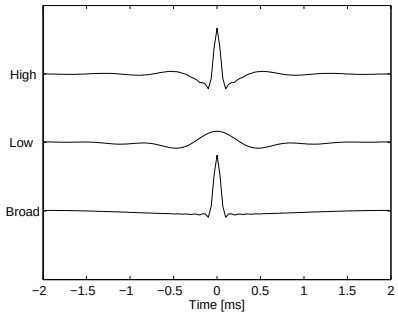



Figure 5.7: The three click stimuli.

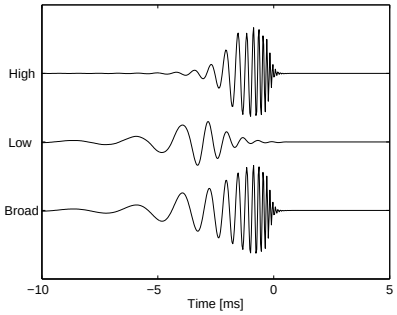



Figure 5.8: The three chirp stimuli, all based on "chirp 3" from Elberling et al. (2010).

### 5.4.1 Test subjects

The ABR measurements were carried out at the Centre for Applied Hearing Research (CAHR), Technical University of Denmark. Ten normal-hearing test subjects (10 left ears) participated in the study. All subjects had normal hearing defined as pure tone thresholds equal to or better than 20 dB HL in the range from 125 Hz to 8 kHz. The subjects were all students between 20-30 years old (2 females and 8 males). The session lasted for maximally 1.5 hours including a short briefing and fitting of electrode cap. Only the left ear was tested.

### 5.4.2 Measurement procedure

The test subject was placed in an electrically and acoustically shielded booth. The signals were presented at 48 kHz sampling frequency through an Etymotic Research ER-2 insert earphone. The recording of the ABR was done using a Medical Equipment

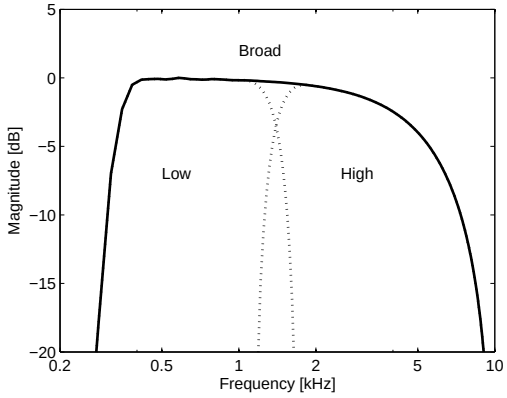



Figure 5.9: Spectra of the different stimuli. The sum of the two hence low- and high-frequency clicks or chirps have the same power spectrum as the broad-band stimulus.

ApS Synamps2, which sampled the recorded signal at 10 kHz. The electrodes were placed at vertex (reference), ipsi-lateral mastoid, and forehead (ground). An impedance between the electrodes below 1 k$\omega$ was achieved for the majority of the test subjects. The post-processing was done using MATLAB. The raw data was averaged, and filtered using a band-pass filter with cut-off frequencies at 100 and 3000 Hz. Wave-V was detected in a time interval from 0 - 7 ms after the offset of the stimulation. The wave-V amplitude was calculated as the difference in amplitude between the maximum amplitude and the minimum amplitude found in the subsequent 2 ms.

## 5.5 Results

Fig. 5.10 shows the mean and one-standard deviation of wave-V amplitudes of the 6 conditions measured. The broad-band click and chirp used in this study are identical to the ones presented by Elberling et al. (2010). They found an averaged click evoked wave-V amplitude of 0.368 ţV and an averaged chirp evoked amplitude of 0.645 ţV. This compares well with the amplitudes measured in this study.

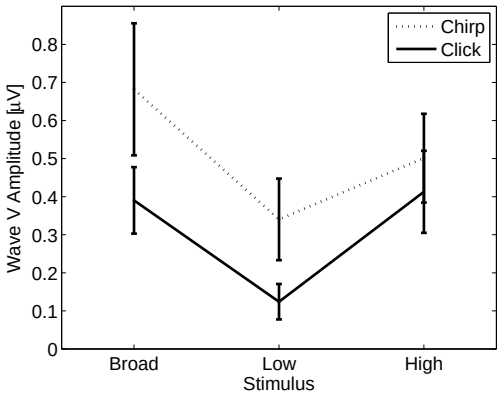



Figure 5.10: Mean ABR Wave-V amplitude and one standard deviation plotted for each stimulus condition.

The mean amplitudes indicate that the chirp stimuli generate larger ABR Wave-V amplitude compared to the click stimuli across all conditions. The high-frequency chirp condition is significantly different from both the broad-band chirp (High $\neq$ Broad: p value = 0.014) and the low-frequency chirp condition (High $\neq$ Low: p value = 0.005), indicating that both high and low frequencies are adding to the measured amplitude. It cannot be rejected that the high-frequency click gives rise to the same amplitude as the broad-band click (High $\neq$ Broad: p value = 0.614) indicating that the broad-band click is entirely determined by the high-frequency contribution. The p-values were calculated using a two-sample t-test.

The difference between the click evoked and chirp evoked wave-V amplitude was calculated for each test subject to reduce the influence of the inter-subject variability. The mean and standard deviation of the improvements from click to chirp are shown in Fig. 5.11. A t-test was applied to analyze the data (see Table 5.1). All three stimuli types show significantly larger amplitudes for chirps over clicks, supporting the hypothesis that the increased synchrony happens over the entire frequency range. It is also shown that the high-frequency improvement was significantly different from the broad-band improvement, and thus the high frequencies cannot be the entire explanation for the larger amplitude measured with a chirp instead of a click. It cannot

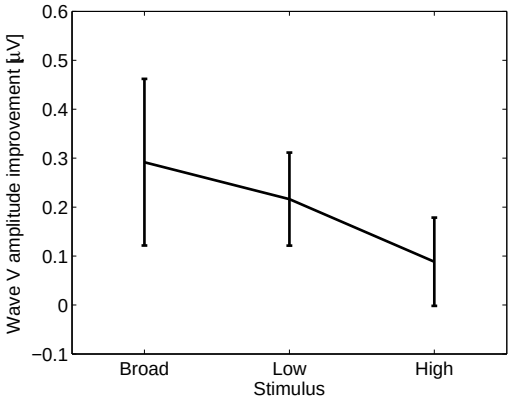



Figure 5.11: Improvement in wave-V amplitude from click to chirp evoked responses. The mean and one standard deviation are plotted.

| Hypothesis | P-value |
|---|---|
| Low > 0 | « 0.001 |
| High > 0 | 0.006 |
| Broad > 0 | « 0.001 |
| Low $\neq$ Broad | 0.237 |
| High $\neq$ Broad | 0.004 |

Table 5.1: Statistical analysis of data in Fig. 5.11. The three upper P-values are calculated using a one sided one-sample t-test. The two lower using a two-sample t-test.

be rejected that the improvement measured with the low-frequency stimuli are equal to the improvement of the broad-band conditions. These results will be further discussed in the discussion section.

## 5.6 Discussion

This study investigated the frequency regions contributing to the chirp ABR Wave-V amplitude. It was found that an increase in ABR wave-V amplitude when stimulating with a chirp stimulus rather than a click, was observed both at lower and higher frequencies, indicating that the increased synchrony of the nervous responses takes

place across the entire frequency range. It was also shown that the high-frequency region cannot explain the improvement from click to chirp when stimulating with the broad-band stimuli. However, the improvements observed at the low-frequency conditions and the broad-band conditions were not significantly different, indicating that the lower frequencies can explain all the improvement from the click to chirp condition. This contradiction in the results, that the high-frequency improvement is significantly larger than zero, and that the low-frequency improvement is not significantly different from the broad-band improvement, would likely be clarified if more test subjects had been used.

Fig. 5.10 shows that high frequencies were the main contributor to the formation of ABR Wave-V amplitudes for both clicks and chirps. This was likely due to the fact that the high-frequency stimuli contains more power, and to the fact that the high-frequency basilar membrane responses have short impulse responses that were inherently better aligned than the longer impulse responses at low frequencies. However, the improvement from click to chirp at high frequencies was small.

In Fig. 5.12 the amplitudes of the low-frequency and high-frequency responses were added for each test subject and compared to the broad-band evoked amplitudes. It is clearly observed that the summed amplitude is larger than the broad-band evoked amplitude. This shows that the auditory pathway behaves nonlinearly. The explanation is that the outer-hair-cells (OHC) amplifies weak sounds more than louder sounds (compression) and the fact that the filtered responses gives rise to spread of excitation on the basilar membrane in the region surrounding the 1500 Hz cut-off frequency. The 1500 Hz region would in the broad-band conditions have been masked. The low level "off-frequency" excitation will be amplified by the OHC and the summed response of the two frequency limited conditions will thus be stronger than the one measured with the broad-band stimulus. The increased amplitudes observed with the summed low and high responses, are though equally large for both click and chirp stimulus. This leads to a very limited effect on the wave-V improvements shown in Fig. 5.11, and the possible uncertainty regarding the unmasked off-frequency effects were thus negligible.

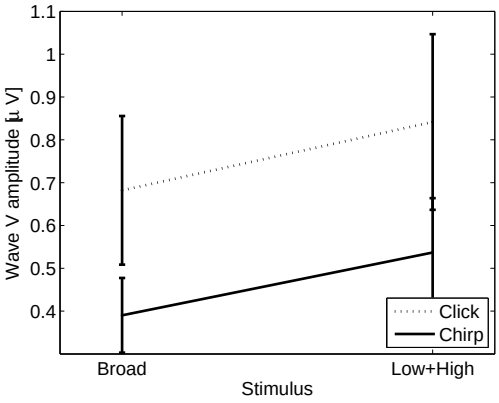



Figure 5.12: ABR Wave-V for filtered stimuli are added for each subject and compared to data for broadband. The mean and one standard deviation are shown.

## 5.7 Conclusion

This study examined the influence of frequency range on chirp evoked ABR at a presentation level of 40 dB HL. It was shown that both low and high frequencies contribute to the increase in wave-V when using a chirp stimulus instead of a click stimulus. This demonstrates that synchronization of nerve fibers occur across the entire frequency range. However, the largest increase in wave-V is observed at lower frequencies.

# 6

# Modeling human tone-burst and click-train evoked ABRs

This chapter is based on the paper called "Modeling human auditory evoked brainstem responses based on nonlinear cochlear processing" (Harte et al., 2010), and describes simulations of tone bursts and click-train evoked ABRs. The simulation of click-train evoked ABRs represents the first step, in this thesis, towards simulating responses to longer-duration stimuli. In the following two chapters the 40-Hz ASSR and speech-syllable evoked ABRs are simulated. Compared to the original publication, the description of the theoretical framework, which already was described in chapter 3, has been taken out from the method section to avoid repetition.

## 6.1 Abstract

The aim of this study was to accurately simulate auditory evoked potentials (AEPs) from various classical stimuli such as clicks and tones, often used in research and clinical diagnostics. In an approach similar to Dau (2003), a model was developed for the generation of auditory brainstem responses (ABR) to transient sounds and frequency following responses (FFR) to tones. The model includes important cochlear processing stages (Zilany and Bruce, 2007) such as BM tuning and compression, inner hair-cell (IHC) transduction, and IHC auditory-nerve (AN) synapse adaptation. To generate AEPs recorded at remote locations, a convolution was made of an elementary unit waveform (obtained empirically) with the instantaneous discharge rate function for the corresponding AN unit. AEPs to click-trains as well as to tone pulses at various frequencies were both modeled and recorded at different stimulation levels

and repetition rates. The observed nonlinearities in the recorded potential patterns with respect to ABR wave latencies and amplitudes could be largely accounted for by level-dependent BM processing as well as effects of short-term neural adaptation. The present study provides further evidence for the importance of cochlear tuning and AN adaptation on AEP patterns and provides a useful basis for the study of more complex stimuli including speech.

## 6.2 Introduction

For sounds which convey information, such as speech and music, much of the information is carried in the changes in the stimulus, rather than in the parts of the sound which are relatively stable. Through the last decades both psychoacoustic and physiological studies have investigated how the auditory system analyzes the temporal modulations of sounds. When various sounds are presented to human subjects, it is possible to record auditory evoked potentials (AEPs) on the surface of the human scalp. Auditory evoked potentials are the summed response from many remotely located neurons recorded via scalp electrodes. They can be recorded from all levels of the auditory pathway, from the auditory nerve, the brainstem up to the cortex. They are typically grouped in terms of time of occurrence after stimulus offset and thus are known as; auditory brainstem responses (ABRs) recorded between 1 and 7 ms after stimulus offset; middle latency responses (MLRs) recorded in the interval 15-50 ms after acoustic stimulus; and auditory late response (ALR) recorded in the interval 75-200 ms after stimulus.

Hearing deficiencies often lead to difficulties in understanding speech, especially in noisy and reverberant environments. Auditory evoked potentials are a powerful tool used to diagnose and assess classical hearing deficiencies. This has led to a trend in the literature of assessing and investigating speech and complex speech-like stimuli with AEPs (e.g. Aiken and Picton, 2008; Akhoun et al., 2008; Chandrasekaran and Kraus, 2010; Lalor and Foxe, 2010). AEPs are relatively well understood for basic stimuli, i.e. transients, tone bursts and tones. However, for more complex stimuli, which include amplitude and frequency modulations as well as sharp on-set and off-set

transients, it is still relatively poorly understood how the various neurophysiological processing along the auditory pathway gives rise to the AEP recorded at surface electrodes. A clearer understanding of how the underlying neurophysiology in the auditory system leads to surface-recorded scalp potentials could help to assess hearing impairment, or to evaluate how well this has been compensated for with an auditory prosthesis (Aiken and Picton, 2008), such as a hearing aid or cochlear implant.

The long-term goal of this study is to model and simulate speech evoked and complex (non-speech) sound evoked AEPs originating in the auditory nerve and brainstem, based on current knowledge of neural auditory signal processing. Dau (2003) developed a model for the generation of early AEPs, including auditory brainstem responses (ABR) to transient sounds like clicks and frequency following responses (FFR) to tones. Both of these AEPs are generated by neurons in the auditory nerve (AN) and subsequent stages along the auditory brainstem. The model included important cochlear processing stages such as basilar-membrane filtering with a compressive feedback loop, inner hair-cell (IHC) transduction, and IHC-AN synapse adaptation. The instantaneous AN discharge rate from the model was convolved with an empirically obtained elementary unit waveform, to simulate AEPs.

In the present paper, the Dau (2003) model is extended to include current advances in AN modeling Zilany and Bruce (2007) and is humanized. The original Dau (2003) model used the Heinz et al. (2001) AN model fitted to experimental cat AN data. Here, the Zilany and Bruce (2007) AN model will be adapted for humans by ensuring that the model has appropriate thresholds, tuning curves, BM traveling wave latencies etc., based on current state-of-the-art knowledge derived from both behavioral and objective measures where possible. This study will present a comparison of the model output with basic transient, tone-burst and click-train data, in an attempt to build up stimulus complexity towards the final goal of speech. Thus it is possible to challenge the model with relatively basic stimuli, before increasing complexity. This study focuses on the role of basilar membrane tuning and the adaptation mechanism of the AN model and looks at the consequences for AEPs generated. Neural adaptation is the phenomenon where the neural output is reduced due to prolonged or repeated stimulation, in each stage of the auditory pathway.

The role of adaptation in AEPs, and more specifically ABRs is important because in clinical practice it is highly desirable to obtain accurate recordings of ABRs quickly, particularly from uncooperative subjects and neonates. Any morphological differences, such as amplitude and latency, from normative data caused by stimulus rate adaptation could interfere with diagnosis. The desire for quicker acquisition time has led to the use of rapid rates of stimulation via so-called pseudo-random binary sequences or maximum length sequences (e.g. Burkard et al., 1990; Jewett et al., 2004). The response to these pseudo-random pulse trains needs to be deconvolved to obtain an estimate of the ABR. The higher rate of the sequence leads to typically smaller ABR amplitudes. This is believed to be a result of neural adaptation.

## 6.3 Methods

### 6.3.1 Model for AEP generation

The structure of the ABR model is shown in Fig. 6.1. Within the overall ABR model, a parallel bank of AN fibers is individually modeled. Each AN fiber is tuned to a specific CF. The number of fibers included is a trade off between computational time and model precision. Throughout this study 500 fibers were used for each simulation, representing a range of 0.1 to 10kHz. The output of the AN model, the instantaneous firing rate of all the AN fibers, is summed and convolved with the unitary response function.

#### A humanized AN model

Zilany and Bruce (2006, 2007)'s AN model was fitted to cat AN data, and has thus been modified to better model human AN response here. The following changes to the original cat AN model were implemented by Bruce and co-workers:

The original cat middle-ear transfer function has been replaced by a human middle ear. This was based on the linear circuit model of Pascal et al. (1998) of human cadavers. The model magnitude response function is shown in Fig. 6.2.

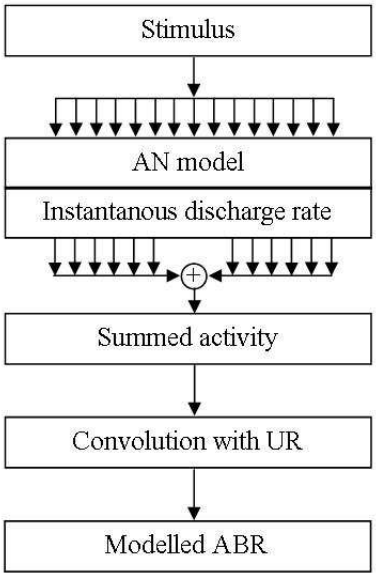



Figure 6.1: Structure of the ABR model. 500 AN fibers tuned to different CFs are individually modeled by the AN model. The summed instantaneous firing rate is then convolved with a unitary response to create the modeled ABR.

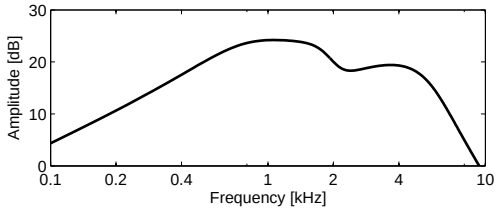



Figure 6.2: Frequency response of the human middle ear implemented in the AN model.

It has been argued that humans have significantly sharper BM mechanical tuning than cats and other experimental animals (Shera et al., 2002). To incorporate this, the model equivalent rectangular bandwidth quality factor, $Q_{\mathrm{ERB}}$, for cochlear tuning was defined to be,

$$Q_{\mathrm{ERB}} = 12.7 \left( \frac{f_c}{1000} \right)^{0.3} \tag{6.1}$$

where $f_c$ is the center frequency of the BM filter. This function was taken from Shera et al. (2002) and is applicable to humans at frequencies at and above 1 kHz. The choice of QERB will be further discussed later.

The tip of a suppression tuning curve is at a slightly higher frequency than the tip of the excitatory tuning curve (Delgutte, 1990). This is implemented in the original Zilany and Bruce (2007) model by shifting the CF of the so-called control path filter by 1.2 mm on the BM. Without sound knowledge of how this mechanism works in humans, the default is retained here. However, a human frequency-place mapping for the BM is needed and has been updated from the original to the human fit from Greenwood (1990):

$$f_c = A(10^{ax} - k) \tag{6.2}$$

where $x$ is the distance on the BM apex in mm, and the constants are; $A$ = 165.4, $a$ = 0.06 and $k$ = 1.

Two additional changes to the Zilany and Bruce (2007) model was made here. In Zilany and Bruce (2007), the synapse gain, which describes the relationship of the inner hair cell potential to the synaptic release rate, varies as a function of CF to ensure that the model thresholds match empirical data from cats. Without such physiological data available, human behavioral monaural absolute thresholds (Killion, 1978) were used to fit the model. Thus, the synapse gain function from Zilany and Bruce (2007) was changed to be;

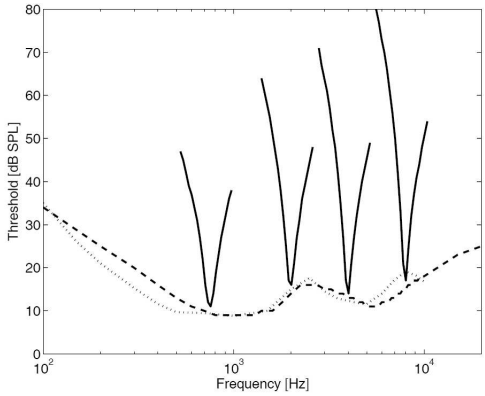



Figure 6.3: Model example tuning curves (solid curves) for representative CFs and simulated (dashed curve) and reference (dotted curve) absolute thresholds.

$$K_{CF} = 0.91 \cdot \min\{4000, 10^{0.1 f_c/10^3 + 0.4}\} \tag{6.3}$$

where the characteristic frequency, $f_c$, is in units of hertz.

Figure 6.3 (solid curves) shows example tuning curves of AN fibers across a range of CFs for the revised AN model. The same procedure from Zilany and Bruce (2007) and Chintanpalli and Heinz (2007) was used to adaptively determine the tuning curves. Absolute thresholds are also shown on the figure as the lower dashed line, as well as the reference behavioral thresholds (dotted curve) from Killion (1978).

Fig. 6.4 shows the $Q_{ERB}$ versus CF measured from the $Q_10$ from the model tuning curves, via the transformation from Ibrahim and Bruce (2010):

$$Q_{ERB} = \frac{Q_{10} - 0.2085}{0.505} \tag{6.4}$$

Also shown in Fig. 6.4 are the $Q_ERB$ from Shera et al. (2002) used to set the BM tuning in the model.

As described above and shown in Fig. 6.3 and 6.4, the AN model tuning properties

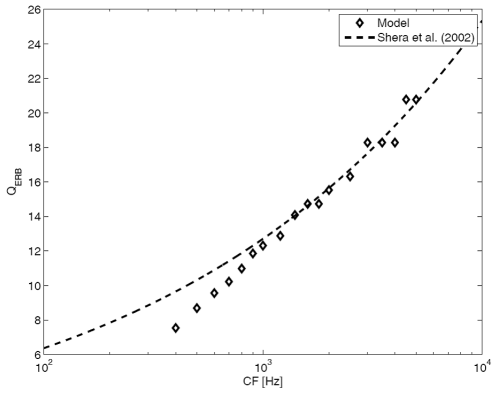



Figure 6.4: $Q_{ERB}$ values vs CF, measured from the model tuning curves and reference from Shera et al. (2002).

are determined by the frequency dependent $Q_{ERB}$ in Eq. 6.1. However, an additional delay function exists in the primary C1 filter path of the AN model. This acts as a so-called signal-front delay (see Ruggero and Temchin, 2007). This has been altered in the present model, to ensure that the model produces overall delays (signal front and traveling wave group delays) similar to the estimated BM delay reported in Shera et al. (2002). To achieve this, each AN impulse response function was determined, the envelope was extracted (via low pass filtered Hilbert envelope), and the latency of the peak of the enveloped recorded. The following logarithmic function was then fitted to the difference between the model output latencies and those reported in Shera et al. (2002):

$$\tau_{CF} = 10^{-3} \cdot max\{0, -10.09 \cdot log_{10}(f_c) + 29.23\} \tag{6.5}$$

By using this additional delay, it is hypothesized that physiologically plausible BM latencies can be approximated in the model. This is vital as it is well known that cochlear processing and delay has a strong influence on recorded brainstem evoked potentials (Dau, 2003; Dau et al., 2000; Wegner and Dau, 2002).

**The unitary response**

The unitary response describes the transformation of the output of the auditory nerve to the potential measured at electrodes placed on the scalp. The unitary response, like in Dau (2003), was obtained by deconvolving an experimentally recorded click ABR with the summed neural activity pattern for the click, generated by the AN model. The deconvolution is an ill posed mathematical problem and has an infinite number of solutions. A stable and probable solution was found by using Tikhonov regularization (Tikhonov, 1963). The calculations were carried out in Matlab using a toolbox provided by Hansen (1998).

### 6.3.2 Tone-burst simulation

Auditory evoked potentials have been used historically to obtain indirect estimates of cochlear delay in humans. Tone-burst evoked ABRs have been studied extensively in the literature as a means of estimating BM delay (Gorga et al., 1988; Neely et al., 1988; Harte et al., 2009). Thus, this was a logical choice of basic stimuli to test if the AN model in the present study adequately modeled cochlear delay. In order to test if the BM delay introduced within the present model is reasonable, a simulation was run using Hanning windowed tone bursts as stimuli, with CFs and durations given in table 1. Levels of 40 to 100 dBpe SPL were used, in 10 dB steps. The choice of stimuli was inspired by the experiments from Norton and Neely (1987) and Serbetcioglu and Parker (1999). The tone-burst durations represent a trade-off between having an equal number of cycles for all frequencies and a relative narrow spread in their spectrum. The organization of frequency along the cochlear partition is roughly logarithmic and tone bursts with a fixed number of cycles result in uniform energy splatter in log-frequency. The stimulus rise time is responsible for the simultaneous neural activation leading to the brainstem responses (Suzuki and Horiuchi, 1981) and to obtain a detectable ABR response. A sharp stimulus onset (i.e., a short rise time) produces a large amount of synchronized neural activity, but also decreases the frequency specificity of the stimulus. Rise times for frequencies of 2 kHz and above include approximately 5 cycles and therefore ranged

| Frequency | Total Length | |
|---|---|---|
| kHz | ms | cycles |
| 0.5 | 10 | 5 |
| 0.75 | 7 | 5.25 |
| 1 | 5 | 5 |
| 1.5 | 5 | 7.5 |
| 2 | 5 | 10 |
| 3 | 3.4 | 10.2 |
| 4 | 2.5 | 10 |
| 6 | 1.7 | 10.2 |
| 8 | 1.25 | 10 |

Table 6.1: Tone burst stimuli used, with length in ms and number of cycles.

from 2.5 to 1.25 ms. Below 2 kHz it was felt that the reduced energy spread, by keeping a fixed number of cycles, would make it almost impossible to record a wave-V response. Therefore, a compromise was struck, similar to Gorga et al. (1988), between the need for rapid stimulus onsets and reduced energy spread in the choice of rise time. The number of cycles in the rise time were reduced to 3.25 at 1.5 kHz and approximately 2.5 for 1.0 kHz.

ABR wave V is the wave with the largest amplitude and hence the most easily detectable. In the simulation, the ABRs for the tone burst stimuli were generated and the wave V latency calculated and plotted against Neely et al. (1988)'s empirically determined model of latency derived from tone burst simulations:

$$\tau_{waveV} = a + bc^{-i}\left(\frac{f_c}{1000}\right)^{-d} \tag{6.6}$$

where $i$ is the tone-burst intensity (divided by 100), $f_c$ is the tone burst center frequency in Hertz, and $a$ = 5 ms, $b$ = 12.9 ms, c = 5:0 and d = 0:413 were fitted constants to Neely et al. (1988)'s data.

### 6.3.3 Experimental methods

A total of four normal hearing test subjects (four female) participated in the experimental part of this study, and were aged between 22-26 years. The experiments were conducted in an electrically and acoustically shielded audiometric booth (IEC 268-13). The basic stimulus used in this experiment was a 5 sample duration impulse played at 44.1 kHz. Five sets of stimuli conditions were presented at a constant inter-epoch rate of $\approx$ 8 Hz (i.e. a duration of 125 ms). The first stimuli set was a single impulse to evoke s standard ABR used to empirically determine the unitary response functions. The remaining sets were trains of impulses with a within-train rate of 40, 80, 190 and 250 Hz. A total of 4000 averages were made per stimulus type and repeated twice (three times for the single impulse condition) to ensure repeatability of results. The stimuli were all presented at a level of 80 dB pe SPL, to ensure reasonable SNR and test subject comfort. The stimuli were generated in MATLAB and A/D conversion made through an RME ADI-8 Pro 24-bit sound card. The levels were set via a TDT PA5 programmable attenuator. The stimuli were presented to the left ear of the test subject via an ER-2 insert earphone. EEG activity was recorded differentially between the vertex and ipsi-lateral mastoid, with the ground electrode placed on the forehead. Silver/silver chloride electrodes were used, and an inter-electrode impedance was maintained below 5kW. EEG activity was recorded on a SynAmps2 amplifier at a sampling rate of 10000 Hz, and band-pass filtered between 0.05 and 2000 Hz. After recording, the EEG-data were epoched and filtered again from 100 to 1500 Hz using a 200 tap FIR filter with zero phase delay. The epochs were averaged using an iterative weighted-averaging algorithm (Riedel et al., 2001).

## 6.4 Results

### 6.4.1 Auditory brainstem response and unitary response

Single transient evoked potentials were averaged across 12000 epochs (all 3 runs) for subject ML and are shown by the dotted curve in Fig. 6.5. The recorded ABR shows

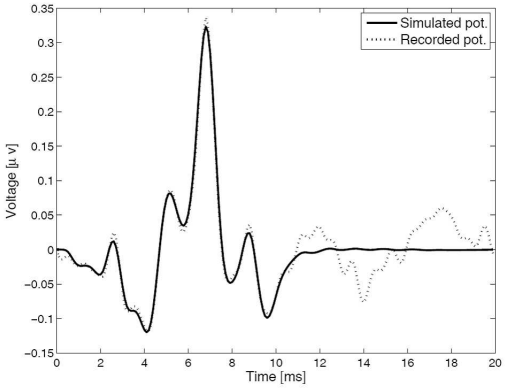



Figure 6.5: Recorded (dotted line) and simulated (solid line) auditory brainstem response to single transient stimuli.

the typical pattern with clear waves I, III, and V at latencies that are consistent with the literature. The wave V peak is the largest occurring at $\approx 6.5$ ms.

Figure 6.6 shows the calculated unitary response obtained from a deconvolution of the recorded potential with the AN model. The unitary response function obtained in the present study is similar to and consistent with Dau (2003). There is significant subject dependence of the unitary response, but the essential morphology remains the same. The interested reader is referred to Dau (2003) for a detailed discussion of the form of the unitary response and comparisons with previous studies.

The simulated AEP obtained from the convolution of the AN model output with the unitary response is indicated by the solid curve in Fig. 6.5. There is a very good agreement between the recorded and the simulated potentials, over the length of the unitary response calculated (10 ms). The unitary response was not calculated for longer durations as this would have included evoked potential contributions higher than the brainstem, which are not of interest in the present study. In the present study, linear superposition is assumed above the level of the AN, and thus the calculated unitary response function given in Fig. 6.6 was used for any input stimulus at any level.

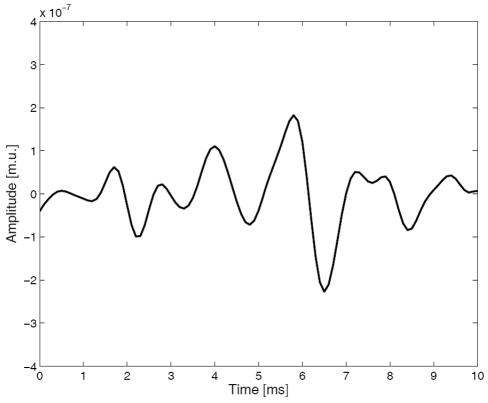



Figure 6.6: Unitary response function, calculated via deconvolving the recorded potential with the output of the AN model.

### 6.4.2 Tone-burst simulation

Figure 6.7 shows the wave V latencies for the ABR model simulations to tone-burst stimuli, with center frequencies from 1 to 8 kHz and excitation levels 40 to 100 dB pe SPL in 10 dB steps. Also shown are dotted lines representing the empirically fitted latency model of Neely et al. (1988) given in Eq. 6.6. Both the simulated ABR and modeled latencies show exponentially decreasing delays as a function of frequency. At the lowest levels of excitation, the simulated ABR latencies have a slope similar to that seen in Neely et al. (1988)'s modeled latencies. This is logical as the AN model tuning and delay was based on Shera et al. (2002)'s stimulus frequency otoacoustic emission delay estimates, made at 40 dB SPL. Further, as excitation levels increase the simulated ABR rate of change of latency with frequency decreases. The overall spread of simulated ABR latencies with level is reasonable at lower frequencies(1-2 kHz), but seems compressed at higher frequencies relative to Neely et al. (1988)'s results.

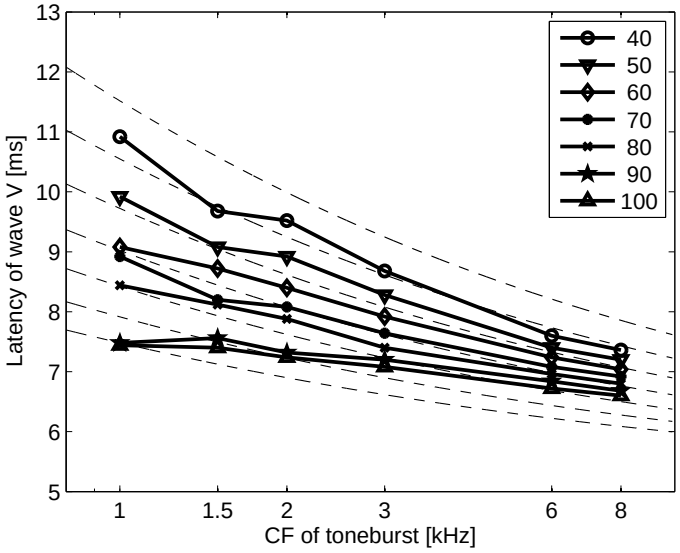



Figure 6.7: Simulated (solid curves) and modeled (dashed curves, based on Eq. 6.6) ABR wave V latencies as a function of tone-burst center frequency and level.

### 6.4.3 Click-train ABR

Figure 6.8 shows the recorded (dot-dashed curve) and simulated (solid curve) ABR to a single click and click-train stimuli with within-train rates of 40, 80, 190 and 250 Hz for one illustrative subject. The noise floor for the recorded ABR is shown by the vertical bar near 0 ms on each trace. The vertical line to the right of the single click ABR indicates the scale on the figure.

As the within-train rate increases the smaller waves that make up the click ABR (waves I, II, III and IV ) become more difficult to distinguish and only the wave V seems to be visible. As the within-train rate increases, the peak amplitudes of the wave V decrease for rates higher than 80 Hz. The first peaks are typically the largest, and these then decrease as rates increase. The modeled ABR seems to accurately predict the recorded ABR at moderate within-train rates of 40 Hz. Wave V amplitude seems unchanged within trains and latencies seem well modeled. As the within-train

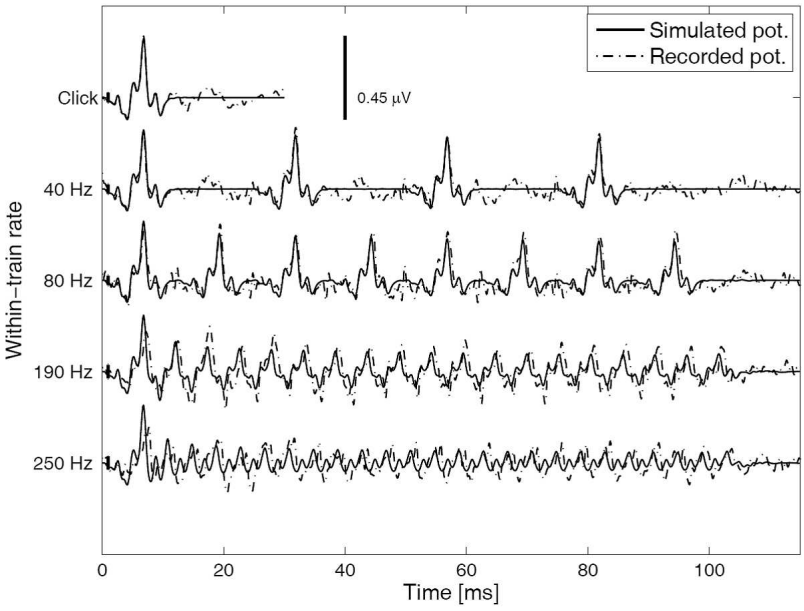



Figure 6.8: Recorded (dot-dashed line) and simulated (solid line) auditory evoked brainstem potentials to click-train stimuli at 40, 80, 190 and 250 Hz within-train rates.

rate increases, the modeled ABR amplitude seems to decrease faster than the recorded ABR. In addition, the timing of the peaks of modeled ABR are faster for higher rates than for the recorded potentials at the same rate. For the highest rate stimuli, the simulated ABR wave V peaks drop in magnitude seemingly exponentially for successive stimuli. The recorded ABR on the other hand tends to have a sharp initial drop in magnitude and does not demonstrate such an exponential decrease. Similar trends were observed for all of the subjects tested, though the magnitudes and timing of the responses demonstrated some subject-dependent variability.

## 6.5 Discussion

### 6.5.1 Frequency-dependent delay

The intrinsic relationship between frequency and travel time in the cochlea is fairly well represented by the AN and the ABR model. Gorga et al. (1988), in the original study on tone-burst evoked ABR wave V latency, did not specify the earphones they used to present the stimuli nor the coupler used to calibrate them. Therefore there is some ambiguity as to the exact levels used by Neely et al. (1988) to model these, and reproduced here in Eq. 6.6. With that in mind, one could not expect an exact fit of the present simulated ABR wave V latencies with those modeled by Eq. 6.6. The range of latencies across level and frequency, should be covered however. As mentioned earlier, the simulated ABR latencies at higher frequencies seem compressed relative to those seen within the literature. This could be an indication that the level-dependent bandwidth is not well implemented in the AN model.

At low excitation levels, the simulated ABR wave V latencies accurately reproduces the latencies across frequency seen in the literature. The frequency dependent delay in the AN model used here arose due to the cochlear tuning, $Q_{ERB}$, incorporated. This was given in 6.1 and the additional delay in 6.5. There is some contention in the literature about accurate estimates of $Q_{ERB}$ in humans (Bentsen et al., 2011). In the present study, $Q_{ERB}$ estimates from Shera et al. (2002) were used. These $Q_{ERB}$ values were obtained by averaging objective (based on stimulus frequency otoacoustic emission group delay) and behavioral (forward masking) estimates. In these $Q_{ERB}$ values, as seen in Fig. 6.4, the auditory filters are very sharp and become effectively sharper as frequency increases. Alternative estimates of $Q_{ERB}$ suggest much broader tuning, and a near frequency independence. These estimates come from objective stimulus frequency otoacoustic emission iso-suppression tuning curves (Keefe et al., 2008), and behavioral simultaneous masking (Glasberg and Moore, 1990).

Ruggero and Temchin (2007) offered an alternative novel estimate of in vivo cochlear delay in humans using post-mortem delay estimates with the post-mortem effects compensated for via comparison with experimental animal data. Bentsen et al.

(2011) showed that Ruggero and Temchin (2007)'s cochlear delay estimates led to $Q_{ERB}$ estimates similar to those obtained with simultaneous masking and stimulus frequency otoacoustic emission iso-suppression tuning curves. If $Q_{ERB}$ were much smaller than those used in the present model (where Ruggero and Temchin (2007)'s were approximately 2.5 times shorter than Shera et al. (2002)), then the latency estimates of the modeled wave-V's seen in Fig. 6.7 would be much shorter. Thus a greater degree of disparity would be seen between the modeled and historically reported latencies. This provides some indirect evidence to support Shera et al. (2002)'s estimates of $Q_{ERB}$.

An alternative source of error lies with the unitary response function. In the present ABR model, the only frequency dependent delay is due to the BM filtering in the AN model. It is implicitly assumed that linear-superposition holds at higher stages in the model, with the frequency- and level-independent unitary response function. If the unitary response function were to be strongly frequency- or level-dependent, then the wave V latencies simulated in Fig. 6.7 would be significantly altered. However, there is good physiological evidence to suggest this is not the case. Wave-V latency is often considered to be composed of the sum of the synaptic delay, synaptic, the neural delay, $\tau_{neural}$, as well as the cochlear delay $\tau_{BM}$ (Neely et al., 1988). The synaptic delay is the time between the inner hair cells activity and the auditory-nerve fibers firing. It is typically around 1 ms (Burkard and Secor, 2002; Kiang, 1975; Kim and Molnar, 1979; Mø ller and Jannetta, 1983) and frequency- and level-independent (Don et al., 1998). The neural conduction time (neural delay) is the time between the auditory-nerve activity and the place generating the ABR wave. Synaptic delay and cochlear delay are both included in the AN model. However, the neural conduction time is not, and is implicitly in the unitary response function. There is no historical neurophysiological evidence to suggest that the neural conduction time is frequency dependent (Don and Eggermont, 1978; Don and Kwong, 2002; Eggermont and Don, 1980). However, it would still be prudent to investigate both the frequency and level dependence of the unitary response function in future studies.

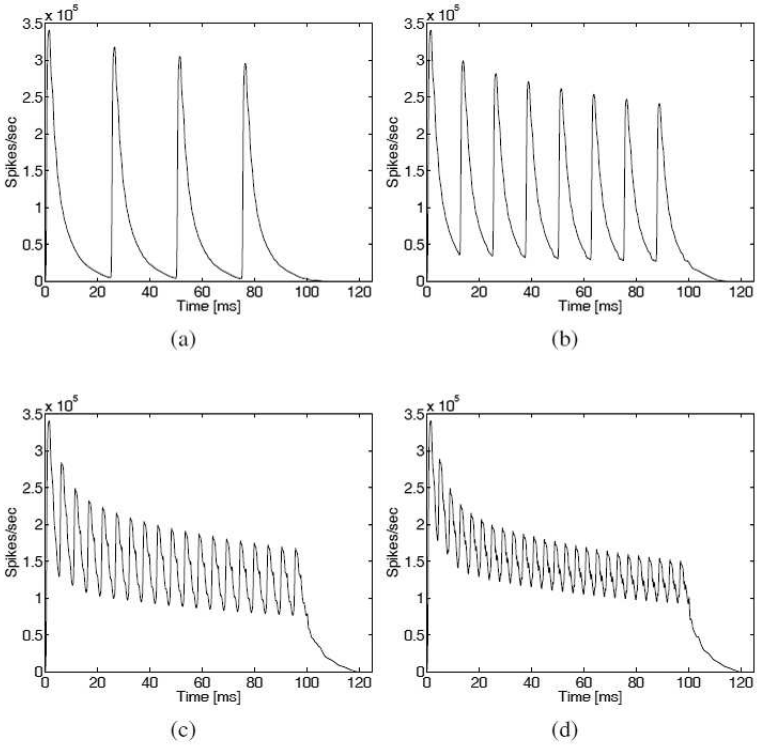



Figure 6.9: Summed auditory nerve model output for within-click train rates of (a) 40 Hz, (b) 80 Hz, (c) 190 Hz and (d) 250 Hz.

### 6.5.2 Click-train ABR and neural adaptation

The simulated ABR were successful at modeling the recorded ABR for within-train rates of 40 Hz, as seen in Fig. 6.8. At these relatively slow rates, little or no neural adaptation was expected. Figure 6.9a shows the output of the summed AN model in the present study, for the 40 Hz within-train rate stimuli.

The model output clearly reverts to baseline (50 spikes/s, AN spontaneous rate) after each click, and the peak of the response for each new stimulus click within the train does not decrease significantly.Thus the stimuli do not interfere with each other

within the AN model. As the within-train rate increases, the ABR wave V tends to dominate the response due to the convolution of smaller peaks and the reduction in amplitude of the spikes in the summed AN model output, as seen in Fig. 6.9. For the higher-rate stimuli the summed AN model output never returns to baseline, and the peak magnitudes reduce. The model does not return to baseline due to the ringing of the filters in the AN model. The reduction in the peak spike rates is linked with adaptation and appears to follow an exponential decrease with each new click. Zilany and Bruce (2007)'s rate adaptation at the synapse between IHC and AN fibers was a purely exponential model, albeit with multiple short and long time constants. Zilany et al. (2009) have suggested a new rate adaptation model incorporating both exponential and power-law dynamics. Incorporating this model revision into the present model might help to improve the under-predicted wave V amplitudes at high rates. This will be investigated in future versions of the ABR model.

### 6.5.3 Outlook

It was stated in the introduction that the role of neural adaptation in AEP recording was important to understand, due to the clinical use of high rate stimuli. In addition to this, there is a trend in AEP studies to use steady state signals, where neural adaptation will play an even greater role. Auditory steady state responses (ASSR) are typically responses to carrier signals with amplitude modulation (AM) imposed on them at different rates. Such ASSRs give excellent frequency specificity as the response will mainly contain energy at the AM from a narrow band of AN fibers at the carrier frequency (John and Picton, 2000). This is obviously an advantage clinically to test auditory function at specific frequencies. Invasive animal studies and magnetoencephalographic (MEG) source analysis studies in humans have shown that the ASSR is generated in different brain regions, depending on the modulation frequency of the stimulus (Kuwada et al., 2002; Schoonhoven et al., 2003). For low rates of AM, around 40 Hz, a number of studies have demonstrated that the ASSR can be predicted from the convolution of single middle-latency and brainstem transient responses with a click train with the appropriate repetition rate (Galambos et al., 1981; Picton et al., 1987; Hari et al., 1989; Gutschalk et al., 1999;

Bohorquez and Oezdamar, 2008). This is further supported by the finding in the present study, that little or no interaction occurs in the AN model for the different clicks in the 40 Hz click train, as seen in Fig. 6.9a. For modulation rates above 80 Hz, ASSRs are typically argued to be generated by neurons in the brainstem that both respond to transient stimuli and are locked to the envelopes of AM tones (John and Picton, 2000; Kuwada et al., 2002; Sininger and Cone-Wesson, 2002). The different within-train rates were chosen in the present study to span the AM rates investigated in the literature. The present study has the potential to help understand the brainstem contribution to ASSRs. This is an advantage as sources due to the brainstem are hard to investigate using classic dipole source modeling (Scherg, 1990), due to the brainstem sources depth and small signal strength.

# 7

# Investigating the potential of auditory steady-state responses to assess loss of cochlear compression

This chapter is based upon the submitted paper Rønne et al. (2012a). It is mainly a study of the possibility of using ASSRs to assess cochlear compression in humans. This is investigated using both a simple analytical model, experimental work with human subjects and simulations using an extended version of the ABR model, called the ASSR model. It can thus be read as an independent study, or as another step in the development and evaluation of the modeling work in this thesis.

## 7.1 abstract

In this study, it is investigated whether the auditory steady state response (ASSR) can be used as a tool to estimate human cochlear compression. First, a simplified analytical model is presented, for amplitude modulated tones passing through a static nonlinear system. The approximate closed-form solution derived from this analysis is used to construct two hypotheses for ASSR level growth as input level and modulation depth are varied. Two experiments are then presented measuring ASSR modulation and level-growth functions in human subjects. Finally, a more complex nonlinear numerical model for ASSR generation is presented. This second model is capable of accurately simulating the complex processing carried out in the auditory periphery, and is used here to evaluate the assumptions of the simple static model and to interpret the experimental ASSR findings. The study demonstrates that both the level- and

modulation growth functions can be used to measure cochlear compression. However, the clear recommendation is to measure level-growth functions due to their larger accuracy and efficiency. A secondary finding, based on the experimental modulation-growth function, is the indication of an effective compression, seemingly independent of cochlear compression. This second compressive mechanism remains unexplained by both the analytical and the numerical ASSR model.

## 7.2 Introduction

The human auditory system is able to perceive root-mean-square (RMS) fluctuations in air pressure from as low as 20 $\mu$Pa, corresponding to a dynamic range of about $10^6$ or 120 dB. To achieve this, the local mechanical vibration of the basilar membrane (BM) in the cochlea, excited at its natural frequency, grows in a nonlinear or compressive fashion with increasing sound pressure level (Ruggero, 1992; Harte et al., 2005). A number of studies and reviews (Sellick et al., 1982; Nuttall and Dolan, 1996; Rhode and Recio, 2000; Robles and Ruggero, 2001) have investigated and reported BM input-output level curves in experimental animals, where the 120 dB input dynamic range is mapped to 30-40 dB output range usable for neural encoding. The compressive input-output level curve for humans is estimated to have linear growth at excitation levels below sound pressure levels (SPL) around 40 dB, i.e. a 10 dB increase in input leads to a 10 dB increase in output. Between 40 and 90 dB SPL, sharp compression is observed, where an increase in 10 dB only leads to about 3 dB increase in the response. Above approximately 90 dB, the level curve tends to become linear again. This is often explained in terms of the active mechanism within the cochlea supplying significant amplification at low excitation levels and saturating at mid-levels. At high levels, this mechanism becomes exhausted and is unable to further contribute to the BM response. This compressive behavior will be termed cochlear compression throughout this study. The local input-output compressive non linearity depends on the integrity of the outer hair cells (OHC) (Ruggero, 1992; Robles and Ruggero, 2001). Damage to OHCs, common in many forms of sensori-neural hearing losses, reduces or completely eliminates the active amplification of

low-level sounds, leading to a linearized input-output level curve. Sensory-neural hearing loss thus often leads to loss of cochlear compression. It is desirable to have an objective physiological metric sensitive to cochlear compression and its loss. Such a measure will necessarily be sensitive to local BM vibration and could be used as a further objective audiometric tool for neonates or uncooperative subjects, where subjective methods are challenging. The auditory steady-state response (ASSR), being a robust objective measure already used clinically for other purposes, could be an interesting and suitable choice for such a metric.

When transient sounds are presented to human subjects, the summed response from many remotely located neurons can be recorded via scalp (non-invasive) electrodes. These auditory evoked potentials (AEPs) can be recorded from all levels of the auditory pathway, from the auditory nerve (compound action potential, CAP); the brainstem (auditory brainstem response, ABR); up to the cortex (cortical auditory evoked potential, CAEP). These classical AEPs are obtained by presenting transient stimuli at slow repetition rates. At more rapid rates, the responses to each stimulus overlap with those evoked by the preceding stimulus to form a steady-state response (Picton et al., 1987). Typically, such auditory steady-state responses (ASSR) are evoked by sinusoidally amplitude modulated (AM) tones (Kuwada et al., 1986; Rees et al., 1986; Picton et al., 1987), and are argued to give excellent frequency specificity as the stimulus only contains energy at the carrier frequency and the side-bands due to the modulation (John and Picton, 2000). The ASSR is therefore typically analyzed in the frequency domain, where the amplitude of the Fourier component at the AM rate is used as the ASSR response magnitude. AM rates of around 40 Hz have been shown to produce the largest ASSR response amplitude (Kuwada et al., 2002, 1986). Although the ASSR has been heavily studied, the effect of cochlear compression on the ASSR is still unclear.

The amplitude of a recorded ASSR is necessarily dependent on cochlear mechanical processing and reflects the variation in level of the amplitude modulated sinusoid used to elicit it. Thus, one might expect to see cochlear compression reflected in the ASSR amplitude, as either the depth of amplitude modulation or the stimulus level is systematically varied. However, the recorded ASSR is a compound potential arising

from the summation of many neural fibers along the auditory pathway. Intuitively, fibers along the tonotopic axis tuned close to the carrier frequency will typically have the largest contribution, and thus one might expect to see evidence of local cochlear compression.

Only a few studies have examined ASSR magnitude as a function of modulation depth for sinusoidally amplitude modulated stimuli (Kuwada et al., 1986; Rees et al., 1986; Picton et al., 1987; Boettcher et al., 2001). The ASSR modulation-growth functions (defined here as the log. ASSR response magnitude plotted as a function of the log. modulation depth) generally seems to grow in a slightly compressive fashion. Typical slopes vary between 0.5 and 0.8 dB/dB as modulation depth is varied. This might support the assertion that they reflect local cochlear compression. However, the degree of compression estimated is significantly less than expected, of the order of 0.2 to 0.3 dB/dB as seen in other physiological estimates of compression (e.g. Prieve et al., 1996; Ruggero et al., 1997; Moore et al., 1999). Unfortunately, there is also significant variation in absolute amplitudes across the historical studies, probably due to variations in electrode placement, excitation level used and the limited number of test subjects used. It is not clear whether the limited compression seen in the ASSR modulation-growth function truly reflects cochlear compression, or some other property of the ASSR generation mechanism. Cochlear compression could also be estimated using ASSR by varying the stimulus level. Typical slopes of compression were historically reported to be $\approx 0.2$ dB/dB (Kuwada et al., 1986; Picton et al., 1987). These slopes are similar to those observed when measuring cochlear compression psychoacoustically, with oto-acoustic emissions (OAE) or in vivo in animals (e.g. Prieve et al., 1996; Ruggero et al., 1997; Moore et al., 1999). It is, however, difficult to establish whether the ASSR level and modulation-growth functions reflect cochlear compression, and not effective compression applied at higher, retro-cochlear, stages of the auditory pathway.

This study develops two models to investigate the role of cochlear compression on ASSR generation. The first, provided in section 7.3, is a highly simplified analytical model, used to explain how amplitude modulated stimuli are processed through simple static nonlinear systems. This is used to derive experimentally testable

predictions on the nature of modulation-growth and level-growth functions. A second, more physiologically plausible, nonlinear numerical model is also developed (section 7.5) by extending an existing model of ABR generation (Dau, 2003; Harte et al., 2010; Rønne et al., 2012) to be able to account for the ASSR. Two experiments, using normal-hearing test subjects, were carried out and reported here (section 7.4), measuring ASSR magnitude growth functions as modulation depth and level are varied. The numerical ASSR model and the simple analytical model make reasonable predictions of the experimental results and are used to argue that local cochlear compression can indeed be estimated using both ASSR modulation-growth and level-growth functions. However, care should be taken with modulation-growth estimates as they are more prone to experimental uncertainty, and it is recommended that level-growth functions be employed in future studies.

## 7.3 Analytical model for AM tones passing through a static nonlinear system

The physiology underlying the generation of auditory steady state responses is complex. A sinusoidal amplitude modulated tone consists of a carrier with two side tones, whose equal frequency separation from the carrier equals the modulation frequency. Cochlear mechanical processing spatially filters the stimulus to yield a place-specific excitation pattern. This will necessarily be subject to cochlear compressive non linearity. The inner hair-cells (IHC) in the cochlea are responsible for mechanoelectrical transduction, and act like a half-wave rectifier and a low-pass filter (Russell and Sellick, 1978). This processing extracts the envelope for stimuli with a high enough carrier frequency, $f_c \gtrsim 1.5-2$ kHz (Palmer and Russell, 1986). Thus, the nonlinearities in the peripheral processing and mechanoelectrical transduction process effectively ensure that AN fibers firing patterns reflect the compressed envelope of an amplitude modulated stimulus. Additionally, the transmission of neural spike trains from the brainstem to surface potentials acts like another low-pass filter stage, effectively ensuring that only the compressed envelope can be recorded.

To illustrate how AM signals are represented after such processing, a simple analytical model is presented. It is explored what happens to a sinusoidally amplitude modulated tone when it is passed through a static compressive nonlinear system. Specifically, an approximate closed-form solution is derived for the amplitude of the first harmonic of the AM frequency, after passing through the nonlinear system. It is argued that this could reflect the experimentally recorded ASSR, and yield testable hypotheses for the experimental part of the study.

The basic stimuli used in the present study are sinusoidally amplitude modulated tones, defined as;

$$s_t = S \cdot \sin(2\pi f_c t) \cdot \left( \frac{1 + m \cdot \sin(2\pi f_m t)}{2} \right) \tag{7.1}$$

where $f_c = 1$ kHz is the carrier frequency, $f_m = 40$ Hz the modulation frequency, $m$ the modulation depth and $S$ defines the overall stimulus level. The subscript $t$ represents a variable with time dependency throughout the paper.

### 7.3.1 Static nonlinear model of compression

Static or memory less nonlinearities are defined such that the current output time series, $y_t$, is a function only of the current input time series, $x_t$, i.e.

$$y_t = f(x_t) \tag{7.2}$$

A simple example of a static nonlinear system is a power-law non linearity, given by

$$y_t = |x_t|^{\alpha} \mathrm{sgn}[x_t] \tag{7.3}$$

where compression is ensured if the power, $\alpha$, is less than unity. The signum function, defined by

$$\mathrm{sgn}[x_t] = \begin{cases} -1, & \text{for } x_t < 0, \\ +1, & \text{for } x_t \geq 0 \end{cases} \tag{7.4}$$

ensures asymmetry in the nonlinear characteristic.

Figure 7.1 illustrates a SAM tone (bottom left) passing through a static compressive power-law non linearity (eqn. 7.3) with compression ratio $\alpha = 1/3$. Also shown is the instantaneous characteristic function (top left), $y(x)$, of the compressive non linearity and the output time series (top right). The variation of the input envelope is mapped to a reduced range in the output, indicted by the dashed lines.

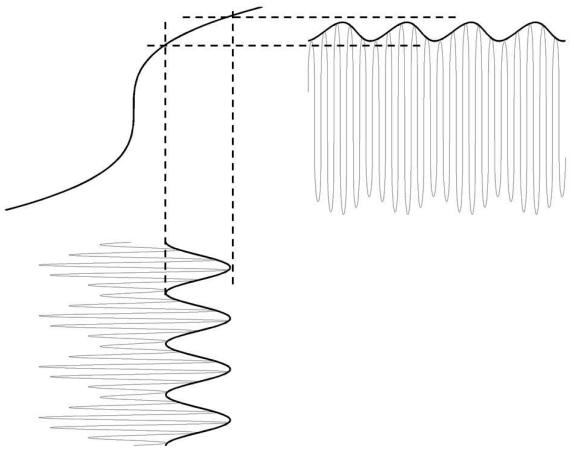


Figure 7.1: Illustration of a SAM tone passing through a static compressive non linearity (with $\alpha = 1/3$) and resulting output time series. The envelopes of the input an output are shown in black curves.

### 7.3.2 Approximate closed-form solution for envelopes processed through a compressive non linearity

It is assumed that the envelope varies at a much slower rate than the carrier frequency. This means that it is possible to treat the envelope and the carrier components of the stimulus passing through the nonlinear characteristic as separate. Assuming that the

stimulus is given by

$$s_t = \chi_t \cdot \sin(2\pi f_c t) \tag{7.5}$$

with the input envelope given by

$$\chi_t = S\left(\frac{1 + m \cdot \sin(\omega_m t)}{2}\right) \tag{7.6}$$

where $m$ is the modulation depth, $\omega_m = 2\pi f_m$ the angular modulation frequency, and S defines the level. Passing this sinusoidally amplitude modulated tone through the instantaneous power-law non linearity, given by Eq. 7.3, and noting that $\chi_t > 0$ then it can be shown that

$$y_t = \chi_t^{\alpha} \cdot |\sin(2\pi f_c t)|^{\alpha} \mathrm{sgn}\left[\sin(2\pi f_c t)\right] \tag{7.7}$$

The two last terms on the right hand side constitute the carrier of the output time series and can be considered a harmonic tone complex of $n\omega_c$(with minimal contribution to the overall envelope as it was assumed $\omega_m << \omega_c$). Thus, in the present analysis only the output envelope, $\eta_t$, will be considered:

$$\begin{aligned} \eta_t &= \chi_t^{\alpha}, \qquad \text{for } \chi_t > 0 \\ \eta_t &= S^{\alpha} \cdot \left(\frac{1 + m \cdot \sin(\omega_m t)}{2}\right)^{\alpha} \end{aligned} \tag{7.8}$$

For a compressive non linearity, $0 < \alpha < 1$. It is clear that the output envelope's dependence on the overall level $S$ is a simple power law.

Special attention is needed for the right-hand term in eqn. 7.8, which is defined in the range from 0 to 1, with

$$\zeta_t^{\alpha} = \left(\frac{1 + m \cdot \sin(\omega_m t)}{2}\right)^{\alpha} \tag{7.9}$$

It is possible to expand $\zeta^\alpha$ in terms of a Taylor series about the arbitrary point $\hat{\zeta}$:

$$\begin{aligned}\zeta^\alpha &= \hat{\zeta}^\alpha \left( \sum_{k=0}^{\infty} \binom{\alpha}{k} \left( \frac{\zeta}{\hat{\zeta}} - 1 \right)^k \right) \\ &= \hat{\zeta}^\alpha \left( 1 + \alpha \left( \frac{\zeta}{\hat{\zeta}} - 1 \right) + \frac{\alpha(\alpha-1)}{2!} \left( \frac{\zeta}{\hat{\zeta}} - 1 \right)^2 + \cdots \right)\end{aligned} \tag{7.10}$$

where $\binom{\alpha}{k}$ represent generalized binomial coefficients, defined as

$$\binom{\alpha}{k} := \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-k+1)}{k!} \tag{7.11}$$

It is possible to represent $\eta_t$, the output envelope, in an alternative form as an infinite sum of harmonics of the fundamental modulation frequency:

$$\eta_t = \sum_{p=0}^{\infty} A_p \sin\left(p\left(\omega_m t + \beta_0\right)\right) \tag{7.12}$$

where $A_p$ are Fourier coefficients, $p$ is the order of the infinite sum, and $\beta_0$ is some phase offset.

Using the method of harmonic balance (Nayfeh and Mook, 1995), each term in the power series in Eq. 7.10 is expanded and factored by $\sin(q\omega_m + \phi_0)$, where $q$ is an integer and represents harmonics of the modulation frequency. Finally combining equations 7.6, 7.8 and 7.10, allows the derivation of an approximate closed-form solution for the first harmonic, $A_1$, corresponding to the Fourier coefficient of $\omega_m$:

$$\begin{aligned}A_1 = \left( \frac{S}{2\hat{\zeta}} \right)^\alpha \Bigg[ & m \left( \sum_{k=0}^{\infty} \binom{\alpha}{k} \frac{k(1-\hat{\zeta})^{k-1}}{\hat{\zeta}^k} \right) \\ & + m^3(\cdot) + m^5(\cdot) + O(m^{2n-1}) + \ldots \Bigg]\end{aligned} \tag{7.13}$$

Thus, $A_1$ is represented as an infinite power series in terms of the modulation depth $m$, comprising only odd-orders $(2n-1)$ of $m$. The terms for the orders of $m$ higher than

1 are not shown here for brevity. Assuming that $m$ is small, i.e. $m << 1$, it is possible to ignore the higher-order terms, such that:

$$A_1 \approx \left(\frac{S\hat{\zeta}}{2}\right)^{\alpha} \cdot m \left(\sum_{k=0}^{\infty} \binom{\alpha}{k} \frac{k(1-\hat{\zeta})^{k-1}}{\hat{\zeta}^k}\right) \tag{7.14}$$

It can be shown that the infinite summation in the right hand set of brackets is equal to $\alpha\hat{\zeta}^{-\alpha}$, and thus

$$A_1 \approx \left(\frac{S}{2}\right)^{\alpha} \cdot m\alpha \tag{7.15}$$

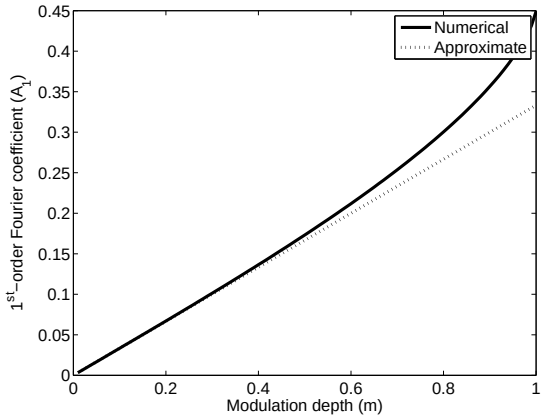



Figure 7.2: Comparison of approximate closed form solution (dotted line) for 1st-Fourier component and numerically simulated (solid line) result, for $\alpha = 1/3$ and $S = 1$.

Eq. 7.15 represents a simple approximate closed-form solution for the response amplitude at the amplitude modulation frequency after passing through a compressive static non linearity. To test this simple model, a numerical simulation was carried out in MATLAB, passing a SAM tone with $S = 1$ through a static non linearity with $\alpha = 1/3$, and taking the Hilbert envelope. Figure 7.2 shows the numerically determined value of the Fourier coefficient (solid curve) at the modulation frequency $\omega_m$ as the modulation depth $m$ was varied between 0 and 1. The approximate closed-form solution of eqn. 7.15 is shown by the dotted curve, and is a good approximation for the true value for small $m$.

According to Eq. 7.15, the amplitude of the first harmonic in the response is dependent on the input level $S$ via a power-law relation. Therefore, given a fixed modulation depth, the slope of $A_1$ as a function of the excitation level on double log axes yields a straight line with slope $\alpha$. This can be used to estimate the degree of compression in the input/output level-curve, i.e. cochlear compression. If the excitation level is fixed, while varying the modulation depth, $A_1$ will yield a straight line with slope $\alpha(S/2)^{\alpha}$, if plotted on a linear axis. Again, this could lead to an estimate of the local compression. On a log-log axis, this simple analytical model predicts a slope of 1. Thus BM compression can be obtained from experimental data via:

1. the slope, $\alpha$, of the ASSR level-growth function, plotted on a double logarithmic scale.

2. via the slope, $\alpha(S/2)^{\alpha}$, of the ASSR modulation-growth function plotted on linear scales. The easiest method to derive $\alpha$ from the slope $\alpha(S/2)^{\alpha}$ is to vary excitation level and derive the parameter estimate, rather than by directly inverting [1].

[1] It is possible to invert eq. 7.15 solving for the compression ratio $\alpha$, by recasting the equation as

$$\alpha \log\left(\frac{S}{2}\right) e^{\alpha \log\left(\frac{S}{2}\right)} = \frac{A_1}{m} \log\left(\frac{S}{2}\right)$$

This has the form $x(\alpha) = w(\alpha)e^{w(\alpha)}$ and its solution for $\alpha$ is given by the Lambert W function (Corless et al., 1996), also known as the product logarithm, i.e.

$$\alpha = \frac{W\left(\frac{A_1}{m} \log\left(\frac{S}{2}\right)\right)}{\log\left(\frac{S}{2}\right)}$$

where $W(\cdot)$, the Lambert W function, is a multi-valued function that can be complex. Care must be taken to pick the appropriate branch of this function for a physically realistic solution and in this application this is not necessarily trivial. Therefore two methods are later proposed to experimentally fit/approximate $\alpha$ from the slope of the ASSR-level growth function (plotted on double log. axes) and via the slope of the modulation growth function (on linear axes).

## 7.4 ASSR experiment: Evoked response growth as a function of modulation depth and stimulus level

### 7.4.1 Methods

#### Subjects

In experiment A, the left ear of eight normal hearing subjects were tested. In experiment B, both ears of a total of ten normal hearing test subjects were tested, yielding a total of 20 data sets. All subjects had hearing thresholds $\leq$ 25 dB HL between 0.5 and 6 kHz in both ears. The experiments were conducted in an electrically and acoustically shielded audiometric booth (IEC 268-13). To control the subjects attention and prevent them from sleeping, they remained supine and watched a silent subtitled movie during the recording session.

#### Stimuli

In both experiments, the subjects were presented with sinusoidally amplitude modulated tones (eqn. 7.1). In experiment A, the stimulus level was varied between 55, 65, 75 and 85 dB SPL, with the modulation depth held constant at $m = 0.75$. These stimulus levels were chosen to be within the expected compressive region of the cochlear input/output function. Experiment B varied $m$ (in Eq. 7.1) between 0.25, 0.5, 0.75 and 1.0. A constant stimulus level of 55 dB SPL was used. The stimuli were calibrated to have identical RMS values. This resulted in actual post calibration modulation depths of 0.3, 0.58, 0.81 and 1.0. On a logarithmic scale, relative to a modulation depth of 1.0, this corresponds to $M_{dB} = -10.41, -4.77, -1.81, 0$ dB, respectively.

The starting and end phases of the stimuli were matched to ensure that it could be repeated continuously without audible discontinuities. Each epoch lasted 375 ms, corresponding to 375 cycles of the carrier and 15 cycles of the modulation frequency. In experiments A and B, a total of 1200 and 2000 averages were made for each test

condition, respectively. All stimuli were generated in MATLAB and playback was made through an RME ADI-8 Pro 24-bit sound card at a sampling frequency of 44.1 kHz. Stimulus levels were set via a TDT PA5 programmable attenuator. The stimuli were presented to the subjects via ER-2 insert earphones.

**ASSR recording and data analysis**

EEG activity was recorded differentially between the vertex and the ipsi-lateral mastoid, with the ground electrode placed on the forehead. Ag/AgCl electrodes were used, and an inter-electrode impedance was maintained below 5kΩ and within 1kΩ of each other. EEG activity was recorded on a SynAmps2 amplifier at a sampling rate of 10 kHz (experiment A) and 5 kHz (experiment B), and band-pass filtered between 0.05 and 500 Hz. After recording, the EEG-data were epoched and filtered again from 10 to 300 Hz, using a 40 tap FIR filter with zero phase delay. The epochs were averaged using an iterative weighted-averaging algorithm (Riedel et al., 2001). The recorded averaged time series were transformed to the frequency domain using a Fast Fourier transform. The amplitude of the complex vector of the 40-Hz component was calculated.

A frequency domain F-ratio test (John and Picton, 2000) was used to detect if an ASSR was present in the recorded signal. The energy at 40 Hz was compared with the background noise, estimated from 7 neighboring spectral bins where no evoked response would be present (Dobie and Wilson, 2001). This yielded an F-distribution with [2,14] degrees of freedom with a critical value of 6.51 at the 1% significance level. Responses were only included in the study if their F-value exceeded the critical value. Data sets from an individual ear were only included if more than 1 data point was accepted. These acceptance criteria resulted in one subject being removed from experiment A, and one ear of one subject being removed from experiment B.

An analysis of covariance (ANOCOVA) was performed on the ASSR modulation-growth functions. The ANOCOVA assumes linear regression. In the log.-log. plots this was obtained by taking the logarithm on both variables. An estimate of the

slope of the best fitted single line and a standard deviation on the slope estimate was obtained. Each ear were treated as a separate data set in the analysis.

### 7.4.2 Experiment A - Results

Averaged ASSR magnitude and standard errors for all 7 subjects are shown in Fig. 7.3 (diamonds), as a function of stimulus level (dB SPL). The spectral magnitude of the 40-Hz component is given relative to 1$\mu$V rms. Error bars with $\pm 1$ standard error are also shown and reflect the large individual differences. For each recording from a given subject, the ASSR magnitude increases monotonically as stimulus level increases.

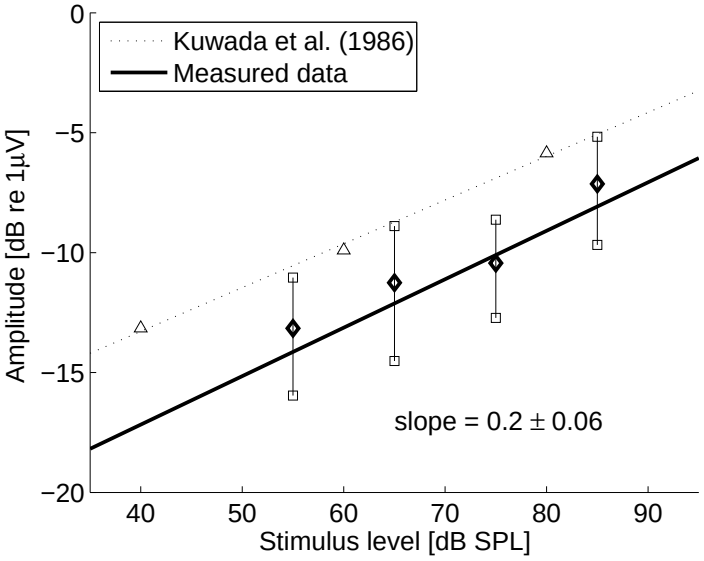



Figure 7.3: ASSR amplitude versus stimulus level averaged over the 7 normal-hearing subjects. Modulation depth was 75%, and the error bars show $\pm 1$ standard error. The estimated slope (compression ratio) is indicated as well as the $\pm 1$ standard deviation on the slope estimate. Also shown are literature data derived from Kuwada et al. (1986) (dotted curve).

An ANOCOVA analysis was carried out on the ASSR magnitudes (solid line in

Fig. 7.3). The slope estimate was found to be 0.20 dB/dB with a standard deviation of 0.06. A slope of 1 would indicate linearity, and a slope of $< 1$ implies compression. The low uncertainty on the slope estimate from the ANOCOVA confirms that the individual differences, indicated by the error bars, were mainly offsets of the overall ASSR amplitude in the individual recordings, rather than variations of the slope. For comparison, Fig. 7.3 also reproduces the data from Kuwada et al. (1986) (triangles). Kuwada et al. (1986) measured ASSR with similar electrode placements, modulation depth, and stimulus modulation- and carrier-frequency. The slope of the level-growth function from Kuwada et al. (1986) obtained by linear regression on the log. variables was 0.18. The same slope (0.18) was found for a similar data set presented by Picton et al. (1987) (not shown on figure). Thus the estimates reported in the present study are similar to historically published ones. The slope of the ASSR level-growth function thus show compression of an amount similar to cochlear compression, as previously discussed in section 7.3.

### 7.4.3 Experiment B - Results

Averaged ASSR magnitude and standard errors for all ten subjects (19 ears) measured at 55 dB SPL are shown in Fig. 7.4 (diamonds), as a function of log. modulation depth relative to 100%. As in the level-growth functions from Fig. 7.3, the magnitudes increase monotonically as modulation depth increases. A direct comparison with historical data is difficult due to differences in stimulus level and calibration, carrier and modulation frequency, electrode placement and, in some cases, a very limited number of subjects. However, the ASSR RMS-amplitudes reported here are in agreement with those reported by Kuwada et al. (1986); Rees et al. (1986); Picton et al. (1987); Boettcher et al. (2001).

The ANOCOVA analysis, carried out on the ASSR magnitudes (solid line in Fig. 7.4), gave a slope estimate of 0.78 dB/dB with a standard deviation of 0.09. For comparison, Fig. 7.4 also reproduces the data from Boettcher et al. (2001) (upwards and downwards pointing triangles). The dotted curve was fitted to log. ASSR amplitudes recorded in response to AM tones, with carrier frequency of 520 Hz,

a modulation frequency of 40 Hz, and a stimulus level of 65 dB SPL. The dashed curve (also from Boettcher et al., 2001) was obtained at a carrier frequency of 4 kHz. Slope estimates obtained by linear regression of the two curves are 0.73 ($f_c$ = 520 Hz) and 0.62 ($f_c$ = 4 kHz). Slope estimates were also derived from Kuwada et al. (1986) and Picton et al. (1987) (not reproduced here to aid clarity) for comparison and were found to be 0.62 and 0.61, respectively. Thus, the estimates reported in the present study are in reasonable agreement with historically published results, even though stimulus conditions varied significantly across studies. The ASSR modulation-growth functions are not consistent with the theoretical predictions from section II. In the theoretical predictions ASSR growth functions had a slope of 1, when plotted on double logarithmic scales.

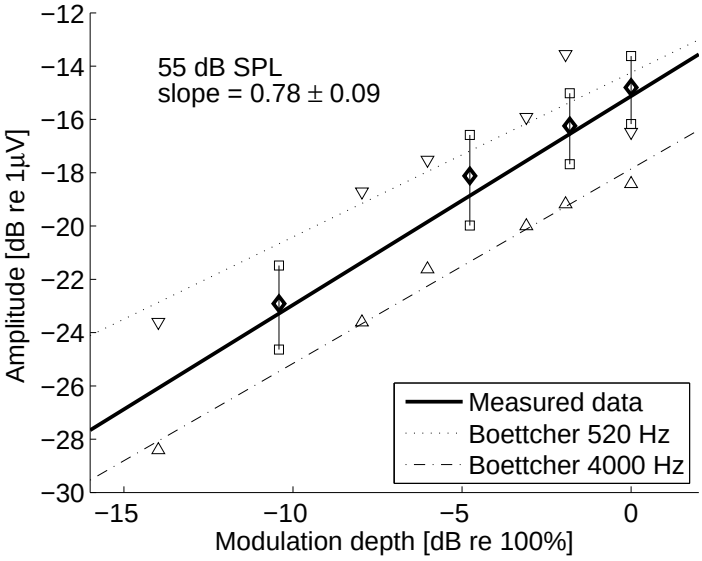



Figure 7.4: ASSR amplitude versus modulation depth averaged over the 10 normal hearing subjects and left and right ears. Stimulation level was at 55 dB SPL, and the error bars show ±1 standard error. The estimated slope (compression ratio) is indicated as well as the ±1 standard deviation on the slope estimate. Also shown are literature data derived from Boettcher et al. (2001) (dotted and dashed curves), see the text for details.

The closed-form solution derived in the section 7.3 predicted that the compression

could be estimated from the slopes of the modulation-growth function, as long as two independent data sets of different stimulus levels were measured. For this reason, three of the subjects included in experiment B were retested and their ASSR modulation-growth functions measured at a higher level of 71 dB SPL. Unfortunately the uncertainty on the slope estimates from the ANOCOVA was almost 50% of its value, probably due to the few test subjects available. Due to this large uncertainty, an estimate of the compression coefficient, $\alpha$, based on the simplified closed form solution, could not be obtained.

### 7.4.4 Experiment summary

In summary, the ASSR level-growth function showed compressive behavior, with a slope estimate in the order of 0.2 dB/dB and thus corresponded well to both ASSR literature slopes, theory predictions and cochlear compression estimated using alternative psychoacoustic or OAE measures. The modulation-growth function, demonstrated a slope of 0.78 dB/dB. On double logarithmic scales, this suggests a power-law relation with $m$ not predicted by a simple instantaneous compressive non linearity (Sec. 7.3). If cochlear compression should have been derived, two estimates of the ASSR modulation-growth function slope (estimated at linear scales) at different excitation levels would have been needed. However, due to the high variability of the ASSR magnitudes, the variation in the slope estimates, for the small number of subjects (three) measured at two levels, rendered this impossible to fit.

The most serious inconsistency between the analytical model predictions and the experimental recordings is the slope of the ASSR modulation-growth function being less than unity. In an attempt to investigate this further, the next section develops a more physiologically plausible numerical model of ASSR generation.

## 7.5 ASSR model

### 7.5.1 Modeling framework

This section derives an ASSR model, which is used to predict how local BM compression is reflected in ASSR magnitudes. The ASSR model is inspired by the work of Goldstein and Kiang (1958), who described evoked responses as a linear convolution of a single fiber instantaneous auditory nerve (AN) discharge rate in response to a given stimulus with an elementary unit waveform, called the unitary response (UR). The UR describes the contributions made to the AEP each time a cell discharges. Following this idea, Dau (2003) proposed a model for the generation of ABRs using the instantaneous discharge rate for single nerve fibers summed across frequency at the level of the AN to create a neural activity pattern. Harte et al. (2010) and Rønne et al. (2012) updated and evaluated an ABR model, following the principles of Dau (2003). This model was shown to be successful in simulating ABR responses to varies stimuli as clicks, tone bursts and chirps. A number of studies (e.g. Galambos et al., 1981; Hari et al., 1989; Plourde et al., 1991; Gutschalk et al., 1999; Bohorquez and Oezdamar, 2008) have demonstrated that the ASSR, at modulation rates around 40 Hz, can be predicted from the convolution of single middle-latency transient responses with a click train with the appropriate repetition rate. Thus, the predominant response in the ASSR is due to the $N_a$ - $P_a$ and $N_b$ - $P_b$ components of the middle-latency response (MLR), originating in the early auditory cortex, and a smaller contribution due to the ABR. Given the success of the ABR model, and the argument that the ASSR can be modeled as a linear superposition of the ABR wave V and the $N_a$ - $P_a$ and the $N_b$ - $P_b$ components of the MLR, an ASSR model was created in this study. The ASSR model was, like Rønne et al. (2012), based on the AN model (Zilany and Bruce, 2007; Zilany et al., 2009) and a linear, subject and stimulus independent UR. The model distinguishes itself from simpler convolutive models (e.g. Sparacino et al., 2004; Bohorquez and Oezdamar, 2008), in the nonlinear front end AN model.

In Fig. 7.5, a schematic diagram of the ASSR model is shown. The ASSR model builds upon the Zilany et al. (2009) auditory nerve (AN) model, which simulated the

instantaneous discharge rate from a single AN fiber tuned to a specific frequency. The AN model, and thus the ASSR model, includes key properties of nonlinear cochlear processing, such as compressive BM filtering, inner hair-cell (IHC) transduction, and IHC-AN synapse adaptation. The ASSR model simulates AN responses from 500 different characteristic frequencies (CFs), in the range from 100 Hz to 16 kHz. The responses were summed to form the neural activity pattern and convolved with a unitary response to produce the simulated ASSR. The stimuli presented to the ASSR model were defined in Pascals and calibrated such that the root-mean-square value equaled 1.

The ASSR model is similar to the Rønne et al. (2012) ABR model. However, three modifications were undertaken. First, the Zilany and Bruce (2007) AN model was replaced with Zilany et al. (2009). This was done, since the latter AN model includes an updated synapse stage that simulates effects of neural adaptation more realistically. This is highly important for longer-duration signals (the Zilany et al. (2009) AN model was "humanized" in an identical manner as done in Rønne et al. (2012)). Second, the response of low spontaneous rate fibers (0.1 spikes/s) of the AN model was simulated, as opposed to the high-spontaneous rate (50 spikes/s) used in Rønne et al. (2012). This change was made as high-spontaneous rate fibers saturate for the relatively high-level and long-duration AM stimuli, and the response is thus likely dominated by low spontaneous rate fibers (Sumner et al., 2002; Zilany et al., 2009). Third, the unitary response (UR) was recalculated to include the contribution from the middle latency response (MLR). As discussed above, the ASSRs generated using a modulation rate of 40 Hz have contributions from neurons in the AN, brainstem and up to the early auditory cortex. By incorporating the MLR into the UR function, the higher-stage contributions could be modeled to a first approximation. The UR was only calculated once as the deconvolution between the summed neural activity pattern produced by the AN model (in response to a 60dB pe SPL click), and a recorded MLR (Harte, 2007) using the identical click stimulus and electrode position as in the recordings presented in Sec. 7.4. Once obtained for the 60 dB pe SPL click, the UR was fixed for all further numerical simulations carried out in this paper.

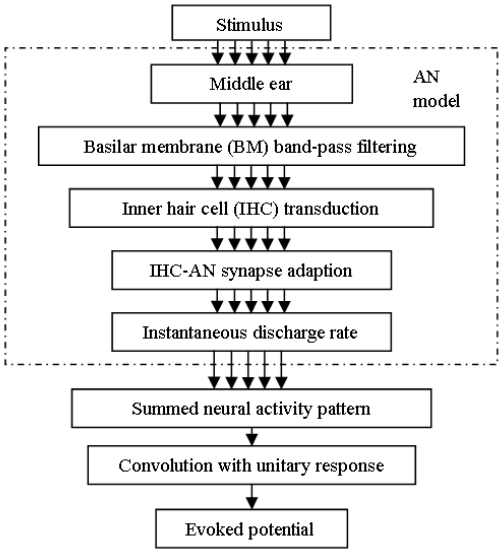



Figure 7.5: Schematics diagram of the ASSR model. A stimulus is presented to the AN model tuned to a single frequency. The signal is then processed through the AN model stages of middle-ear filtering, BM filtering, IHC transduction and IHC-AN synapse. The sum of 500 individual simulation with the AN model tuned to different frequencies produces the summed neural activity pattern. This pattern is then convolved with the UR to produce the ASSR.

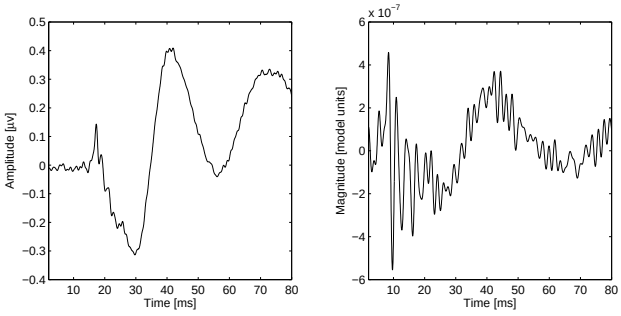



Figure 7.6: Left: A 60 dB pe SPL click evoked MLR (data from Harte (2007)). Right: The unitary response. Derived as the deconvolution of the click evoked MLR and the summed neural activity pattern obtained as the summed responses of the humanized AN model given the identical stimulus.

### 7.5.2 Simulations

ASSRs were simulated as a function of the stimulus modulation depth ($m$ = 0.25, 0.5, 0.75 and 1) and stimulus level (15 dB SPL to 95 dB SPL in steps of 10 dB). The ASSR components were derived from the amplitude of the 40-Hz component in the spectrum of the simulated ASSR time series. Figure 7.7 (left) shows modulation-growth functions simulated at varying stimulus levels (10 to 75 dB). A regression line was fitted to the 55 dB SPL curve and a slope estimate of 1.04 was obtained. This is close to a linear slope of 1 as predicted by the static non linearity, but deviates from the experimentally measured slope of 0.78. Figure 7.7 (right) shows the level-growth function for the 75% modulated ASSRs. It is observed that the nonlinear model produces a slope of 0.48 in the compressive region above 35 dB SPL stimulus level, and a close-to-linear slope below this stimulus level.

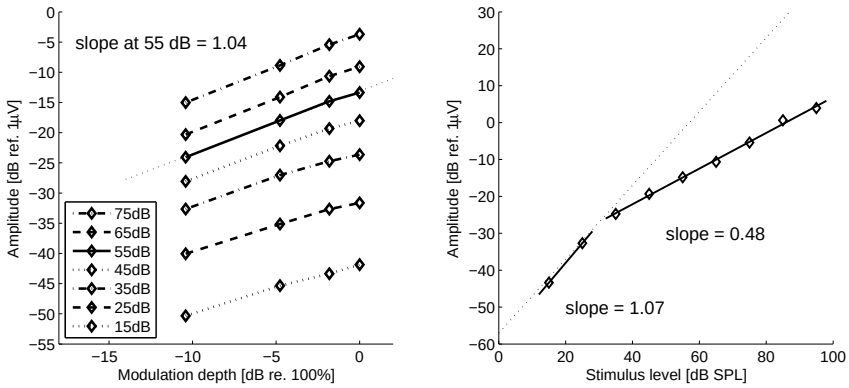



Figure 7.7: Left: Simulated ASSR amplitudes of the 40Hz component as a function of stimulus modulation depth and stimulus level, plotted on log.-log. axes. Compressive growth, of similar magnitude as expected from cochlear compression is observed as function of level. Whereas linear processing is observed as a function of modulation depth. Right: Simulated ASSRs as function of stimulus level. The modulation depth is kept at 75% and the stimulus level is varied from 15 to 95 dB SPL in steps of 10dB.

In Fig. 7.8 (left panel), the same simulated results are shown on linear scales. For each modulation-growth function, a regression line was fitted and a slope estimate obtained. In section 7.3 it was shown that, for a static non linearity, the compression ratio, $\alpha$, can be estimated from two adjacent modulation-growth functions ($n$ =1,2).

Based on Eq. 7.15, the slope of the modulation-growth function, $k_n$, can be described as:

$$k_n = \left(\frac{S_n}{2}\right)^{\alpha} \cdot \alpha \tag{7.16}$$

where $S$ is the stimulus level. Assuming that the compression affecting two different modulation-growth functions is the same, an estimate of the compression ratio can be found as:

$$\alpha = \frac{log_{10}(\frac{k_2}{k_1})}{log_{10}(\frac{S_2}{S_1})} \tag{7.17}$$

The right panel of Fig. 7.8 shows compression ratios calculated based on Eq. 7.17 and two adjacent slope estimates from Fig. 7.8. The abscissa represents the average stimulus level, such that the compression ratio derived based on the 55 dB SPL and 65 dB SPL slopes are plotted at 60 dB SPL. The compression ratios amounts to 1 at low levels, and decreases towards the dotted line representing the compression ratio, CR = 0.48, found in Fig. 7.7 (right).

### 7.5.3 ASSR model discussion

The ASSR model includes a dynamic compression function which differs significantly from the simple static compression function used in the theoretical model, introduced in Sec. 7.3. Additionally, the numerical model includes other key stages in auditory processing important for the generation of evoked potentials, such as IHC transduction and IHC-AN synapse adaptation. The numerical ASSR model simulates contributions to the ASSR from 500 parallel channels reflecting AN fibers across the tonotopoic axis. For channels with center frequencies close to the carrier frequency of the AM stimulus, the response was compressive. For off-frequency channels, the contributions showed linear growth. The numerical model is capable of describing far more details than the simple analytical treatment in Sec. 7.3.

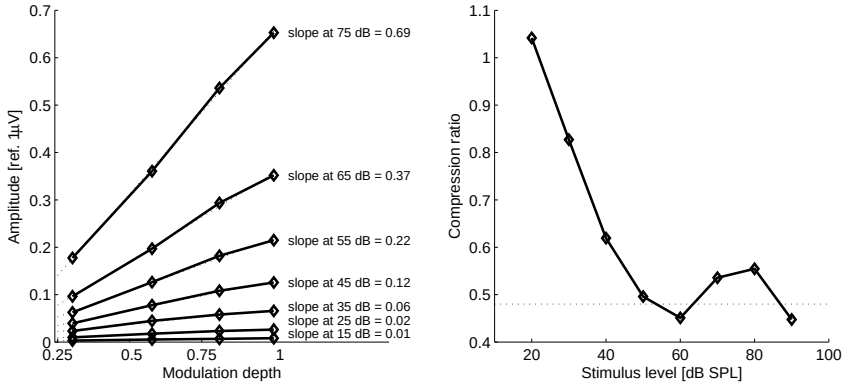



Figure 7.8: Left panel: Simulated ASSR amplitudes of the 40Hz component as a function of stimulus modulation depth plotted on linear scales. The different curves show the results of different stimulus levels. Right panel: CRs calculated using Eq. 7.17 and data from two adjacent curves in the left panel. The stimulus level on the abscissa represents the average stimulus level for two adjacent curves. The CR is 1 for low stimulus levels, and close to the 0.48 (dotted line) corresponding to the slope in Fig. 7.7 (right panel), at higher stimulus level.

The numerical model showed that realistic cochlear mechanical filters, IHC mechanoelectrical transductions and IHC-synapse adaptation resulted in the same ASSR modulation growth functions and level-growth functions as a static compressive nonlinear function processing the amplitude-modulation of the stimulus. Plotted on log.-log. axes, the model predicts a modulation-growth function slope of 1.04 which is nearly linear and thus does not reflect cochlear compression, whereas the simulated level-growth function in the compressive region above 35 dB SPL shows a compression ratio of 0.48. In Sec 7.3, it was argued that a compression ratio could be derived for the modulation-growth functions plotted on linear axes (Fig. 7.8). The numerical simulations supported the theoretical model and found a compression ratio close to 0.48. This method of estimating the compression ratio, from two slope estimates from two stimulus levels, is fine for the numerical simulations here, as the results were entirely noise free. Any variation or uncertainty on the slope estimates, $k_{1,2}$, as one would see in experimental data, would be increased as the ratio is taken. Thus, this is not considered to be the preferred method for obtaining estimates of compression.

The simulated cochlear compression ratio was found to be 0.48, i.e. considerably larger than the experimentally measured ratio of 0.2 dB/dB. To investigate the cause of this, an additional simulation was made (not shown) with only the 30 fibers closest to the stimulus frequency, i.e. the frequency region of 868 Hz to 1158 Hz. This eliminated off-frequency contributions, which would be expected to have linear growth (Rhode and Recio, 2000). The simulation yielded a level-growth function with linear growth below 35 dB and compressive growth above 35 dB with a slope of 0.19. A similar compression of 0.20 can be observed for simulated single fiber response to a pure-tone stimulus level growth. This follows the experimental findings better where a compression ratio of 0.20 was estimated. Thus, the numerically simulated broad band level-growth function does not strictly show local cochlea compression, but rather exhibits a slope that is the result of a mixture of on-frequency compression and off-frequency linearity. In human measurements, a similar effect might be expected. However, it is unknown to what extent the off-frequency contributions linearize the human level growth. Further, the human cochlear compression has been measured using both OAEs and psychoacoustics both showing compression ratios between 0.2 and 0.3. Thus, while the numerical model seems to be capable of capturing the key physiological generator mechanisms for ASSRs, it does not correctly model the contributions across different nerve fibers precisely. The numerical model seems to give more weight to linear off-frequency contributions than is observed in experimental data. Further work to look at the model nonlinear mechanical filters sharpness of tuning (or Q-factor); and/or contributions from parallel high-, medium- and low-spontaneous rate fibers (only low-spontaneous rate were simulated here), might shed light on this disparity.

It is important to emphasize that the ASSR model shows the same compression obtained using either of the two techniques developed in section 7.3. This supports the hypothesis that it is cochlear compression that are measured using these techniques.

## 7.6 Overall discussion

### 7.6.1 Summary

This study investigated the potential for ASSRs to estimate cochlear compression. Based on a closed-form solution of how the envelope of an AM signal would be affected when processed by a static nonlinear compressive system, two testable hypotheses of how to obtain estimates of cochlear compression were made. First, the compression ratio can be obtained as the slope of the ASSR level-growth function plotted on double log. axes. Second, the compression ratio can be obtained from the slopes of two modulation-growth functions measured at two different levels plotted on linear scales. A numerical model of ASSR generation was also presented; it consisted of a phenomenological AN model capable of accurately describing the outer and inner ear; nonlinear cochlear mechanical filtering, IHC processing (half-wave rectification and low-pass filtering) and IHC-AN synapse adaptation. The output of the AN model was convolved with an empirically derived unitary response function, used to model auditory pathway processing and propagation of cell discharge potential to the recording electrodes. The numerical model demonstrated that local cochlear compression could be estimated by the two methods inspired from the simple analytic model, provided care is taken to limit off-frequency contributions to the ASSR.

Two experiments were carried out. In one experiment, ASSR level-growth functions were measured in a total of 8 subjects, and a compression ratio of 0.20 was obtained. In another experiment, using 10 subjects and a total of 20 ears, modulation-growth functions were measured. A slight compression of 0.78 was observed, when plotted on double logarithmic axes. This is not consistent with the analytical and the numerical model both predicting a linear behavior in this condition. The modulation-growth functions of three subjects were additionally measured at a higher stimulus level. However, the variability was too large to derive a meaningful compression ratio based on the second hypothesis.

The numerical ASSR model predicted a compression ratio of 0.48 for both the level-growth function and the modulation-growth functions. It was found that the

reason for the the decreased amount of compression in the simulations results from the contributions of the off frequency fibers where the signal is processed linearly, such that the overall resulting response becomes less compressive than in the region around the stimulus frequency.

### 7.6.2 Best practice for estimating cochlear compression using ASSR

This study demonstrated that cochlear compression can be estimated using ASSR, by either measuring level- or modulation-growth functions. However, estimating cochlear compression from ASSR modulation-growth functions requires double the number of measurement points as two slope estimates have to be obtained. Any experimental recording of a physiological parameter will necessarily be noisy, i.e. have an associated uncertainty. To estimate the compression ratio, the ratio of the two modulation-growth functions needs to be taken. This has the effect of adding the two variances or mean-square errors for the individual slope estimates to give the uncertainty on the compression ratio. This makes the estimate of compression ratio implicitly more uncertain than via the level-growth functions. This was confirmed by the difficulty in deriving a useful compression ratio from the modulation-growth experiments. The clear recommendation is thus to measure level-growth functions if cochlear compression is to be estimated from the ASSR. A single measurement point can be measured using 1200 epochs of each 375ms length, giving a measurement time per stimulus level of 7.5 minutes. An estimation of a compression ratio can thus be done by measuring ASSRs at 3 or 4 different levels, and would require less than 30 minutes of recording time. This is still a lengthy procedure and does not lend itself as a clinically viable option at this stage.

### 7.6.3 Is cochlear compression reflected in experimental logarithmic modulation-growth functions?

Plotting the experimental modulation-growth functions on double log. axes (Fig. 7.4) demonstrated a small degree of compression, with a slope of $0.78 \pm 0.09$. This is at odds with the simple theoretical predictions and the results from simulations with the physiologically inspired numerical model, both predicting linear growth, i.e. a slope of 1.0. The experimental modulation-growth functions were obtained from only 10 normal-hearing subjects, so this disparity could be ascribed to experimental uncertainty. However, the ANOCOVA fitting of the data yielded a low uncertainty on the slope estimate of only $\pm 0.09$.

It could be argued that the small degree of compression seen in the experimental data might arise from a compressive stage in auditory processing independent of local cochlear compression. If one considers the AN model employed to give an accurate description of peripheral processing and non linearity, then the IHC-AN synapse or early brainstem might contain the additional stage. An effect that could give rise to such an independent compression could be the modulation gain (e.g. Joris and Yin, 1992; Frisina et al., 1996; Joris et al., 2004; Malone et al., 2010).

Joris and Yin (1992) measured the ability of cat AN fibers to synchronize their firing to AM stimuli. They normalized the synchrony by the modulation depth employed to derive a modulation gain function. Using a stimulus level of 49 dB SPL, Joris and Yin (1992) found a modulation gain of $\approx 9$ dB at 10% modulation depth, monotonically decreasing to $\approx 2$ dB at 100% modulation depth. No exact physiological mechanism was suggested as being responsible for the gain. In the AN model employed here, Zilany et al. (2009) demonstrated that it is capable of simulating the modulation gain from Joris and Yin (1992) for the cat. There is no way of ensuring that this is correctly modeled in humans for the present study. Malone et al. (2010) described how the gain was increased in the rostral field and even further increased in the auditory cortex, indicating that the ascending auditory pathway privileges low amplitude modulation depths, and indicating that higher stages of the auditory pathway also influence the modulation gain. Joris and Yin (1992) showed that synchrony and, consequently, the

modulation gain are also stimulus level dependent in a non monotonic way. The non monotonic stimulus level dependency and the increasing magnitude with ascending place in the auditory pathway lead to the conclusion that the underlying mechanism might be independent from cochlear compression. However, the modulation gain does not seem to be the major cause of the small degree of compression seen in the modulation-growth function. This conclusion is based on the fact that the simulations using the ASSR model do not show a compressive slope even though the modulation gain is modeled by the underlying AN model. The reason could be that the modulation gain in all literature studies (and the AN simulation) were measured in single nerve fibers of different species of animals. It has not been studied how the modulation gain would be expected to affect a real ASSR, which naturally consists of the response of numerous on- and off-frequency tuned fibers.

The apparent compression observed in the logarithmically analyzed modulation-growth function is thus still unexplained. It might be reflecting a compression independent of the regular cochlear compression, i.e. at a retro-cochlear stage, which is not reflected in the model provided here.

## 7.7 Conclusion

This study evaluated the potential of using ASSR as a tool to estimate cochlear compression. Two different methods were evaluated, from measurements of the modulation- and level-growth functions. To evaluate these methods, three different approaches were taken, a simple analytical model based on a static non linearity, experimental measurements and a numerical nonlinear ASSR model. The two modeling approaches illustrated that both level and modulation growth functions could be used to estimate cochlear compression. However, the level-growth function was found to be superior as it requires less measured data and has less uncertainty. The level-growth function was experimentally measured in seven subjects and a compression ratio of 0.20 was found, corresponding to compression ratios found in literature using both ASSR and psychoacoustic measures. Additionally, the measured modulation-growth function, when plotted on double logarithmic scales, showed a

small degree of compression, contradictory to the model predictions. It was argued that this was evidence for an effective compressive stage independent of cochlear compression.

# 8

# Modeling auditory evoked brainstem responses to speech syllables

This chapter presents work that, in cooperation with coauthors James Harte and Torsten Dau, is in preparation for submission to the Journal of the Acoustical Society of America.

## 8.1 Introduction

Auditory evoked potentials (AEP) have been used to assess the neural encoding of sound both for clinical and research purposes. Various types of stimuli have been considered, such as transients like clicks, chirps and tone-bursts (e.g. Jewett and Williston, 1971; Neely et al., 1988; Dau et al., 2000, and chapter 3); steady-state signals such as amplitude modulated (AM) tones (e.g. Galambos et al., 1981; Picton et al., 1987; Rees et al., 1986, and chapter 7), but also more complex signals like speech (e.g., Warrier et al., 2004; Agung et al., 2006; Swaminathan et al., 2008; Chandrasekaran and Kraus, 2010). Most studies have focused on the auditory brainstem response (ABR) as they are less affected by attention and sleep than potentials with origin at higher neural stages. The ABR has also been observed to be unaffected by training. However, a number studies have recently investigated and found plasticity[1] of the ABR, both considering short term training effects (e.g. Russo et al., 2005; Song et al., 2008) and long-term experience effects (e.g. Krishnan et al., 2005; Johnson et al., 2008a, see Chandrasekaran and Kraus (2010) for

[1] physiological changes of the nervous system due to e.g. learning

review). Russo et al. (2005) recorded brainstem responses to the stimulus-syllable /da/ in learning-impaired children. The responses of the learning-impaired children were recorded before and after an eight week period containing 35-40 one-hour sessions of auditory training. The authors showed that the correlation between the ABR to the clean /da/ syllable and the response to /da/ in noise, improved for the learning-impaired children over this relatively short training period, thus demonstrating plasticity in the brainstem. This result suggested that features of the brainstem-response might reflect the ability to comprehend speech and speech in noise. Johnson et al. (2008a), Hornickel et al. (2009) and Skoe et al. (2011) measured brainstem responses to the synthetically created syllable-stimuli /ba/, /da/ and /ga/, in normal and learning-impaired children. Both groups of children were reported to have normal audiometric thresholds and ABR wave-V latencies. Hornickel et al. (2009) measured stop consonant differentiation scores, comparing the latencies of the major peaks of the three ABRs evoked by the different syllables, and reading abilities and speech-in-noise perception. They reported a correlation between the stop consonant differentiation score and the two behavioral measures, such that large differences between peak-latencies (large consonant differentiation score) correlated with good performance in the speech-in-noise test and the test of reading ability. Hornickel et al. (2009) argued that this result showed plasticity in the brainstem, as the group with the good behavioral performance had undergone long-term learning and that the better performance was an indication of that learning had affected both the behavioral performance and the electrophysiological brainstem recordings. The observed differences, between the learning-impaired and the normal-learning subject groups in the ABR measures of stop-consonant differentiation scores, were thus argued to be the result of efferent (top-down) neural processes, and not the result of peripheral auditory afferent processing.

Johnson et al. (2008) presented similar syllable-evoked ABR recordings from 22 normal-hearing children. They measured the latency of the major peaks for each of the three syllable-evoked ABRs and found that, although the three recorded time-series were much alike, the peaks of the time-series response to /ga/ had shorter latencies than the peaks of /da/ which again had shorter latencies than /ba/. The three syllables only differed in the frequency content of the second formant, $f_2$, and the

third formant, $f_3$. Hornickel et al. (2009) and Skoe et al. (2011) used almost identical stimuli. Due to the difference in the frequency content of the syllables and due to the tonotopic mapping of frequencies to places on the BM, the peaks of the ABR responses were represented early for the /ga/ ($f_2$ = 2480 Hz), later for the /da/ ($f_2$ = 1700 Hz) and latest for the /ba/ ($f_2$ = 900 Hz). The underlying processes accounting for the findings of Johnson et al. (2008) thus appears to be afferent (bottom-up). However, since the stimuli were similar, any efferent processing that affected the recordings from Hornickel et al. (2009) should also have affected the Johnson et al. (2008) recordings. Skoe et al. (2011) developed a "cross-phaseogram" from the time-varying cross-power-spectral-density between two ABR recordings. When analyzed in time-frames, the outcome was a spectrogram-like representation of the phase-lag as a function of time and frequency. It allowed for a more detailed investigation of which part of the stimuli caused the peak-latency difference observed by Johnson et al. (2008).

A crucial stage in simulating ABR latencies is the cochlear filter stage and its tuning within the model (Rønne et al., 2012, and 3). Broad cochlear filter tuning, often associated with loss of OHC functionality, is believed to lead to shorter wave-V latencies (e.g. Elberling, 1976; Folsom, 1984). However, in subjects with an audiometric threshold within "normal hearing" (<20 dB HL) there is still a considerable variation in tuning. In a recent study Elberling et al. (2010) showed that the traveling-wave delay is highly individual. The traveling wave delay is also believed to be dependent on the cochlear tuning, and it can be shown (see calculation in section 8.2.5) that subjects with broader tuning in a group of normal-hearing subjects can have Q-values that are less than half the Q-values of subjects with sharper but still normal tuning. The possible consequence of different filter tuning on the simulations of the syllable-evoked phase-shifts will be investigated in this study.

In the present study, a phenomenological ABR model was developed based purely on bottom-up afferent processing. The developed ABR model was similar to the model of Rønne et al. (2012) (developed in chapter 3); however, the AN model used to create the summed activity pattern was updated from Zilany and Bruce (2007) to

Zilany et al. (2009), as the IHC-AN synapse adaptation of the latter AN model is more precise for long-duration syllable-stimuli.

Using the ABR model to simulate syllable-evoked ABRs, two questions were addressed in the study: a) Can the ABR model, being purely afferent, simulate key features of the syllable-evoked responses, and b) can the difference in the recorded cross-phaseogram between normal and learning-impaired children (Skoe et al., 2011) be explained by potential cochlear tuning differences between the groups? To evaluate the first question Skoe et al. (2011)'s cross-phaseograms was used to assess three hypotheses that can be deduced from experimental observations made by Johnson et al. (2008). First, differences in the frequency content of $f_2$ between the syllable-stimuli, should results in components of the evoked-ABRs being differently delayed due to the tonotopic mapping. This should be seen as phase-shifts in the cross-phaseograms. Second, as the differences in $f_2$ diminish over the course of the response, the phase-shifts observed in the cross-phaseogram should vanish completely when steady state is reached. Third, due to the phase-locking properties of the IHCs (upper limit of phase-locking), neural encoding consists largely of phase-locking to frequencies below $f_2$. This leads to phase-locking to the envelope rather than the fine-structure at and above the $f_2$ frequencies. This should result in phase-shifts observed in the cross-phaseogram at frequencies well below the $f_2$.

The second question will be addressed by changing the tuning of the model and evaluating the simulations based on models with broad versus sharp tuning, however still representing limits of normal hearing. The cross-phaseograms will be used to evaluate whether a systematic change in the phase-shift between the syllable-evoked ABRs can be obtained by altered tuning such that, for instance, broad tuning systematically leads to smaller phase-shifts between the syllable-evoked ABRs.

## 8.2 Method

### 8.2.1 ABR model

Figure 8.1 shows the structure of the ABR model used in this study. The model was similar to the model of Rønne et al. (2012) (see also chapter 7). However, the AN model used to compute the summed activity pattern was updated such that the Zilany et al. (2009) AN model was used instead of the Zilany and Bruce (2007). This update was made as the Zilany et al. (2009) has an improved IHC-AN stage producing more realistic adaptation properties. As the syllable-stimuli are of longer duration, a precise adaptation is beneficial. The change of the AN model required a recalculation of the unitary response (UR). Fig. 8.2 shows the UR (based on standard cochlear filter tuning) calculated similar in Rønne et al. (2012) as the deconvolution of a 95.2 dB peSPL grand average click-evoked ABR recording (Elberling et al., 2010; Rønne et al., 2012) and the summed activity pattern obtained by simulating the response to an identical click-stimulus.

The simulated ABRs were at the output filtered with a 2nd order band-pass filter with cutoff frequencies at 70 Hz and 2 kHz. These filter settings were identical to the output filters of Hornickel et al. (2009) and Skoe et al. (2011).

### 8.2.2 Stimuli

Synthetic /ba/, /da/ and /ga/ syllables (Hornickel et al., 2009; Skoe et al., 2011) was used, that only differ in the frequency content of the second formant, $f_2$, of the first 60 ms, corresponding to the consonant part of the stimuli. The second formants decrease in the [ga] stimulus from 2480 Hz, in the [da] from 1700 Hz and increased in the [ba] stimulus from 900 Hz, reaching a steady-state frequency (corresponding to the /a/ part of the syllable) of 1240 Hz in all 3 stimuli. The /a/ vowel-part of the syllables was the same for the three syllables, consisting of the formant frequencies $f_0$ = 100 Hz, $f_1$ = 720 Hz, $f_2$ = 1240 Hz, $f_3$ = 2500Hz, $f_4$ = 3300 Hz, $f_5$ = 3750 Hz and $f_6$ = 4900 Hz. All three stimuli were calibrated to have a root-mean-square (RMS) level of 1, and

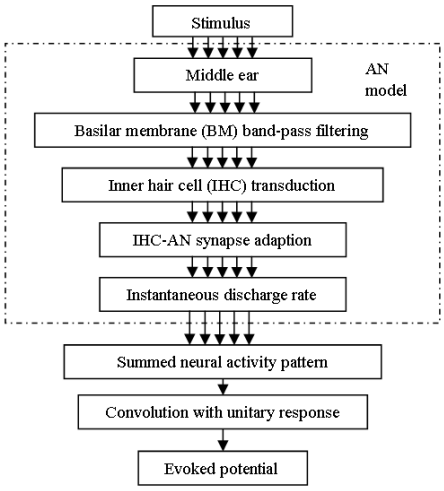



Figure 8.1: Schematic structure of the ABR model. 500 AN fibers tuned to different CFs are individually simulated by the AN model. The summed activity pattern is convolved with a unitary response and represents the simulated ABR to a given stimulus.

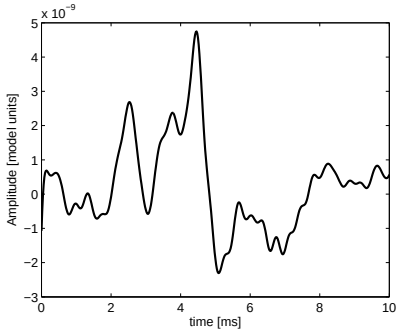



Figure 8.2: The unitary response, calculated as the deconvolution of the summed activity pattern and the Elberling et al. (2010) grand average click. The cochlear tuning of the AN model is the Shera et al. (2002) tuning.

were presented to the model at a level corresponding to 80 dB SPL, which was also used int the study by Skoe et al. (2011).

### 8.2.3 Cross-phaseogram

Skoe et al. (2011) proposed a cross-phaseogram to illustrate the phase-differences and thus the time delays between two ABR recordings. The first step in the procedure was to divide the two recordings into time frames of 20 ms, starting with the first frame at t = -40 ms. Each successive frame started 1 ms later than the previous one, creating an overlap of 19 ms. A Hanning window of 20 ms length, including 10 ms onset and 10 ms offset ramps was applied to each frame, resulting in a 3 dB main lobe width of 141 Hz. The cross power spectrum density, i.e. the power spectrum density of the cross correlation, was computed between each pair of frames from the two recordings. An artificial frequency resolution of 4 Hz was obtained by zero padding, effectively acting as a smoothing operation. Finally, the unwrapped phase (in radians) was extracted and plotted as a function of time (midpoint of the 20 ms frames) and frequency.

Skoe et al. (2011) also proposed the average phase-shift to simplify the cross-phaseogram into a single number that could be compared to other measures, such as the psychoacoustic speech-in-noise performance. The average phase-shift (in $\pi$ radians) was calculated on the formant transition period (15 to 60 ms) of the syllable-evoked ABR in the frequency range of 70 to 1100 Hz.

### 8.2.4 Weighted cross-phaseogram

The cross-phaseogram weights time-frequency bins with little activity as high as bins with much activity. This limits the use of the cross-phaseogram as it is impossible to distinguish between time-frequency bins of presumable little importance due to low activity from bins of major importance due to large activity. A weighted cross-phaseogram is therefore suggested in this section. The first step in the procedure was to derive the energy in similar time-frequency bins as those chosen in the Skoe et al. (2011) cross-phaseogram (Fig. 8.5). Each of the two syllable-evoked ABRs were thus

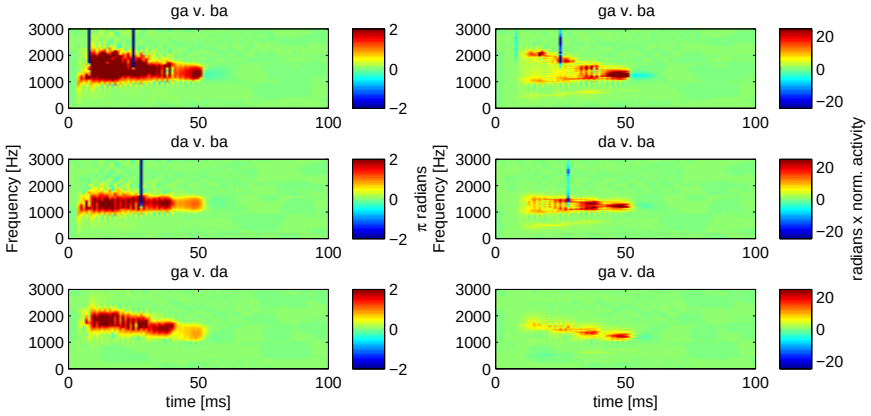



Figure 8.3: The three left panels show cross-phaseogram representations of the three comparisons between the syllable-stimuli. Warm colors indicate that the syllable mentioned first in the respective title phase-leads the other. The time axis refers to the center point of the 20 ms time frame. The three right panels, show weighted cross-phaseograms of the same stimuli-comparisons.

divided into 20 ms frames with 19 ms overlap, and the fast Fourier transform (fft) was calculated with a frequency resolution of 4 Hz. The two resulting matrices were summed and normalized with the average bin activity. This matrix was then multiplied bin-per-bin with the original cross-phaseogram. The reason for the normalization of the activity matrix was to create a weighted cross-phaseogram that highlights the phase-shifts and does not just express the overall activity.

Fig. 8.3 displays both the cross-phaseograms (left) and the weighted cross-phaseograms (right) for the different stimulus pairs. Each time-frequency bin represents the corresponding phase lead (warm colors) or lag (cold colors) of the first syllable-stimulus in the title over the second. The period from 15 to 60 ms shows the formant transition period, the period after 60 ms the steady state part of the response (Skoe et al., 2011). Both sets of figures (left and right panels) show that the phase-shifts between the stimuli are in the frequency region above 1 kHz.

### 8.2.5 Variability of cochlear filter tuning

Cochlear filter tuning and BM delay are inherently related (Folsom, 1984; Eggermont, 1979; Bentsen et al., 2011), such that broader filters lead to shorter delays. Elberling and Don (2008) measured derived-band latencies from a total of 81 normal-hearing subjects (hearing thresholds < 15 dB HL), at four different band center frequencies (bCF; 710, 1400, 2800 and 5700). ABR wave-V latency and a inter-subject standard deviation (SD) were derived. The BM delay was achieved by subtracting the wave I-V delay (4.1 ms) and the synaptic delay (1 ms), see table 8.1. A representation of the variation of cochlear filter tuning in normal-hearing subjects can be obtained from the mean latencies $\pm$ 1 standard deviation. The stimulus of Elberling and Don (2008) was a click presented at approximately 90 dB peSPL.

Eggermont (1979) derived a theoretical relation between the cochlear filter tuning, $Q_{10}$, and the average number of cycles in the impulse response up to the latency (minus 1 ms of synaptic delay) of the derived band CAP, $N_{av}$;

$$N_{av} = \frac{0.5}{\pi^2}\left(\frac{5(1+\gamma)(2+\gamma)}{12\gamma}Q_{10} - 1\right)\left(2 + ln\frac{5(1+\gamma)(2+\gamma)}{12\gamma} + lnQ_{10}\right) \tag{8.1}$$

where $N_{av}$ can be calculated as $(CF/1000) * \tau_{CF}$, where $\tau$ is the BM latency of at the $CF$. In table 8.1 $N_{av}$ values derived from the mean latencies and from the latencies $\pm 1$ standard deviation are shown ($\pm 2$ SD were also calculated but not shown in the table due to clarity). $\gamma = 2$ is representative of a normal cochlea (Eggermont, 1979), and $Q_{10}$ values can thus be calculated based on the $N_{av}$ values from table 8.1 and equation 8.2.5. To convert the $Q_{10}$ values into $Q_{ERB}$ values, the conversion from Ibrahim and Bruce (2010) was applied:

$$Q_{ERB} = \frac{Q_{10} - 0.2085}{0.505} \tag{8.2}$$

| bCF (Hz) | Mean latency | SD | $N_{av}$ | $N_{av}$ (-SD) | $N_{av}$ (+SD) |
|---|---|---|---|---|---|
| 5700 | 1.17 | 0.32 | 6.7 | 4.8 | 8.5 |
| 2800 | 1.86 | 0.40 | 5.2 | 4.1 | 6.3 |
| 1400 | 2.93 | 0.56 | 4.1 | 3.3 | 4.9 |
| 710 | 4.57 | 0.79 | 3.2 | 2.7 | 3.8 |

Table 8.1: Derived-band latencies and a one standard deviation (SD) from Elberling et al. (2010). The 1 ms synaptic delay has been subtracted from the latencies. The number of cycles in the impulse response up to the bCF latency, $N_{av}$, for the mean latencies and for the mean latency $\pm$ the standard deviation is also shown.

Fig. 8.4 shows the $Q_{ERB}$ values derived from Elberling and Don (2008)'s measured delays $\pm$1 SDs and $\pm$2 SDs. The $Q_{ERB}$'s calculated the mean delays corresponds well with the Shera et al. (2002) estimates of tuning (solid curve). New tuning-curve estimates were obtained from the $\pm$1 SD and $\pm$2 SD based Q-estimates, by multiplying the Shera et al. (2002) estimates by a constant offset. The broader tuning-estimates were obtained by multiplying Shera et al. (2002)'s tuning estimates by 0.80 and 0.60, the sharper tuning-estimates by 1.15 and 1.28. The four suggested tuning curves were implemented in the ABR model. For each simulated condition, a new UR was calculated. The URs were almost identical to the ones presented in Fig. 8.2 and are thus not shown explicitly here.

## 8.3 Results

Figure 8.5 presents cross-phaseograms and weighted cross-phaseograms derived from each of the three possible combinations of the simulated ABRs. Figure 8.6 reproduces the cross-phaseograms presented in Skoe et al. (2011). These results can thus be compared to the simulated cross-phaseograms (left panels of Fig. 8.5). Table 8.2 shows the average phase-shifts obtained in Skoe et al. (2011) and the corresponding values obtained from the simulations presented in Fig. 8.5. Both experimental results and simulations show the largest phase-shift between /ga/ and /ba/, which also differs most in their frequency spectrum. Also, the data and the simulations both show that the phase-shift between /ga/ and /da/ is smaller than the phase-shift between /da/ and /ba/.

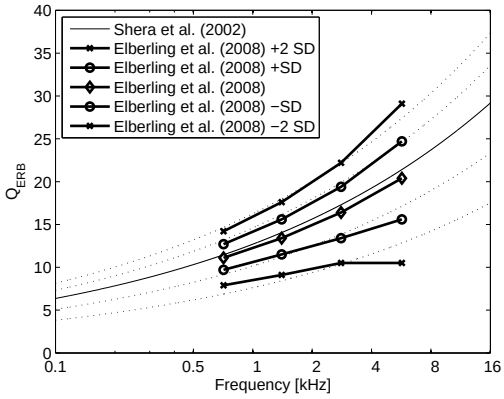



Figure 8.4: $Q_{ERB}$'s calculated based on Elberling and Don (2008)'s measured derived band latencies (diamonds). In circles and triangles, $Q_{ERB}$ estimates based on Elberling and Don (2008)'s measured latencies $\pm 1$ SD an $\pm 2$ SD. Also shown is the Shera et al. (2002) tuning (solid line) which is implemented in the standard ABR model. The alternative tuning curves (dotted lines) are fitted to the Elberling and Don (2008) based tuning ($\pm 1$ SD and $\pm 2$ SD) and also implemented in the model.

| | Skoe et al. (2011) | Simulations | Simulations (weighted) |
|---|---|---|---|
| /ga/-/ba/ | $0.317 \pm 0.040$ | 0.353 | 3.040 |
| /da/-/ba/ | $0.288 \pm 0.031$ | 0.243 | 2.163 |
| /ga/-/da/ | $0.208 \pm 0.028$ | 0.141 | 1.660 |

Table 8.2: Average phase-shifts of Skoe et al. (2011) recordings (left column), simulated average phase-shifts (center column), and weighted average phase-shifts (right column). The average is taken across the region from 15 to 60 ms, and from 70 to 1100 Hz.

The cross-phaseogram in Fig. 8.5 show that the /ga/ phase leads both /da/ and /ba/ (warm colors in the formant transition period of panel 1 and 3), and that /da/ phase leads /ba/ (warm colors in panel 2). Further, the only difference between stimuli was the frequency content of $f_2$, and the observed phase-shifts in the cross-phaseograms can thus be argued to be caused by the stimuli-frequency differences. This is also confirmed by table 8.2 presenting average phase-shifts of the consonant period, where it is seen that /ga/ phase leads /da/ that phase leads /ba/. Further, Fig. 8.5 illustrates that the simulated phase-shifts clearly diminishes over time, and that the phase-shifts are vanished at steady state (>60ms). This shows that the memory of the peripheral non-linearity's, e.g. the IHC-AN synapse adaptation, is short compared to the duration

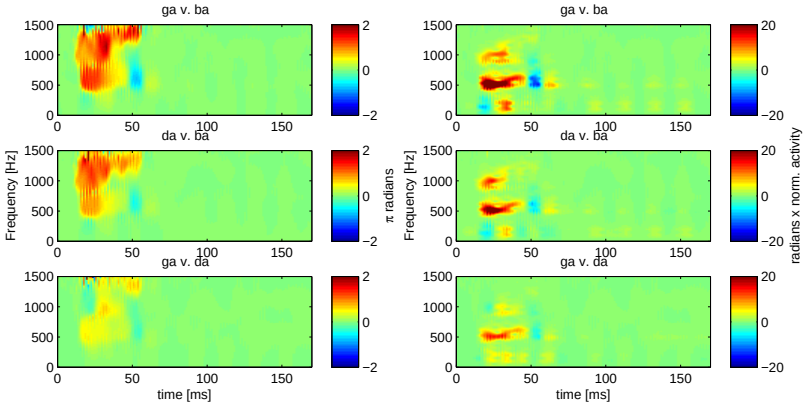



Figure 8.5: Cross-phaseogram (left panels) and weighted cross-phaseogram (right panels) representations of the three comparisons between the syllable-evoked ABRs. Warm colors indicate that the syllable mentioned first in the respective title phase-leads the other. The time axis refers to the center point of the 20 ms time frame. It can be observed that the largest phase-shift is found in the /ga/ v. /ba/ plot, and the least phase-shift is found between /ga/ and /da/.

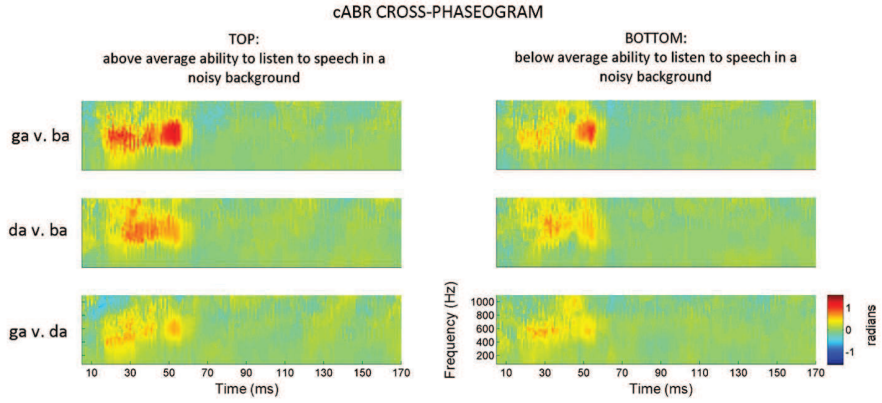



Figure 8.6: Cross-phaseograms from Skoe et al. (2011) of the three comparisons between the syllable-evoked ABRs. Left panels, are calculated based on the top performing group of subjects in a hearing in noise test (HINT). Right panels, presents the worst performers. Note that the frequency range is different from the frequency range presented in Fig. 8.5. © Journal of Neuroscience Methods.

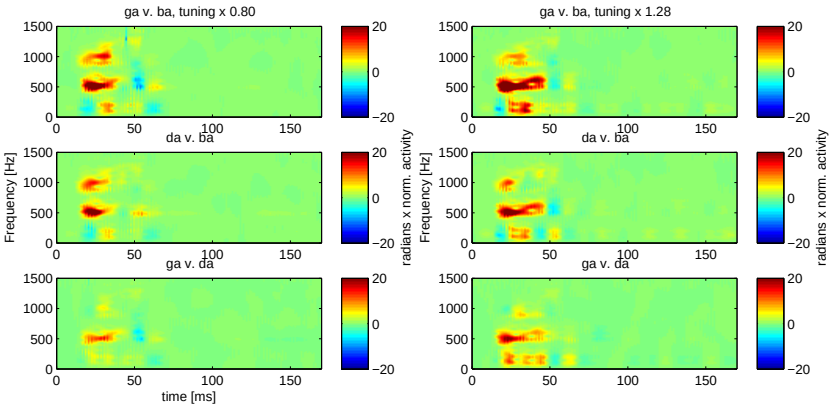



Figure 8.7: Weighted cross-phaseograms for each of syllable combinations, for both broad (x 0.80) and sharp (x 1.28) tuning.

of the stimuli. In Fig. 8.5 it can also be observed that there are phase-shifts up to approximately 1500 Hz, i.e. both below and in the second formant frequency range. However, the weighted cross-phaseograms of Fig. 8.5 (right panels) does not show components at these frequencies, indicating that the high-frequency phase-shifts reflect time-frequency bins with very little activity, and thus potentially little importance. The main trend is thus that the $f_2$-frequency-difference between stimuli, results in phase-differences at frequencies well below the $f_2$. The causes for this finding in the simulations are discussed later.

Figure 8.7 shows weighted cross-phaseograms of the syllable pairs, for simulations of a relatively sharp (x 1.28) and relatively broad (x 0.80) tuning. It can be seen (more bins with warm colors) that the phase-shift is larger for the sharp tuning. In Fig. 8.8, weighted average phase-shifts for all syllable comparisons and all five different tuning-curve implementations are shown. Although the growth of the phase-shift with increasing tuning amount is not monotonic, a trend is observed, where sharp tuning leads to larger phase-shifts. This confirms that the state of the auditory periphery affects the cross-phaseogram and weighted average phase-shifts. The implications for the Hornickel et al. (2009) and Skoe et al. (2011) studies are discussed further below.

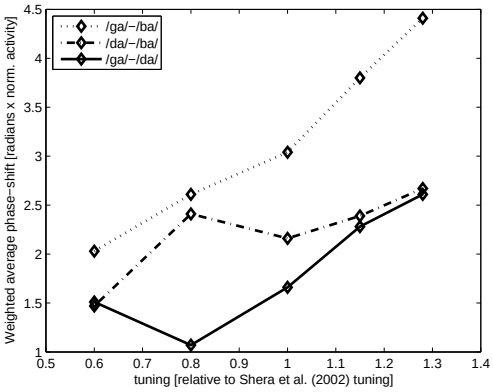



Figure 8.8: Weighted average phase-shifts for each of the syllable combinations, for both broad (0.60 and 0.80), standard (1.00) and sharp (1.15 and 1.28) tuning.

## 8.4 Discussion

### 8.4.1 Unweighted versus weighted cross-phaseogram

The cross-phaseogram and the average phase-shifts was developed by Skoe et al. (2011) and has proven to be a valuable tool for investigating phase-shifts between different frequency components of the recorded (or simulated) ABR. However, the equal weighting of all time-frequency bins limits the value of the average phase-shift Skoe et al. (2011), since a bin with little activity will hardly influence the ABR generation. In fact, a time-frequency bin with little energy is likely to be dominated by measurement noise, and the average measure might thus emphasize noise.

In the simulations presented in this study, noise is not included. This makes a comparison between simulations and data in the terms of the average phase-shift difficult, as a systematic phase-shift at bins with little activity will be included in the simulated average phase-shift, whereas such a phase-shift is likely to be influenced or masked by measurement noise in the data-derived average phase-shift. This could be solved by adding noise to simulations. However, this would imply that the model would no longer be deterministic which has not been considered in the present study.

### 8.4.2 Explaining the presence of phase-shifts below the second formant

In section 8.3 it was shown that second formant differences between stimuli, result in phase-differences at frequencies well below $f_2$. Johnson et al. (2008) argued that this is due to the phase-locking properties of the IHCs (upper limit of phase-locking), and that neural encoding consists largely of phase-locking to frequencies below $f_2$. This leads to phase-locking to the envelope rather than the fine-structure at and above the $f_2$ frequencies. However, the IHC stage of the AN model (effectively modeling the upper limit of phase-locking) consists of a nonlinearity and a low-pass filter with a cut-off frequency at 3 kHz. It is thus unlikely that the IHC stage should be the cause of the simulated phase-shifts at frequencies below $f_2$ in the simulations.

Figure 8.9 visualizes the simulated response to the syllable /da/ in an AN-UR-spectrogram. Each horizontal line represents the output from one AN model, i.e. the response to the stimulus at the respective model CF, convolved with the UR. A summation across CFs will thus yield the simulated ABR (see section 6 for introduction to the AN-UR-spectrogram). It can be observed that most of the energy in the simulations is centered at the onset response and the frequency regions of 100, 200 and 500 Hz (latter one highlighted by the ellipse). It can be seen in Fig. 8.9 that phase-locking clearly occurs in the frequency range up to 1 kHz (this can be observed as the number of peaks at, e.g. 500 Hz is 5 peaks per 10 ms, i.e. the corresponding periodicity). The response at larger CFs exhibits primarily a periodicity corresponding to the fundamental frequency, $f_0$=100Hz, i.e. the envelope of the response (highlighted by the arrows in Fig. 8.9).

To fully explain the presence of phase-shifts below $f_2$ the stimulus and model has to be analyzed step by step. The syllable-stimuli formants (e.g. $f_2$) are modulated at the rate of the fundamental frequency ($f_0$ = 100Hz) and its higher harmonics. Thus, at the characteristic places on the BM of the $f_2$ frequencies, a signal with an $f_2$ carrier frequency modulated by an $f_0$ (+ harmonics) modulation frequency will be processed. Further, the stimulus-level was high (80 dB SPL) causing upwards spread of excitation. The left panels of Fig. 8.10 shows the single channel response at the

output of the filter stage (see Fig. 3.2 for diagram of AN model), tuned to CF = 2405 Hz, in response to the /ga/ stimulus. The time-series shows a periodic signal and its spectrum (shown below it) clearly shows frequency components separated by $f_0$. Further, is it seen that the energy is centered on the CF, but also that upwards spread of excitation results in this channel being excited by contributions from lower frequencies. The IHC stage applies physiologically inspired half-wave rectification and low-pass filtering. The output of the IHC stage is shown in the right panels of Fig. 8.10 for the same CF channel and stimulus. It is seen that the half-wave rectification creates low-frequency energy, as inter-modulation and harmonic distortion products. However, the majority of energy is still centered on the CF. The synapse adaptation stage, that occurs after the IHC stage (see Fig. 3.2 for AN model), has no significant effect on the spectrum of the single channel response. However, the UR that is convolved onto the single channel response effectively acts like a low-pass filter. The left panels of Fig. 8.11 show the UR and the spectrum of the UR. As a convolution is effectively the same as a multiplication in the frequency domain, the UR is effectively acting as a low pass filter with the frequency response corresponding to the spectrum shown in the lower left panel. Thus the resulting simulated single channel potential (shown in the right panels) is limited to low frequencies. The 2405 Hz fiber will thus contribute with frequency components at low frequencies, which will carry the traveling wave delay (and thus phase) of the CF of the fiber. The frequency-differences between stimuli at the $f_2$ frequencies will thus be depicted as phase-differences at the $f_0$ and corresponding harmonic frequencies in the phaseograms.

The outcome measure predicted by the hypotheses of Johnson et al. (2008), that phase-shifts should be found at low frequencies, was thus found in both Skoe et al. (2011)'s experimental analysis and in this study's simulations. However, the simulations showed that the predicted phase-shifts were mainly caused by a combination of upwards spread of excitation and the effective low-pass filtering applied by the UR. Further, it was shown not to be caused by the upper limit of phase-locking, as hypothesized by Johnson et al. (2008).

Note, the UR represents the contributions made from local potentials in the AN and the brainstem to the far-field potential recorded at the electrodes on the scalp of the

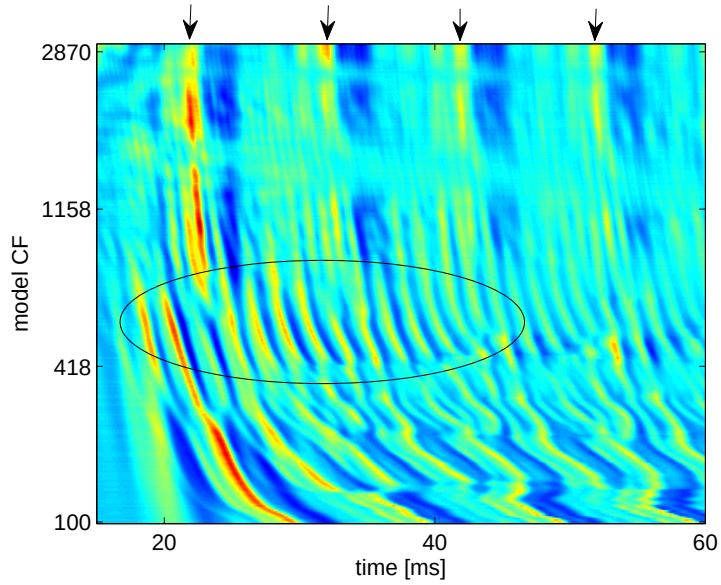



Figure 8.9: AN-UR-spectrogram visualizing the components in frequency range from 100 to 3000 Hz that adds up to form the simulated /da/ evoked ABR. It is created by convolving each of the simulated AN fibers responses with the UR. The ellipse highlights the region with the most activity. At frequencies above approximately 1 kHz, the single fiber response tracks the envelope, i.e. the fundamental-frequency periodicity of 10 ms (indicated by the arrows), rather than the fine-structure of the signal.

subject. The peaks of the UR and the time between them, thus describes ascending places along the auditory pathway were local potentials are generated, that contributes to the ABR potential. The UR is thus not representing the neural encoding in the brainstem but rather the times after onset where a contribution to the surface potential is made. In this study, the UR is seen to limit the transmission of the neurally encoded signal to the recorded surface potential. This is the consequence of the effective low-pass filtering that again is the consequence of the distance between the major peaks, and thus neural generators, of the UR. The effective low-pass filtering is also limiting the utility of this kind of electrophysiology in investigating neural encoding of sound.

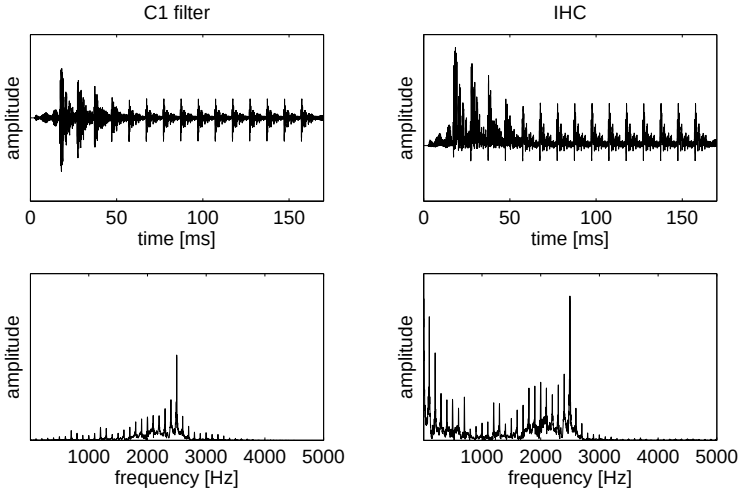



Figure 8.10: Left panels: Time-series and spectrum of the C1 filter output. Right panels: Time-series and spectrum of the IHC stage output. The CF of the fiber was 2405 Hz and the stimulus was /ga/.

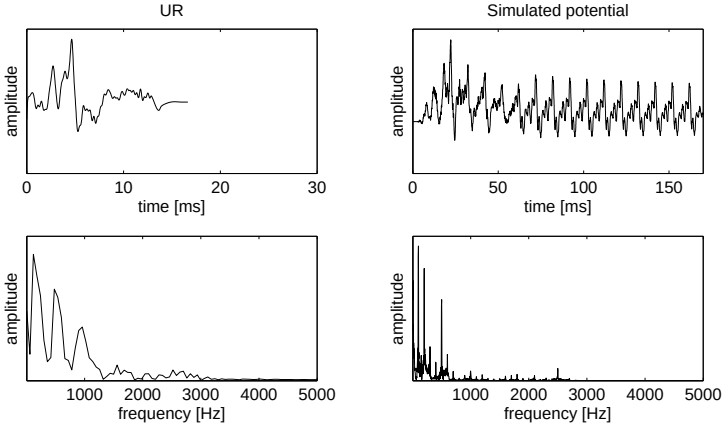



Figure 8.11: Left panels: Time-series and spectrum of the UR. Right panels: Time-series and spectrum of the single fiber response of the model (AN model output convolved with UR). The CF of the fiber was 2405 Hz and the stimulus was /ga/.

### 8.4.3 Limitation of simulating high spontaneous rate fibers

A deviation between simulations and data is the absolute amplitude of the simulated ABRs (not shown). The simulated peak-to-trough amplitude is approximately 0.1 $\mu v$, whereas the measured data in both Johnson et al. (2008) and Hornickel et al. (2009) indicates amplitudes around 0.5 $\mu v$. The reason for the under prediction is the choice of simulating the responses of high spontaneous rate fibers. The high stimulus-level of 80 dB SPL results in saturated fiber responses for high spontaneous rate fibers (Sumner et al., 2002). This saturation reduces the overall amplitude of the response. However, the phase-information in the ABR was the point of interest in this study, not the amplitude of the response, and the choice was therefore to simulate high spontaneous rate fibers, as these has been shown to be mainly responsible for the onset of signal-components. However, if other speech-evoked ABRs with an amplitude-based outcome measure, were to be simulated this limitation would need to be addressed. A possible solution would be to simulate a mixture of both high and low spontaneous rate fibers, to predict both the amplitude and the phase-information as accurately as possible.

### 8.4.4 Implications of changing cochlear tuning on Skoe et al. (2011) conclusions

Hornickel et al. (2009) and Skoe et al. (2011) found correlations between learning-impairments of children, and recorded cross-phaseogram phase-shifts (peak latencies in Hornickel et al., 2009) between syllable-evoked ABRs, such that a small average phase-shift was an indication of learning-impairment. A basic assumption of Hornickel et al. (2009) was that the two groups of hence normal and learning-impaired children have equally good peripheral hearing. Hornickel et al. (2009) argued that this was the case as all subjects had audiometric thresholds below 20 dB HL and had normal ABR wave-V latencies. The wave-V latency was measured as an indication of the state of the cochlear tuning, as broad cochlear tuning are believed to lead to shorter wave-V latencies (e.g. Elberling, 1976; Folsom, 1984). However, in a recent study Elberling and Don (2008) showed that the traveling wave delay was highly

individual. The traveling wave delay is also thought to be dependent on the BM tuning, and it was in the present study suggested that the broadest BM tuning in a group of audiometric-wise normal-hearing subjects can have a Q-value that is less than half the Q-value of the sharpest BM tuning. Given the possible variation of "normal" BM tuning an alternative explanation for the Hornickel et al. (2009) results can be hypothesized. A broad cochlear tuning leads to shorter peak-latencies for all three stimuli. Further, do the traveling-wave delay decrease logarithmically with increasing stimulus frequency (e.g. Neely et al., 1988; Elberling et al., 2010). A broad tuning would thus lead to a decreased difference between the ABR peaks, and thus a smaller phase-shift. Phase-shift differences similar to the one Skoe et al. (2011) finds between the groups of normal and learning-impaired children, could thus be hypothesized to also be found when measuring ABRs to two normal-hearing groups but with different cochlear tuning.

The results from this modeling study showed that there is indeed a relation between filter tuning and weighted averaged cross-phaseogram values, where sharper tuning leads to larger phase-shifts. Although this relation was not strictly monotonic it do indicate that the phaseograms are sensitive to changes in the auditory periphery. Whether this finding offers an alternative explanation for the results of Hornickel et al. (2009) and Skoe et al. (2011) are, however, questionable. That would require the assumption that the group of learning-impaired children, had significantly overall broader cochlear tuning than the normal children. Although this hypothesis is not unlikely, this study cannot by any chance verify such a claim. That would require a major study, where the cochlear tuning of learning-impaired and normal subjects were measured carefully, and correlated with weighted average phase-shifts. The conclusion of this part of this study is thus, that the huge spread of normal-hearing cochlear-tuning, in the simulations, leads to a huge spread in weighted average phase-shifts. Skoe et al. (2011) showed that average phase-shifts was related to learning-impairment. Further, did Skoe et al. (2011) conclude that the correlation between learning-impairment and average phase-shifts show plasticity of brainstem. This conclusion was based on the assumption that the state of the auditory periphery was equal (i.e. normal hearing) in both groups. However, this study has indicated, that the cochlear tuning of the normal-hearing subjects could have an effect on the average

phase-shift, and do thus challenge the underlying assumption of the conclusions from Hornickel et al. (2009) and Skoe et al. (2011). Further, this study has shown that the use of audiograms and click-evoked ABR wave-V latencies are unlikely to be precise enough to claim that the cochlear tuning are similar between two groups.

## 8.5 Summary and conclusion

This study evaluated the performance of an ABR model to simulate ABR responses to three synthetic syllables. The ABR model was shown to predict phase-shifts between the responses to the three syllable stimuli. It was also shown that the model accounts for these phase-shifts which diminish over time, as the spectral differences between the stimuli also decrease, and that there are no differences in the steady-state part of the responses. The model also correctly described that the frequency-region of the response that were mainly phase-shifted was well below the frequency-region that differed between the three stimuli. Based on the simulations it was shown that this phase-shift was mainly due to upwards spread of excitation and effective low-pass filtering applied by the UR and not the consequence of the upper limit of phase-locking as hypothesized by Johnson et al. (2008). Furthermore, it was shown that altering the cochlear tuning influenced the simulated phase-shifts, illustrating that the state of the auditory periphery is crucial when analyzing responses based on the cross-phaseogram. The results suggests that the assumption of Hornickel et al. (2009) and Skoe et al. (2011), that the peripheral hearing was similar between their two groups of test subjects, might be flawed and the following conclusion, that the larger phase-shifts for the non-learning-impaired children was the consequence of plasticity, might thus be wrong.

# 9

# General discussion

## 9.1 Summary

In this Ph.D. thesis, AEP models based on a convolutive approach were developed, where the response of a nonlinear peripheral model were convolved with a linear UR. The peripheral model simulated single-fiber responses to a given stimulus. The response from 500 individually tuned fibers were summed to form the summed activity pattern, i.e. the activity at the distal end of the AN. This summed activity pattern was then convolved with a linear UR, representing the contributions made to the formation of the far field potential (AEP), from ascending places along the neural auditory pathway in response to the events in the summed activity pattern. The UR thus represented the impulse response of the transmission from the activity at the distal end of the AN to the electrodes attached to the scalp of a test subject. This approach made use of the assumptions that the UR was independent of test subject and stimulus, and unaffected by nonlinear neural processing. Two different peripheral models were used. The Zilany and Bruce (2007) AN model to simulate transiently evoked responses to clicks, tone bursts and chirps, and the Zilany et al. (2009) AN model to simulate evoked responses to amplitude modulated tones and speech syllables. Both AN models were originally fitted to cat data. The Zilany and Bruce (2007) model was humanized by Ian Bruce and colleagues, such that the frequency-dependent cochlear tuning was fitted to the human tuning estimates of Shera et al. (2002), and the middle ear stage was replaced by the human model of Pascal et al. (1998). This humanization was also applied to the Zilany et al. (2009) model. The difference between the two models was the more advanced IHC-AN synapse adaptation stage included in the

Zilany et al. (2009) model. This more precise adaptation was argued to be important when longer-duration stimuli like amplitude modulated tones or syllables were used.

The ABR model developed in Rønne et al. (2012) was based on the Zilany and Bruce (2007) AN model and a UR covering the first 10 ms of neural processing, i.e. including the ABR wave I-VII. The ABR model was shown to predict the frequency dependence of tone-burst wave-V latencies and the amplitude of wave-V's evoked by clicks and chirps at different stimulus-levels and chirp sweeping rates. However, the ABR model under estimated the stimulus-level dependence of wave-V latencies. An alternative ABR model, using the DRNL model as peripheral model, was also considered (see Rønne et al., 2011) to investigate whether the under estimation of the level-dependence of click-latencies was bound to the structure of the AN model. However, the DRNL-based ABR model was not found to improve predictions. The models capability to simulate ASSRs was also evaluated. This was done as part of the Rønne et al. (2012a) study, where the possibility of using ASSRs to assess cochlear compression was evaluated both experimentally and in simulations. The ASSR model was based on the Zilany et al. (2009) AN model and a UR covering the first 80 ms of processing, thus including the middle-latency response (MLR) components. The model was shown to be able to predict the main trends of ASSRs to a wide range of stimulus levels and modulation depths. However, the model failed to predict the slight compression observed in the experimentally measured modulation-growth function (Rønne et al., 2012a). The model accounted for on-frequency level-growth compression similar to what would be expected. However, when simulating responses from all 500 fibers, the mixture of on- and off-frequency contributions provided a weaker compression than experimentally measured. In chapter 8 an ABR model was developed based on the Zilany et al. (2009) AN model and a UR covering the first 10 ms of the neural processing, with the purpose to evaluate whether the model could simulate responses to complex stimuli. This model was used to simulate the response to speech syllables. One key prediction was that the phase-shifts between two ABRs evoked by two different syllables were correctly accounted for frequencies significantly lower than the frequency content that differed between the two syllables. This simulation was explained as resulting from upwards spread of excitation and the effective low-pass filtering applied by the UR. The effect

of variation of cochlear tuning within what could be expected from a group of normal hearing test-subjects was also investigated. Here it was found that sharper tuning generally led to larger phase-shifts. Based on the assumption that the peripheral hearing was equal between groups, Skoe et al. (2011) argued that the difference in recorded phase-shift between two groups of normal and learning-impaired children was caused by plasticity of the brainstem. However, the conclusion of the simulations from the present study was that the variation in normal-hearing tuning is large enough to cause significant phase-shifts, and the underlying assumption of Skoe et al. (2011)'s conclusion might thus be incorrect.

This thesis also comprised two experimental studies. One of them investigated whether the higher amplitude of an ABR evoked by a rising chirp compared to a click was mainly a consequence of the better alignment of the low-frequency (<1500 Hz) versus the high-frequency (>1500 Hz) components. Although both regions were found to contribute to the ABR, the region with the largest additional contribution to the chirp-evoked compared to the click-evoked ABR was the low frequencies. In the other experimental study, it was investigated whether the ASSR could be used to assess human cochlear compression. The conclusion was that both the level-growth function and the modulation-growth function could be used to obtain an estimate of cochlear compression. However, the modulation-growth function required the double amount of data and had inherently more noise associated. Thus, the clear recommendation was to use the level-growth function in future work both clinically and in research. One interesting finding was the slight compression observed when plotting the modulation-growth function on double logarithmic scales. According to the developed analytical model and the ASSR model, no compression should have been observed this way. The result remained unexplained.

## 9.2 Revisiting assumptions of the convolutive approach to modeling

The modeling work of this study was built upon the convolutive approach assuming linear superposition, where a nonlinear summed activity pattern was convolved with a linear UR. The UR was assumed to be independent of stimulus-type (level, frequency and fluctuations), independent of subjects, unaffected by top-down efferent processing as training, and unaffected by bottom-up nonlinear neural processing.

A UR with level- and frequency-dependence, as proposed by Chertoff (2004) has already been discussed in section 3.6.3 and 4.5. However, the UR could also be thought to be dependent on temporal fluctuations/modulations of the stimulus. In chapter 7, it was reported that a slight compression was observed when recording ASSR modulation-growth functions and plotting them on double logarithmic scales. A slope of 1 was predicted by the ASSR model, but a slight compression was observed experimentally (slope = 0.78). It was suggested that the modulation gain, reported by Joris and Yin (1992) for single-fiber cat AN responses, could be the cause of the compression, as the modulation gain describes how synchrony[1] is increased in the neural representation of the AN. However, the modulation gain was included in the AN model and could thus not explain the found compression (Zilany et al., 2009). Joris et al. (1994) reported a further increased neural synchronization in the AVCN compared with the synchronization in the AN fibers. They argued that this was due to the convergence of inputs from two or more AN fibers on an AVCN cell that require coincident input spikes before firing (Joris et al., 1994). Malone et al. (2010) showed that the synchronization is further increased at ascending places along the auditory pathway. An increased synchronization represents nonlinear processing and thus is not described by the linear UR. Future work could be to implement a neural stage where the increased synchronization could be accounted for in the framework of the present AEP model. Such an extra neural stage could potentially improve the

[1] Synchrony measures how densely nerve-firing is clustered around the peaks of the envelope response (Malone et al., 2007)

simulations, such that the slight compression found in the logarithmically plotted ASSR modulation-growth functions could be explained.

Another basic assumption underlying the linear UR of the present AEP model is that the model is independent of test subject and independent of time. In chapter 3, subject independence was investigated and all simulations were rerun using individually estimated UR functions from three different subjects. This resulted in small changes to the overall simulated response amplitudes and introduced an individual latency offset. However, the shape of the UR and the distance between peaks remained the same, as expected. This investigation was though only interested in the first 5 ms of the UR, i.e. up to wave-V. Furthermore, all subjects were young normal-hearing adults. A test of whether higher neural stages, potentially affecting wave-V and higher generation sites, differed between individual subjects was never conducted. Such a potential neural difference could arise from brainstem plasticity, i.e. physiological changes to the brainstem processing due to learning. A potential effect of plasticity was described by Hornickel et al. (2009) and Skoe et al. (2011), where degrees of learning impairment were found to correlate with electrophysiological ABR phase-shift measures. It was argued that the reason for the correlation was that the normal-learning children were better trained and thus showed plasticity of the brainstem, i.e. that auditory training had resulted in physiological changes of the brainstem. If this was true, a general across-subject UR would not be reflective of the individual differences in neural processing. However, plasticity is the effect of long-term learning in the range from weeks to several years, and the consequence is thus that the UR does not necessarily need to be non-linear to simulate these differences. Rather, the consequence is that the UR should be calculated for each individual subject and could benefit from frequent recalculations (to anticipate plasticity of the brainstem over time).

## 9.3 Limitations of the present AEP model

The AEP model of this study was shown to be limited with regards to two different sets of simulations. The first was the level-dependence of click-evoked ABR latencies where the predicted slope of the latency-growth function was -0.015 ms/dB compared

to the slope of -0.05 ms/dB found in literature. The second limitation was the ASSR level-growth compression, yielding a compression ratio of 0.48 dB/dB compared to experimentally measured compression ratios of 0.2 dB/dB. The underestimation of the click-latencies was investigated in Rønne et al. (2011), with focus on the influence of the auditory periphery. It was found that the major contributor to click-latencies was the tuning of the cochlear filters and, to a lesser degree, the IHC-AN synapse adaptation. Therefore, the conclusion from that study was that the filter tuning at high stimulus levels and high stimulus frequencies might have been incorrect.

The under-estimated ASSR compression ratio was found in Rønne et al. (2012) to be a consequence of on-frequency compression and off-frequency linearity. The on-frequency compression was shown to have a compression ratio of 0.2, i.e. similar to the experimentally recorded compression. However, when mixed with off-frequency linear contributions, the mixture demonstrated compression with a ratio of 0.48. Three suggestions for this disparity were made: 1) The filter tuning of the model could be imprecise, such that the mixture of on- and off frequency contributions were wrong. An updated implementation would result in either a stronger on-frequency compression or a suppression of off-frequency contributions, for instance by making the filter skirt roll-off sharper. 2) The potentially increased synchrony in the AVCN could also affect the cochlear compression measured by ASSR, as the neural synchrony has been shown to be stimulus-level dependent. The increased synchrony is though not monotonically dependent on the stimulus-level, and can thus not be a major contributor to a simulated compression that is too low over the entire compressive stimulus-level region. 3) In the ASSR study, only low-spontaneous rate fibers were simulated. This was done as high-spontaneous rate fibers were shown to be saturated at most stimulus levels. However, a saturation represents effectively an extreme compression. An appropriate mixture of low- and high-spontaneous rate fibers, could thus potentially increase the on-frequency compression, such that that the mixture of on- and off-frequency contributions would be changed, and an effectively higher compression could be obtained. Additional simulations showed that simulating the response of high-spontaneous rate fibers led to a saturated on-frequency response as expected. However, as the level of the saturated on-frequency responses were low (due to the saturation), the off-frequency contributions (which were not saturated)

were inherently given more relative weight. The resulting mixture of extreme on-frequency compression and linear off-frequency contributions with higher weight resulted in a compression ratio very close to the originally simulated compression ratio of 0.48. This does thus likely neither provide an explanation for the weak simulated compression.

Common for the two main limitations of the AEP model is thus the uncertainty about the implemented cochlear tuning. It has not been within the scope of this Ph.D. to update the cochlear filters, it has rather been the scope to investigate the limitations of the current knowledge and the present model. However, a future study should focus on getting the filter tuning accurately modeled. At present, uncertainties remains regarding the tuning ($Q$-values) at high stimulus-levels and high stimulus-frequencies and further regarding the slope of the filter skirts, i.e. the part of the filter description not included in the $Q_{10}$ value.

### 9.3.1 Modeling high- versus low-spontaneous rate fibers

Throughout this study, either high or low-spontaneous rate fibers have been modeled. At no point has a mixture of low- and high-spontaneous rate fibers been attempted. The low-spontaneous rate fibers show slow recovery after stimulation whereas high-spontaneous rate fibers recover faster (Relkin and Doucet, 1991), making the high-spontaneous rate fibers important when simulating timing and onset responses. Further, the high-spontaneous rate fibers show saturating response characteristics for increasing stimulus level, whereas the low-spontaneous rate fibers show a linear growth (Winter et al., 1990). It thus seems evident that low-spontaneous rate fibers are responsible for encoding high stimulus-level signals, whereas the high-spontaneous rate fibers encode low stimulus levels and onsets of signals. Thus, to be able to simulate all aspects of AEPs evoked by fluctuating stimuli, like AM signals or syllables, the inclusion of a mixture of low and high spontaneous-rate fibers is needed. A starting point for a future inclusion of low- and high-spontaneous rate fibers would be to determine an appropriate ratio of the number of hence low- and high-spontaneous rate fibers to include in the model, and secondly to ensure that the two types of fibers

have appropriate sensitivity. Thus, the summed activity pattern would consist of 500 channels, each consisting of the sum of a low and high spontaneous rate fiber response.

## 9.4 Perspectives

### 9.4.1 ASSRs as an objective predictor of cochlear compression

Rønne et al. (2012a) investigated the potential use of the ASSR to assess cochlear compression. It was found that measuring compression via the level-growth function was possible on a group basis for normal-hearing subjects. The measurement of compression at one CF took approximately 30 minutes. Han et al. (2006) showed that ASSRs can be recorded at four different CFs simultaneously. Therefore, ASSRs could potentially be a fairly fast (< 30 min) method to get a broad overview (at four CFs) of the cochlear compression. However, there are still important questions that need to be addressed before such a method would be ready for clinical usage. First, Rønne et al. (2012a) showed that compression could be assessed on a group basis, but it was never shown that the method also was reproducible and accurate on an individual subject level, which is crucial if the method should be applied in clinical diagnostics. Second, it was neither shown that the method works with hearing-impaired subjects. In hearing-impaired subjects with resulting broader filters, the hearing threshold will typically also be elevated. Further, it is difficult to raise the stimulus-level as the test needs to be restricted to the compressive region of the cochlear I/O function of (approximately 40 to 90 dB SPL). Therefore, the ASSR recording will be carried out closer to threshold. This could lead to a weaker neural signal and thus more noise-prone recordings. It should therefore be tested whether the ASSR can be recorded on individual hearing-impaired subjects as well. Finally, the question is what the information of the state of the cochlear compression in a hearing-impaired subject can be used for in technical application. Currently, no hearing aid or cochlear implant manufacturer uses such information in their fitting procedures. Therefore studies on how to use the information should also be undertaken in the future. A reproducible

ASSR test of individual local cochlear compression would be a major benefit to both the research community and the outside world.

### 9.4.2 Electrophysiological correlate of speech perception

In chapter 8, Skoe et al. (2011)'s cross-phaseogram was introduced as a method to visualize the difference in ABR recordings between two syllable-evoked ABRs. This cross-phaseogram analysis could be highly interesting for research and clinical purposes, specifically, if it could be used as an electrophysiological correlate of speech intelligibility. This would be the case if it was shown that the weighted average phase-shift correlate with speech-in-noise test results, for a wide variety of stimuli and subjects. Hornickel et al. (2009) and Skoe et al. (2011) have indicated for a very specific set of stimuli, /ba/, /da/ and /ga/ syllables, that this could be the case. However, a series of studies has to be carried out to assess, the sensitivity of the measure, how general the measure is and, finally, how the measure is influenced by hearing impairment, before it can be claimed that cross-phaseogram and the average phase-shift is an electrophysiologically correlate of speech intelligibility. On a short time scale, a first study to carry out could be to determine whether the cross-phaseogram can be generalized to also account for differences between other syllable pairs. It could be hypothesized that the cross-phaseogram is a distance measure between two syllables and, thus, that the larger the average phase-shift is the easier distinguishable would two syllables be. An outcome measure could be a correlate between a psychoacoustic test giving a syllable confusion matrix, and the ABR-based averaged phase-shifts. Further, tests with a series of synthetic syllables, forming a range of stimuli that are morphing from one syllable into another (e.g. Stephens and Holt, 2011) could be interesting. Here the hypothesis that the cross-phaseogram is a distance measure could be tested directly. Furthermore, the sensitivity and repeatability should be tested such that it is investigated whether the cross-phaseogram can be used to assess individual intelligibility.

### 9.4.3 AEP model improvements

The present AEP model is capable of simulating many features of AEPs evoked by both complex and simple stimuli. However, there are still many types of responses this model cannot simulate accurately. Improvements of the model would be highly beneficial for the research community as it would allow the testing of our understanding of the underlying physiology behind AEPs evoked by more complex stimuli.

Suggestions for future improvements of the model have already been made in this thesis, to make the current simulations more accurate. It was suggested that the cochlear tuning might be imprecise and that the simulations of the cochlear compression using ASSR as well as the click-evoked ABR wave-V latency could benefit from an update of this cochlear tuning. Such an update would require reliable data and thus a thorough investigation of tuning at high stimulus levels and high stimulus frequencies, as well as an investigation of the slope of the filter-skirts. Another suggestion was to include a mixture of high- and low-spontaneous rate fibers. This could make the model capable of accurately simulating both the phase and the amplitude of syllable-evoked ABRs. Finally, it was suggested to include an AVCN stage to increase the AM synchrony and thus the modulation gain. This AVCN stage should only influence the components of the UR associated with an onset delay of more than 3-5 ms. This would complicate the AEP modeling as a non-linear stage would be added. Amongst the complications would be that the deconvolutive approach to estimate the UR would become invalid.

The AEP model could also be developed to include higher neural stages. This could be important if complex speech-like stimuli were to be considered. As a starting point the modeling work by Dugue et al. (2010) could be used. Dugue et al. (2010) measured evoked potentials in epileptic patients where the electrodes were implanted in the primary auditory cortex. These data were compared to modeling work based on the DRNL model. Dugue et al. (2010) extended the DRNL model, such that the chopper neurons from the DRNL model, were combined in a coincidence detector argued to simulate the inferior colliculus. These stages were followed by stages simulating the medial geniculate body, the thalamic reticular nucleus and the

primary auditory cortex. The model was shown to be able to account for the temporal-modulation transfer-function data. The model is, however, not directly comparable to the modeling work of this thesis, as the data used to fit the model were recorded from electrodes inside the scalp. Some kind of unit function associated to each of the neural stages should thus also be developed to be able to simulate the scalp-recorded AEPs.

A final improvement of the model would be to simulate the responses from hearing-impaired subjects. This would be highly relevant for studies where clinically relevant stimuli were to be developed. A starting point could be to consider the hearing-impairment related to the loss of OHC functionality. The implementation of OHC loss in the AN model has already been attempted for the cat-fitted version (Zilany et al., 2009). However, the outcome measure were single-fiber AN responses and not scalp-recorded AEPs. Zilany and Bruce (2007) could though inspire a fairly easy implementation of hearing-impairment in the form of broader tuning due to loss of OHC functionality in the AEP model. Whether such an implementation of OHC loss would be sufficient to simulate AEP responses from hearing impaired subjects is unknown, and an evaluation of the capabilities of the hearing-impaired AEP model should thus be carried out.

# References


Agung, K., Purdy, S. C., McMahon, C. M. and Newall, P. (2006), 'The use of cortical auditory evoked potentials to evaluate neural encoding of speech sounds in adults', *Journal of the American Academy of Audiology* **17**(8), 559–572.

Aiken, S. J. and Picton, T. W. (2008), 'Envelope and spectral frequency-following responses to vowel sounds', *Hearing research* **245**(1-2), 35–47.

Akhoun, I., Gallego, S., Moulin, A., Menard, M., Veuillet, E., Berger-Vachon, C., Collet, L. and Thai-Van, H. (2008), 'The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme vertical bar ba vertical bar in normal-hearing adults', *Clinical Neurophysiology* **119**(4), 922–933.

Anderson, S., Parbery-Clark, A., Yi, H.-G. and Kraus, N. (2011), 'A Neural Basis of Speech-in-Noise Perception in Older Adults', *Ear and hearing* **32**(6), 750–757.

Bentsen, T., Harte, J. M. and Dau, T. (2011), 'Human cochlear tuning estimates from stimulus-frequency otoacoustic emissions', *J. Acoust. Soc. Am.* **129**(6), 3797–3807.

Berger, H. (1929), 'Electroencephalogram in humans', *Archiv fur psychiatrie und nervenkrankheiten* **87**, 527–570.

Boettcher, F., Poth, E., Mills, J. and Dubno, J. (2001), 'The amplitude-modulation following response in yound and aged human subjects', *Hearing research* **153**, 32–42.

Bohorquez, J. and Oezdamar, O. (2008), 'Generation of the 40-Hz auditory steady-state response (ASSR) explained using convolution', *Clinical Neurophysiology* **119**(11), 2598–2607.

Burkard, R. and Secor, C. (2002), Overview of auditory evoked potential., *in* 'Handbook of Clinical Audiology', Lippincott, Williams, and Wilkins., pp. 233–248.

Burkard, R., Shi, Y. and Hecox, K. (1990), 'A comparison of maximum length and legendre sequences for the derivation of brain-stem auditory-evoked responses at rapid rates of stimulation', *J. Acoust. Soc. Am.* **87**(4), 1656–1664.

Carney, L. (1993), 'A model for the responses of low-frequency auditory-nerve fibers in cat.', *J. Acoust. Soc. Am.* **93**(1), 401–417.

Carney, L., McDuffy, M. and Shekhter, I. (1999), 'Frequency glides in the impulse responses of auditory-nerve fibers', *J. Acoust. Soc. Am.* **105**(4), 2384–2391.

Chandrasekaran, B. and Kraus, N. (2010), 'The scalp-recorded brainstem response to speech: Neural origins and plasticity', *Psychophysiology* **47**(2), 236–246.

Chertoff, M. (2004), 'Analytic treatment of the compound action potential: Estimating the summed post-stimulus time histogram and unit response', *J. Acoust. Soc. Am.* **116**(5), 3022–3030.

Chertoff, M., Lichtenhan, J. and Willis, M. (2010), 'Click- and chirp-evoked human compound action potentials', *J. Acoust. Soc. Am.* **127**(5), 2992–2996.

Chintanpalli, A. and Heinz, M. G. (2007), 'Effect of auditory-nerve response variability on estimates of tuning curves', *J. Acoust. Soc. Am.* **122**(6), 203–209.

Collura, T. (1993), 'History and evolution of electroencephalographic instruments and techniques', *Journal of clinical neurophysiology* **10**(4), 476–504.

Corless, R., Gonnet, G., Hare, D., Jeffrey, D. and Knuth, D. (1996), 'On the Lambert W Function', *Adv. Comput. Math.* **5**, 329–359.

Dau, T. (2003), 'The importance of cochlear processing for the formation of auditory brainstem and frequency following responses', *J. Acoust. Soc. Am.* **113**(2), 936–950.

Dau, T., Wegner, O., Mellert, V. and Kollmeier, B. (2000), 'Auditory brainstem responses with optimized chirp signals compensating basilar membrane dispersion', *J. Acoust. Soc. Am.* **107**(3), 1530–1540.

Davis, H., Davis, P., Loomis, A., Harvey, E. and Hobart, G. (1939), 'Electrical reactions of the human brain to auditory stimulation during sleep', *Journal of neurophysiology* **2**(6), 500–514.

deBoer, E. (1975), 'Synthetic whole-nerve action potentials for the cat', *J. Acoust. Soc. Am.* **58**, 1030–1045.

Delgutte, B. (1990), 'Physiological-mechanics og psychophysical masking - observations from auditory-nerve fibers', *J. Acoust. Soc. Am.* **87**(2), 791–809.

Dobie, R. and Wilson, M. (2001), 'A comparison of t test, f test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids', *J. Acoust. Soc. Am.* **100**, 2236–2246.

Don, M. and Eggermont, J. (1978), 'Analysis of click-evoked brain-stem potentials in man using high-pass noise masking', *J. Acoust. Soc. Am.* **63**(4), 1084–1092.

Don, M. and Kwong, B. (2002), Auditory brainstem response: Differential diagnosis., *in* 'Handbook of Clinical Audiology', Lippincott, Williams, and Wilkins., pp. 274–297.

Don, M., Ponton, C., Eggermont, J. and Kwong, B. (1998), 'The effects of sensory hearing loss on cochlear filter times estimated from auditory brainstem response latencies', *J. Acoust. Soc. Am.* **104**(4), 2280–2289.

Dugue, P., Le Bouquin-Jeannes, R., Edeline, J.-M. and Faucon, G. (2010), 'A physiologically based model for temporal envelope encoding in human primary auditory cortex', *Hearing research* **268**(1-2), 133–144.

Eggermont, J. (1979), 'Narrow-band AP latencies in normal and recruiting human ears', *J. Acoust. Soc. Am.* **65**(2), 463–470.

Eggermont, J. and Don, M. (1980), 'Analysis of the click-evoked brain-stem potentials in humans using high-pass noise masking. II. Effect of click intensity', *J. Acoust. Soc. Am.* **68**(6), 1671–1675.

Elberling, C. (1976), 'High frequency evoked action potentials recorded from the ear canal in man', *Scandinavian audiology* **5**, 157–164.

Elberling, C., Callø, J. and Don, M. (2010), 'Evaluating auditory brainstem responses to different chirp stimuli at three levels of stimulation', *J. Acoust. Soc. Am.* **128**(1), 215–223.

Elberling, C. and Don, M. (2008), 'Auditory brainstem responses to a chirp stimulus designed from derived-band latencies in normal-hearing subjects', *J. Acoust. Soc. Am.* **124**(5), 3022–3037.

Elberling, C. and Don, M. (2010), 'A direct approach for the design of chirp stimuli used for the recording of auditory brainstem responses', *J. Acoust. Soc. Am.* **128**(5), 2955–2964.

Elberling, C., Don, M., Cebulla, M. and Stuerzebecher, E. (2007), 'Auditory steady-state responses to chirp stimuli based on cochlear traveling wave delay', *J. Acoust. Soc. Am.* **122**(5, Part 1), 2772–2785.

Elberling, C., Kristensen, S. G. B. and Don, M. (2012), 'Auditory brainstem responses to chirps delivered by different insert earphones', *J. Acoust. Soc. Am.* **131**(3), 2091–2100.

Fobel, O. and Dau, T. (2004), 'Searching for the optimal stimulus eliciting auditory brainstem responses in humans', *J. Acoust. Soc. Am.* **116**(4, Part 1), 2213–2222.

Folsom, R. (1984), 'Frequency specificity of human auditory brain-stem responses as revealed by pure-tone masking profiles', *J. Acoust. Soc. Am.* **75**(3), 919–924.

Frisina, R., Karcich, K., Tracy, T., Sullivan, D., Walton, J. and Colombo, J. (1996), 'Preservation of amplitude modulation coding in the presence of background noise by chinchilla auditory-nerve fibers', *J. Acoust. Soc. Am.* **99**(1), 475–490.

Fromm, B., Nylen, C. and Zotterman, Y. (1935), 'Studies in the mechanism of the Wever and Bray effect.', *Acta Otolaryngologica* **22**, 477–486.

Galambos, R., Makeig, S. and Talmachoff, P. (1981), 'A 40 Hz auditory potential recorded from the human scalp', *Procedings of the national academy of sciences of the United States of America-Biological sciences* **78**(4), 2643–2647.

Geisler, C., Frishkopf, L. and Rosenblith, W. (1958), 'Extracranial responses to acoustic clicks in man', *Science* **128**(3333), 1210–1211.

Glasberg, B. R. and Moore, B. C. (1990), 'Derivation of auditory filter shapes from notched-noise data', *Hearing research* **47**(1-2), 103 – 138.

Goldstein, M. and Kiang, N. (1958), 'Synchrony of neural activity in electric responses evoked by transient acoustic stimuli', *J. Acoust. Soc. Am.* **30**(2), 107–114.

Gorga, M., Kaminski, J., Beauchine, K. and Jesteadt, W. (1988), 'Auditory brainstem responses to tone bursts in normally hearing subjects', *Journal of speech and Hearing research* **31**(1), 87–97.

Greenwood, D. (1990), 'A cochlear frequency-position function for several species - 19 years later', *J. Acoust. Soc. Am.* **87**(6), 2592–2605.

Gutschalk, A., Mase, R., Roth, R., Ille, N., Rupp, A., Hahnel, S., Picton, T. and Scherg, M. (1999), 'Deconvolution of 40 Hz steady-state fields reveals two overlapping source activities of the human auditory cortex', *Clinical neurophysiology.* **110**(5), 856–868.

Hall, J. (1992), *Overview of Auditory Evoked Responses: Past, Present, and Future*.

Han, D., Mo, L., Liu, H., Chen, J. and Huang, L. (2006), 'Threshold estimation in children using auditory steady-state responses to multiple simultaneous stimuli', *Orl-journal for oto-rhino-laryngology and its related specialities* **68**(2), 64–68.

Hansen, P. C. H. (1998), 'Regularization tools. a matlab package for analysis and solution of discrete ill-posed problems,', *http://www2.imm.dtu.dk/ pch/* .

Hari, R., Hamalainen, M. and Joutsiniemi, S. (1989), 'Neuromagnetic steady-state response to auditory stimuli', *J. Acoust. Soc. Am.* **86**(3), 1033–1039.

Harte, J. (2007), 'Constrained ica for the analysis of high stimulus rate auditory evoked potentials', *7th International Conference on Independent Component Analysis and Signal Separation, 9-12 September 2007, London, UK* .

Harte, J., Elliott, S. and Rice, H. (2005), 'A comparison of various nonlinear models of cochlear compression', *J. Acoust. Soc. Am.* **117**, 3778–3786.

Harte, J., Pigasse, G. and Dau, T. (2009), 'Comparison of cochlear delay estimates using otoacoustic emissions and auditory brainstem responses', *J. Acoust. Soc. Am.* **126**(3), 1291–1301.

Harte, J., Rønne, F. and Dau, T. (2010), 'Modeling human auditory evoked brainstem responses based on nonlinear cochlear processing', *Proceedings of 20th International Congress on Acoustics* .

Heinz, M., Zhang, X., Bruce, I. and Carney, L. (2001), 'Auditory nerve model for predicting performance limits of normal and impaired listeners', *ARLO* **5**(3), 91–96.

Hornickel, J., Skoe, E., Nicol, T., Zecker, S. and Kraus, N. (2009), 'Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception', *Proceedings of the national academy og sciences of the United States of America* **106**(31), 13022–13027.

Ibrahim, R. A. and Bruce, I. C. (2010), Effects of Peripheral Tuning on the Auditory Nerve's Representation of Speech Envelope and Temporal Fine Structure Cues, *in* LopezPoveda, EA and Palmer, AR, ed., 'Neurophysiological bases if auditory perception', Med Elect; Hear Life, pp. 429–438. 15th International Symposium on Hearing, Salamanca, Spain, Jun 01-05, 2009.

Jewett, D. (1970), 'Volume-conducted potentials in response to auditory stimuli as detected by averaging in cat', *Electroencephalography and clinical neurophysiology* **28**(6), 609–&.

Jewett, D., Caplovitz, G., Baird, B., Trumpis, M., Olson, M. and Larson-Prior, L. (2004), 'The use of QSD (q-sequence deconvolution) to recover superposed, transient evoked-responses', *Clinical neurophysiology.* **115**(12), 2754–2775.

Jewett, D. and Williston, J. (1971), 'Auditory-evoked far fields averaged from scalp of humans', *Brain* **94**, 681–&.

John, M. and Picton, T. (2000), 'Human auditory steady-state responses to amplitude-modulated tones: phase and latency measurements', *Hearing research* **141**(1–2), 57–79.

Johnson, K., Nicol, T., Zecker, S., Bradlow, A., Skoe, E. and Kraus, N. (2008), 'Brainstem encoding of voiced consonant-vowel stop syllables', *Clinical neurophysiology* **119**(11), 2623–2635.

Johnson, K., Nicol, T., Zecker, S. and Kraus, N. (2008a), 'Developmental plasticity in the human auditory brainstem', *Journal of neuroscience* **28**(15), 4000–4007.

Joris, P., Bergevin, C., Kalluri, R., McLaughlin, M., Michelet, P., van der Heijden, M. and Shera, C. (2011), 'Frequency selectivity in Old-World monkeys corroborates sharp cochlear tuning in humans', *Procedings of the national academy of sciences of the United States of America* **108**(42), 17516–17520.

Joris, P., Carney, L., Smith, P. and Yin, T. (1994), 'Enhancement of neural synchronization in the anteroventral cochlear nucleus .1. Responses to tones at the characteristic frequency', *Journal of Neurophysiology* **71**(3), 1022–1036.

Joris, P., Schreiner, C. and Rees, A. (2004), 'Neural processing of amplitude-modulated sounds', *Physiological reviews* **84**(2), 541–577.

Joris, P. and Yin, T. (1992), 'Responses to amplitude-modulated tones in the auditory-nerve of the cat', *J. Acoust. Soc. Am.* **91**(1), 215–232.

Junius, D. and Dau, T. (2005), 'Influence of cochlear traveling wave and neural adaptation on auditory brainstem responses', *Hearing research* **205**(1-2), 53–67.

Keefe, D. H., Ellison, J. C., Fitzpatrick, D. F. and Gorga, M. P. (2008), 'Two-tone suppression of stimulus frequency otoacoustic emissions', *J. Acoust. Soc. Am.* **123**(3), 1479–1494.

Kiang, N. (1965), *Discharge patterns of single fibers in the cat's auditory nerve*, Cambridge, Mass., M.I.T. Press.

Kiang, N. (1975), *Stimulus representation in the discharge patterns of auditory neurons.*, in, The Nervous System. Volume 3: Human Communication and Its Disorders. Raven Press, New York.

Kiang, N. (1990), 'Curious oddments of auditory-nerve studies', *Hearing research* **49**(1-3), 1–16.

Killion, M. (1978), 'Revised estimate of minimum audible pressure - where is the "missing 6 dB"', *J. Acoust. Soc. Am.* **63**(5), 1501–1508.

Kim, D. and Molnar, C. (1979), 'Population study of cochlear nerve-fibers - comparison of spatial distribution of average-rate and phase-locking measures of responses to single tones', *J. Neurophysiol.* **42**(1), 16–30.

Krishnan, A., Xu, Y., Gandour, J. and Cariani, P. (2005), 'Encoding of pitch in the human brainstem is sensitive to language experience', *Cognitive brain research* **25**(1), 161–168.

Kuwada, S., Anderson, J., Batra, R., Fitzpatrick, D., Teissier, N. and D'Angelo, W. (2002), 'ources of the scalp-recorded amplitude-modulation following response', *Journal of the American Academy of Audiology* **13**(4), 188–204.

Kuwada, S., Batra, R. and Maher, V. (1986), 'Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones', *Hearing research* **21**(2), 179–192.

Lalor, E. C. and Foxe, J. J. (2010), 'Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution', *European journal of neuroscience* **31**(1), 189–193.

Lopez-Poveda, E. and Meddis, R. (2001), 'A human nonlinear cochlear filterbank', *J. Acoust. Soc. Am.* **110**(6), 3107–3118.

Malone, B. J., Scott, B. H. and Semple, M. N. (2007), 'Dynamic amplitude coding in the auditory cortex of awake rhesus macaques', *Journal of neurophysiology* **98**(3), 1451–1474.

Malone, B. J., Scott, B. H. and Semple, M. N. (2010), 'Temporal Codes for Amplitude Contrast in Auditory Cortex', *Journal of neuroscience* **30**(2), 767–784.

Meddis, R. (2006), 'Auditory-nerve first-spike latency and auditory absolute threshold: A computer model', *J. Acoust. Soc. Am.* **119**(1), 406–417.

Meddis, R., O'Mard, L. and Lopez-Poveda, E. (2001), 'A computational algorithm for computing nonlinear auditory frequency selectivity', *J. Acoust. Soc. Am.* **109**(6), 2852–2861.

Melcher, J. and Kiang, N. (1996), 'Generators of the brainstem auditory evoked potential in cat .3. Identified cell populations', *Hearing research* **93**(1-2), 52–71.

Mø ller, A. and Jannetta, P. (1983), 'Interpretation of brain-stem auditory evoked-potentials - results from intracranial recordings in humans', *Scand. Aud.* **12**(2), 125–133.

Moore, B., Vickers, D., Plack, C. and Oxenham, A. (1999), 'Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism', *J. Acoust. Soc. Am.* **106**(5), 2761–2778.

Murray, J., Cohn, E., Harker, L. and Gorga, M. (1998), 'Tone burst auditory brain stem response latency estimates of cochlear travel time in Meniere's disease, cochlear hearing loss, and normal ears', *American journal of otology* **19**(6), 854–859.

Nayfeh, A. and Mook, D. (1995), *Nonlinear Oscillations*, Wiley Classics Library Edition.

Neely, S., Norton, S., Gorga, M. and W., J. (1988), 'Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli', *J. Acoust. Soc. Am.* **83**(2), 652–656.

Norton, S. and Neely, S. (1987), 'Tone-burst-evoked otoacoustic emissions from normal-hearing subjects', *J. Acoust. Soc. Am.* **81**(6), 1860–1872.

Nuttall, A. and Dolan, D. (1996), 'Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig', *J. Acoust. Soc. Am.* **99**, 1556–1565.

Palmer, A. and Russell, I. (1986), 'Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells', *research* **24**(1), 1–15.

Pascal, J., Bourgeade, A., Lagier, M. and Legros, C. (1998), 'Linear and nonlinear model of the human middle ear', *J. Acoust. Soc. Am.* **104**(3, Part 1), 1509–1516.

Picton, T., Skinner, C., Champagne, S., Kellett, A. and Maiste, A. (1987), 'Potentials-evoked by the sinusoidal modulation of the amplitude or frequency of a tone', *J. Acoust. Soc. Am.* **82**(1), 165–178.

Plourde, G., Stapells, D. R. and Picton, T. W. (1991), 'The human auditory steady-state evoked potentials', *Acta Otolarygol. (Stockh). Suppl.* **491**, 153–160.

Prieve, B., Gorga, M. and Neely, S. (1996), 'Click- and tone-burst-evoked otoacoustic emissions in normal-hearing and hearing-impaired ears', *J. Acoust. Soc. Am.* **99**(5), 3077–3086.

Puria, S. (2003), 'Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions', *J. Acoust. Soc. Am.* **113**(5), 2773–2789.

Recio, A. and Rhode, W. (2000), 'Basilar membrane responses to broadband stimuli', *J. Acoust. Soc. Am.* **108**(5, Part 1), 2281–2298.

Rees, A., Green, G. and Kay, R. (1986), 'Steady-state evoked-responses to sinusoidally amplitude-modulated sounds recorded in man', *Hearing research* **23**(2), 123–133.

Relkin, E. and Doucet, J. (1991), 'Recovery from prior stimulation .1. Relationship to spontaneous firing rates of primary auditory neurons', *Hearing research* **55**(2), 215–222.

Rhode, W. and Recio, A. (2000), 'Study of mechanical motions in the basal region of the chinchilla cochlea', *J. Acoust. Soc. Am.* **107**(6), 3317–3332.

Ribeiro, F. M. and Carvallo, R. M. (2008), 'Tone-evoked ABR in full-term and preterm neonates with normal hearing', *International journal of audiology* **47**(1), 21–29.

Richter, U. and Fedtke, T. (2005), 'Reference zero for the calibration of audiometric equipment using 'clicks' as test signals', *International Journal og Audiology* **44**, 478–487.

Riedel, H., Granzow, M. and Kollmeier, B. (2001), 'Singlesweep- based methods to improve the quality of auditory brainstem responses. part ii: Averaging methods', *Z. Audiol.* **40**, 62–85.

Robles, L. and Ruggero, M. (2001), 'Mechanics of the mamalian cochlea', *Physiological reviews* **81**, 1305–1352.

Rønne, F. and Gøtsche-Rasmussen, K. (2011), Low-frequency versus high-frequency synchronisation in chirp-evoked auditory brainstem responses, *in* 'Proceedings of International Symposium on Auditory and Audiological Research', pp. 275–282.

Rønne, F., Harte, J., Elberling, C. and Dau, T. (2011), Modeling the level-dependent latency of the auditory brainstem response, *in* 'Proceedings of Forum Acousticum'.

Rønne, F., Harte, J., Elberling, C. and Dau, T. (2012), 'Modelling auditory evoked brainstem responses to transient stimuli', *J. Acoust. Soc. Am.* **131**, 3903–3913.

Rønne, F., Harte, J., Møllenbach, S. and Dau, T. (2012a), 'Investigating the potential of auditory steady-state responses to assess loss of cochlear compression', *J. Acoust. Soc. Am.* **submitted**.

Ruggero, M. (1992), 'Responses to sound of the basilar membrane of the mammalian cochlea', *Curr. Opin. Neurobiol.* **2**, 449–456.

Ruggero, M. A. and Temchin, A. N. (2007), 'Similarity of traveling-wave delays in the hearing organs of humans and other tetrapods', *JARO-Journal of the association for research in otolaryngology* **8**(2), 153–166.

Ruggero, M., Rich, N., Recio, A., Narayan, S. and Robles, L. (1997), 'Basilar-membrane responses to tones at the base of the chinchilla cochlea', *J. Acoust. Soc. Am.* **101**(4), 2151–2163.

Russell, I. and Sellick, P. (1978), 'Intracellular studies of hair cells in mammalian cochlea', *Journal of physiology-London* **284**(NOV), 261–290.

Russo, N., Nicol, T., Zecker, S., Hayes, E. and Kraus, N. (2005), 'Auditory training improves neural timing in the human brainstem', *Behavioural brain research* **156**(1), 95–103.

Scherg, M. (1990), Threshold prediction using auditory brainstem response and steady-state evoked potentials with infants and young children., *in* 'Auditory Evoked Magnetic Fields and Electric Potentials', Karger Basel., pp. 40–69.

Scherg, M. and von Cramon, D. (1985), 'A new interpretation of the generators of BAEP waves I-V: results of a spatio-temporal dipole model', *Electroencephalography and clinical neurophysiology* **62**(4), 290–299.

Scherg, M. and von Cramon, D. (1985a), 'Two bilateral sources of the late AEP as indentified by a spatio-temporal dipole model', *Electroencephalography and clinical neurophysiology* **62**(1), 32–44.

Schoonhoven, R., Boden, C., Verbunt, J. and de Munck, J. (2003), 'A whole head MEG study of the amplitude-modulation-following response: phase coherence, group delay and dipole source analysis', *Clinical neurophysiology.* **114**(11), 2096–2106.

Sellick, P., Patuzzi, R. and Johnston, B. (1982), 'Measurement of basilar membrane motion in the guinea pig using the mössbauer technique', *J. Acoust. Soc. Am.* **72**, 131–141.

Serbetcioglu, M. and Parker, D. (1999), 'Measures of cochlear travelling wave delay in humans: I. Comparison of three techniques in subjects with normal hearing', *Acta Oto-Laryngologica* **119**(5), 537–543.

Serpanos, Y., O'Malley, H. and Gravel, J. (1997), 'The relationship between loudness intensity functions and the click-ABR wave V latency', *Ear and hearing* **18**(5), 409–419.

Sharma, A. and Dorman, M. (1999), 'Cortical auditory evoked potential correlates of categorical perception of voice-onset time', *J. Acoust. Soc. Am.* **106**(2), 1078–1083.

Shera, C. A., Guinan, Jr., J. J. and Oxenham, A. J. (2010), 'Otoacoustic Estimation of Cochlear Tuning: Validation in the Chinchilla', *JARO-Journal of the association for research in otolaryngology* **11**(3), 343–365.

Shera, C., Guinan, J. and Oxenham, A. (2002), 'Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements', *Proceedings of the National Academy of Sciences of the United States of America* **99**(5), 3318–3323.

Shore, S. and Nuttall, A. (1985), 'High-synchrony cochlear compound action-potentials evoked by rising frequency-swept tone bursts', *J. Acoust. Soc. Am.* **78**(4), 1286–1295.

Sininger, Y. and Cone-Wesson, B. (2002), Threshold prediction using auditory brainstem response and steady-state evoked potentials with infants and young children., *in* 'Handbook of Clinical Audiology', Lippincott, Williams, and Wilkins., pp. 298–322.

Skoe, E., Nicol, T. and Kraus, N. (2011), 'Cross-phaseogram: Objective neural index of speech sound differentiation', *Journal of neuroscience methods* **196**(2), 308–317.

Søndergaard, P. L., Culling, J. F., Dau, T., Goff, N. L., Jepsen, M. L., Majdak, P. and Wierstorf, H. (2011), Towards a binaural modelling toolbox, *in* 'Proceedings of the Forum Acousticum 2011'.

Song, J. H., Skoe, E., Wong, P. C. M. and Kraus, N. (2008), 'Plasticity in the adult human auditory Brainstem following short-term linguistic training', *Journal of cognitive neuroscience* **20**(10), 1892–1902.

Sparacino, G., Nale, A., Santarelli, R. and Arslan, E. (2004), 'Method for the deconvolution of auditory steady-state responses', *Medical & biological engineering & computing* **42**(4), 569–576.

Stephens, J. D. W. and Holt, L. L. (2011), 'A standard set of American-English voiced stop-consonant stimuli from morphed natural speech', *Speech communication* **53**(6), 877–888.

Sumner, C., Lopez-Poveda, E., O'Mard, L. and Meddis, R. (2002), 'A revised model of the inner-hair cell and auditory-nerve complex', *J. Acoust. Soc. Am.* **111**(5, Part 1), 2178–2188.

Sumner, C., Lopez-Poveda, E., O'Mard, L. and Meddis, R. (2003), 'Adaptation in a revised inner-hair cell model', *J. Acoust. Soc. Am.* **113**(2), 893–901.

Suzuki, T. and Horiuchi, K. (1981), 'Rise time of pure-tone stimuli in brain-stem response audiometry', *Audiology* **20**(2), 101–112.

Swaminathan, J., Krishnan, A. and Gandour, J. T. (2008), 'Pitch encoding in speech and nonspeech contexts in the human auditory brainstem', *Neuroreport* **19**(11), 1163–1167.

Tan, Q. and Carney, L. (2003), 'A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide', *J. Acoust. Soc. Am.* **114**(4, Part 1), 2007–2020.

Tikhonov, A. (1963), 'Solution of incorrectly formulated problems and regularization method', *Dóklady Akademii NAUK SSSR* **151**(3), 501–&.

Warrier, C., Johnson, K., Hayes, E., Nicol, T. and Kraus, N. (2004), 'Learning impaired children exhibit timing deficits and training-related improvements in auditory cortical responses to speech in noise', *Experimental Brain Research* **157**(4), 431–441.

Wegner, O. and Dau, T. (2002), 'Frequency specificity of chirp-evoked auditory brainstem responses', *J. Acoust. Soc. Am.* **111**(3), 1318–1329.

Westerman, L. and Smith, R. (1988), 'A diffusion-model of the transient-response of the cochlear inner hair cell synapse', *J. Acoust. Soc. Am.* **83**(6), 2266–2276.

Wever, E. and Bray, C. (1930), 'Action currents in the auditory nerve in response to acoustical stimulation', *Proceedings of the national academy og sciences of the United States of America* **16**, 344–350.

Winter, I., Robertson, D. and Yates, G. (1990), 'Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibers', *Hearing research* **45**(3), 191–202.

Withnell, R. (2001), 'Brief report: The cochlear microphonic as an indication of outer hair cell function', *Ear and hearing* **22**(1), 75–77.

Zhang, X., Heinz, M., Bruce, I. and Carney, L. (2001), 'A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression', *J. Acoust. Soc. Am.* **109**(2), 648–670.

Zilany, M. S. A. and Bruce, I. C. (2006), 'Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery', *J. Acoust. Soc. Am.* **120**(3), 1446–1466.

Zilany, M. S. A. and Bruce, I. C. (2007), 'Representation of the vowel (epsilon) in normal and impaired auditory nerve fibers: Model predictions of responses in cats', *J. Acoust. Soc. Am.* **122**(1), 402–417.

Zilany, M. S. A., Bruce, I. C., Nelson, P. C. and Carney, L. H. (2009), 'A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics', *J. Acoust. Soc. Am.* **126**(5), 2390–2412.

# Contributions to Hearing Research

**Vol. 1:** *Gilles Pigasse*, Deriving cochlear delays in humans using otoacoustic emissions and auditory evoked potentials, 2008.

**Vol. 2:** *Olaf Strelcyk*, Peripheral auditory processing and speech reception in impaired hearing, 2009.

**Vol. 3:** *Eric R. Thompson*, Characterizing binaural processing of amplitude-modulated sounds, 2009.

**Vol. 4:** *Tobias Piechowiak*, Spectro-temporal analysis of complex sounds in the human auditory system, 2009.

**Vol. 5:** *Jens Bo Nielsen*, Assessment of speech intelligibility in background noise and reverberation, 2009.

**Vol. 6:** *Helen Connor*, Hearing aid amplification at soft input levels, 2010.

**Vol. 7:** *Morten Løve Jepsen*, Modeling auditory processing and speech perception in hearing-impaired listeners, 2010.

**Vol. 8:** *Sarah Verhulst*, Characterizing and modeling dynamic processes in the cochlea using otoacoustic emissions, 2010.

**Vol. 9:** *Sylvain Favrot*, A loudspeaker-based room auralization system for auditory research, 2010.

**Vol. 10:** *Sébastien Santurette*, Neural coding and perception of pitch in the normal and impaired human auditory system, 2011.

**Vol. 11:** *Iris Arweiler*, Processing of spatial sounds in the impaired auditory system, 2011.

**Vol. 12:** *Filip Munch Rønne*, Modeling auditory evoked potentials to complex stimuli, 2012.

The auditory evoked potential (AEP) is an electrical signal that can be recorded from electrodes attached to the scalp of a human subject, when a sound is presented. The signal is believed to reflect neural activity in response to the acoustic stimulation, and is as such well established as a tool to objectively assess the hearing of humans. However, the physiological generation mechanisms of AEPs are a complicated interaction between linear and nonlinear cochlear and neural processes, and are not well understood. The purpose of this thesis was to develop a phenomenological model that could predict key features of recorded AEPs. The model provides an opportunity to investigate the influences of the different stages along the auditory pathway upon the generation of AEPs.

This thesis describes the development of an AEP model capable of simulating click-, tone-burst-, chirp- and syllable-evoked auditory brainstem responses, and auditory steady-state responses (ASSRs). Further, the thesis describes how the ASSR can be used to assess human cochlear compression most effectively.

**DTU Electrical Engineering**
Department of Electrical Engineering

Ørsteds Plads
Building 348
DK-2800 Kgs. Lyngby
Denmark
Tel: (+45) 45 25 38 00
Fax: (+45) 45 93 16 34
www.elektro.dtu.dk

ISBN 978-87-92465-73-3