Downloaded from orbit.dtu.dk on: Dec 18, 2017

Technical University of Denmark

A Model for the representation of Speech Signals in Normal and Impaired Ears

Christiansen, Thomas Ulrich; Poulsen, Torben

Publication date:
2004

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Christiansen, T. U., & Poulsen, T. (2004). A Model for the representation of Speech Signals in Normal and Impaired Ears.

DTU Library
Technical Information Center of Denmark

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

A Model for the Representation of Speech Signals in Normal and Impaired Ears

Thomas Ulrich Christiansen

31st October 2003

This thesis was submitted in partial fulfilment of the requirements for obtaining the degree of Doctor of Philosophy (Ph.D.) in Electronics and Communication at the Technical University of Denmark (DTU)
The work was done from August 2000 through October 2003 at the Acoustic Technology (AT) section, DTU under the supervision of Torben Poulsen (AT).
The public defense was held March 26, 2004.
The work was funded by the Oticon Foundation.
Printed in Denmark by Digitalservice, DK-2970 Hørsholm.
Acoustic Technology report no 84.
ISBN 87-91184-42-8

Thomas Ulrich Christiansen

Resume

En model af den menneskelige auditive periferi gående fra det ydre øre til hørenerve udvikledes. Modellen består af følgende komponenter: det ydre øres overførselsfunktion, mellemørets overførselsfunktion, model af basilar membran hastighed, model af de indre hårcellers receptor potentiale, model af sandsynlighed for frigivelse af neurotransmitter fra indre hårceller, hørenervens refraktionsperiode.

Modellen bygger på tidligere publicerde modeller, men parametre for basilar membran modellen og modellen for sandsynlighed for frigivelse af neurotransmitter fra indre hårceller blev tilpasset data fra psykofysiske og fysiologiske eksperimenter for normalt hørende og hørehæmmede.

De psykofysiske data stammede fra tre postmaskerings eksperimenter. Den såkaldte "temporal window model" blev afprøvet og den blev fundet i stand til at redegøre for data med undtagelse af lavfrekvent stimulus. Det blev foreslået at "temporal window model" burde være frekvensafhængig.

Sensorineuralt høretab blev simuleret som en kombination af ydre- og indre hårcelletab. Procentdelen af døde indre hårceller blev beregnet udfra en ny metode, som relaterer tærskler for individuelle fibre i hørenerven til den absolutte psykofysiske høretærskel.

Sluttelig blev en model for hørenervens fiber population foreslået både for normalt hørende og for hørehæmmede.

Abstract

A model of human auditory periphery, ranging from the outer ear to the auditory nerve, was developed. The model consists of the following components: outer ear transfer function, middle ear transfer function, basilar membrane velocity, inner hair cell receptor potential, inner hair cell probability of neurotransmitter release and auditory nerve fibre refractoriness.

The model builds on previously published models, however, parameters for basilar membrane velocity and inner hair cell probability of neurotransmitter release were successfully fitted to model data from psychophysical and physiological data for normal hearing and impaired hearing.

The psychophysical data consisted in forward masking data from three studies. The "temporal window model" was tested and found to account for the data, except for low frequency stimulus. It was suggested that the temporal window should be frequency dependent.

Impaired hearing was modelled as a combination of outer- and inner hair cell loss. The percentage of dead inner hair cells was calculated based on a new computational method relating auditory nerve fibre thresholds to behavioural thresholds.

Finally, a model of the entire auditory nerve fibre population was proposed for normal and impaired hearing.

Acknowledgments

Although this thesis has my name on it, it would not have been possible to complete it without the kind help of the long list of people below.

I would like to thank The Oticon Foundation for the grant enabling me to work on the present thesis. In this connection I would also like to thank Claus Elberling for encouraging me to apply for the grant despite my slightly exotic background as a computer scientist with a minor in linguistics.

I want to thank Juhl P. for encouraging comments and for proof reading my proceedings contributions.

My research stay at the University of Essex marked an important six months in the project. During these six month Professor Raymond Meddis conducted several master classes in his usual informal yet very instructive fashion. I owe it to Ray that I got command of the DRNL and got access to data without which modelling makes little sense.

From the Essex group I would also like to thank Chris Sumner for answering every single question I asked almost instantaneously despite being a continent away, and for letting me use his DSAM/MatLab interface code. Steve Holmes kept spirits high in the office and provided first aid when simulations broke down; thanks for that. I would like to thank Lowel O'Mard for quickly resolving any DSAM/AMS queries pleasantly and quickly.

I owe a great deal of gratitude to Enrique Lopez-Poveda. Initially for encouraging me to go Essex, but also for patiently helping me with many Matlab questions and also for letting me use his code. Finally, I must thank Enrique for giving me access to the forward masking data he worked so hard for.

Thanks also goes to Robert Wickesberg for granting me access to detailed spike train data. Martin Hansen provided me with code which I ended up not using,

but thanks anyway Martin. I am sure I will make good use of it in a not too distant future. I would also like to thank Oliver Fobel for letting me look over his shoulder for MatLab and LaTeXcode.

I want to thank Stephan Ewert for always giving competent answers to my sometimes very basic modelling question. Moreover, Stephan moved my computer, papers and books to my fourth floor apartment when I hurt my leg and had to use crutches, just two weeks before my submission deadline. Thanks, Stephan.

I would like to thank Martin Vestergaard for countless discussion ranging from thesis format to types of hearing losses. Last, but not least, I would like to thank my supervisor Torben Poulsen. I appreciate your supportive style which has made the thesis work extremely enjoyable – even if the focus shifted slightly in the course of the project.

Preface

In 1993 I worked on the project 'Speech Recognition using Neural Networks' as a part of my Master Degree in computer science. This project raised a basic question in my mind: How is speech represented in the underlying system enabling humans to perform so well in speech recognition in comparison to computers?

Years later, in 1999, I contacted Claus Elberling, Head of Oticon's Research Unit at Eriksholm. He encouraged me to apply for a Ph.D. grant from the Oticon foundation with Associate Professor Torben Poulsen, Acoustic Technology, Technical University of Denmark as supervisor. We discussed the subject of my Ph.D., and arrived at 'A Model for the Representation of Speech Signals in Normal and Impaired Ears', the eventual title of the thesis.

We were fortunate enough to receive the grant and in pursuit of the answer to my previous question I started working on my Ph.D. August 2000.

Thus it was quite easy to decide a two part structure of the project. The first part should deal with the underlying encoding of sound signals in the auditory periphery of normal and impaired hearing. The second part should investigate the encoding of speech.

As the project progressed, it became more and more apparent that modelling peripheral processing would be the most important part of the thesis. In fact, peripheral modelling constitutes the first six out of eight chapters. As a consequence less time and effort has gone into examining speech encoding. I believe that the present thesis does justice to its title, albeit not quite the way it was originally planned.

Lyngby, 31st of October — Thomas Ulrich Christiansen

Contents

1 The Outer- and Middle Ear Model **1**

1.1 Characteristics of the outer ear 1

1.2 Characteristics of the middle ear 6

1.3 Implementing the outer- and middle ear model 9

2 The Basilar Membrane Model **11**

2.1 Cochlear mechanics 12

2.1.1 Basic anatomy 12

2.1.2 Basilar membrane vibration 16

2.1.3 The role of the hair cells 21

2.1.4 Sensorineural hearing loss 23

2.2 Taxonomy of cochlear models 23

2.2.1 Hydromechanical models 24

2.2.2 Micromechanical models 25

2.2.3 Phenomenological models 26

2.3 The chosen basilar membrane model 27

2.3.1 Introducing the DRNL filter 27

2.3.2 The fixed DRNL parameters 28

2.3.3 The free DRNL parameters 30

2.3.4 Examples of DRNL properties 31
2.3.5 From animal data to a human filter bank 34
2.3.6 Fitting DRNL parameters using pulsation threshold data 35
2.4 Other basilar membrane models 36
2.4.1 Transmission line models 36
2.4.2 The gammatone filter 37
2.4.3 Level dependent gammatone filters 38
2.4.4 Comments to basilar membrane models 40

3 Modifying the Basilar Membrane Model 43
3.1 Theoretical background for modelling forward masking 43
3.1.1 The basic psychoacoustics of forward masking 44
3.1.2 Modelling forward masking 47
3.2 Modelling actual forward masking experiments 52
3.2.1 Forward masking with varying masker frequencies 52
3.2.2 Forward masking with varying probe and gap durations . 55
3.2.3 Forward masking with varying frequencies and gap durations 59
3.2.4 Interdependence of nonlinearity and temporal window in the temporal window model 70
3.3 Final basilar membrane model for normal hearing 72
3.4 Final basilar membrane model for impaired hearing 74
3.5 Summary and conclusion 77

4 The Inner Hair Cell and Auditory Nerve Models 81
4.1 Inner hair cell electrophysiology 81
4.2 Basic auditory nerve anatomy 84
4.3 Auditory nerve discharge properties 85
4.3.1 Spontaneous activity and thresholds 86

4.3.2 Rate/intensity functions 88
4.3.3 Adaptation 90
4.3.4 Temporal properties of auditory nerve fibres 91
4.4 Modelling the inner hair cell and auditory nerve complex 92
4.4.1 Inner hair cell receptor potential 93
4.4.2 Inner Hair Cell and Auditory Nerve Complex Model . . . 95

5 Modifying the Inner Hair Cell and Auditory Nerve Complex Model 99
5.1 Auditory Nerve Responses to Trains of Clicks 100
5.1.1 Modelling the Data 102
5.1.2 Results of simulations 103
5.2 Final model for the inner hair cell and auditory nerve complex . 106
5.2.1 Normal hearing 106
5.2.2 Impaired hearing 107

6 The Complete Model 113
6.1 Finalising the model 113
6.2 Final model summary 115
6.2.1 Complete model for normal hearing 116
6.2.2 Complete model for impaired hearing 116

7 Representation in the auditory nerve 119
7.1 Representation in the auditory nerve 120
7.2 Developing the rate diagram 120
7.3 Rate diagrams for speech signals in quiet 122
7.4 Rate diagrams for speech signals in noise 126
7.5 Discussion of speech representation 130

8 Conclusions and perspectives **131**
8.1 Summary of the results 131
8.2 Future projects 132
8.2.1 Modelling forward masking 132
8.2.2 Models of impaired hearing 133
8.2.3 Speech decoding from the auditory nerve 133
8.3 Conclusion 134

Bibliography **135**

A DRNL parameters **147**

B BNAM article **149**

C Danavox article **157**

D Guided trial-and-error method **175**

E Forward masking simulation results **179**
E.1 Forward masking with varying masker frequencies 180
E.2 Forward masking with varying probe and gap durations 181
E.3 Forward masking with varying frequencies and gap durations . . 182

F Inner hair cell and auditory nerve complex parameters **185**

Chapter 1

The Outer- and Middle Ear Model

This chapter deals with the anatomy and functional aspects of the outer- and middle ear. Moreover, it briefly discusses some of the measuring techniques employed in obtaining the physical characteristics of the outer- and middle ear. The end goal is the design of a digital filter describing the physical characteristics of the outer- and middle ear.

Since the main focus of this thesis is on processing in the inner ear and auditory nerve the chapter is kept short. It is in three sections. The first section discusses the outer ear. The second section deals with the middle ear and the third section describes the design of the digital filter.

1.1 Characteristics of the outer ear

The outer ear, pinna or auricle is a groove filled, oval shape of cartilage (c.f. Figure 1.1). The biggest groove is the concha. The approximately 25 mm external auditory canal, the meatus, leads to the ear drum, the tympanic membrane, (c.f. Figure 1.2). The wall of the first 15-18 mm of the meatus is soft made up of cartilage and skin – the inner 7-10 mm wall is made of bone. The tympanic membrane consists of fibrous tissue which seals the middle ear from the outside air.

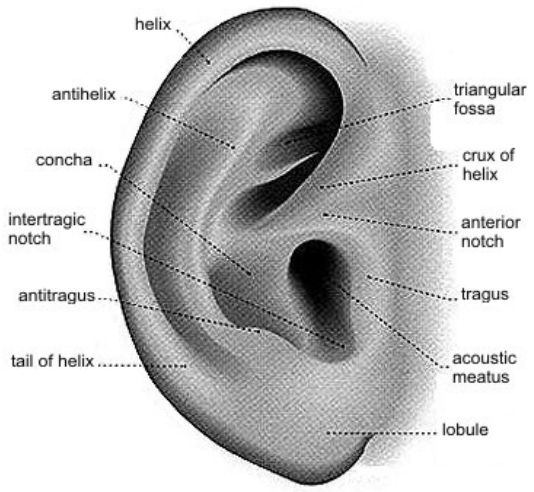


Figure 1.1: Taken from www.fauxpress.com/kimball/med/sensory/

The human pinna is not directly attached to muscles as is the case for some mammals. For this reason the pinna is locked in position relative to the head. The head can be moved relative to the torso and sound source.

The head-related transfer function (HRTF) is the signal spectrum at the source divided by the spectrum at the tympanic membrane. Its characteristics are formed by the torso, head, pinna and meatus diffraction and reflection properties.

The most fundamental HRTF is a free field, frontal, zero angle elevation sound incidence measurement with the head and torso in "normal position" (e.g. Pralong and Carlile [1994]). Varying the sound incidence in the horizontal plane, the azimuth, and the vertical plane it is possible to chart the HRTFs. Normally the head and torso are in "normal position" for these types of measurements.

Sometimes the HRTF is specified as consisting of two components: a) transfer function that is common to all source locations, such as contribution from the recording microphone and from meatus resonances, and b) transfer function depending on source location. The latter is called the directional transfer function and is determined by the torso, head and pinna.

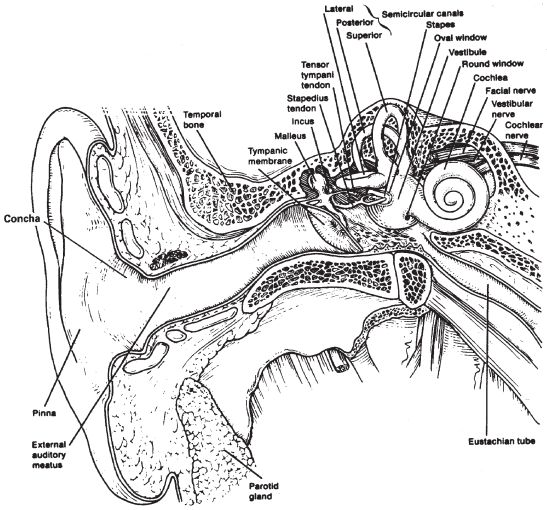


Figure 1.2: Redrawn from Pickles [1982]

Usually the measurements are performed on a KEMAR model of torso and head equipped with pinnae made from molds of test subject ears. There are many good practical reasons for using a KEMAR: ear canal seems to change considerably with movement of the jaw and head, the delicacy of the tympanic membrane make microphone positioning difficult and last but not least the KEMAR does not require breaks. The downside is that the material of the imprint of the ear is not identical to the human pinna and tympanic membrane.

As such the HRTF is a measure of purely passive mechanical properties as opposed to the active[1] mechanisms in the middle and inner ear.

Recording HRTFs is not a trivial matter since placing the recording microphone in the meatus affects the sound field. Moreover, the microphone position presents a problem in that theoretically it should be placed as close to the tympanic membrane as possible in order to avoid longitudinal standing waves. However, the complex acoustic field close to the tympanic membrane means that more distal positions in the canal are often used. Since the transfer function depends heavily on the position of the microphone, transfer functions should be

[1]The word “active” is used here in the sense “physiologically evoked” i.e. through muscle activity or hair cell contraction

specified along side with the microphone position as a reference. Pralong and Carlile [1994] used a position 2 mm from the tympanic membrane.

An example of sound distribution in the meatus is shown in Figure 1.3. This figure indirectly shows HRTF dependence on microphone position.

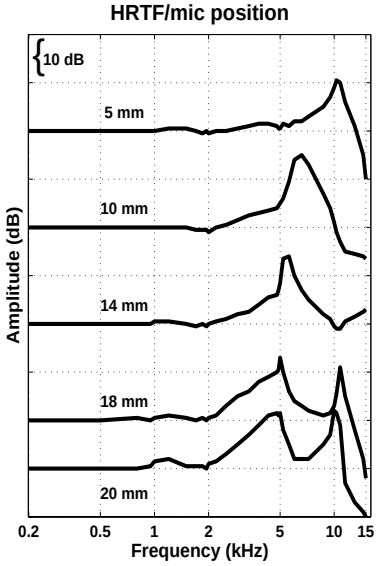


Figure 1.3: The transfer function from a reference point in the meatus to the tympanic membrane. The parameter is the distance from the reference point to the tympanic membrane. The figure indicates the dependence of transfer function on microphone location. Re-plotted from Mehrgardt and Mellert [1977]

Comparing HRTFs for various sound incidences the microphone position is not important in that the *differences* of HRTFs is of interest.

Inter-subject variability of HRTFs is very high especially for high frequencies due to the naturally occurring variability in pinna shapes [Pralong and Carlile, 1996].

An example of an HRTF for frontal incidence is shown in Figure 1.4.

In psychoacoustic experiments the stimulus is frequently presented in headphones. Therefore creating a faithful presentation of the stimulus at the tympanic membrane requires a headphone transfer function (HpTF) in addition to the HRTF. To be more exact, in order to recreate a given stimulus at a given

place in the meatus the signal is convolved with the HRTF and the inverse HpTF (both HRTF and HpTF should be measured to the same reference point). The HpTF of course depends on the headphones, but also on the test subject. As for the HRTF intersubject variability of the HpTF is significant in the 4-10 kHz region.

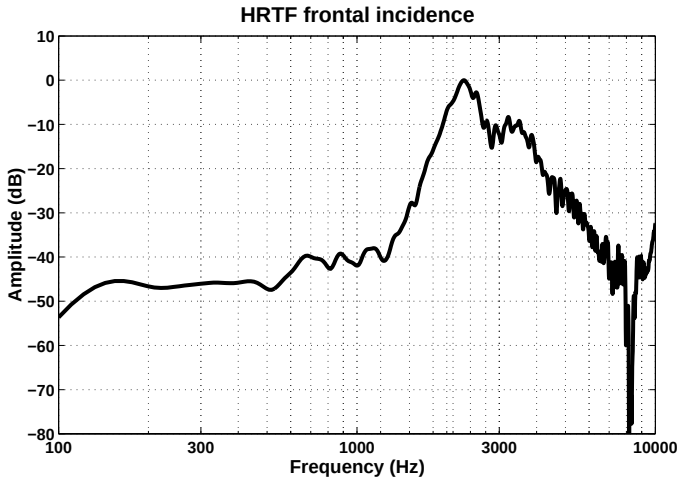


Figure 1.4: HRTF for frontal incidence. Redrawn from Florentine and Buus [1981]

An example of an HpTF is shown in Figure 1.5.

Since the psychoacoustic experiments in this thesis do not involve directional transfer functions it is sufficient to take the HpTF into account.

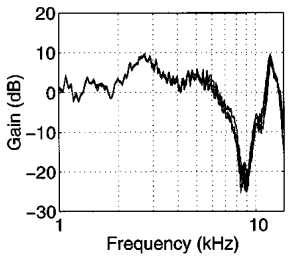


Figure 1.5: HpTf. Re-plotted from Pralong and Carlile [1996]

1.2 Characteristics of the middle ear

The middle ear contains the three smallest bones in the body the hammer, anvil and stirrup more formally the malleus, incus and stapes, together referred to as the ossicles. The tympanic membrane is attached to the malleus at the manubrium (the hammer handle). The malleus is in turn attached rigidly to the incus. When the tympanic membrane is pushed inward the malleus and incus move in a rotation like fashion shown in Figure 1.6. The force is propagated to the stapes which in turn delivers it to the inner ear fluids via the oval window of the cochlea.

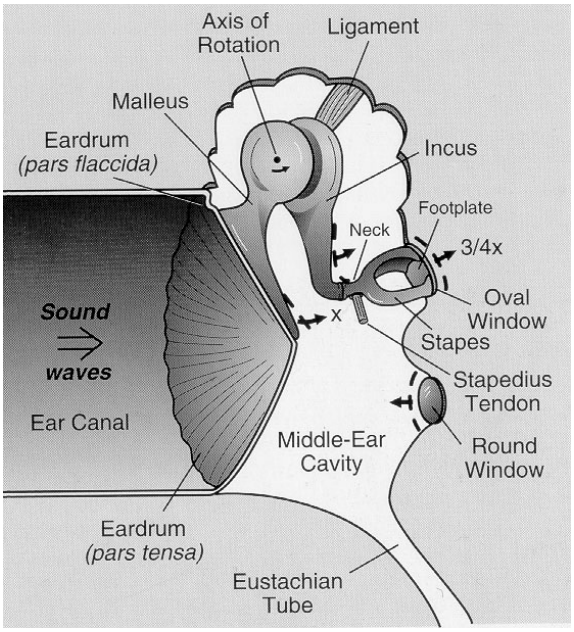


Figure 1.6: The middle ear. Redrawn from Geisler [1998]

Attached to the stapes is the stapedial muscle, the smallest muscle in the human body. It is activated by sound levels above approximately 80 dB SPL. The muscle pulls the stapes slightly out of alignment with the incus (perpendicular to the axis of vibration) lowering the efficiency of sound propagation to the inner ear. This reflex called the stapedial reflex. The manubrium of the malleus is

also attached to a muscle, tensor tympani[2], which has a similar effect to the stapedial muscle. Less is known about the workings of the tensor tympani, but in some mammals it is activated together with the stapedial muscle in response to intense sounds. The term middle ear reflex or acoustic reflex refers to the activation of the stapedial muscle and tensor tympani.

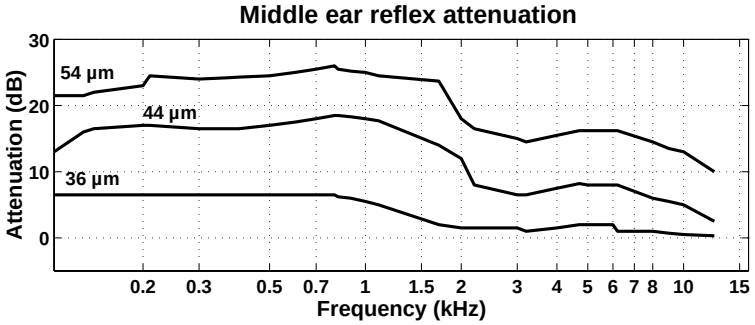


Figure 1.7: The effective attenuation provided by the middle ear reflex. Re-plotted from Geisler [1998]

The acoustic reflex provides a "gain control" protecting the inner ear from loud sounds. The latency of the reflex is approximately 10 ms and the effective attenuation is approximately 0.6 dB per dB increase in stimulus level above the reflex threshold. Although the reflex does attenuate loud sounds it does not prevent damage to the inner ear, be it from long duration noise or sudden impulse sound.

The acoustic reflex is also triggered immediately before vocalisation in humans. The "reflex" is elicited at lower levels when triggered by vocalisation (approximately 70 dB SPL) and the muscle contractions occur immediately before the actual vocal fold vibration. Figure 1.7 shows the attenuation caused by the middle ear reflex.

The main evolutional justification for the ossicles is not, however, the "gain control" just described. The ossicles perform impedance matching between air and the inner ear fluids. Without the mechanism most of the sound would simply be reflected from the inner ear fluids. The middle ear achieves the matching through three interdependent lever actions: 1) The area of contact of the malleus

[2]The name "tensor tympani" is used because when activated the manubrium is pulled inward and so is the tympanic membrane so in effect activation affects the tympanic membrane hence tensor tympani

manubrium to the tympanic membrane is 17 times greater than the area of contact between the stapes and the oval window [Yost, 2000]. 2) The displacement of the malleus is 1.3 times greater than the displacement of the stapes [Yost, 2000] and 3) Because of its conical shape the tympanic membrane buckles when pushed inward. This doubles the force exerted on the malleus relative to the force exerted on the tympanic membrane [Tonndorf and Khanna, 1972; Yost, 2000]. All together the middle ear enhances the sound pressure at the tympanic membrane by a factor of $17 \times 1.3 \times 2 = 44$ or approximately 30 dB, which is comparable to the loss that would have occurred due to the change of media from air to fluid without the middle ear.

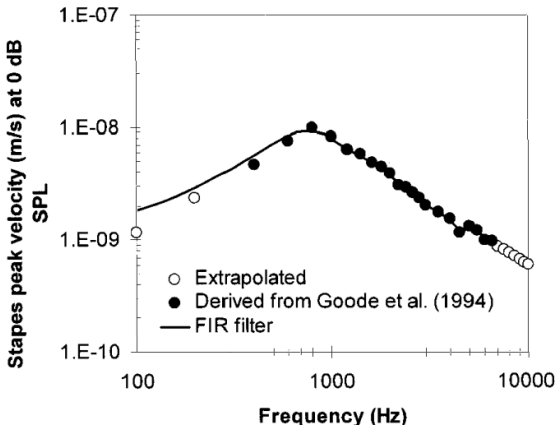


Figure 1.8: The frequency response of the middle ear. Redrawn from Lopez-Poveda and Meddis [2001] based on Goode et al. [1994]

This factor, and the underlying factors, are all approximations since the seemingly simple task of anatomically measuring them is not so simple after all [Tonndorf and Khanna, 1972]. Moreover, the impedance matching is frequency dependent. Two main factors influence the frequency characteristics of the middle ear: the mass of the ossicles and the spring-like behaviour of the entire middle ear. The mass dependence gives the middle ear a low pass filter characteristic. The spring-like influence originating from the ligaments and muscles combined with the compression of air in the middle ear cavity, gives a high pass filter characteristics. These two influences add to give the middle ear its bandpass like filter shape shown in Figure 1.8 with a centre frequency around 1 kHz (c.f. eg. [Goode et al., 1994; Puria et al., 1997; Aibara et al., 2001]).

1.3 Implementing the outer- and middle ear model

Because of the simplicity of the models employed here, it is convenient to combine the outer- and middle ear models into one filter, which will be referred to as "pre-emphasis filter" i.e., the combined HpTF and middle ear transfer function.

Two types of pre-emphasis filters are applied: Average data filters and individualised filters. The filters are applied according to the task at hand e.g. modelling the performance of a specific test subject in a psychoacoustic experiment the pre-emphasis filter for that person is used.

The average data filter was constructed according to Lopez-Poveda and Meddis [2001] which in turn is developed based on HpTF data from Pralong and Carlile [1996] (c.f. Figure 1.5) and middle ear data from Goode et al. [1994] (c.f. Figure 1.8).

Since the middle ear part of the average pre-emphasis filter was modelled based on Goode et al. [1994] no phase response data were available. So while modelling the magnitude spectrum in the pre-emphasis filter the phase response was not modelled. For this reason and for simplicity a linear phase response was chosen thus ignoring the phase data for the outer ear.

The pre-emphasis filter was implemented as a 512 tab, finite impulse response filter. The tabs were calculated by applying the Matlab inverse fourier transform to the combined magnitude spectra for the outer- and middle ear data shown in Figures 1.5 and 1.8 thus assuming a linear phase.

The work of Puria et al. [1997]; Aibara et al. [2001] provides phase response of the middle ear and thus it is now possible to build a pre-emphasis filter with a realistic magnitude and phase response. It remains to be seen how much difference such an improved pre-emphasis filter would provide. Moreover, even though impressive as the measurements of the middle ear are, they are post-mortem, in-vitro measurements thus neglecting physiological aspects as well as dynamical aspects of the middle ear cavity. Nevertheless, it seems worth while applying the phase data in order to build as accurate a pre-emphasis filter as possible. Very recent research questions the traditional views of the human middle ear bandpass characteristics and suggest a much flatter characteristics Ruggero and Temchin [2003].

The frequency response of the final average pre-emphasis filter is shown in Figure 1.9.

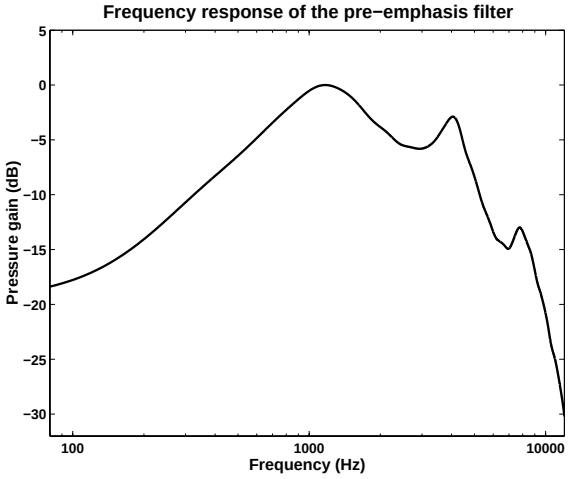


Figure 1.9: The frequency response of the pre-emphasis filter normalised so that the maximum pressure gain is 0 dB

Chapter 2

The Basilar Membrane Model

The waveform in the air is propagated through the outer- and middle ear to the inner ear fluids. As we saw in Chapter 1 the outer ear, middle ear bones and muscles transform the waveform. The inner ear fluids and physiology also transform the waveform, yet it does so in a radically different way.

Since the revolutionary measurements by Békésy [1960] the importance of basilar membrane vibration for mammalian transduction of sound has been recognised and scrutinised by countless scholars. This chapter focuses on the vibration of the basilar membrane and modelling efforts hereof.

Many parts of cochlear anatomy and physiology affect the movement and should be taken into consideration when describing basilar membrane vibration. In addition to the basilar membrane's own properties, outer hair cell motility and tectorial membrane movement are two prominent examples of aspects of cochlea physiology affecting basilar membrane vibration. The next section, entitled Cochlear mechanics, briefly deals with this topic.

The section "Taxonomy of basilar membrane models" categorises the modelling efforts from literature. Following this is a description of the basilar membrane model employed throughout this thesis. Finally, alternative models are briefly discussed in Section 2.4.

2.1 Cochlear mechanics

Cochlear mechanics literature is immensely rich and to give an exhaustive or even detailed account here would be out of proportions. However, in order to understand some of the limitations of the simplified models it is appropriate to discuss the basic concepts of cochlear mechanics here.

This is done in four sections. First, the anatomy is discussed, followed by observations of basilar membrane vibration. The role of the hair cells is discussed in Section 2.1.3 and finally Section 2.1.4 briefly defines sensorineural hearing loss.

2.1.1 Basic anatomy

Sound is propagated to the cochlea through the oval window via the bones in the middle ear. The cochlea is a tubular structure coiled like a snail. It is divided into three fluid filled chambers along its length by Reissner's membrane and the basilar membrane (c.f. Figure 2.1).

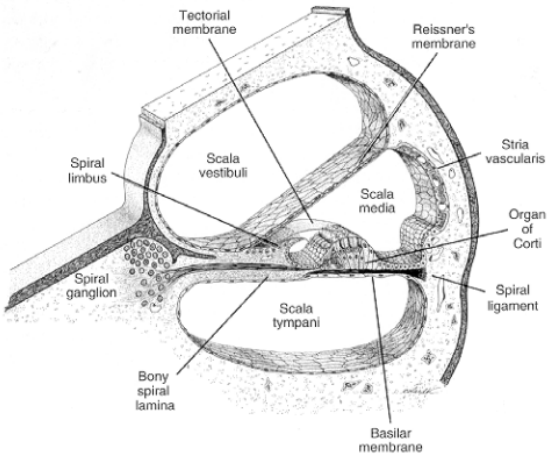


Figure 2.1: The three scali and the organ of corti. Redrawn from Geisler [1998]

The three dimensions are usually referred to as 1) Longitudinal dimension (meaning in the direction from base to apex), 2) Vertical dimension (meaning in the direction from the scala tympani to scala vestibuli), and 3) Radial

dimension (meaning in the direction from the spiral ganglion to the spiral ligament, c.f. Figure 2.1). This "coordinate system" is relative to the physiological structures. In particular the longitudinal dimension is coiled as the cochlea has just over $2\frac{1}{2}$ turns (c.f. Figure 2.2).

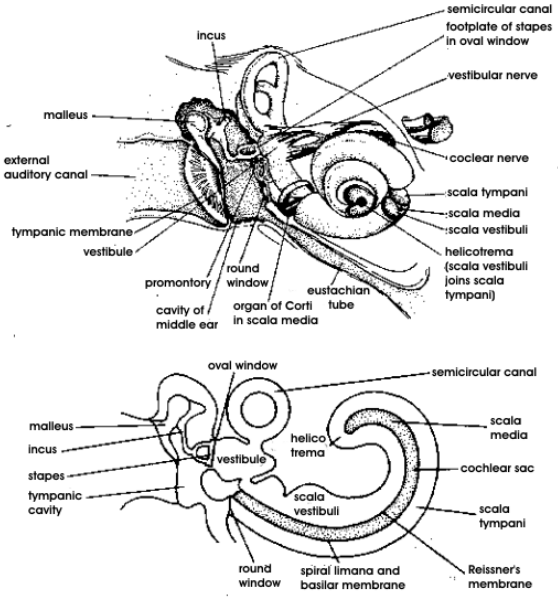


Figure 2.2: The cochlea seen from the outside. Redrawn from Yost [2000]

Reissner's membrane is so thin and flexible that it is considered mechanically transparent – it has no influence on acoustic waves. Its purpose is to electrically insulate and physically separate the perilymph in the scala vestibuli from the endolymph in scala media.

The basilar membrane is approximately 35 mm in length. It is attached to the bony spiral lamina on the side of the modiolus (spiral ganglion in Figure 2.1) and to the spiral ligament on the other side. The basilar membrane is at its narrowest and thickest at the base, and thus widens and gets thinner towards the apex. As a result the stiffness varies by a factor of 100 along its length, with the stiffest part at the base [Greenberg, 1988; Naidu and Mountain, 1998]. It appears to be made from unstructured fibre material containing collagen and fibronectin.

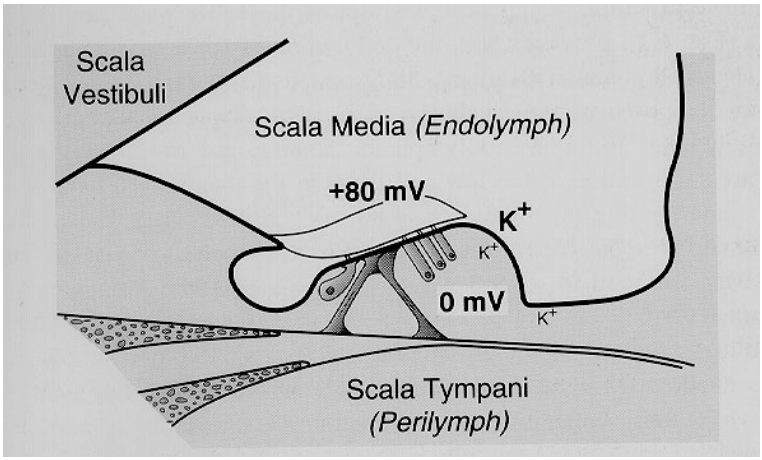


Figure 2.3: The distribution and electrical properties of perilymph and endolymph. Redrawn from Geisler [1998]

The basilar membrane carries the Organ of Corti. Together the organ of Corti and the basilar membrane are also referred to as the cochlear partition. The organ of Corti supports three rows of outer hair cells and one row of inner hair cells (c.f. Figure 2.3).

The outer hair cells are closest to the outside of the cochlea, hence the name. Humans have approximately 12,000 outer hair cells and approximately 3,500 inner hair cells. Both hair cell types have stereocilia ("hairs") protruding from their top. The stereocilia are arranged in stair-like configuration (c.f. Figure 2.4). On the inner hair cells the stereocilia form rows – on outer hair cells they form a "w" or "v"-shape seen from above.

Note how the stereocilium decreases its diameter dramatically close to its base. This enables the rigid stereocilia to move more easily in response to sheering forces. The stereocilia are interconnected by tip links. The tip links are believed to control passage of potassium and calcium ions into the stereocilia [Assad et al., 1991].

Two types of nerve fibres are attached to the base of the hair cells. Outer hair cells have predominantly efferent nerve fibres, whereas inner hair cells predominantly have afferent nerve fibres. Efferent nerve fibres *receive* nerve impulses from the brain – afferent nerve fibres *send* nerve impulses to the brain. We will return to how inner hair cells generate nerve impulse in Chapter 4.

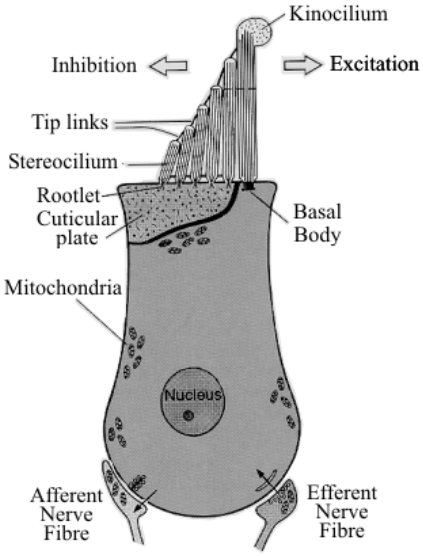


Figure 2.4: The structure of a "generic vertebrate hair cell". Mammalian hair cells do not have the special type of stereocilia called kinocilia shown in the Figure [Pickles, 1982]. Redrawn from Geisler [1998].

Reissner's membrane separates the scala vestibuli from the scala media, which in turn is separated form scala tympani by the cochlear partition. Scala vestibuli and scala tympani are connected at the apex by the helicotrema (c.f. Figure 2.5). These two scali contain perilymph whereas scala media contain endolymph. To be more precise perilymph fills the organ of corti up to the reticular lamina (the top of the hair cells) as shown in Figure 2.3.

Perilymph resembles other extracellular fluids of the body as it has a low concentration of potassium ions (K^+), a high concentration of sodium ions (Na^+) and a electrical potential equal to that of the vascular system.

Although endolymph is an extracellular fluid it is unique in that it has a high concentration of potassium ions and a low concentration of sodium ions. Moreover, scala media holds an electrical potential of +60 to +100 mV relative to the vascular system. This is referred to as the endocochlear potential. Its significance is discussed in Section 2.1.3.

The inner hair cells are flask shaped; the outer hair cells have an elongated cylindric shape. The inner hair cells are locked in position by supporting cells. In contrast, the outer hair cells are only supported by Deiters' cells at their bases enabling them to move much more freely (c.f. Figure 2.3).

There is one row of inner hair cells along the cochlear partition whereas the outer hair cells are arranged in three rows.

The tectorial membrane is a gelatinous structure lying on top of the hair cells. It appears to be attached to the cilia of the outer hair cells while not touching the inner hair cell cilia.

2.1.2 Basilar membrane vibration

When the stapes propagates sound to the inner ear a local pressure change occur in the scala vestibuli fluid. This creates a fluid pressure wave which propagates along the scala vestibuli, through the helicotrema and back to the base in the scala tympani (c.f. Figure 2.5) where the round window is pushed outward. Moreover, the pressure wave exerts pressure on the basilar membrane which is pushed towards scala tympani. The bulk of the energy transmitted from scala vestibuli to the scala tympani is transmitted across the basilar membrane and hence through scala media.

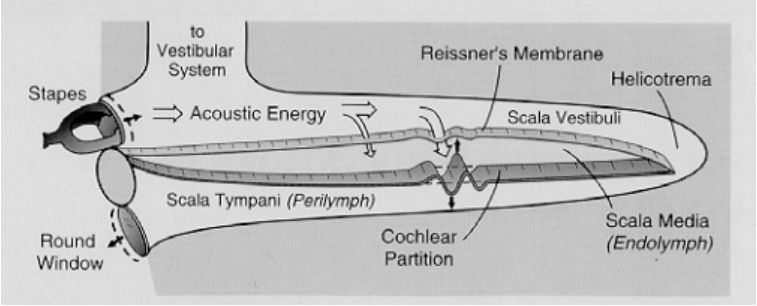


Figure 2.5: Basilar membrane vibration in the uncoiled cochlea. Redrawn from Geisler [1998]

When the basilar membrane moves the organ of corti moves with it vertically. As a result the place where the basilar membrane is attached to the spiral lamina becomes a pivot point creating a radial, sheering force (c.f. Figure 2.6).

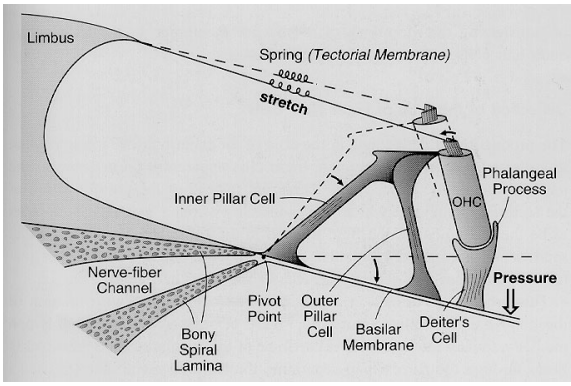


Figure 2.6: Organ of corti movement and connection to the influence of the tectorial membrane on outer hair cells

The interaction of the pressure wave in the fluid and the mechanical response of the basilar membrane results in a complex movement which, to this very day, is not fully understood. A more detailed account of the theoretical issues involved in cochlear mechanics is beyond the scope of this thesis. Rather, the remainder of this section is dedicated to observations of net basilar membrane vibration. Good accounts of cochlear mechanics can be found in Yates [1995] and Geisler [1998].

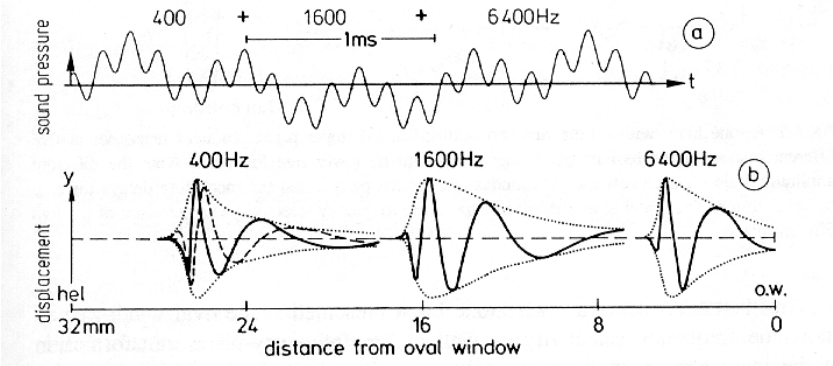


Figure 2.7: Net basilar membrane vibration and envelope. Redrawn from Zwicker and Fastl [1990]

The mechanical properties, i.e. the varying stiffness, of the basilar membrane means that any given place, in the longitudinal dimension, has a "preferred frequency" to which it responds more than other frequencies. This frequency is referred to as the characteristic frequency for the given place. In other words a frequency to place mapping occurs at the basilar membrane with the basal end responding to high frequencies, and the apical end responding to lower frequencies. Although it is tempting to use wordings like "the basilar membrane resonates at characteristic frequency" care should be taken since the word "resonance" is normally applied when describing purely physical systems. In the case of the basilar membrane physical properties act *together* with a whole range of physiological processes complicating the matter considerably. So using the word resonance in this connection for convenience is acceptable albeit a gross simplification.

Figure 2.7 shows a schematic drawing of basilar membrane response to pure tones. Note that the response is not symmetrical around the place of the characteristic frequency. The amplitude drops steeply on the apical side.

Measuring basilar membrane vibration was first done by von Békésy using visible light in a stroboscopic technique [Békésy, 1960]. The Mössbauer technique was used by Rhode [1971] producing remarkable results. Twenty years later laser interferometry, yet a more accurate method, was used by Ruggero and Rich [1991]. The latest in measuring technique is applying scanning laser interferometry making it possible to take snapshots of *a range* of the basilar membrane, not just a point [Ren, 2002].

The measuring techniques frequently provide basilar membrane vibration in terms of vibration velocity rather than vibration amplitude. Throughout this thesis vibration velocity is employed. Converting from displacement to velocity can be achieved using the simple formula

$$v = 2\pi f D \tag{2.1}$$

where v is velocity, f is frequency and D is displacement[Meddis et al., 2001].

Figure 2.8 shows the response of at a particular place on the basilar membrane to different pure tones. Note the broadening of frequency range to which the basilar membrane responds with level. Also, note the fact that the characteristic frequency, 9 kHz in the figure, decrease with level.

Figure 2.9 shows the response of at a particular (basal) place on the basilar membrane to different levels of tones. This curve is frequently referred to as the

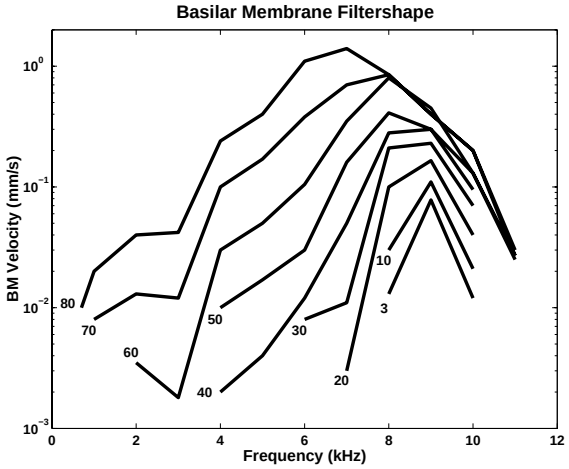


Figure 2.8: Measurements of basilar membrane velocity as a function of stimulus frequency for a site close to the oval window of the chinchilla. The parameter designates the level in dB SPL. Re-plotted from Ruggero and Rich [1991]

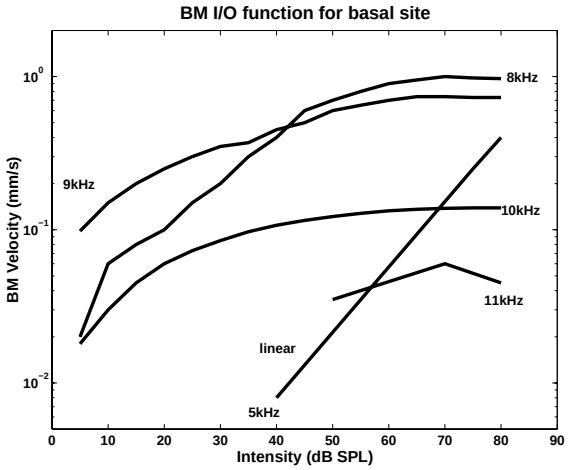


Figure 2.9: Measurements of basilar membrane velocity as a function of stimulus intensity for a site close to the oval window of the chinchilla. The parameter is stimulus frequency. Re-plotted from Ruggero [1992]

input/ouput curve. Note how tones in the vicinity of the (9 kHz) characteristic frequency show a "compressive nonlinear" response. "Nonlinear" in the sense that an increase in the stimulus intensity level by 20 dB does not result in a 20 dB increase in the velocity response. "Compressive" in the sense that the slope is less than unity[1]. Interestingly, the response to frequencies well below best frequency is linear as indicated for the 5 kHz tone. Had it not been for the compression the basilar membrane excursion for a stimulus level of 90 dB would have been 10,000 times that of a stimulus of 10 dB, which according to Equation 2.1 yields $\frac{10{,}000 \times 0.1\ mm/s}{2\ \pi\ \times\ 9{,}000 Hz} = 18\mu m$. This is more than three times the height of the hair cell and would mean that the sensory cells would be in danger of physical damage.

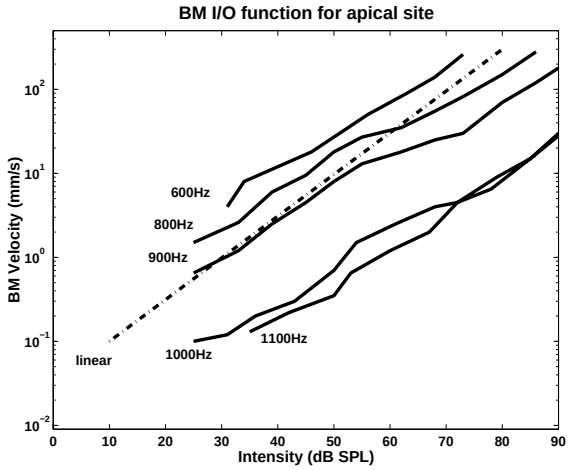


Figure 2.10: Measurements of basilar membrane velocity as a function of stimulus intensity for a site close to the chinchilla apex. The parameter is stimulus frequency. Re-plotted from Rhode and Cooper [1996]

Figure 2.10 shows the response of at a particular apical place on the basilar membrane to different levels of tones, c.f. Figure 2.9. The apical measurements show very little compression, the slopes of the curves are close to unity, in comparison with a basal site.

[1]Note the abscissa scale is given in dB SPL which means that "unity slope" is 20 dB per velocity decade

This concludes the brief introduction to basilar membrane vibration characteristics. Interesting topics like the phase characteristics and suppression were not discussed in order to keep the section short.

2.1.3 The role of the hair cells

Figure 2.6 illustrates how the sheering forces come about. Moreover, the spring effect thought to exist between the outer hair cells and the tectorial membrane is illustrated. The movement of the stereocilia themselves is seen shown in Figure 2.11. This figure also shows how the space between the tectorial membrane and the reticular lamina provides a radial fluid flow in which the stereocilia sway.

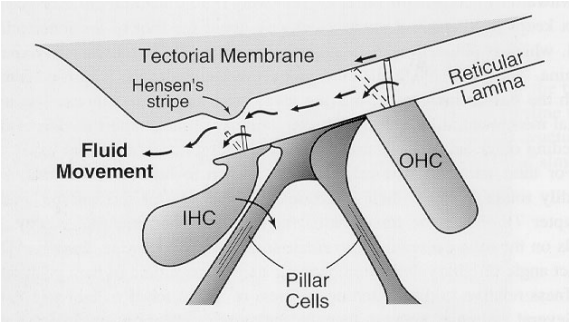


Figure 2.11: Inner- and outer hair cell stereocilia movement in the fluids between the reticular lamina and the tectorial membrane. Redrawn from Geisler [1998]

The sheering motion of the stereocilia causes ion channels to be opened as illustrated in Figure 2.12. The ion channel opening mechanism is believed to be mechanical because of the speed at which it occurs ($< 50\mu s$) [Geisler, 1998]. The influx of ions depolarises the hair cell.

Since the endolymphatic space has a potential of approximately +80 mV and the potential inside the hair cell is less than -40 mV (outer hair cells $\approx$ -60 mV and inner hair cells $\approx$ -40 mV), the potential difference across the apical end of the hair cell is very large – in fact the largest difference known in the human body. This means that opening ion channels allows a high current to flow. Since all the individual cilia contribute to this process this transduction mechanism is extremely sensitive and quick: 10 nm tip deflection results in approximately 1 mV voltage change [Geisler, 1998].

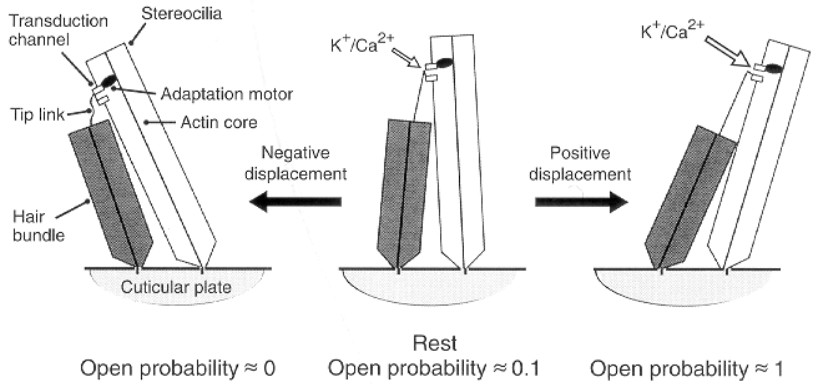


Figure 2.12: Tip links opening ion channels

Comparing the mechanical power absorbed by the hair cell, due to cilia friction, with the electrical power transduced into the hair cell shows that the mechanical power is amplified many times as the signal is transduced. The factor is approximately 100 at 160 Hz [Geisler, 1998]. The added power originates from the voltage differences across the cell membrane which is sustained by homeostatic mechanisms[2].

For the inner hair cell, depolarisation initialises the process whereby nerve impulses are generated. This process is discussed in greater detail in Chapter 4.

For the outer hair cell the depolarisation has an additional effect: it makes the cell contract to a shorter and thicker state [Brownell et al., 1985]. Conversely hyperpolarisation makes the cell elongate. The protein prestin, placed on the basolateral sides of the outer hair cell membrane, responds to the transmembrane voltage change by contracting or expanding [Zheng et al., 2000].

The length change is 30 nm/mV [Santos-Sacchi and J., 1988] giving a maximum of 1-2 μm). It appears to be sufficient in accounting for the amplitude boost of basilar membrane vibration at low levels. The speed of the contraction is a different matter. Even though outer hair cell length change is capable of following voltage changes up to at least 79 kHz [Frank et al., 1999] the voltage change cannot be conveyed to the hair cell this rapidly due to the capacitance

[2]Homeostasis is the physiological processes maintaining an equilibrium. In this case the difference in transmembrane potential which is maintained by stria vascularis. This contraption is sometimes referred to as “the battery of the ear”

of the basolateral membrane. It rolls of at approximately 6 dB/octave above 1 kHz [Russell et al., 1986] making the reverse transduction almost negligible at high frequencies. This is in stark contrast to the basilar membrane measurements and thus poses a problem for this somatic electromotility explanation of active amplification in the cochlea.

An alternative conjecture places the outer hair cell motor mechanism in the stereocilia [Hudspeth and Gillespie, 1994]. The motor is thought to be a result of actin-myosin interaction with the calcium current, the result being a shearing force *from* the stereocilia on the basilar membrane. Traditionally, the argument is the other way around: Basilar membrane displacement leads to a shearing force on the stereocilia. In theory, stereocilia movement would provide a better account of the active mechanism, however, no direct observations supporting this is possible at present[Yates, 1995].

2.1.4 Sensorineural hearing loss

Sensorineural or neurosensory hearing losses are caused by problems in the hair-cells. Malfunctioning or loss of outer hair cells accounts for mild to moderate hearing losses while severe to profound hearing losses also involve malfunctioning of inner hair cells.

The outer hair cell is the single most prominent source of nonlinearity in the peripheral hearing system. They provide the large dynamic range of the auditory system, and therefore malfunctioning outer hair cell results in a reduced dynamic range. Sensorineural hearing loss thus means a loss of dynamic range and sensitivity. As we shall see later this has implications for frequency selectivity, loudness, temporal integration and many other related auditory properties.

Hence, understanding hair cell function and malfunction is fundamental to hopes of improving understanding of hearing in general and hearing loss in particular.

2.2 Taxonomy of cochlear models

This section provides an overview of the types of cochlear models employed at present. The goal is to provide the background for the model used in this thesis.

The term cochlear model is most frequently used for models describing the micro- and/or hydromechanics of the inner ear. For reasons of simplicity the

focus is kept on basilar membrane vibration. Figure 2.13 presents an attempt of grouping the various types of cochlear models.

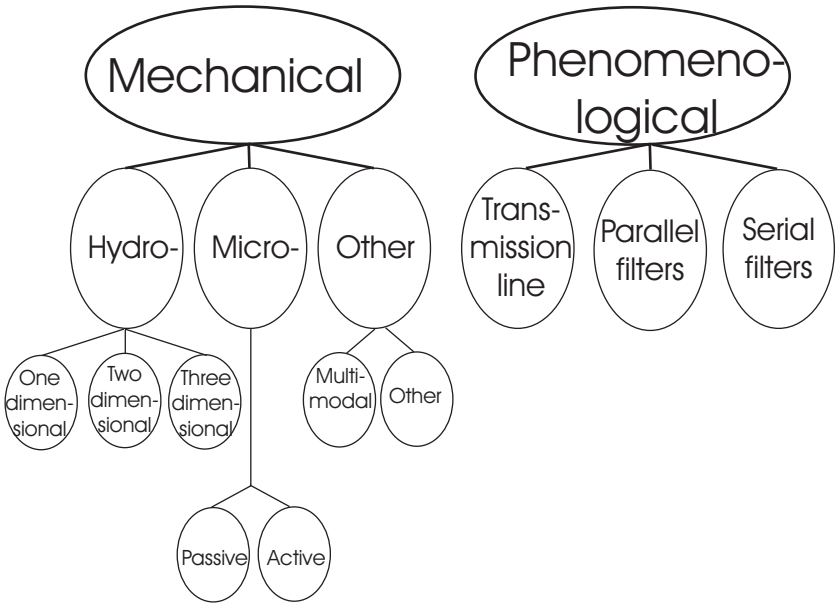


Figure 2.13: An attempt at categorising cochlear models. For details please refer to the text

2.2.1 Hydromechanical models

The hydromechanical models focus on the fluids in the cochlear encasement and their interaction with the organ of Corti. Hence they are also known as hydromechanical models. The basilar membrane itself is reduced to a single mass with very simple viscoelastic properties.

One-dimensional models in Figure 2.13, provide response along the longitudinal dimension. Two-dimensional models also incorporate vertical displacement. Three-dimensional models include a description of the radial dimension. Frequently, one-dimensional models are expressed in terms of the pressure difference between the upper- and lower- scalae at a given place on the basilar membrane.

The ratio between the mass of the scala fluids and the mass of the basilar membrane decreases from base to apex, as the scali narrows and the basilar

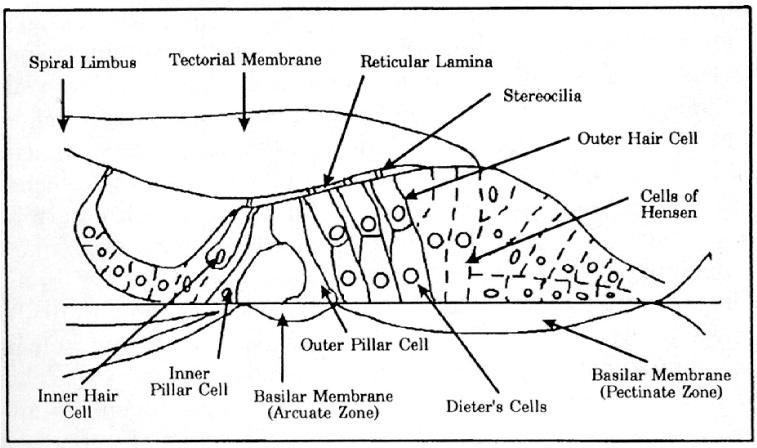


Figure 2.14: Schematised drawing of the cross section of one cochlear turn showing a magnified view of the Organ of Corti. Redrawn from Mountain and Hubbard [1996]

membrane gets wider and thicker. The two-dimensional models take this into account yielding a more accurate description. In particular the steepness of high frequency fall-off is better modelled.

The arcuate and pectinate zones (see Figure 2.14) show differences in compliance. Three-dimensional models have been proposed to take this fact into account. Historically, the success of three-dimensional models has been hindered by the considerable analytical complexity involved. Furthermore, computer power has been a limiting factor.

The discovery of evoked otoacoustic emissions [Kemp, 1978] has lead to the emergence of active hydromechanical models – models that mimic the active production of kinetic energy in the cochlea. This subdivision of the hydromechanical models is not shown in Figure 2.13. Alone the hydromechanical models fail to explain significant experimental results [Mountain and Hubbard, 1996], and today hydromechanical models primarily serve a complementary role to the micromechanical models.

2.2.2 Micromechanical models

Micromechanical models focus on the mechanical properties of the many components constituting the organ of Corti rather than the fluids in the cochlea.

The complex anatomy of the organ of Corti makes modelling of the mechanical properties an intricate matter subject to ubiquitous speculation.

Micromechanical models are categorised in terms of their "degrees of freedom" rather than their dimensionality. The degrees of freedom are, in turn, determined by the level of detail at which the mechanics of the anatomy is modelled. The degrees of freedom are not shown in Figure 2.13. The micromechanical models can either be passive or active. Passive models presuppose that the energy entering the cochlea dissipates whereas active models presuppose that the cochlea generates energy which in turn dissipates.

Multimodal models allow more than one wave propagation mode. They are either hydromechanical or micromechanical. Multimodal, micromechanical models are "state-of-the-art" in cochlear modelling e.g. Sen and Allen [1999].

Extensive and detailed experimental animal- as well as human data is necessary for further progress in the area to occur. Thus progress hinges on improved measuring techniques. Particularly mid-frequency human cochlear responses are in demand, but also simultaneous measurements of pressure and volume velocities would be valuable for shedding light on the active cochlear mechanisms [Mountain and Hubbard, 1996]. A recent step in this direction was achieved when longitudinal patterns of 1 mm *sections* of the basilar membrane was recorded with a scanning laser [Ren, 2002].

Other mechanical models have been realised, e.g. physical replica of the basilar membrane mounted inside a duct.

2.2.3 Phenomenological models

Modelling basilar membrane response can also be performed at a "conceptual" level, i.e. signal processing schemes can be applied in order to produce the desired response irrespective of the intrinsic physiology and mechanics. An approach similar to the one-dimensional model described above is an arrangement of cascaded filters each tuned to a lower frequency than the preceding filter. The travelling wave of the cochlea is modelled as the signal travels along the cascaded filters. The transmission line model proposed by Giguère and Woodland [1994] is a well-known example.

An alternative approach is to present the signal to a filter bank i.e., filters in parallel where each filter represents one place on the basilar membrane. The Dual Resonance Non Linear (DRNL) filter bank presented in this thesis falls in this latter category of cochlear models.

2.3 The chosen basilar membrane model

The basilar membrane model used throughout this thesis is the DRNL. The explanation in the following is based on work performed at the hearing group in the Psychology Department, University of Essex [Meddis et al., 2001; Lopez-Poveda and Meddis, 2001; Sumner et al., 2002; Holmes, 2002; Sumner et al., 2003] and personal experience of the author in connection with the research stay with the hearing group.

The basilar membrane used in this thesis is introduced in Section 2.3.1. This is followed by two sections describing the parameters of the model – the first describes the parameters of the model which have been fixed to published values [Lopez-Poveda and Meddis, 2001], the second describes the parameters investigated in the thesis.

2.3.1 Introducing the DRNL filter

The DRNL is a filter arrangement that models basilar membrane response for a given place along the basilar membrane. It accepts input in terms of stapes velocity, in units of m/s, and the output is transversal, i.e. vertical dimension, velocity of the basilar membrane, in units of m/s.

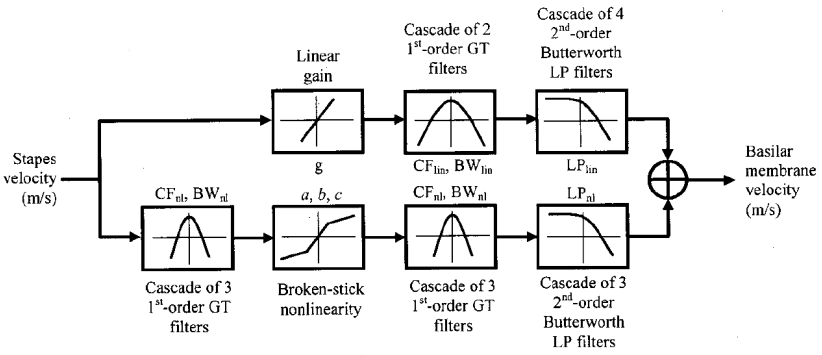


Figure 2.15: The fundamental DRNL filter setup. Top row shows the linear processing path. The lower row shows the nonlinear path

The DRNL has two processing paths: One linear and one nonlinear (c.f. Figure 2.15). The output is the sum of the two processing paths. The linear path

consists of a linear gain, a gammatone filter [Patterson et al., 1991] and a low pass filter in cascade. The nonlinear path consists of a gammatone bandpass filter, a compression function followed by a second gammatone filter. The bandpass filters themselves are cascaded gammatone filters. The low pass filters consists of four cascaded second order Butterworth filters (e.g. Steiglitz [1996]). Since the centre frequency of the linear path deviates slightly from the centre frequency of the nonlinear path the filter shows "dual resonance" hence the name.

The nonlinear function used is a "broken-stick" given by:

$$\mathrm{y}(t) = \mathrm{sgn}(\mathrm{x}(t)) \min(a|\mathrm{x}(t)|,\, b|\mathrm{x}(t)|^c) \tag{2.2}$$

where $\mathrm{x}(t)$ is the input time signal, $\mathrm{y}(t)$ is the output, $\mathrm{sgn}(x)$ is a function which is 1 for $x > 0$ and -1 for $x \leq 0$, $\min(x, y)$ is a function defined as the smallest of x and y, a, b and c are parameters of the model (see below for details).

2.3.2 The fixed DRNL parameters

This section describes the "fixed" parts of the DRNL, i.e. the parameters not varied from previously published parameters [Lopez-Poveda and Meddis, 2001].

The low pass filter in the linear path was constructed from cascading four second order Butterworth filters which each have 6 dB down points at the centre frequency of the gammatone filters of the linear path.

Similarly the low pass filters of the nonlinear path was constructed from cascading three low pass second order Butterworth filters with 6 dB down points at the centre frequency of the gammatone filters of the nonlinear path.

The low pass filters could have been chosen to be subject to fitting in terms of number of filters in the cascade and initially they were indeed (c.f. Lopez-Poveda and Meddis [2001] Table I and Table II where the low pass filter cascade in the linear path differ). However, it was decided to fix the low pass filters as described, in order to limit the number of free parameters. Initial investigation justified this decision [Meddis et al., 2001] and further support was given by the current version of the DRNL [Lopez-Poveda and Meddis, 2001].

The three gammatone filters used were each constructed from a cascade of three gammatone in cascade resembling the auditory filter shapes [Hartmann, 1997].

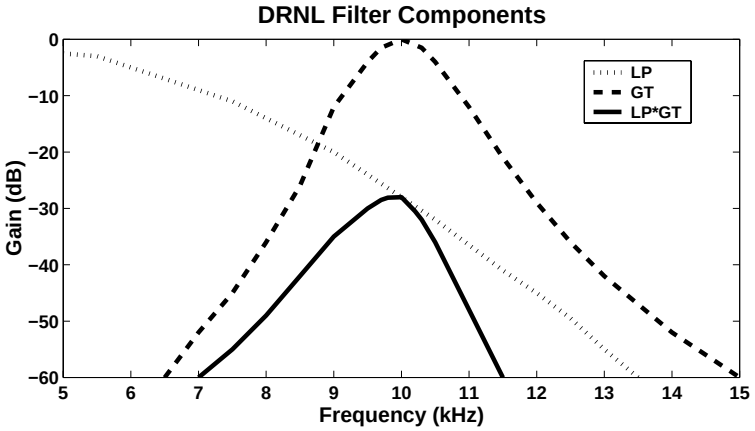


Figure 2.16: The frequency response of the low pass and gammatone filters used in the DRNL. The low pass filter shown is cascaded four times, the gammatone is cascaded three times corresponding to the linear path of the DRNL. The transfer functions for the low pass filter in the nonlinear path is not shown, but differs only in that it is cascaded three times rather than four times

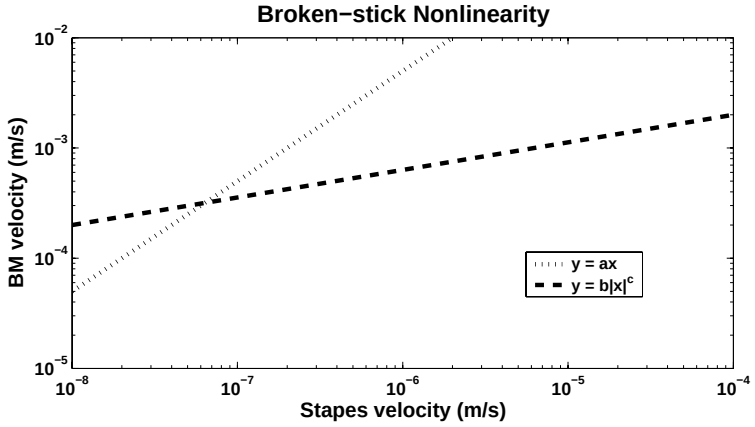


Figure 2.17: The compression mechanism as implemented in the DRNL. The parameters in the legend refers to Equation 2.2

Figure 2.16 shows the various filter functions involved which make up the DRNL. The implementation of the compression is illustrated by Figure 2.17.

The parameter c controls the slope of the nonlinear part of the broken-stick nonlinearity (together with b, c.f. Figure 2.17 and Equation 2.2). This parameter was fixed to the value 0.25 by Meddis et al. [2001] who concluded that satisfactory results could be achieved with this value although Plack et al. [2002] used 0.16.

The output of the model was chosen to be basilar membrane vibration velocity for the given place on the basilar membrane. The alternative, basilar membrane displacement, could have been used, but it was found that compression threshold changed less with respect to frequency when using the velocity measure [Meddis et al., 2001]. Moreover, modern measurements tend to provide velocities rather than displacements.

2.3.3 The free DRNL parameters

This section describes the DRNL parameters that are modified from published values. These parameters will be referred to as "free parameters". Not allowing all DRNL parameters to vary is productive in two ways:

1. The number of parameters in the DRNL is quite high. Fixing some parameters dramatically constrains parameter space.
2. It enables comparisons between studies.

The potential downside, that the DRNL will not accurately model the data, has not been a serious problem thus far.

Adjusting the parameters as described in the following, allows the DRNL model to account for specific experimental data. Below is a list of the DRNL-parameters which are allowed to vary when modelling experimental data.

$\mathbf{CF_{lin}}$ The centre frequency of the linear filter path.

$\mathbf{BW_{lin}}$ The filter bandwidth for the linear filter path given as the 3 dB down points.

g The gain factor of the linear path. This scalar is used for balancing the contribution of the linear path against the contribution from the nonlinear path. In Figure 2.15 this is called linear gain.

$\mathbf{CF_{nl}}$ The centre frequency of the nonlinear filter path which is slightly higher than CF_{lin}.

$\mathbf{BW_{nl}}$ The filter bandwidth for the nonlinear filter path given as the 3 dB down point of the three cascaded gammatone filters.

a The slope of the linear part of the broken-stick nonlinearity. Together with b this parameter determines the gain of the nonlinear path (c.f. Figure 2.17.

b The "slope" of the nonlinear part of the broken-stick nonlinearity. Together with c this parameter controls the slope and indirectly the intercept of the nonlinear part of the broken-stick nonlinearity.

An example of DRNL parameter values is given in Appendix A. It is important to note that each of these parameters has one value *for each place* along the basilar membrane. So for a given place on the basilar membrane each variable has a value. This set of parameter values defines the model response for this one place only. Defining the response for the entire basilar membrane, i.e. defining a DRNL *filter bank*, is described in Section 2.3.5.

2.3.4 Examples of DRNL properties

The DRNL filter arrangement may seem obscure and unfounded in view of the complex physical and physiological mechanisms described in this chapter. The justification for the DRNL is that it models basilar membrane vibration quite well and that it is computationally effective [Meddis et al., 2001]. The remainder of this section provides examples of the key properties of the DRNL supporting this claim.

In order to evaluate how powerful the DRNL is Meddis et al. [2001] performed a number of comparisons to animal data. The comparisons were done for:

Basic basilar membrane input/output function The results for the 800 Hz best frequency for the chinchilla, 10 kHz and 18 kHz best frequency for the guinea pig where modelled successfully. The notches in basilar membrane response reported in animal data at 100 dB SPL was also modelled, at least for the 10 kHz best frequency.

Basilar membrane phase response The phase lag is greater in the model than in the animal data for high intensities and frequencies just above best frequency. Improvement might be achieved by means of broader filters in the nonlinear path.

Two tone suppression Two tone suppression was not found at the apical site (800 Hz) as expected. Strong two tone suppression was successfully modelled for the basal sites (10 kHz and 18 kHz).

Basilar membrane impulse response There are some discrepancies between animal data and model impulse response. The culprit is speculated to be the imperfection of clicks produced by loudspeakers in the animal studies. A simulation of DRNL response to "imperfect clicks" supports the speculation [Meddis, 2002].

Distortion products The distortion products $2f_1 - f_2$ and $3f_1 - 2f_2$ were tested for $f_1 = 12.5$ kHz and $f_2 = 14$ kHz for 50 dB SPL stimulus at 10 kHz best frequency . The results were comparable to the measurements available.

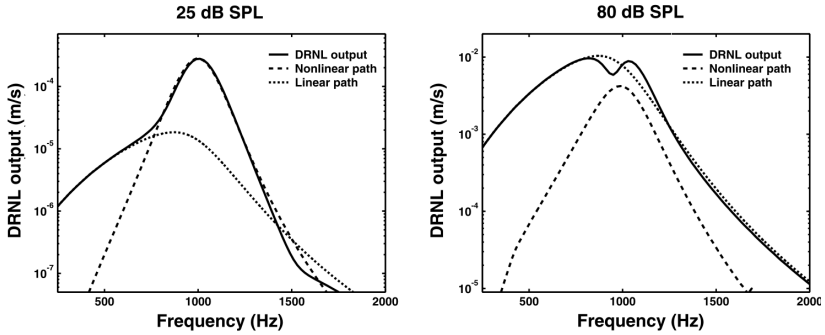


Figure 2.18: Change in DRNL filter shape and centre frequency with level

Yet another key property of the basilar membrane is its level dependant filter shape illustrated in Figure 2.8. The DRNL models this property as indicated in Figure 2.18. The DRNL filter response broadens with level as does the response of the basilar membrane.

In order to look at the change in filter shape quantitatively, a comparison of psychophysically determined filter shapes with DRNL filter shapes is shown in

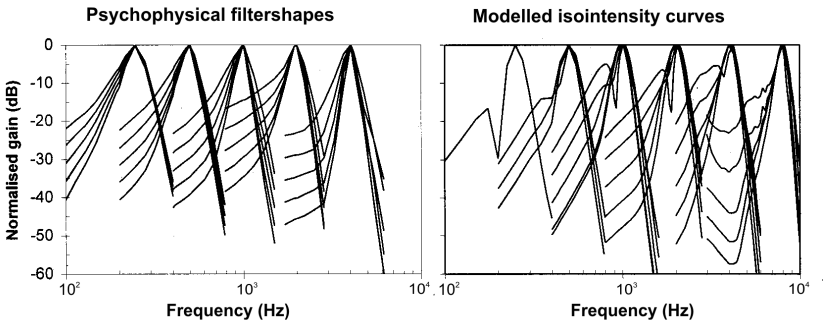


Figure 2.19: Comparison of iso-intensity curves to psychophysically measured data. The data for the psychophysical measured data in the panel to the left is taken from Baker et al. [1998]. The modelled data for the panel to the right is taken from Lopez-Poveda and Meddis [2001]

Figure 2.19. The main difference between the psychophysical data and the model data is the distinct notch below characteristic frequency in the DRNL data. The DRNL data is consistent with physiological data [Rhode and Recio, 2000]. The discrepancy from simultaneous masking data to the psychophysical data can be explained in the following way: In a simultaneous masking experiment the subject is simultaneously monitoring several auditory filters always selecting the optimal filter for probe detection. Hence the auditory filter in the notch region will not be utilised for probe detection since an adjacent filter provides an optimal probe-to-masker representation. This explanation resolves the apparent conflict between psychophysical and physiological data.

Finally, the "basal ward" shift in characteristic frequency is an important property of the basilar membrane. This effect, by which the characteristic place for a given frequency at low levels moves towards the oval window at higher levels, is implicitly shown in Figure 2.8. The figure shows data for a *specific point* on the basilar membrane. That point responds best to 9 kHz tones at low levels and to 7 kHz tones at higher levels – the characteristic frequency decreases with level for a given point on the basilar membrane. In other words there is an "apical ward" shift in characteristic place for a given frequency with increasing level or conversely there is a "basal ward" shift in characteristic frequency for a given place with increasing level.

That this is modelled qualitatively can be seen from Figure 2.18. The extent of the characteristic frequency change is believed to be in the region of a factor of 1.1 to 1.4 between characteristic frequency at 65 dB SPL and 95 dB SPL [Mc-

Fadden and Yama, 1983]. This constraint is observed when adjusting DRNL parameters, but not subject to further investigation at present.

2.3.5 From animal data to a human filter bank

The basic construction of the DRNL was based on animal data. This section presents a description of how the *human* DRNL filter bank was developed [Lopez-Poveda and Meddis, 2001].

The DRNL as described above simulates the vibration at a given point on the basilar membrane. Generalising the idea to provide filter responses for the entire basilar membrane requires a bank of DRNL filters, one for each point on the basilar membrane where the response is to be simulated. This in turn requires parameter values for the each of the parameters in in Appendix A, Table A.1.

Lopez-Poveda and Meddis [2001] describes such a filter bank based on human data from a psychoacoustic experiment. First the individual DRNL parameters were fitted for six different frequencies corresponding to six points on the basilar membrane. For each of the seven free parameters (c.f. Section 2.3.3) a function of the form shown in Equation 2.3 was established.

$$p_{\mathrm{DRNL}}(f) = 10^{(p_0 + m \log_{10}(f))} \tag{2.3}$$

where, f is frequency, $p_{\mathrm{DRNL}}(f)$ is parameter value at frequency f, p_0 is the linear regression intercept, m is the linear regression slope.

For instance, to calculate the bandwidth of the nonlinear path (BW_{nl}) for the 10 kHz place one needs the values of m and p_0 for this parameter. These values are -3.1930 x 10^{-2} and 7.7426 x 10^{-1} respectively (c.f. Appendix A and later in this section) yielding a bandwidth of the nonlinear path (BW_{nl}) equal to $10^{(-0.03193+0.77426x \log_{10}(10000))} = 1346$. In this way extrapolation and interpolation enables generalisation of the data obtained from a few points to extend to all of the basilar membrane.

In fact, the f in Equation 2.3 was substituted with CF_{NL} so that only six parameters needed to fit the regression line. This correspond to defining the characteristic frequency of a given place of the basilar membrane as the frequency where the response is greatest at low levels. It is convenient and does not affect the capabilities of the DRNL.

Pulsation threshold data for characteristic frequencies 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz were collected [Plack and Oxenham, 2000a]. The thresholds were measured for signal levels ranging from 25 dB SPL to 85 dB SPL in steps of 5 dB. From this, the DRNL parameters in Table A.1 were determined for the six frequencies so as to best fit the experimental data [Lopez-Poveda and Meddis, 2001].

The next step was then to fit the p_0 and m linear regression coefficients as best possible. So replacing the individual direct parameter values from Table A.1 with the two linear regression coefficients p_0 and m provides a description of basilar membrane vibration for any given point. The terms DRNL and DRNL filter bank will be used interchangeably in the remainder of the thesis. The resulting filter bank parameters are shown in Table A.1.

2.3.6 Fitting DRNL parameters using pulsation threshold data

This section provides an example of a psychoacoustic experiment that can be used for building a human DRNL filter bank. It was actually the first method to be employed [Lopez-Poveda and Meddis, 2001].

The pulsation threshold technique first introduced by Houtgast [1972], can be described as follows: An interrupted sound is perceived as continuous if another sound fills the interruption period with sufficient energy. This effect is called the continuity effect [Elfner and Caskey, 1965]. Take the example of a pure tone signal (T) and a lower frequency pure tone masker (M). T is alternated with M. Consider the place at the basilar membrane with best frequency equal to that of T. If the basilar membrane response to M at this place is equal to or greater than the response to T, the stimulus will be perceived as continuous. For a given frequency and level of T, the pulsating threshold is determined as the level of M at which the perception changes from pulsating to continuous. Frequently a 0.6 T-to-M frequency ratio is used [Plack and Oxenham, 2000a] in this type of experiment.

If T and M are sufficiently far apart in frequency it can be assumed that the basilar membrane at the best frequency of T responds compressive nonlinearly to any increase in the level of T, e.g. an increase of 20 dB in the level of T results in (much) less than 20 dB increase in basilar membrane response. Moreover, the basilar membrane responds linearly to any increase of the level of M at this

place, e.g. a 20 dB increase in the level of M yields a 20 dB increase in response. Hence basilar membrane compression can be estimated as follows:

For a given level the pulsation threshold is determined. An increase in the level of T is applied and the pulsation threshold is determined for this level. The compression ratio is simply the T-to-M level ratio. By reiterating the method for multiple levels and frequencies the shape of the masking function can be estimated.

The data obtained using this method is valuable for two reasons 1) it gives an estimate of *human* basilar membrane compression and 2) it does so for the mid-frequency range. An alternative approach, direct measurement in nonhuman cochleae, is only possible at the apical and basal turns at present [Plack and Oxenham, 2000a] and thus only provide nonhuman data outside the mid-frequency range. The accuracy of the compression ratio determined using pulsation threshold is quite good, although care must be taken in order to avoid the caveats inherently associated with psychoacoustic experiments.

2.4 Other basilar membrane models

A great number of phenomenological basilar membrane models have been proposed over the past three decades [Lyon, 1982; Seneff, 1988; Shamma, 1988; Jenison, 1991; Carney, 1993; Kollmeier et al., 1993; Giguère and Woodland, 1994; Dau et al., 1996; Irino and Patterson, 1997; Meddis et al., 2001; Heinz et al., 2001a; Zhang et al., 2001]. Rather than a review of all the models two models have been selected to illustrate the diversity of the approaches employed namely the transmission line models and the gammatone filter models.

2.4.1 Transmission line models

In transmission line models, much like the one dimensional models in Figure 2.13, the pressure difference between scala media and scala tympani is simulated by a number of segments [Lyon, 1982; Giguère and Woodland, 1994]. These segments represent equidistantly placed sites along the basilar membrane. Each segment has a "passive" component corresponding to the passive mechanical properties of the basilar membrane and an "active" component corresponding to the outer hair cell feedback of mechanical energy to the basilar membrane. Transmission line models thus reflect the travelling wave directly.

Giguère and Woodland [1994] implemented the passive components of the model as digital filters derived from the laws of mechanics combined with measurements of the hydromechanical and micromechenical entities involved.

The mechanical force generation of the outer hair cells is assumed to consist of a nonlinear frequency independent transduction of vertical basilar membrane displacement into a receptor potential and a reverse transduction force generated by this receptor potential. This mechanism is modelled as a voltage source saturating at high amplitudes.

The basilar membrane and outer hair cell models proposed by Giguère and Woodland are components in a model aiming at modelling the transmission of sound from the free field to the auditory nerve. Although workings of the model is described in detail, no elaborate evaluation of the model has, to the knowledge of the author, been published.

The model accounts for otoacoustic emissions. Moreover, the efficient implementation makes the model a viable choice for inner ear modelling. As such it is a useful tool for studying the behaviour of the auditory periphery in general.

2.4.2 The gammatone filter

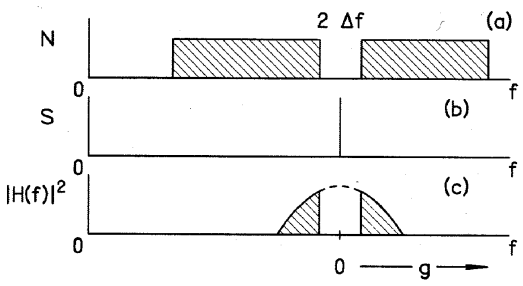


Figure 2.20: Illustration of notched noise experiment designed for deriving auditory filter shapes. Redrawn from Hartmann [1997]

The roex(p) filter was proposed by Patterson [1976]. It is based on notched noise masking experiments estimating the transfer function of a hypothesised auditory filter centred at a given frequency f_c. The experiment is illustrated in Figure 2.20. Masking thresholds are measured yielding the notch width as a function of signal power. From this the transfer function for a given cen-

tre frequency can be estimated. The parameterised functional form suggested by Patterson and Nimmo-Smith [1980] is shown in Equation 2.4.

$$|H(f)|^2 = (1 + pg)\, e^{-pg} \tag{2.4}$$

where p is parameter describing the width of the filter and g is the normalised deviation of frequency f to centre frequency f_c given by $g = |\frac{f - f_c}{f_c}|$.

The roex(p) filter does not have a well-defined impulse response and is thus not easy to implement. For this reason the gammatone filter is often used for simulating basilar membrane filtering [Patterson et al., 1991]. The impulse response $h(t)$ of the gammatone filter is given in Equation 2.5.

$$h(t) = b^{\eta} t^{(\eta-1)}\, e^{-2\pi bt} \cos(2\pi f_c t + \phi) \tag{2.5}$$

where η is the order of the filter, b is a decay factor and f_c is the centre frequency of the filter.

The amplitude response of a fourth order gammatone filter is very close to the roex(p) filters. Further, the impulse response provides a good match to auditory nerve fibre responses from physiological studies, e.g. Carney and Yin [1988].

2.4.3 Level dependent gammatone filters

The gammatone and the roex(p) transfer functions are symmetrical around centre frequency on a linear frequency scale, in contrasts to the results of other notched noise experiments [Moore et al., 1990; Glasberg and Moore, 1990]. These studies show that the high frequency slopes are steeper than the low frequency slopes. Moreover, the asymmetry increases with level.

One approach remedying this deficiency, the gammachirp filter, was proposed by Irino and Patterson [1997]. The impulse response of the gammachirp filter is given by:

$$h(t) = b^{\eta} t^{(\eta-1)}\, e^{-2\pi bt} \cos(2\pi f_c t + c \ln t + \phi) \tag{2.6}$$

where η is the order of the filter, b is a decay factor, f_c is the centre frequency of the filter and c is an additional parameter used for implementing level dependence.

The level dependent filter shape is achieved by fitting the "new" parameter c to data from notched noise experiments. The impulse response shown in Equation 2.6 is identical to the gammatone impulse response except for the term $c\ln(t)$. Figure 2.21 shows the transfer function of the gammachirp filter.

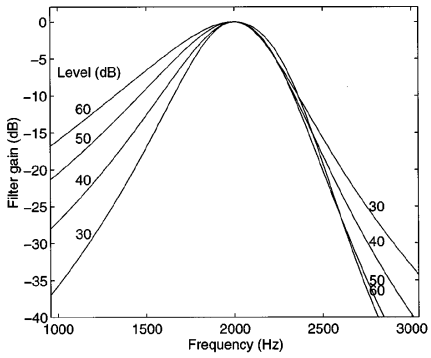


Figure 2.21: The transfer function of the gammachirp. Redrawn from Irino and Patterson [1997]

A different approach dealing with the level dependent filter shapes was proposed by Carney [1993]. The approach is based on a non-linear feedback mechanism reflecting the role of the outer hair cells (c.f. Figure 2.22).

In this case the gammatone filter impulse response is parameterised slightly differently:

$$h(t) = (t-\alpha)/\tau^{\eta-1}\, e^{-(t-\alpha)/\tau} \cos(\omega_{cf}(t-\alpha)) \tag{2.7}$$

where α is onset delay, τ is envelope decay time constant, $\omega_c f$ is the radian centre frequency and η is the order of the filter.

Briefly, the feedback loop effectively changes parameter τ of Equation 2.7 to model the broadening of filters with level. The filter asymmetry is controlled by the low-pass filter in the feed back stage. To align the responses in time according to the travelling wave, the signal is delayed.

This approach is just a part of a model aimed at auditory nerve responses and has exclusively been evaluated in this context.

The description of gammatone filters have focused on describing individual filters i.e., to model basilar membrane vibration for a single place along the basilar

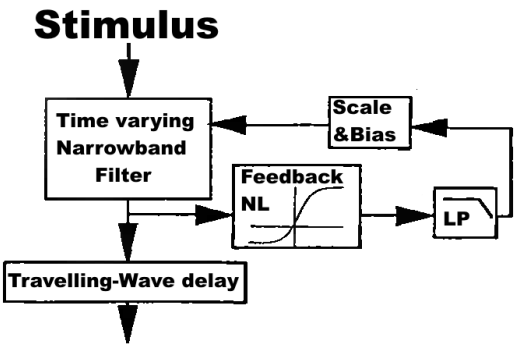


Figure 2.22: Schematic diagram of the feedback model used in Carney [1993]

membrane, one filter is constructed. In this way any given point on the basilar membrane can be modelled. With a gammatone filter bank the full basilar membrane vibration can be modelled limited only by the resolution of the filter bank.

Gammatone filters and gammatone filter banks are perhaps the most widely used model of basilar membrane vibration at present. Its well documented, easy implementable and computationally efficient properties are the reasons for its success.

2.4.4 Comments to basilar membrane models

The few examples of phenomenological basilar membrane models here are not exhaustive, but illustrates some of the proposed methods.

Two recent methods must be mentioned in this connection, although they are not discussed in detail. The model proposed by Dau et al. [1996] presents a quantitative model of the "effective" signal processing in the auditory periphery. It does so based on adaptation loops, which is a novel concept.

The model suggested by Heinz et al. [2001b] and Zhang et al. [2001] is physiologically based, but implemented as a computer model. The model is comprehensive in that it systematically accounts for large data sets.

The models discussed here have all been constructed from a theoretical baseline paired with data from physiological or psychoacoustic experiments. The result

is a mathematical description of basilar membrane vibration covering particular phenomena.

The next natural step in developing more general models would be to include more phenomena from multiple studies. The main problem arises when attempting to do this, is inter-subject and inter-study variability. For instance in order to account for data in the pulsation threshold experiment from Section 2.3.6 assumptions were made about the outer- and middle ear transfer functions. This information cannot be expected to be available from the study – and it is not available. It is possible, in this example, to model the data with a set of DRNL parameters and some assumptions about the outer- and middle ear, but it is not clear how general the derived parameters are. This fundamental problem is not specific to the DRNL, rather all models that claim to be general, i.e. capable of explaining several phenomena, face this challenge. Rather than discarding these types of models, however, this challenge should be addressed.

Chapter 3

Modifying the Basilar Membrane Model

This chapter describes how the DRNL parameters described in the preceeding chapter are modified in order to match a range of forward masking experiments. The eventual goal of this chapter is to arrive at a tractable, quantitative model description of the basilar membrane.

Section 3.1 describes the fundamentals of forward masking and how to model it. Next, a description of actual modelling efforts is given in Section 3.2. The two following sections summarises the results in that the final DRNL parameters for normal hearing and impaired hearing are presented. Finally, Section 3.5 summarises the issues presented in the chapter.

3.1 Theoretical background for modelling forward masking

This section describes the background for modelling forward masking. It does so by first introducing the basic psychoacoustics of forward masking. Next, a description of the methods employed in modelling forward masking is given. This description focuses on one method namely the so-called temporal window model.

3.1.1 The basic psychoacoustics of forward masking

Non-simultaneous masking or temporal masking are terms used to describe phenomena where time effects play the major role. The study of non-simultaneous masking combines aspects from temporal integration of loudness and frequency selectivity. Both of these topics are essential within hearing research and have been the topic of numerous theories and models over the years [Viemeister and Wakefield, 1991; Glasberg and Moore, 1990; Moore et al., 1997; Zwicker et al., 1957; Buus et al., 1999]). A general review of loudness and frequency selectivity is beyond the scope of this thesis. Instead this section concentrates on forward masking.

In temporal masking a short signal typically called probe (or sometimes just "signal") is presented non-simultaneously to a longer duration masker. Two types of non-simultaneous masking exist: forward masking (the masker is presented before the probe) and backward masking (the masker is presented after the probe). Figure 3.1 shows the basic stimulus structure applied in forward masking experiments.

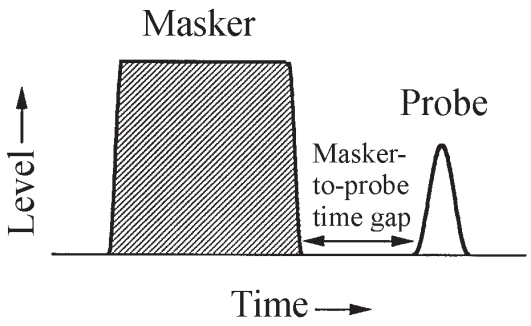


Figure 3.1: Illustration of the stimulus forward masking experiment

Backward masking is poorly understood partly because subjects improve performance with practise [Moore et al., 1988]. Therefore backward masking experiments require highly trained subjects making these experiments lengthy. More importantly even in trained subjects the effect is small [Oxenham and Moore, 1994]. In forward masking, in contrast, the effect is clear. The following will concentrate on forward masking.

Some facts about non-simultaneous masking.

1. Forward masking increases as the time between masker and probe is decreased.
2. Regardless of masker level no forward masking occurs after 100-200 ms [Moore, 1997] (obviously discounting physical damage to the hearing system).
3. The rate of recovery from forward masking increases with masker level [Moore, 1997].
4. Forward masking is influenced by the relationship between masker and probe frequencies.
5. The relationship between forward masker level and masked threshold is not simple, i.e. a 10 dB increase in masker level does not necessarily produce a 10 dB increase in masker level.
6. The amount of forward masking increases with duration of the masker. The duration effect is effective up to between 50 ms [Fastl, 1976] and 200 ms [Zwicker, 1984]. This is sometimes referred to as the masker duration effect. The term "time-intensity trade" is also used in this connection.
7. The masker duration effect increases with level [Plack and Oxenham, 2000b].
8. Forward and backward masking are not additive [Wilson and Carhart, 1971] in the sense that if a forward and backward masker were equally effective by themselves, applying them together with the probe, does not increase masking by 3 dB. This phenomenon is sometimes called excess masking.

The fact that the relationship between masker level and masked threshold is not simple can be explained by compression taken place at the basilar membrane. Typically the masker is in the compressed region of the basilar membrane input/output function. Therefore increasing the masker level by 6 dB will mean an increase in response of (much) less than 6 dB. If the probe is in the linear region of the basilar membrane input/output function the response increases linearly and hence the threshold increase is much less than 6 dB.

That the masker duration effect increases with level can also be explained by compression. Typically the probe is in the low level, linear region of the basilar membrane input/output function, and the masker is in the compressed region.

Increasing masker duration in this configuration will entail a proportional increase in probe level. At higher levels, in contrast, the probe might reach the compressed region of the basilar membrane input/output function. Increasing the masker duration in this configuration will require a larger than proportional increase in probe level.

Excess masking can be explained in terms of compression and temporal integration. Consider the example where an equally effective forward and backward masker are presented separately for the same probe level. Probe detection is believed to be achieved at a certain probe-to-masker energy ratio. In this example this ratio is the same for the forward and backward maskers. Presenting the forward and backward masker together, would mean the probe-to-masker energy is simply halved. Simply doubling the level of the probe would not restore the detectable probe-to-masker energy ratio since the probe is independently compressed.

The origin of forward masking

Forward masking is believed to originate from a combination of three sources: the basilar membrane, the auditory nerve and "a more central location". Firstly, even after cessation of the stimulus the basilar membrane vibrates for some time eliciting signals to the higher levels of the hearing system. In this way the basilar membrane plays a significant role in forward masking. Secondly, generally adaptation occurs in the auditory nerve even for stimulus of very short duration (e.g. Meddis [1986]). However, this effect has been deemed too small to be considered contributing significantly to forward masking [Relkin and Turner, 1988; Turner et al., 1994]. Moreover, adaptation alone cannot account for a range of forward masking phenomena quantitatively. The role of adaptation has recently been revisited [Oxenham, 2001] with the conclusion that it cannot be ruled out that adaptation, in the auditory nerve or elsewhere, plays a role in forward masking. Thirdly, the term "more central location" indicates that the exact origin is not known at present. As we shall see in the following section this "more central location" can be modelled as a black box based on behavioural studies.

Simultaneous masking versus non-simultaneous masking

The main reason for using non-simultaneous masking for estimating basilar membrane compression is that the simultaneous masking paradigms inherently

have the problem of suppression. When a masker is presented simultaneously with a tone they interact causing the nonlinear phenomenon of suppression to occur. In many situations the suppression effects can be ignored safely in simultaneous masking. However, for estimating basilar membrane compression simultaneous masking experiments have shown great discrepancies to physiological estimates. For instance the compression ratio was estimated to approximately 2.5 in van der Heijden and Kohlrausch [1995] whereas Yates et al. [1990] and Yates [1990] arrived at 4 - 5. One possible explanation for the discrepancy could be suppression [Oxenham and Plack, 1997].

In forward masking suppression effects can be completely eliminated by using pure tone maskers. On the other hand, other problems like spectral splatter and off-frequency listening are introduced. These issues have to be catered for by carefully designing the experiment.

3.1.2 Modelling forward masking

This section describes how forward masking can be modelled focusing on the temporal window model [Moore et al., 1988]. The temporal window model is, as most models of forward masking, a "persistence" model in that it hypothesises that forward masking is due to persistence of neural activity. The most successful example of an alternative modelling approach is the feedback loops presented in Dau et al. [1996]; Dau and Püschel [1996].

Temporal integration

Traditionally the time-intensity trade has been described by a temporal integration or accumulation process which frequently is given by a convolution integral of the type:

$$\mathrm{y}(t) = \int_{-\infty}^{t} \mathrm{h}(t-\tau)\,\mathrm{x}(\tau)\,d\tau \tag{3.1}$$

where $\mathrm{y}(t)$ is the function describing the temporal integration for time t, x is the signal as a function of time τ, h is a monotonically increasing weighting function. An alternative way of viewing the same temporal integration process as described in Equation 3.1 is simply as low pass filtering applying a filter with impulse response $h(-t)$.

This view clearly neglects the influence of peripheral processing and thus more recent models of temporal masking incorporate knowledge about the auditory periphery. In these models the input to the temporal integrator is taken from the output of the auditory periphery model. The auditory periphery model can be a simple power law or a complex physiological, mechanical or electrical model.

Temporal masking models applying the setup described above have traditionally been focusing on describing h in Equation 3.1. The models of the auditory periphery has been relatively simple, e.g. Oxenham and Moore [1994], consisting of applying a compression coefficient of 0.25 to the signal. Other models are slightly more advanced in that they divide the input/output function into a linear and a nonlinear region, e.g. Plack and Oxenham [1998]; Oxenham [2001].

Plack et al. [2002] suggested utilising the DRNL as the basilar membrane model while maintaining the temporal window model proposed by Oxenham and Moore [1994]. In this approach, which is adopted in this thesis, it is important to observe that the DRNL models *basilar membrane velocity*, which is at odds with the fact that the temporal window operates in the intensity domain. Simply squaring the DRNL output eliminates this problem. An additional complication of the approach is the fact that transduction only occurs for stereocilia shearing in one direction. Simply half-rectifying the DRNL output eliminates this problem.

The temporal window

The temporal window model is shown in its general, time-domain form in Equations 3.2 and 3.3 where $W(t)$ corresponds to $h(t)$ in Equation 3.1. Figure 3.2 shows the weighting function graphically.

$$\mathrm{W}(t) = (1-w)e^{t/T_{bp}} + we^{t/T_{bs}}, t \leq 0 \tag{3.2}$$

$$\mathrm{W}(t) = e^{-t/T_a}, t > 0 \tag{3.3}$$

where t is time relative to the centre of the window, W is the weighting function describing the shape of the temporal window, T_a is the time constant determining the slope of the window at time *a*fter the peak, T_{bp} and T_{bs} are the time constants *b*efore the peak, determining the slope close to the *p*eak and *s*kirt of

the window respectively, and w is the weighting factor determining the relative contributions of T_{bp} and T_{bs}.

The convention used here for the time axis is that $t < 0$ describes forward masking and $t > 0$ describes backward masking. Calculating the value of the conjectured internal variable for temporal masking is thus done by calculating the DRNL response for the relevant frequencies, to the stimulus and convolving the result with function W from Equations 3.2 and 3.3.

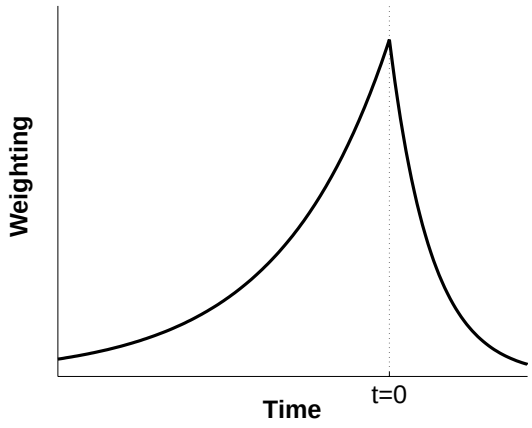


Figure 3.2: Illustration of the temporal window

The temporal model has more recently been used by Plack and Oxenham [1998]; Plack et al. [2002], where the set of parameters taken from Oxenham and Moore [1994] subject AO was used. The parameters are shown in the first column (TW 1) of Table 3.1. TW 2 corresponds to "Fit 2" of Oxenham [2001] and TW 3 corresponds to "Fit 3" of Oxenham [2001].

Combining Equations 3.2 and 3.3 with one column of constants from Table 3.1 gives instances of the function h in Equation 3.1.

Although the temporal window model in its generic form (Equations 3.2 and 3.2) is considered "general" the constants in Table 3.1 must be specified in order to apply the model. Oxenham [2001] argues that these constants depend on the model of the periphery. In other words the auditory periphery nonlinearity and the temporal window are interdependent and should be specified together.

The three sets of the temporal window model constants showed in Table 3.1 are the ones used in this thesis.

Table 3.1: Temporal window model weights and time constants

	TW 1	**TW 2**	**TW 3**
w	0.025	0.206	0.170
T_a	3.5	3.5	3.5
T_{bp}	4	3.1	4.6
T_{bs}	29	21.0	16.6

Determining thresholds with the temporal window model

The output of the temporal window is the time course of the value of a conjectured internal variable representing "signal strength". In the implementation described here a DRNL filter bank consisting of many frequency channels is applied as the model of the auditory periphery resulting in a spectro-temporal representation of a given signal. Modelling a two-alternative-forced-choice forward masking paradigm experiment, this spectro-temporal representation can be applied in the following way:

Thresholds predicted by the model is hypothesised to be determined by the ratio k between value of the internal variable calculated for the masker alone and the value of the internal variable calculated for masker and probe, at some "optimum place in the spectro-temporal representation".

In the case of pure tone maskers and pure tone probes it is assumed that the optimum place in the spectral domain corresponds to the probe frequency. So in this case the DRNL filter bank is reduced to one channel only. The optimum time is simply the time where the ratio is the biggest c.f. Figure 3.3.

There is no obvious way of estimating the ratio k which is the criterion for the threshold. For this reason k is often allowed to act as a "free parameter" enabling fitting of the thresholds. However, it seems reasonable to use the same k-value for similar conditions.

In order to be able to account for the threshold of the probe in the absence of the masker the additional concept of a noise floor is introduced. This means that a constant is added to the internal variable (calculated as the temporal window output). This constant, the noise floor, is determined by the threshold of the probe in the absence of the masker as described in Equation 3.4.

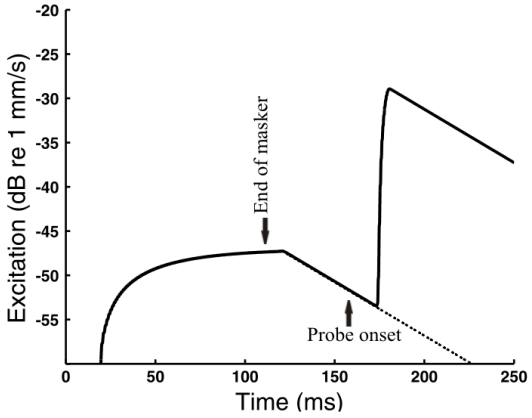


Figure 3.3: Illustration of the modelled internal representation in a forward masking experiment with 50 ms time gap. The solid lines shows the response to masker-and-probe whereas the dotted line shows the response to the masker alone. The arrows mark stimulus events (cessation of masker and onset of probe). For illustration purposes this example shows a probe far above threshold

$$R = \frac{\max(y(s_{thr}))}{k-1} \tag{3.4}$$

where R is the noise floor, $y(s_{thr})$ is the output of the temporal window model as described above in response to the probe at absolute threshold and k is the chosen decision ratio.

Briefly, modelling the results from a forward masking experiment is done in the following way: A threshold ratio k is chosen. From this k and the absolute threshold of the probe, i.e. probe without masker, the internal noise floor is calculated. For each condition in the experiment, the modelled threshold is determined by calculating the output to the temporal window, adding the noise floor and calculating the maximum ratio of the masker versus masker and probe. When this value reaches k the threshold predicted by the model is determined.

Potential weaknesses of the temporal window model

Although the temporal window models is, in many respects, a powerful model it does have weaknesses. The following section describes the most significant of these weaknesses.

It is assumed that the internal representation of the masker decays over time at the same rate for all frequencies. This rate is described in the temporal window model constants as shown in Table 3.1. This assumption seems reasonable, but no direct evidence supporting its validity is known to the author.

For simplicity it is assumed that the internal variable is determined by the basilar membrane vibration at *probe frequency only* (as in Plack et al. [2002]) combined with temporal integration. This seems justified by the fact basilar membrane response to the probe is at its maximum at the probe frequency. However, other probe detection methods, like detection by spectral splatter, cannot be precluded. Also, the basal ward shift of characteristic frequency with level is neglected. Finally, any influence that adaptation might have is ignored.

3.2 Modelling actual forward masking experiments

The previous section provided the tools for implementing models of forward masking experiments. This section describes three such simulations of forward masking experiments, 1) Forward masking with varying masker frequency, 2) Forward masking with varying probe and gap duration and 3) Forward masking with varying frequencies and gap durations.

The eventual goal is to arrive at new and improved DRNL parameters. To this end only data from the latter of the three experiments were used. The other experiments, described in sections 3.2.1 and 3.2.2, provided experience with modelling forward masking and guidance in choosing temporal window parameters.

3.2.1 Forward masking with varying masker frequencies

The psychoacoustic experiment

The first experiment modelled was the forward masking experiment described in Oxenham and Plack [1997]. Briefly, 4 ms sinusoidal probes[1] gated with 2 ms raised cosine ramps, i.e. no steady-state portion, were presented 2 ms after 104 ms maskers. The maskers were gated with 2 ms raised cosines ramps at both ends. To prevent off-frequency listening a background high pass or low

[1] All probe and masker in this section durations are given as zero-amplitude times

pass noise was presented simultaneously. The noise was not implemented in the model since off-frequency listening was presumed to be eliminated by only considering out put of one channel. The probe was always a 6 kHz sinusoid whereas maskers were either 3 kHz or 6 kHz sinusoids. The level of the probe was varied from 40 dB SPL to 90 dB SPL in steps of 5 dB. For each level, the corresponding thresholds were determined for both masker frequencies, i.e. the highest masker level for which the probe could be detected was determined. Average data across three normal hearing listeners was collected using a two-alternative-forced-choice scheme.

Modelling of the experiment

The simulation of the experiment was done as outlined in section 3.1.2. Parameters used for the DRNL was taken from Lopez-Poveda and Meddis [2001], Table III, without any modifications. The outer ear and middle ear transfer functions were also derived from Lopez-Poveda and Meddis [2001], Figure 2. In Plack et al. [2002] the HpTF and the characteristics of the middle ear are not discussed. The pre-emphasis filter is important in this modelling context for two reasons 1) It "calibrates" the basilar membrane nonlinearity in that it moves the input/output function along the abscissa depending on the gain applied and 2) The frequency response of the pre-emphasis filter at 3 kHz relative to 6 kHz determines the intercept of the 3 kHz masker thresholds (c.f. Figure 3.4). It can only be speculated that the pre-emphasis filter employed by Plack et al. [2002] is "reasonable" and most likely with a flat frequency response.

The decision device criterion used was determined as in Plack et al. [2002]. The optimal value for the criterion variable k was 2.9 dB in comparison to 1.41 obtained by Plack et al. [2002]. However, any k-value between 2 and 3 dB provides a reasonable fit to the data.

Results and comments

The result of the modelling efforts using the DRNL parameters from Lopez-Poveda and Meddis [2001], c.f. Appendix A, is shown in Figure 3.4. Temporal window parameters from Oxenham and Moore [1994] were employed (c.f. Table 3.1).

The model prediction of the data for the 3 kHz masker is 2 – 4 dB too low for probe levels from 40 to around 75 dB. This corresponds to the compressed region for the masker. This could be interpreted as an inaccurate model representation

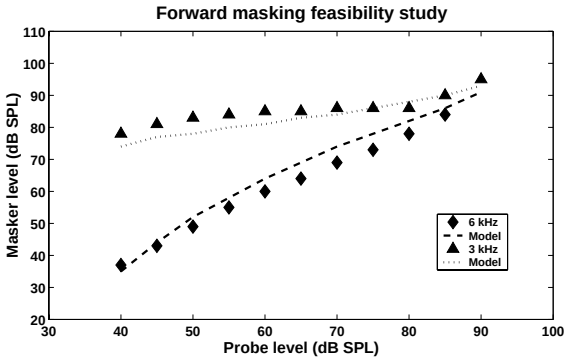


Figure 3.4: Forward masking study with varying masker frequencies. The total RMS errors is 4.01 dB

of compression in that the slope of the line from the model data is too steep. A complementary interpretation of the discrepancy could be offered by the pre-emphasis filter in that the 3 kHz component is too strongly represented, or conversely the 6 kHz component is too weakly represented.

The model prediction of the on-frequency masker data is better, but still some points are 2 – 3 dB away from the measured data points. Again, the slope appears to be slightly steeper than warranted by the measured data indicating a slight overestimation of the compression coefficient. However, considering the limited number of points in the plot these interpretations should be considered tentative.

The model clearly explains the data qualitatively and to a fair extent also quantitatively considering that the standard deviation of the original study is comparable to the 2 – 4 dB distance from the measured data points to the simulated data points. Moreover, considering the fact that this study was done without any adjustment to the parameters from Lopez-Poveda et al. [2002] and also keeping in mind that these parameters are actually an approximation to a human filter bank, for which no data points for the 6 kHz place were employed, the match is quite remarkable.

This simulation indicates that the proposed method and model is capable of explaining "basic" forward masking data, i.e. masker level as a function of probe level for maskers on-frequency and an octave below at 6 kHz probe frequency and a fixed time gap.

The fact that the model parameters for the 6 kHz place is not optimal for this simulation suggest that some fitting of the basilar membrane parameters would improve the results.

In conclusion the obtained model prediction encourages the investigation of more probe frequencies, more masker frequencies, i.e. more probe-to-masker frequency combinations, and more time gaps. This is exactly the conditions examined in Lopez-Poveda et al. [2002], addressed in Section 3.2.3.

Investigating temporal window constants

Table 3.2 shows how different temporal window shapes affect the simulation. The resulting sum of squared errors are all very close for the three shapes of the temporal window indicating that the shape is not important for simulating this experiment. TW 1, TW 2 and TW 3 are as defined in Table 3.1

Table 3.2: Criterion and RMS error for simulation of Plack et al. [2002]

	TW 1	**TW 2**	**TW 3**
k	2.9	1.5	1.3
RMS error	4.01	4.17	4.05

Given that the time gap is fixed in this experiment and the very short time gap of 2 ms, it is not surprising that the temporal window shape is not crucial to the simulation.

3.2.2 Forward masking with varying probe and gap durations

The original study

The second experiment modelled was the forward masking experiment carried out by Oxenham [2001]. Briefly, the masker consisted of a 200 ms broadband Gaussian noise (0 - 7 kHz) gated with 2 ms raised cosine ramps at a fixed spectrum level of 40 dB re 20 μPa. The probe was a 4 kHz sinusoid. Thresholds were measured for the following offset-offset intervals: 4, 6, 9, 12, 22, 52 and 102 ms. Within each of these offset-offset times, the same range of signal durations were used, provided that no overlap in time of masker and probe occurred. For example thresholds were measured for probe durations of 4, 6, 9 12 and 22 ms

for the 22 ms offset-offset interval. In the text the probes are referred to by their half-amplitude duration (2, 4, 7, 10, 20, 50 and 100 ms). A three-interval forced-choice, two-down, one up adaptive method was used.

In Oxenham [2001] the psychoacoustic experiment was modelled in the following way: The stimulus was represented by the *amplitude envelope* which was passed through a nonlinearity described by:

$$L_{out} = \begin{cases} 0.78L_{in}, & L_{in} < 35 \text{ dB SPL} \\ 0.16L_{in} + 21.7, & L_{in} \geq 35 \text{ dB SPL} \end{cases} \tag{3.5}$$

where L_{in} is the input amplitude level in dB SPL and L_{out} is the level of the basilar membrane vibration velocity in dB, which is an *intensity measure* [Plack and Oxenham, 1998].

The masker amplitude level was set corresponding to the level passing through the auditory filter centred at 4 kHz with equivalent rectangular bandwidth of 456 Hz according to Glasberg and Moore [1990].

The output from the nonlinearity passed through the temporal window. The best results are achieved with the "Fit 3" window of Table 3.1 which is the temporal window parameters specifically fitted to the nonlinearity from Equation 3.5. Since there is no Gaussian noise involved in this implementation of the experiment, it is deterministic. This means that the procedural aspects of the psychoacoustic experiment can be ignored in the modelling efforts.

Modelling the experiment

The implementation of the experiment using the DRNL was done in a similar, yet slightly different way. The difference between the method used in Oxenham [2001] and the method employed here is the representation of the stimulus. Since the DRNL produces actual basilar membrane velocity as output, it is not possible to present the stimulus as the amplitude envelope. Instead an equivalent stimulus to the actual time signal, described in the following, was developed.

The criterion as described in Section 3.1.2 only considers output from one of the DRNL filters in the filter bank, namely the filter centred around the probe frequency. However, filtering a Gaussian noise at the probe frequency results in large, random fluctuations in the output. These fluctuations would be spread across the duration of the masker. Further, comparing two Gaussian noise maskers, the fluctuations might exceed the chosen criterion ratio k as described

in Section 3.1.2. This would mean that rather than determine forward masking thresholds, the model would find thresholds for detecting differences in Gaussian noises at probe frequency. This has absolutely nothing to do with the desired forward masking threshold.

Limiting the search for ratio differences between masker and masker+probe to times after the masker is not enough, since the random fluctuations might peak towards the end of the masker. Alternatively, frozen noise could have been used, but then the simulated thresholds depend on the frozen noise. Moreover, implementing a method using real Gaussian noise would make the simulations nondeterministic, i.e. depending on the noise, in which case the psychoacoustic procedure becomes important. It is not obvious how to devise a method for simulating the three-interval forced choice procedure applied in the study – on the contrary it raises some fundamental questions as to how humans perform this task... Even if such a method could be devised, it would require many iterations in order to eliminate the stochastic element introduced by applying Gaussian noise[2].

A better solution to the problem, which is in line with the amplitude envelope method used by Oxenham [2001], is to simulate the noise by a pure tone at probe frequency. The level of this tone is calculated as the level of the noise passing through the DRNL at probe frequency.

The total level of the 40 dB spectrum level, 7 kHz bandwidth noise is 78 dB SPL ($40 + 10 \log 7000 = 78$ dB SPL [Hartmann, 1997], Equation 3.62). The level of the noise that passes through the auditory filter at 4 kHz with an equivalent rectangular bandwidth of 456 Hz [Glasberg and Moore, 1990] is $40 + 10 \log 456 = 67$ dB SPL corresponding to the envelope amplitude method used in Oxenham [2001]. A 70 dB SPL 4 kHz sinusoid filtered by the DRNL filter centred at 4 kHz, with parameters from Lopez-Poveda and Meddis [2001], yields an equivalent amount of excitation as the noise passed through the auditory filter. The noise masker was thus simulated by the 4 kHz sinusoid, on frequency masker. The probe was a 4 kHz sinusoid. This makes the simulation deterministic and thresholds can be found quickly.

[2]One of the advantages of *modelling* psychoacoustic experiments rather than carrying them out for real is that in most cases the procedural and stochastic aspects of psychophysical experimentation can be circumvented

Results and comments

Figure 3.5 compares the simulated results obtained using the DRNL and the measured results from the original study.

The simulated thresholds are close to the measured thresholds. The average difference is 1.84 dB, which is considerably better than the 4.05 dB reported for the "first try" by Oxenham [2001]. Fitting both peripheral nonlinearity and the temporal window parameters Oxenham [2001] achieves a fit with an RMS error of 1.34 dB[3]. Although the fit achieved using the DRNL is not as good as the fit in Oxenham [2001] it is still quite good. Especially considering that it is achieved without any fitting at all – the DRNL parameters are taken straight from Lopez-Poveda and Meddis [2001] and the temporal window parameters are taken straight from Oxenham and Moore [1994].

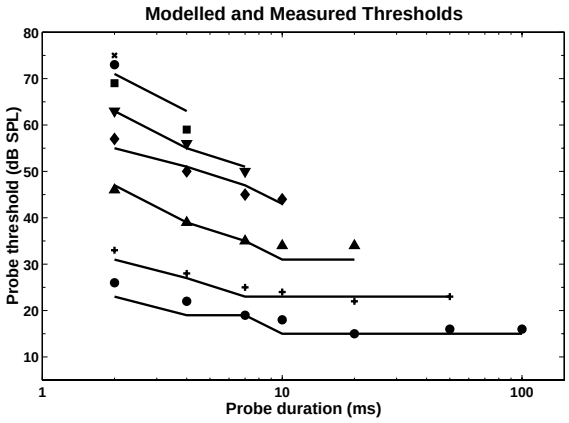


Figure 3.5: Comparison of modelled forward masking thresholds to measured thresholds from Oxenham [2001] as a function of probe duration. Measured data points are plottet as symbols, modelled data is shown as solid lines, except for the 4 ms offset condition where an **x** is used. The parameter is the masker-probe offset-offset time interval. These time intervals are from top to bottom of the graph 4, 6, 9, 12, 22, 52 and 102 ms. The RMS error or average euclidian distance between simulated and measured data points is 1.84 dB

.

[3]The 1.34 dB is calculated from the sum of squared errors value of 50.7 by dividing by the number of data points (28) and taking the square root. This is at odds with the fact that Oxenham [2001] counts the absolute threshold in quiet making the number of data points 29. However, the difference is very small

One of the conclusions from Oxenham [2001] is that the peripheral nonlinearity and the temporal window constants should be matched, i.e. a given peripheral nonlinearity requires a given set of temporal window constants in order to give accurate predictions.

Table 3.3: Simulation of Oxenham [2001] for different temporal windows.

	TW 1	**TW 2**	**TW 3**
k	3.25	1.30	2.00
RMS error	1.84	2.61	3.38

In order to examine if the DRNL has a better matching set of temporal window constant the TW 2 and TW 3 constants from Table 3.1 were examined. The results are shown in Table 3.3. It is clear that the TW 1 constants are better for simulating this experiment. TW 1 was also better for the experiment by Plack et al. [2002]). This leads to the preliminary conclusion that TW 1 matches well with the DRNL parameters employed here. A detailed discussion of the interdependence of the nonlinearity and the shape of the temporal window can be found in Section 3.2.4.

3.2.3 Forward masking with varying frequencies and gap durations

The experiment

In the forward masking experiment in this section [Lopez-Poveda et al., 2002] the test subjects first had their audiogram made. Further, their thresholds to the probe signal in quiet was determined.

Their forward masking thresholds were then determined by varying the *masker level* for a range of conditions while keeping the probe level at 14 dB SL. The audiogram was used to determine the physical level of the probe.

The conditions were based on three variables probe frequency, masker frequency and time gap between masker and probe. Table 3.4 shows the values of these three variables. The third row shows the masker frequency as the ratio to probe frequency. For example, the lowest masker frequency for the 500 Hz probe would be 250 Hz – the highest 800 Hz. The three variables were varied independently, i.e. the total number of conditions was $10 \times 5 \times 6 = 300$.

Table 3.4: Conditions used in [Lopez-Poveda et al., 2002].

Time gap values	Probe frequency values	Prob-to-masker frequency ratio
10 ms	500 Hz	× 0.5
20 ms	1000 Hz	× 0.7
30 ms	2000 Hz	× 0.9
40 ms	4000 Hz	× 1.0
50 ms	8000 Hz	× 1.1
60 ms		× 1.6
70 ms		
80 ms		
90 ms		
100 ms		

The sinusoidal probe had a duration of 8 ms including 4 ms cosine onset ramp and 4 ms offset ramp (i.e. no steady-state part). The sinusoidal maskers had a duration of 108 ms including 4 ms cosine onset ramp and 4 ms cosine offset ramp.

The procedure was a two alternatives forced choice similar to the one used by Plack and Oxenham [1998]. The thresholds for all the 300 conditions listed in 3.4 were determined for three test subjects. Each threshold was determined based on 16 turn points. Each threshold was determined three times. For further details refer to Lopez-Poveda et al. [2002].

The experiment as outline here is special in that the probe level is not varied (as one of the only variables!). As we shall see the reason for this is to ensure that the probe is in the linear section of the basilar membrane input/output function.

Forward masking simulation results

To get a first impression of how well forward masking thresholds are simulated by the DRNL and temporal window model, the case where probe and masker frequency are identical is considered. The results are presented in Figure 3.6.

The RMS errors clearly exceed an acceptable level although the thresholds are qualitatively matched in the simulation. The slope for the simulated lower frequencies (500 Hz and 1 kHz) thresholds appear to be slightly off in comparison

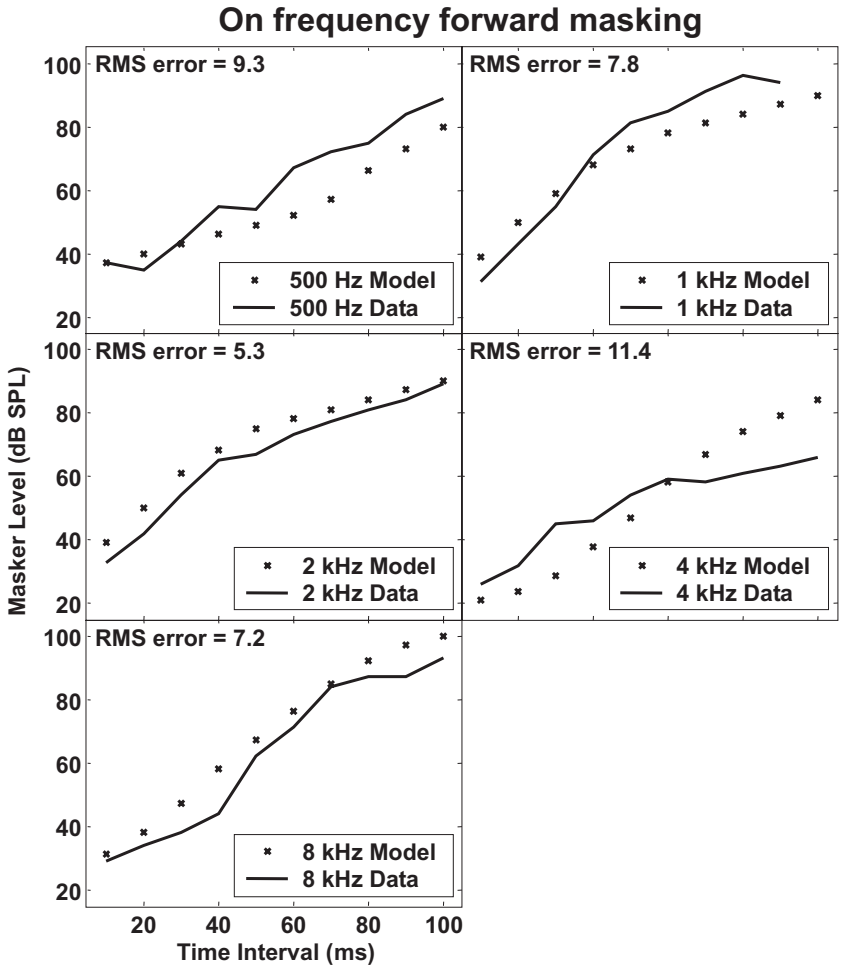


Figure 3.6: Comparison of forward masking results for the on frequency results. The criterion value k was varied across frequency. The TW 3 temporal window provided the best fit

to the measured data, whereas the slopes for the 2 kHz and 8 kHz seem to be correctly modelled. The 4 kHz match is particularly bad with a complete mismatch of the slope. A possible explanation will be explored in Section 3.2.4.

Although the match shown here is not sufficient to claim a quantitative explanation of the data, it does provide a qualitative explanation. Moreover, these results were obtained without any alterations to previously published DRNL parameters.

Table 3.5 shows the modelled results summarised by the RMS error in dB for each of the probe frequencies (rows) and for each of the sets of temporal window constants (columns). Each RMS error represents data for 6 frequency ratios per 10 time intervals i.e. 60 data points. The criterion value k was kept constant within each combination of centre frequency and set of temporal window constants. The DRNL parameters were taken from Lopez-Poveda and Meddis [2001]

Table 3.5: RMS error for modelled results

	TW 1	TW 2	TW 3
500 Hz	15.4	16.9	10.7
1 kHz	12.9	13.3	10.6
2 kHz	10.8	10.1	9.3
4 kHz	11.0	11.0	12.3
8 kHz	8.5	7.8	8.4

The overall best fit is provided by the TW 3 temporal window constants and this window will be used in the remainder of this section. The discussion of temporal window constants is deferred to Section 3.2.4.

The simulated results for the entire data set, for the subject ELP, is shown in Figure 3.7. The RMS errors corresponding to Figure 3.7 are shown in Table 3.6. It shows a breakdown of the RMS error in dB for the combination of the masker-to-probe frequency ratios (columns) and probe frequencies (rows). The N/A (not applicable) fields indicate that the simulated thresholds were unrealistically high (above 105 dB SPL) and therefore not considered. The RMS errors are far from an acceptable level for every combination of centre frequency and temporal window constants.

For the low centre frequencies the thresholds for the lower frequency maskers (frequency ratios 0.5, 0.7, 0.9 and 1) seem to show the correct slope. Conversely, for the higher centre frequencies the thresholds for the higher frequency maskers

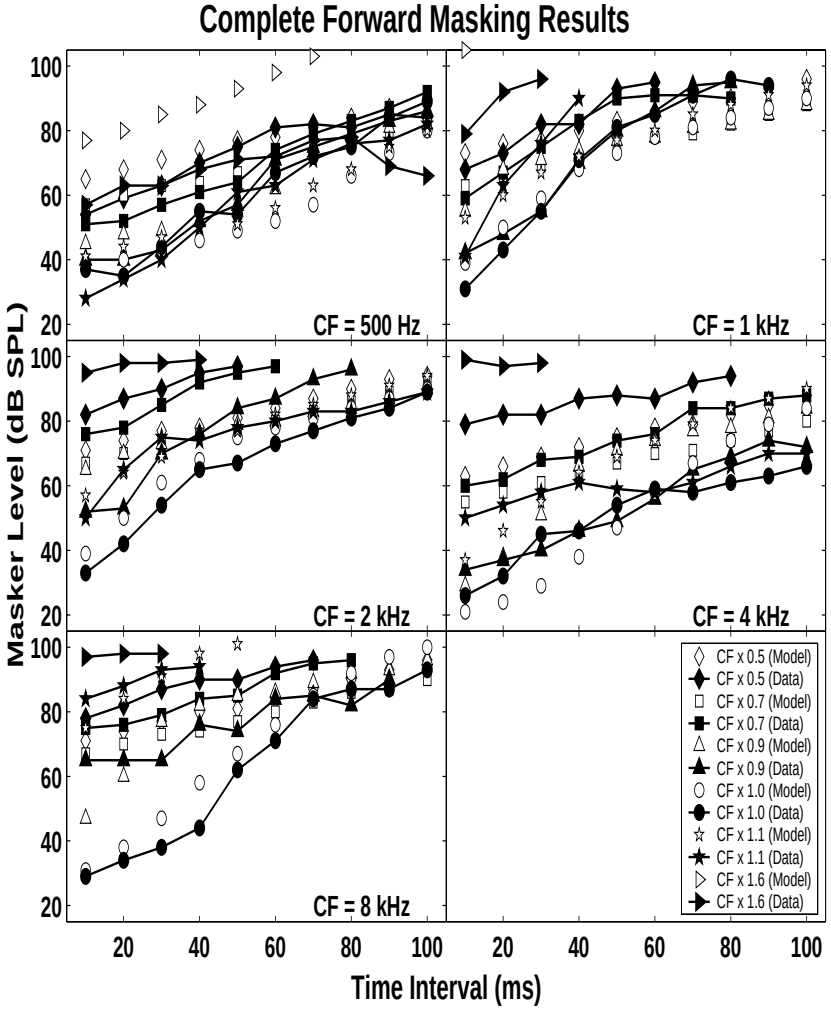


Figure 3.7: Comparison of all the measured forward masking thresholds to the simulated thresholds . The criterion value k was varied across frequency. The TW 3 temporal window provided the best fit

appear to match the measured thresholds. Table 3.6 shows the RMS for the combination of masker and probe frequencies.

Table 3.6: Detailed modelled results

	x 0.5	x 0.7	x 0.9	x 1	x 1.1	x 1.6
500 Hz	5.87	6.47	4.81	9.30	7.77	22.89
1 kHz	6.38	9.85	12.51	7.77	11.81	26.00
2 kHz	14.17	13.89	9.91	5.31	4.12	N/A
4 kHz	13.55	7.96	13.32	11.35	13.94	N/A
8 kHz	9.06	9.27	9.30	7.23	5.40	N/A

Although the previously published DRNL parameters showed acceptable or even good results in the two simulated studies previously discussed in sections 3.2.1 and 3.2.2 the parameters are inadequate in explaining the data from the more comprehensive range of conditions in Lopez-Poveda et al. [2002]. For this reason the DRNL parameters will be modified in order to simulate the data from the study more accurately. The detailed discussion of the results presented in this section is thus deferred and collated with the discussion of the improved DRNL parameters in Section 3.2.3. First, however, the method for improving the DRNL parameters is discussed.

Fitting DRNL parameters using forward masking data

Although Lopez-Poveda and Meddis [2001] was based on pulsation threshold data from Plack and Oxenham [2000a] containing much data, it can be argued that Lopez-Poveda et al. [2002] provides an even better basis for deriving DRNL parameters in that it contains even more data. Even though it is difficult to compare a pulsation threshold study to a forward masking study, for modelling purposes it seems reasonable to consider the number of data points and their contribution towards improving the model. Seen from this perspective the two advantages of Lopez-Poveda et al. [2002] over Plack and Oxenham [2000a] is that the former investigates more masker-to-probe frequency ratios and more masker-to-probe time intervals. Eventually, the results should be reconciled, however, there are many intermediate steps before this can be achieved. Below is an outline of these intermediate steps.

1. Fit the DRNL parameters to the forward masking data for each of the subjects. This is only done for the centre frequency represented in the data (i.e. 0.5, 1, 2, 4 and 8 kHz).

2. Estimate an average across subjects for each of the derived/modified DRNL parameters.
3. Derive a filter bank from the five centre frequencies.
4. Compare and adjust the DRNL parameters to the parameters derived using the pulsation threshold data.
5. Reiterate

Unfortunately, it was not possible to carry out all of the above intermediate steps in the course of the project. To derive at least one set of DRNL parameters it was decided to only use data from one test subject. Further, items four and five from the list were not implemented.

Fitting forward masking data for one subject

The basic fitting method is essentially a trial-and-error method: The thresholds are determined as described in Section 3.1.2. One or more DRNL parameters are then modified. The new thresholds and the RMS error are calculated. The RMS error is compared to the previous best RMS error. If the RMS error is smaller the new parameters are adopted. It was soon realised that this trial-and-error method was inadequate in that it was inefficient.

Fundamentally, the trial-an-error procedure has three problems:

1. The procedure does not specify how the DRNL parameters should be changed in order to improve the thresholds.
2. The 8-dimensional function (7 parameters for each frequency, c.f. Section 2.3.3) constitute a large search space with potentially many local optima. There are no obvious pointers as to what the search space looks like.
3. Even with a more efficient implementation of computing the masked threshold, computation time is still too long to allow exhaustive search of the parameter space.

There seems little one can do to fix the latter of the three problems. Approaching the task from an optimisation point of view would provide guidance from

the comprehensive literature in this research area (e.g. Press et al. [2002]). However, it would also mean a shift of focus from hearing research to optimisation research. Moreover, it is not clear if the results would be better using optimisation methods. For these reasons it was decided to improve the trial-and-error method by addressing the two first problems from the list above. A more transparent way of relating suggested DRNL parameter improvements to the cost function, i.e. RMS error, was developed.

The details of the method is described in Appendix D

Presentation of the improved results

After applying the parameters with the method described in Appendix D the thresholds shown in Figure 3.8 were obtained. The corresponding RMS errors are shown in Table 3.7.

Table 3.7 shows the modelled results summarised by the RMS error in dB for each of the probe frequencies (rows) and for each of the sets of temporal window constants (columns). Each RMS error represents data for 6 frequency ratios per 10 time intervals, i.e. 60 data points. The criterion value k was kept constant within each combination of probe frequency and set of temporal window constants. The DRNL parameters were fitted using the method described in Appendix D.

Table 3.7: RMS error for modelled results

	TW 1	TW 2	TW 3
500 Hz	12.8	10.9	9.4
1 kHz	12.0	18.4	9.4
2 kHz	6.9	6.3	5.9
4 kHz	7.6	12.6	7.0
8 kHz	6.0	4.9	4.9

In comparison to the results from Table 3.5 the improvements are clear. Every single RMS error is improved in comparison to its counterpart. The improvement is from a few dB to around five dB. Having said that, the RMS errors are still too big to claim that the measured thresholds are accounted for by the model.

As was the case for the original DRNL parameters the TW 3 temporal window provides the better results.

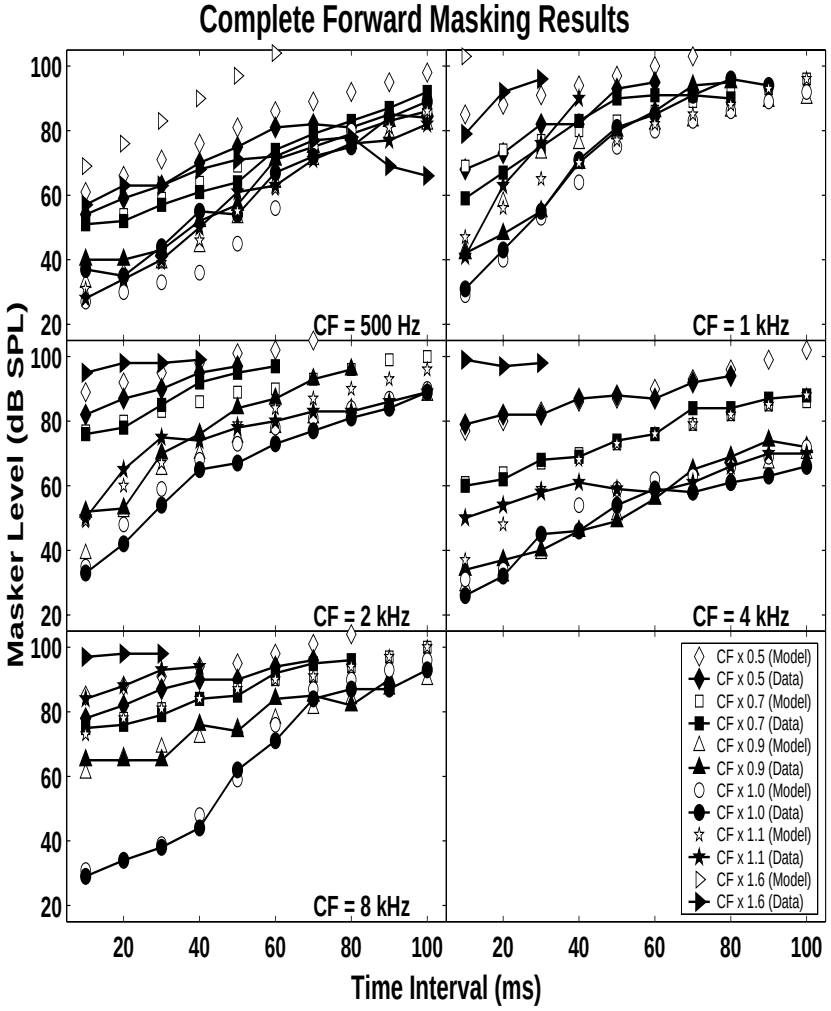


Figure 3.8: Comparison of all the measured forward masking thresholds to the simulated thresholds . The criterion value k was varied across frequency. The TW 3 temporal window provided the best fit

Another observation is that the higher probe frequency the lower RMS error. The is consistent across the three different temporal windows examined.

Table 3.8 shows the RMS errors broken down into the individual probe frequencies (rows) and probe-to-masker frequency ratios (columns). Immediately below each RMS error the corresponding mean standard deviations for the *measured* data point is shown. The N/A (not applicable) fields indicate that the simulated thresholds were unrealistically high (above 105 dB SPL) and therefore not considered. The mean standard deviations were not presented in Lopez-Poveda et al. [2002], but were made available by the authors.

Table 3.8: Breakdown of the RMS error in dB

	x 0.5	x 0.7	x 0.9	x 1	x 1.1	x 1.6
500 Hz	8.51	3.27	5.17	9.24	3.08	21.97
Std Dev	5.01	4.54	5.09	4.13	4.03	5.78
1 kHz	11.40	5.52	9.22	5.93	12.30	24.00
Std Dev	3.78	3.14	3.44	3.15	8.65	3.15
2 kHz	4.98	4.65	9.23	4.12	5.20	N/A
Std Dev	2.96	3.03	3.35	2.83	3.27	N/A
4 kHz	1.77	2.12	4.39	4.99	13.80	N/A
Std Dev	3.54	3.32	4.00	3.84	3.30	N/A
8 kHz	5.11	2.12	3.57	3.44	10.78	N/A
Std Dev	3.23	2.65	4.13	3.00	3.56	N/A

Comparing RMS error to the standard deviation of the *measured* data point seems a reasonable method for assessing the quality of the model prediction. Using this assessment method the derived DRNL parameters cannot be said to account for all the measured thresholds in the cases where the probe-to-masker frequency ratio is above one (the 500 Hz and 2 kHz probe being notable exceptions for a ratio of 1.1).

Further, the DRNL parameters account poorly for the maskers an octave below the probes – except for a 4 kHz probe frequency.

The modelled thresholds show a tendency to better match the measured thresholds for shorter time intervals.

Discussion of the improved results

The fact that the higher probe frequencies provide the better results is hardly surprising in view of the two previously modelled studies (c.f. sections 3.2.1 and 3.2.2). These studies provided stimuli of either 3, 4, 6 kHz or broadband Gaussian noise. That the peripheral nonlinearities and the temporal window model does not directly transfer to account for the data for probe frequencies of below 4 kHz might seem surprising. However, the modelling efforts pursued here tried to match the data by modifying the peripheral nonlinearity only. Modifying *both* the nonlinearity *and* the temporal window shape might have given better results. Theoretically it should be possible, however, to observe the constraint that the peripheral nonlinearity should "match" the temporal window [Oxenham, 2001] by fitting the peripheral nonlinearity only.

A possible explanation for the fact that the thresholds for the extreme masker-to-probe ratios are not modelled convincingly might be the outer-/ middle ear model employed. While the headphone-to-eardrum transfer function was carefully measured and modelled the middle ear model was based on data from Goode et al. [1994] which serves as a generic model. Any discrepancies between this generic middle ear model and the actual subject middle ear transfer function will thus particularly affect the thresholds for extreme masker-to-probe frequency ratios. Consider, for example, the probe frequency 500 Hz and masker frequency 250 Hz. The simulated thresholds are higher than the measured thresholds. If the middle ear filter had a shallower roll-off towards lower frequencies this would increase the masking effect and thus lower the required masker level at thresholds. A similar argument holds for 1 kHz probe frequency and 500 Hz masker.

Middle ear transfer functions do show variation supporting the argument above [Puria et al., 1997; Aibara et al., 2001]. In this connection it is important to keep the criterion value k in mind. A higher k value means that the model requires a larger level difference between excitation from the probe and excitation from the masker, i.e. the larger k-value the lower the thresholds. Moreover, changing the middle ear transfer function moves the compression knee-point. So improving the match of the simulated data to the measured data by modifying the middle function affects the entire fitting procedure.

Another possible factor in explaining the poor results for low frequencies is that the middle ear reflex is more likely to influence results at low frequencies. The masker levels are in the range where the middle ear reflex is active and the middle ear reflex is not implemented in the model.

Off-frequency listening is more likely at low frequencies because of ringing of the narrower auditory filters. This fact can explain discrepancies between the measured data and the simulated data since off-frequency listening was not modelled.

Moreover, the basal ward shift of characteristic frequency with level is not modelled, which might be a problem since the masker levels are quite high for many of the conditions.

Finally, one of the basic assumptions of this version of the temporal window model, that adaptation does not play a role in forward masking, might not turn out to hold.

3.2.4 Interdependence of nonlinearity and temporal window in the temporal window model

The temporal window model accounts for forward masking data by means of the peripheral nonlinearity and a integrator or temporal window. To optimise this model the integrator parameters must be fitted to the nonlinearity hypothesised [Oxenham, 2001]. In the previous sections three such fits were used. The conclusion was that the TW 3 constants matched the nonlinearity best.

To explain why this is the case one has to consider the nonlinearities to which the TW 1, TW 2 and TW 3 constants were derived.

$$L_{out} = 0.5L_{in} \tag{3.6}$$

$$L_{out} = \begin{cases} 0.78L_{in}, & L_{in} < 35 \text{ dB SPL} \\ 0.16L_{in} + 21.7, & L_{in} \geq 35 \text{ dB SPL} \end{cases} \tag{3.7}$$

$$L_{out} = \begin{cases} L_{in}, & L_{in} < 35 \text{ dB SPL} \\ 0.25L_{in} + 26.25, & L_{in} \geq 35 \text{ dB SPL} \end{cases} \tag{3.8}$$

where L_{out} is the level of the basilar membrane output and L_{in} is the level of basilar membrane input.

The TW 1 constants [Oxenham and Moore, 1994] were derived with the nonlinearity corresponding to Equation 3.6. The TW 2 constants [Oxenham, 2001] were derived with the nonlinearity corresponding to Equation 3.7. The TW 3 constants [Oxenham, 2001] were derived with the nonlinearity corresponding to Equation 3.8.

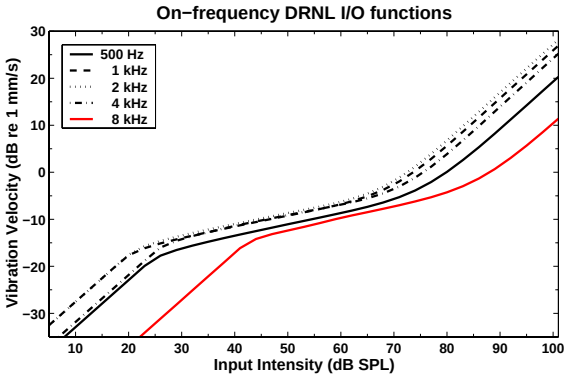


Figure 3.9: The DRNL I/O functions for five characteristic frequencies with on-frequency stimulus. The filter bank parameters from Lopez-Poveda and Meddis [2001] were used. The simulation includes the pre-emphasis filter used in the simulation. In comparison a "flat" outer- and middle ear model was presumably used in Oxenham and Moore [1994] and Oxenham [2001]

In comparison the DRNL has I/O functions shown in Figure 3.9. For the 4 kHz data this can be approximated by the function shown in Equation 3.9, where the output reference is normalised so that 0 dB SPL input corresponds to 0 dB output.

$$L_{out} = \begin{cases} L_{in}, & L_{in} < 22 \text{ dB SPL} \\ 0.28L_{in} + 15.84, & 22 \text{ dB SPL} \leq L_{in} < 65 \text{ dB SPL} \\ 0.69L_{in} - 10.81, & L_{in} \geq 65 \text{ dB SPL} \end{cases} \tag{3.9}$$

where L_{out} is the level of the basilar membrane output and L_{in} is the level of basilar membrane input.

By directly comparing the DRNL I/O function in Equation 3.9 to the three auditory nonlinearities described in Equations 3.6 to 3.8 it is clear that up to 25 dB SPL the TW 2 nonlinearity is closer to the DRNL. From 25 to 65 dB SPL TW 3 nonlinearity is the closest to the DRNL. The TW 1 nonlinearity is closest above 65 dB SPL. These observations point towards the conclusion that the TW 3 constants should be used when modelling the nonlinearity with the DRNL. This conclusion is in agreement with the conclusion based on the actual simulations, although TW 1 was clearly better for simulating Oxenham and Plack [1997] and Oxenham [2001].

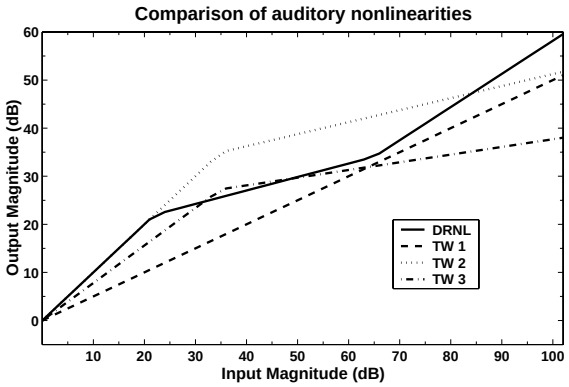


Figure 3.10: Comparison of the nonlinearities used for deriving temporal window constants

Obviously, deriving yet a new set of temporal window constants for the DRNL would have been optimal, however for time reasons this was not done.

One important consequence of the fact that the temporal window constant should match the peripheral nonlinearity [Oxenham, 2001] is, that since the peripheral nonlinearity is a function of frequency, in principle, so should the temporal window be. This was investigated by Plack and Moore [1990] with a simpler nonlinearity than the DRNL. In spite of the more sophisticated nonlinearity the conclusion, that temporal window shape is a function of frequency, still appears to hold.

3.3 Final basilar membrane model for normal hearing

According to the fitting procedure suggested on page 64 data for more subjects should be fitted and an average filter bank derived. Also, ideally the derived DRNL parameters should be verified against the original pulsation threshold study, For time reasons this has not been done.

With the provisions discussed above the DRNL parameters shown in Table 3.9 constitute the basis of constructing a human DRNL filter bank. The parameters were derived in Section 3.2.3 and provided the best fit in to the study.

Table 3.9: The DRNL parameters as determined in Section 3.2.3

	Centre Frequency (Hz)	500	1000	2000	4000	8000
Linear-path-parameters	CF_{lin} (Hz)	462	965	1925	3900	7750
	BW_{lin} (Hz)	130	240	355	316	891
	g	316	63	79	3162	891
Nonlinear-path-parameters	CF_{nl} (Hz)	500	1000	2000	4000	8000
	BW_{nl} (Hz)	103	175	562	891	501
	a	5184	7558	9627	11220	43584
	b	0.05	0.02	0.02	0.056	0.071

Table 3.10 shows the linear regression line coefficients for the DRNL human filter bank derived from the parameters presented in Table 3.9. The parameters were derived using linear regression as described in Section 2.3.5. These parameters constitute the human DRNL filter bank for normal hearing used in the remainder of this thesis.

Table 3.10: Derived linear regression line coefficients

	Regression coefficients	p_0	m
Linear-path-parameters	CF_{lin} (Hz)	-6.8193 x10^{-2}	1.0151 x 10^{0}
	BW_{lin} (Hz)	5.3444 x10^{-1}	5.9507 x 10^{-1}
	g	-3.2290 x10^{-2}	8.6403 x 10^{-1}
Nonlinear-path-parameters	BW_{nl} (Hz)	2.4926 x 10^{-1}	6.9124 x 10^{-1}
	a	1.8371 x 10^{0}	6.7133 x 10^{-1}
	b	-2.2442 x 10^{0}	2.4972 x 10^{-1}

3.4 Final basilar membrane model for impaired hearing

Sensorineural hearing loss results in a loss of sensitivity in that soft sounds are not heard. This in turn entails a reduction of the dynamic range of hearing. The reduction in dynamic range leads to an abnormal growth of loudness referred to as recruitment. Expressed in terms of auditory signal processing a person with a sensorineural hearing loss has a less compressive basilar membrane input/output function.

The loss of sensitivity can be modelled in the DRNL by reducing the output from the filter. This can be achieved by reducing the contributions from the linear as well as the nonlinear paths. The reduction from the linear path is reduced by reducing the linear gain factor g (c.f. Table 3.10).

Reduction in the dynamic range can be simulated in the DRNL by adjusting the compression scale parameters a and/or b (c.f. Table 3.10) in the nonlinear path. The method described here, reducing parameters a, b and g was used in Lopez-Poveda and Meddis [2001]. However, since parameter g represents the gain effective at high stimulus levels where impaired hearing and normal hearing converge, it seems unrealistic that g is affected.

The problem remains, that simulating a moderately severe hearing loss, as defined in Stach [1998], requires further reduction in sensitivity than is possible by modifying compression scale parameters a and b alone. Similarly, moderately severe hearing losses typically consist in outer *and* inner hair cell malfunctioning.

Hence inner hair cell loss should be simulated for these types of hearing losses. A simple model of inner hair cell loss was implemented since the model employed in this chapter has no explicit inner hair cell component. This simple model consisted in simply reducing the output of the basilar membrane by a given number of dB. This reduction was frequency dependant and corresponding to the conjectured contribution to the hearing loss, from inner hair cells. Chapter 5 discusses details of inner hair cell loss modelling.

The model of hearing impairment should simulate the fact that both inner- and outer hair cell loss is a function of characteristic frequency. For the outer hair cells this is simple. Because the DRNL filter bank parameters are a function of centre frequency, changing the linear regression parameters for a and b can be done in such a way as to provide a simulated hearing loss which varies with frequency. A limitation of the DRNL as implemented here, is that the regression

line parameters can only describe a monotonic function of frequency for any given DRNL parameter. As we shall see this is sufficient for the purpose here.

Since the model of inner hair cell loss is simply subtracting a given number of dB from basilar membrane output, this number can simply be defined as a function of centre frequency. Partitioning the hearing loss into two parts one attributable to inner hair cell loss and the other to outer hair cell loss has previously been used [Moore and Glasberg, 1997].

While a forward masking study similar to Lopez-Poveda et al. [2002] for hearing impaired persons would have been preferred in order to derive a DRNL filter bank for hearing impaired listeners, time has not permitted such a study. Instead a moderately severe hearing loss was simulated based on subject JK's left ear data from Oxenham and Plack [1997]. The data consists of an audiogram and forward masking data for the experiment presented in Section 3.2.1. The resulting linear regression DRNL parameters are shown in Table 3.11. This filter bank will serve as the representative of impaired hearing throughout this thesis and is as such not atypical of presbycusis, c.f. Schuknecht [1964].

In comparison to Table 3.10, on which the parameters are based, only the nonlinear path parameter a has been modified to simulate the hearing loss. This modification leads to elevated thresholds as shown in the audiogram in Figure 3.11.

Table 3.11: Derived linear regression line coefficients for impaired hearing

	Regression coefficients	p_0	m
Linear-	CF_{lin} (Hz)	-6.8193 x 10^{-2}	1.0151 x 10^{0}
path-	BW_{lin} (Hz)	5.3444 x 10^{-1}	5.9507 x 10^{-1}
parameters	g	-3.2290 x 10^{-2}	8.6403 x 10^{-1}
Nonlinear-	BW_{nl} (Hz)	2.4926 x 10^{-1}	6.9124 x 10^{-1}
path-	a	5.3000 x 10^{0}	-1.0000 x 10^{0}
parameters	b	-2.2442 x 10^{0}	2.4972 x 10^{-1}

The simulated audiogram in Figure 3.11 shows the moderately severe hearing loss used as the representative of impaired hearing in this thesis. It is based on the parameters from Table 3.11 and the assumption that the absolute thresholds corresponds to a given basilar membrane vibration velocity.

The audiogram only shows the loss of sensitivity associated with sensorineural hearing loss. It is not possible to asses how the frequency selectivity is affected only based on the audiogram. In comparison Figure 3.12, which is based in the

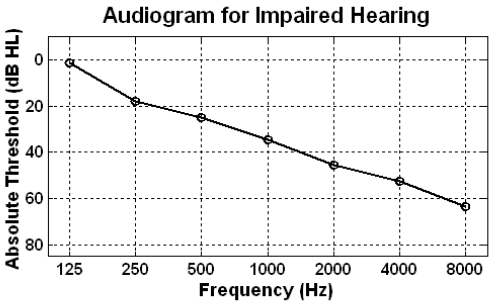


Figure 3.11: The solid line shows the audiogram for impaired hearing as modelled in this thesis. The broken lines show the loss caused by outer hair cells i.e. the contribution modelled by the DRNL. The audiogram was constructed based on the assumption that a given basilar membrane vibration velocity corresponds to the absolute threshold for a given characteristic frequency. Thus a velocity of 2/100 mm/s was assumed to correspond to the threshold. For each frequency a number of dB was subtracted representing hearing loss due to inner hair cell loss. This was done so the resulting audiogram matches that of subject JK's left ear in Oxenham and Plack [1997]. The ISO ISO/DIS 389-8 standard for reference equivalent thresholds was used as an approximation for converting from sound pressure level to hearing level

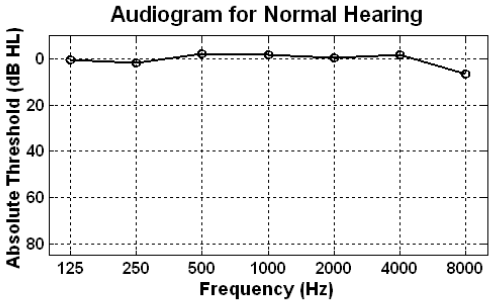


Figure 3.12: The audiogram for normal hearing as modelled in this thesis. The audiogram was constructed based on the assumption that a given basilar membrane vibration velocity corresponds to the absolute threshold for a given characteristic frequency. Thus a velocity of 2/100 mm/s was assumed to correspond to the threshold. The ISO ISO/DIS 389-8 standard for reference equivalent thresholds was used as an approximation for converting from sound pressure level to hearing level

DRNL parameters from Table 3.10, shows the audiogram for a normal hearing person.

In order to justify the DRNL parameters for impaired hearing shown in Table 3.11 the experiment from Oxenham and Plack [1997] was simulated for subject JK's left ear. The results are shown in Figure 3.13.

Because the threshold for the probe in quiet is 77 dB SPL, very few data points were measured. The desire to model a "typical" sensorineural hearing loss limits the data available markedly. This fact and the fact that no two hearing losses are the same present two of the fundamental problems in modelling hearing loss.

3.5 Summary and conclusion

This chapter presented modelling of three different forward masking experiments. Two of these experiments were used to verify that the idea of using the DRNL as the nonlinear stage in a temporal window framework is indeed a viable hypothesis for modelling forward masking as suggested by Plack et al. [2002].

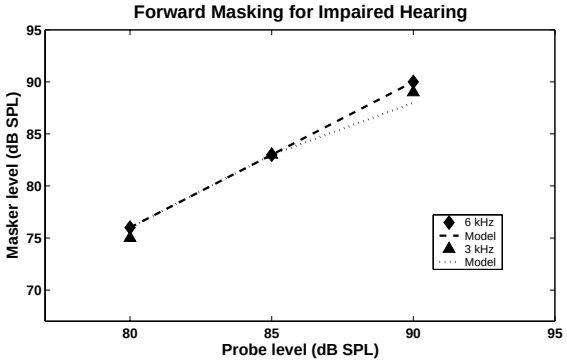


Figure 3.13: The forward masking experiment simulated for the left ear of subject JK in Oxenham and Plack [1997]. The absolute threshold for the probe in quiet was 77 dB SPL so only six data points were obtained. The original study shows no significant differences between thresholds for the 3 kHz masker and 6 kHz masker

The third experiment, described in Section 3.2.3, were used for fitting DRNL parameters, assuming a fixed temporal window shape. The results were not completely satisfactory in that conditions where the masker and probe frequencies were different were generally not well accounted for. Moreover, low stimulus frequencies were not accounted for by the DRNL combined with a fixed temporal window shape. However, conditions with equal masker and probe frequencies, which were high frequencies, were accounted for satisfactorily.

Several possible explanations for the problems encountered can be offered.

- In the off-frequency conditions, masker frequency is different from the probe frequency, the outer- and middle ear plays a role.
- The middle ear reflex is not modelled
- Peripheral adaptation is not modelled
- Detection is assumed to be based on the characteristic place on the basilar membrane for probe frequency only. With increasing level the characteristic place moves. This is not modelled.
- The assumption that temporal window shape is independent of frequency is dubious [Plack and Moore, 1990].

- At low frequencies ringing in the auditory filter influence probe detection more than at high frequencies where the filters are broader and thus show less ringing
- Less is known about peripheral nonlinearity at low frequencies, so the DRNL parameters might be unrealistic.

Further investigation is needed in order to discern the importance of each of these potential explanations. This was deemed beyond the scope of this thesis. Instead DRNL parameters from the successful simulations are used in the following.

Chapter 4

The Inner Hair Cell and Auditory Nerve Models

The final stage in the transduction process takes place in the inner hair cell or, to be more precise, in the stereocilia of the hair cell. In response to deflection the stereocilia open ion gates causing a change in transmembrane potential – transduction. At the base of the inner hair cell the depolarisation of the hair cells causes the release of neurotransmitter into the synaptic cleft. This in turn causes the auditory nerve fibres to elicit action potentials.

Since the vast majority of afferent auditory nerve fibres project from inner hair cell this chapter focuses on the role of the inner hair cells and auditory nerve fibres in forward transduction.

Section 4.1 briefly describes the inner hair cell electrophysiology. After this, the basic anatomy of the auditory nerve is presented in Section 4.2. Section 4.3 presents fundamental discharge properties of the auditory nerve. Finally, Section 4.4 present the model used in this thesis.

4.1 Inner hair cell electrophysiology

The potassium and calcium ions entering the inner hair cells through the stretch induced opening of ion channels (c.f. Figure 2.12), constitute a current affecting the charge and the transmembrane potential of the inner hair cell. The current

is referred to as the receptor current or transduction current. Similarly, the transmembrane potential is called the receptor potential or transduction potential. The potassium constitute the majority of cations entering the hair cell and is the main carrier of the transducer current. The role of the calcium ions is described below and in Section 4.4.2.

Because the apical part of the hair cell is bathed in endolymph and the basal part is bathed in perilymph, the "transmembrane potential" differs from the basolateral[1] part of the hair cell to the apex of the hair cell. The electrochemical gradient for the potassium ions differs accordingly, i.e. the ions are forced into the cell at apical channels and leave the cell at basolateral channels, as is the case in ordinary cells, c.f. Figure 4.1. So while the force driving the ions across the apical cell membrane is the *electrical* gradient, the force driving the ions across the basolateral cell membrane is a combination of the electrical- and chemical gradients, i.e. the electrochemical gradient.

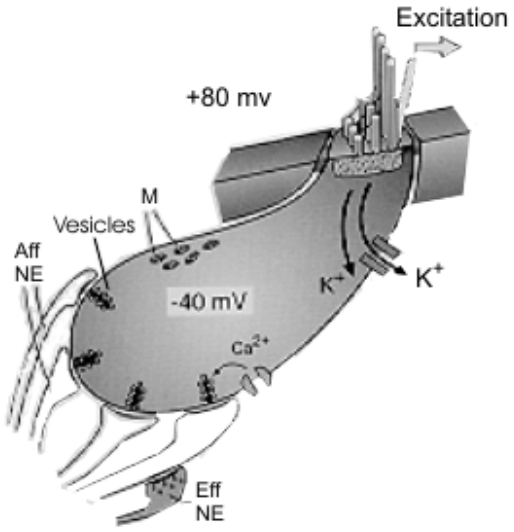


Figure 4.1: Inner hair cell electrophysiology. Redrawn from [Geisler, 1998]

[1]The inner hair cell is "lying on its side" (c.f. Figure 4.1) and therefore the term "basolateral" is used, rather than simply "the basal part"

Even in the absence of stimulus some ion channels are open allowing a standing current to flow due to the potential difference between endolymph and the hair cell. The transducer current is highly asymmetrical in response to ciliary deflection in the two directions. This is due to the asymmetric opening of the ion channels, i.e. ion channels open when stereocilia are displaced towards the taller cilia.

The receptor potential is equally asymmetric in comparison to cilia deflection. An example of guinea pig inner hair cell receptor potential behaviour is shown in Figure 4.2. The change in receptor potential affects the cell membrane capacitance in that an increase in receptor potential entails an increase in capacitance. In Section 4.4.1 a model for this mechanism is presented.

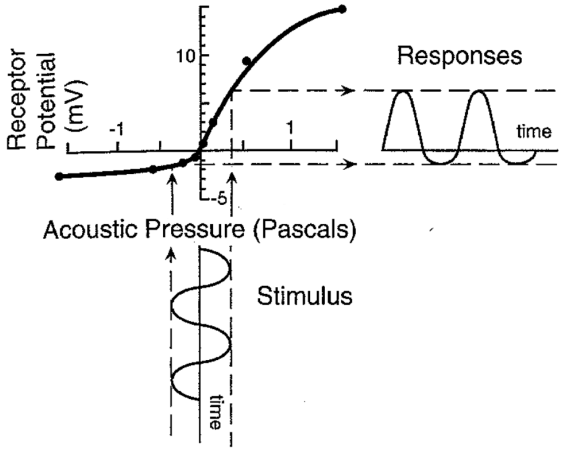


Figure 4.2: Inner hair cell receptor potential input/output function and response to a pure tone at approximately 84 dB SPL. Redrawn from [Geisler, 1998]. Original data from [Russell et al., 1986]

Three types of ionic channels exist: 1) Tension gated, voltage gated and ligand gated channels.

Although virtually all measurements of transducer gating and transducer channels originate from non-mammalian vertebrates, it is presumed that the data is generally applicable to mammalian inner hair cells [Dallos, 1996]. Likewise, it is presumed that *outer* hair cell receptor potential in non-mammalian vertebrates closely resembles that of mammalian outer hair cell receptor potential. Outer hair cell receptor potential is less asymmetric with respect to cilia displacement

than inner hair cell receptor potential. The issue of outer hair cell receptor potential is not discussed any further in this thesis.

The influx of calcium has two main consequences: 1) it opens the calcium gated potassium channels and 2) it increases the likelihood of vesicles release. The opening of potassium channels increases cation influx thus helping transduction. Vesicles contain neurotransmitter. Each vesicle contains a small number of neurotransmitter quanta, possibly only one. The vesicles are docked on the synaptic bar, a type of vesicle pod, which is attached to the cell membrane inside the cell (c.f. Figure 4.2). When the calcium mediated increase in probability of vesicle release surpasses a given threshold, the vesicle is released. It fuses with the cell membrane and diffuses out into the synaptic cleft, where it is picked up by neurotransmitter receptors in the auditory nerve fibres. It is believed that this "quantal" release of one vesicle leads to one discharge (one spike) in one auditory nerve fibre.

The chemical identity of the neurotransmitter has not been firmly established, but it is believed that it is glutamate or some closely related neurotransmitter. It appears that whatever it might be it affects glutamate receptors.

4.2 Basic auditory nerve anatomy

Auditory nerve fibres are divided into two categories: afferent fibres and efferent fibres. Afferent fibres send spikes towards the brainstem, efferent fibres send spikes *from* nuclei in the brainstem.

In the following focus will be on afferent fibres. The afferent fibres can further be divided into two types: Type I (radial) and Type II (outer spiral) fibres. The radial fibres constitute around 90 % of all the afferent fibres. They exclusively innervate inner hair cells. Each radial fibre innervate one or two inner hair cells. Since humans have approximately 3,500 inner hair cells and approximately 30,000 afferent nerve fibres, simple division tells us that each inner hair cell is innervated by eight nerve fibres. However, considering that some fibres innervate more than one fibre, it is generally assumed that approximately ten nerve fibres innervate each inner hair cell. The radial fibres have a insulating myelin sheath.

Less is known about the outer spiral fibres. They are unmyelinated and thinner than the radial fibres. They innervate more hair cells, some of which are outer hair cells. The outer spiral cells get their name from the way they project from the habenula perforata (c.f. Figure 4.3). They cross the tunnel of corti and make

contact with the first row of outer hair cells. The continue to run towards the apex thus following a "spiral" route amongst the "outer" hair cells – hence the name "outer spiral" fibres.

The myelin sheath of the radial fibres enables recording from them. However, the sheath only covers the part of the axon in the direction from the habenula perforata towards, and along, the spiral ganglion. For this reason recordings of auditory nerve fibres are taken from the spiral ganglion and exclusively from radial fibres. Detailed anatomical investigation of auditory nerve fibres can be found in [Liberman, 1980, 1982b; Liberman et al., 1990].

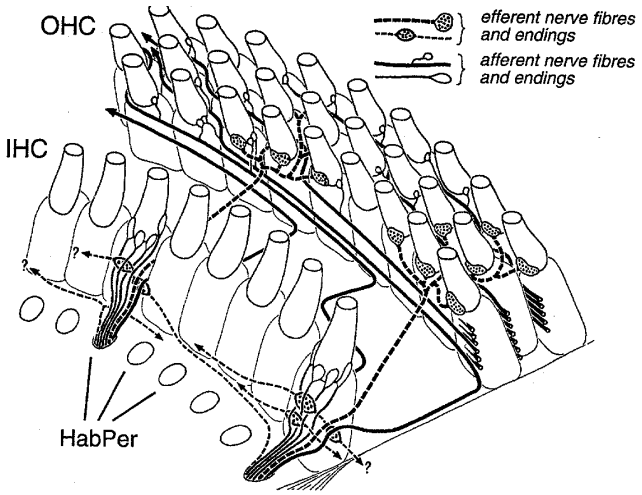


Figure 4.3: Auditory nerve fibre projection. Redrawn from [Geisler, 1998]

The place of contact between the hair cells and the nerve fibres are called synaptic terminals or synapses.

4.3 Auditory nerve discharge properties

This section discusses various discharge properties of the auditory nerve. First, an account of spontaneous activity and thresholds of auditory nerve fibres is given in Section 4.3.1. Next, the rate/intesity functions are presented in Sec-

tion 4.3.2. The concept of auditory nerve fibre adaptation is introduced in Section 4.3.3. Finally, Section 4.3.4 briefly discusses the dichotomy of rate/place representation versus temporal representation of sound stimulus in the auditory nerve.

4.3.1 Spontaneous activity and thresholds

Due to the resting potential, the probability of vesicle release is above zero even in the absence of stimulus. This cause a certain amount of spike generation. This spontaneous activity ranges from close to zero spikes per second to more than 300 spikes per second. Due to the absolute refractory period (see Figure 4.4) the theoretical maximum firing rate is 1000 spikes per second, but combined with the relative refractory period firing rates above a few hundred spikes per second is rarely seen.

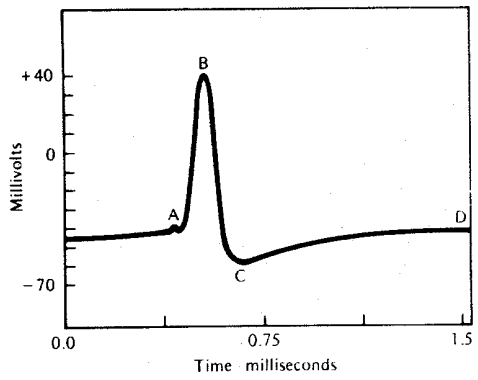


Figure 4.4: The time-course of a typical auditory nerve fibre discharge – the spike. A marks the beginning of the depolarisation, B marks the beginning of hyperpolarisation, C marks the undershoot on the way back to the resting potential, D marks the return to the resting potential. The time between A and is the total refractory period. The time between B and C is called the absolute refractory period. The time between C and D is called the relative refractory period. Redrawn from Yost [2000]

Categorising auditory nerve fibres based on their behavioural properties is an obvious way approaching the massive representation of sound in the auditory nerve. Two of the basic properties readily measurable in the auditory nerve are the rate of firing and the threshold of the individual fibre. The former being the simplest, the latter requiring a definition of threshold.

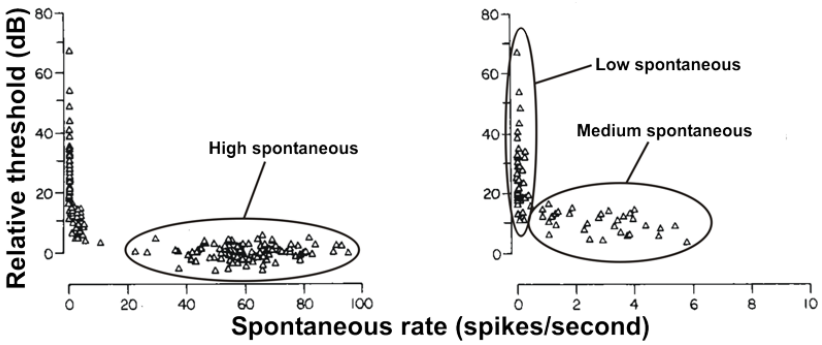


Figure 4.5: The traditional grouping of auditory nerve fibres into high-, medium- and low spontaneous rate fibres. Right panel zooms in on spike rates from 0 to 10 spikes per second. Redrawn from [Liberman, 1978]

Recordings in cat show a bimodal distribution of auditory nerves based on spontaneous rate [Kiang, 1965; Liberman, 1978, 1982a]: one group has spontaneous firing rates below 18 spikes per second the other group above 18 spikes per second (c.f. the left of Figure 4.5). The thresholds of high spontaneous rate fibres lie within 20 dB in good preparations [Evans, 1972], whereas the spread in the lower spontaneous rate group is much larger (c.f. Figure 4.5). Moreover, the high spontaneous rate fibres display a lower threshold than the lower spontaneous rate fibres. This has lead to the suggestion that auditory nerve fibres be categorised into three groups based on their spontaneous rate: high spontaneous rate fibres (more than 18 spikes per second), medium spontaneous rate fibres (between 0.5 and 18 spikes per second) and low spontaneous rate fibres (less than 0,5 spikes per second). Although this categorisation refers to spontaneous rate only, it reflects the variation in thresholds within the three groups in that high- and medium spontaneous rate fibres have little variation in threshold, whereas low spontaneous rate fibres exhibit large variations.

The distribution of spontaneous activity and threshold does not seem to vary significantly with characteristic frequency of the fibres.

The high spontaneous rate fibres constitute 60 percent, medium spontaneous rate 25 percent and low spontaneous rate fibres 15 percent of all afferent fibres [Liberman, 1978].

In addition to morphological/anatomical correlates (e.g. [Liberman, 1982b]), the three way division of nerve fibres correlates with functional aspects, some of which will be discussed in the following sections.

4.3.2 Rate/intensity functions

The rate/intensity function is yet a more detailed description of auditory nerve fibres than the threshold and spontaneous rate, both of which can be determined from the rate/intensity function. Typical examples for high-, medium- and low spontaneous rate fibres are shown in Figure 4.6.

Microelectrodes are used to acquire rate/intensity functions from animals. The stimulus normally used is a pure tone with the same frequency as the characteristic frequency of the fibre. This section exclusively deals with this type of rate/intensity functions. Responses to more complex stimuli are deferred to Chapter 7.

The rate/intensity functions are typically based on responses to presentations of 50-100 ms pure tones averaged across several presentations, and separated by appropriate periods of silence. Frequently the stimulus level is varied pseudo-randomly reducing biases caused by artifacts such as the preceding condition or physical state of the animal.

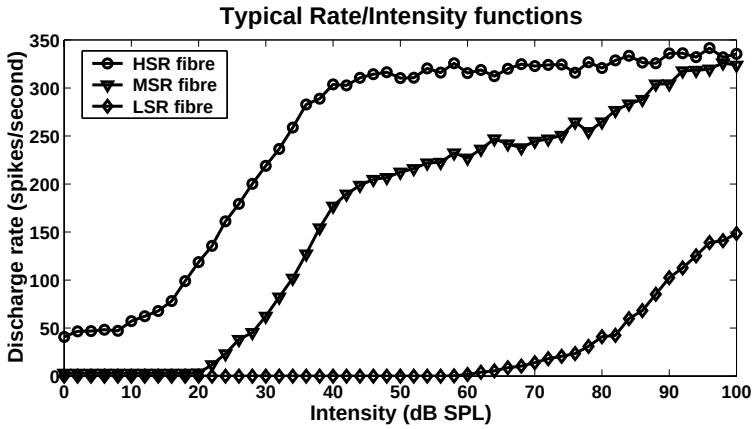


Figure 4.6: Typical rate/intensity functions for high-, medium- and low spontaneous rate fibres. This figure is drawn based on simulations implemented with fibre parameters taken from Table II, fibre IDs HSR, MSR and L_2 in [Sumner et al., 2002]. The characteristic frequency of the fibre was 8 kHz corresponding to the stimulus frequency

In addition to the threshold and spontaneous rate of the fibre, the rate/intensity function provides information about saturation level of the fibre, i.e. the stimulus

intensity at which a further increase in stimulus intensity results in little or no increase in firing rate.

Investigating saturation levels more closely, it is evident that there are three "saturation types", which can be distinguished based on the shape of their rate/intensity functions above threshold: 1) Normal saturation, i.e. maximum discharge rate within 30 dB of threshold, c.f. Figure 4.6 HSR, 2) Sloping saturation, i.e. initially rapid growth, leading to slower growth, but not saturating; c.f. Figure 4.6 MSR and 3) Straight, i.e. approximately constant increase in firing rate with sound pressure level. Although the examples shown here suggest that there is a one-to-one correspondence between saturation type and spontaneous firing rate type, this is controversial. Sachs and Abbas [1974] found high spontaneous rate fibres with sloping saturation, whereas Liberman [1988] did not.

In contrasts, it is generally agreed that low spontaneous rate fibres have higher thresholds than high spontaneous rate fibres. The dynamic range of auditory nerve fibres is correlated with the spontaneous discharge rate of the fibre. High spontaneous rate fibres have the narrowest dynamic range (20-30 dB) and the low spontaneous rate fibre have the widest dynamic range, up to around 60 dB in terms of stimulus level. It might be argued that looking at the driving force of spike generation, e.g. basilar membrane vibration velocity, is more appropriate than considering stimulus level. This gives the dynamic range of around 20 dB. Following this conjecture, that all auditory nerve fibres have this dynamic range, and combining it with the fact that high spontaneous rate fibres have low thresholds, it is tempting to say that auditory nerve fibres mainly differ in spontaneous rates and thresholds. It follows that the variations in saturation type and dynamic range (here dynamic range in terms of stimulus level) can be explained from the compressive nonlinearity of the basilar membrane [Yates et al., 1990].

The distribution of auditory nerve fibres according to threshold and dynamic range raises an interesting question: Is it possible to account for the 120 dB dynamic range of human loudness perception by means of a firing rate measure? Just counting the number of fibres is not adequate [Relkin and Doucet, 1997]. The majority of fibres saturate at a relatively low level, and for this reason intensity discrimination performance should decrease with level This is not the case. Counting the spikes in a cleverer fashion *would* work: The high spontaneous rate fibres would provide the intensity coding at lower levels, and the lower spontaneous rate fibres would gradually take over as intensity is in-

creased. It has been shown that, in theory and by statistical methods, this is indeed a possibility [Viemeister, 1988].

As this section has shown the grouping of auditory nerve fibres according to their spontaneous rate, though apparently superficial and crude, does provide additional information about auditory nerve fibre properties.

4.3.3 Adaptation

The previous section covered the changes in *average* firing rate in response to pure tones. The averaging was carried out over a relatively long period of time (50-100 ms). Looking at the detailed time-course of the discharge patterns of auditory nerve fibre the basic response pattern shown in Figure 4.7 emerges.

Low spontaneous rate fibres qualitatively show the same adaptation behaviour as the high spontaneous rate fibre shown in Figure 4.7. The main difference being that the recovery process is up to 10 times longer for low spontaneous rate fibres.

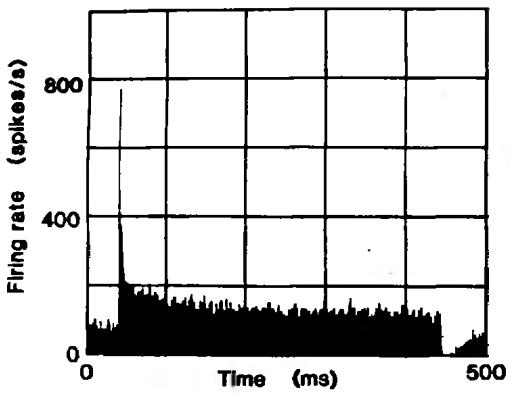


Figure 4.7: Responses from a high spontaneous rate fibre in gerbil. A 400 ms pure tone starting at 50 ms was presented. The figure shows the response in a post-stimulus time histogram. From 0-50 ms spontaneous activity is seen, A rapid on-set response is elicited immediately after stimulus onset. From 50-200 ms a gradual decrease in firing rate takes place (adaptation). The discharge rate drops dramatically at stimulus off-set (450 ms). It actually overshoots the spontaneous rate level. From 450 ms and onwards recovery to the normal spontaneous rate takes place. Redrawn from [Westerman and Smith, 1987]

Primary auditory neuron adaptation can be explained by the so-called reservoir model in which the release of neurotransmitter at any one synapse is described as a tank (reservoir) with a valve. The tank holds a certain amount of neurotransmitter. The valve controls the release of neurotransmitter and is equivalent to the transmembrane potential. A constant influx of neurotransmitter and an overflow keeps the neurotransmitter level in the tank within "operational range". The probability of generating a spike is a combination of the transmembrane potential and the amount of neurotransmitter available for release. More complex models include multiple interconnected reservoirs and more detailed descriptions of the path taken by the neurotransmitter (e.g. [Meddis, 1986]). Since fundamental knowledge of neurotransmitter release is still missing, the details of such models, while interesting, remain speculative. We will return to models of the synapse in Section 4.4.2.

The transmembrane potential controls the release of neurotransmitter. However, examining neurotransmitter release more closely it appears that chemical agents affect the spontaneous rate differently than the "driven" rate. This, in turn, suggests more than one underlying mechanism for neurotransmitter release [Guth et al., 1991]. However complex and poorly understood these underlying mechanisms are, it is believed that the principal component involves calcium ions, which mediate neurotransmitter release in all synapses in the brain.

4.3.4 Temporal properties of auditory nerve fibres

The previous sections has concentrated on firing *rates* in a variety of situations: absence of stimulus, stimulus at high and low levels. Moreover, a rate based suggestion of how intensity is coded in the human auditory system was presented. Temporal coding provides an alternative explanation to rate coding of not just intensity coding, but the encoding of a all properties of sound.

This dichotomy, rate versus temporal encoding of sound properties, forms the basis for discussing the brainstem mechanisms at play. It appears that rather than competing explanations, the hypothesised rate versus temporal encoding of sound are, in many cases, *complementing*. This section deals with temporal properties of auditory nerve fibres.

Post-stimulus time histograms show the detailed time course of the response to a given stimulus (c.f. Figure 4.7). The time resolution is usually in the sub millisecond range. The temporal relationship between the stimulus waveform and the response waveform is crucial. A measure of this relationships is the synchronisation index as proposed by Johnson [1980]. The measure is also known

as the vector strength [Goldberg and P., 1969]. This measure corresponds to the normalised magnitude of the Fourier component at stimulus frequency, i.e. the magnitude of the Fourier component divided by the total number of spikes. A convenient way of calculating this is by means of Equation 4.1.

$$s_i = \frac{\sqrt{(\sum \cos\theta_i)^2 + (\sum \sin\theta_i)^2}}{n} \tag{4.1}$$

where s_i is the synchronisation index for frequency f_i, n is the number of spikes and θ_i is the phase lag of the individual spike relative to the period $1/f_i$ (calculated as $\theta_i = 2\pi t_n/f_i$, where t_n is the spike time). The summation is done for all spikes.

The synchronisation index will be one if all spikes occur at the same phase – if the spikes are uniformly distributed across the period the synchronisation index will be zero.

4.4 Modelling the inner hair cell and auditory nerve complex

This section deals with modelling of the functional aspects of the physiology discussed earlier in this chapter. Focus will be on the models employed in this thesis, which is based on Sumner et al. [2002].

A viable alternative for an auditory nerve model was recently proposed in Heinz et al. [2001b]; Zhang et al. [2001]. This model has a more detailed account of particularly outer hair cells.

It was considered beyond the scope of this thesis to evaluate and compare the effectiveness of various models in the literature. Instead the choice of model was based on accessability and documentation.

The inner hair cell and auditory nerve model consists of three components:

Inner hair cell Model that describes the receptor potential. The input is stapes velocity – the output is receptor potential.

Synapse Model that describes the probability of firing of a single neuron. The input is receptor potential – the output is the probability of firing for a single fibre.

Spike generation Model that describes the generation of discrete spikes. The input is the probability of firing for a single fibre – the output is the discrete spike pattern.

Each of these will be discussed in detail below. The reason for modelling these particular stages is that physiological data is available for inner hair cell receptor potential as well as for spike generation. The intermediate step, called "synapse" above, is the hypothesised, mathematical description linking the two.

4.4.1 Inner hair cell receptor potential

Shamma proposed a model of the inner hair cell receptor potential based on data from guinea pig hair cells and bullfrog sacculus [Shamma, 1986]. Since hair cells of vertebrates appear to be very similar the model is assumed to be generally applicable i.e. also applicable to human hair cells.

Sumner et al. [2002] proposed an improvement to the model incorporating a more detailed model of the apical conductance based on Corey and Hudspeth [1983]. The improved model reflects the number of open channels as a three-state Boltzmann function rather than the two-state version originally proposed (see text below and Equation 4.3). The improved model reflects the data for guinea pig better, while giving the same results for the bullfrog.

Briefly, the model of the inner hair cell receptor potential is subdivided into three submodels:

1. Cilia displacement
2. Transduction conductance of the apical part of the hair cell
3. Transmembrane potential of the basolateral part of the hair cell

More precisely the ciliary deflection as a function of basilar membrane velocity is expressed in Equation 4.2.

$$\tau_c \frac{du(t)}{dt} + u(t) = \tau_c C_{cilia} v(t) \tag{4.2}$$

where τ_c is a time constant, C_{cilia} is a coupling gain factor, $u(t)$ is the cilia displacement as a function of time and $v(t)$ is basilar membrane velocity as a function of time. Equation 4.2 shows that for low frequencies the cilia move in

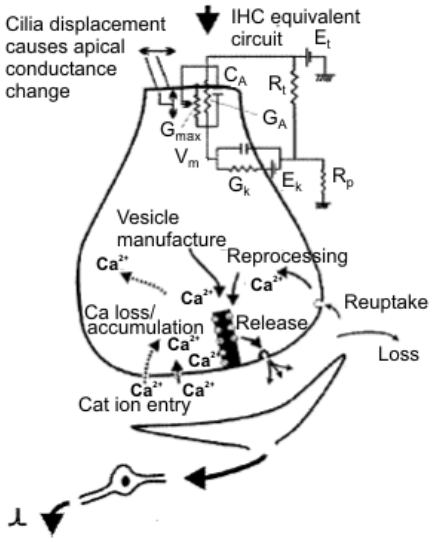


Figure 4.8: Illustration of the electrical and ion processes in a hair cell. The variables are discussed in the text. Redrawn from [Sumner et al., 2002]

phase with basilar membrane velocity and for high frequencies the cilia move in phase with displacement.

The apical conductance is defined by Equation 4.3

$$G(u) = G_a + G_{cilia}^{\max}(1 + e^{-\frac{u(t)-u_0}{s_0}})(1 + e^{-\frac{u(t)-u_1}{s_1}})^{-1} \tag{4.3}$$

where $G(u)$ is the apical conductance as a function of cilia displacement, $G_{cilia}^{\max}$ is the transduction conductance with all channels open, G_a is the passive conductance in the apical part of the cell membrane, s_0, u_0 s_1 and u_1 are constants determining the shape of the underlying three state Boltzmann function.

Finally the receptor potential is defined in Equation 4.4.

$$C_m \frac{dV(t)}{dt} + G(u)(V(t) - E_t) + G_k(V(t)) = E_k + \frac{E_t R_p}{R_t + R_p} \tag{4.4}$$

where C_m is the cell capacitance, $V(t)$ is receptor potential (or intracellular hair cell potential), $G(u)$ is the conductance as a function of cilia displacement, E_t

is the endocochlear potential, basal potential, R_t and R_p are resistances (c.f. Figure 4.8.

Roughly speaking the whole of the inner hair cell receptor potential model acts like a first-order low-pass filter. The model is not the topic of investigation in this thesis and will be not discussed in further detail. The parameters used for the inner hair cell receptor potential throughout this thesis are the parameters originally proposed in [Sumner et al., 2002]. These parameters are given in Appendix F Table F.1.

The above updated model has been tested successfully in modelling the effective role of the inner hair cells on transduction [Sumner et al., 2002, 2003; Shamma, 1986]. An alternative to this model is the inner hair cell model proposed by Rattay et al. [1998]. It also focuses on the effective role of inner hair cells.

4.4.2 Inner Hair Cell and Auditory Nerve Complex Model

In the previous sections we arrived at a mathematical description of the inner hair cell receptor potential in response to sound. This section will take this response one step further so as to provide a mathematical description of the response in the auditory nerve.

The premise of the model is that the receptor potential opens calcium channels close to the synapse. The influx of calcium mediates vesicle release hence local calcium concentration determine discharge probability. From the probability of release an actual distribution of auditory discharges in time, a spike train, can be computed. Computing the spike train from the receptor potential requires the computation of six intermediate steps:

1. Calcium current
2. Fraction of open calcium channels in the vicinity of the synapse
3. Calcium concentration
4. Neurotransmitter levels in reservoirs
5. Probability of vesicle release
6. Spike generation

Calcium current

The basolateral calcium current is defined in Equation 4.5.

$$I_{Ca}(t) = G_{\mathrm{Ca}}^{\max} m_{I_{Ca}}^{3}(t)(V(t) - E_{Ca}) \tag{4.5}$$

where I_{Ca} is the calcium current, $m_{I_{Ca}}$ is the fraction of open calcium channels (see Equations 4.6 and 4.7), $G_{\mathrm{Ca}}^{\max}$ is the maximum calcium conductance in the vicinity of the synapse, $V(t)$ is the receptor potential and E_{Ca} is the reversal potential for calcium.

Fraction of open calcium channels in the vicinity of the synapse

The fraction of steady-state, open calcium channels and in the vicinity of the synapse $m_{I_{Ca},\infty}$ is given by:

$$m_{I_{Ca},\infty} = \frac{1}{1 + e^{V(t)\gamma_{Ca}}/\beta_{Ca}} \tag{4.6}$$

where $m_{I_{Ca},\infty}$ is the fraction of open calcium channels in steady-state, $V(t)$ is the receptor potential, γ_{Ca} and β_{Ca} are calcium current constants.

The general case of the fraction of open calcium channels ($m_{I_{Ca}}$) and the steady-state case $m_{I_{Ca},\infty}$ are linked through Equation 4.7.

$$m_{I_{Ca},\infty} = m_{I_{Ca}}(t) + \tau_{I_{Ca}} \frac{dm_{I_{Ca}}(t)}{dt} \tag{4.7}$$

where $\tau_{I_{Ca}}$ is a time constant.

Calcium concentration

The calcium concentration $[Ca^{2+}]$ is calculated as a function of calcium current $I_{Ca}(t)$

$$\tau_{[Ca]} \frac{d[Ca^{2+}](t)}{dt} + [Ca^{2+}] = I_{Ca}(t) \tag{4.8}$$

where $\tau_{[Ca]}$ is a time constant.

The probability of the release of neurotransmitter is proportional to the cube of the calcium concentration, which in turn is proportional to the neurotransmitter release rate. The release rate is computed from Equation 4.9

$$k(t) = \max(([Ca^{2+}]^3(t) - [Ca^{2+}]^3_{thr}(t))z, 0) \tag{4.9}$$

where $k(t)$ is the neurotransmitter release rate, $[Ca^{2+}]$ is the calcium concentration, $[Ca^{2+}]_{thr}$ is a calcium concentration threshold constant and z is a scalar for converting from calcium concentration to release rate.

Neurotransmitter levels in reservoirs

The release rate from Equation 4.9 combined with the number of vesicles ready for release determines the actual amount of neurotransmitter released into the synaptic cleft.

In order to compute the number of vesicles ready to be released, a model involving three reservoirs is employed. The three reservoirs are the *immediate store*, *synaptic cleft* and *reprocessing store*. These are identical to the stores used in [Meddis, 1986]. The immediate store holds the neurotransmitter ready for release, the synaptic cleft holds the neurotransmitter already released and the reprocessing store prepares neurotransmitter for the immediate store. These reservoirs are interconnected and Equations 4.10, 4.11 and 4.12 describe how. In addition vesicles containing neurotransmitter are manufactured continuously.

$$\frac{dq(t)}{dt} = N(w(t), x) + N(M - q(t), y) - N(q(t), k(t)) \tag{4.10}$$

$$\frac{dc(t)}{dt} = N(q(t), K(t)) - c(t)l - c(t)r \tag{4.11}$$

$$\frac{dw(t)}{dt} = c(t)r - N(w(t), x) \tag{4.12}$$

where $k(t)$ is the transmitter release rate, $q(t)$ is the number of vesicles in the immediate store, $w(t)$ is the number of vesicles in the reprocessing store, $c(t)$ is the number of vesicles in the cleft, M is the maximum number of vesicles in the immediate store, y is a replenishment rate scaling factor, $((M - q(t))y$ is the number of freshly manufactured vesicles), l designates the fraction of vesicles

lost from the synaptic cleft, r designates the fraction of vesicles returned to the reprocessing store from the synaptic cleft, x is the rate at which the reprocessing store produces vesicles for the immediate store and $N(n, p)$ is a quantal, stochastic probability function yielding the number of released vesicles as a function of available vesicles n and rate of release p. In practice the probability of release $p\ dt$ is computed for each simulation epoch.

This three reservoir models explains the auditory nerve adaptation characteristics shown in Figure 4.7 by means of presynaptic transmitter depletion.

Spike generation

The release of a single neurotransmitter quantum into the synaptic cleft is believed to be sufficient to produce an action potential, provided the fibre is not in a refractory state. From Equations 4.2 through 4.12 transmitter release events is calculated. Equation 4.13 takes into account absolute and relative refractoriness (c.f. Figure 4.4).

$$p(t) = \begin{cases} 0, & \text{for } t - t_l \ < R_A \\ 1 - c_r e^{-(t-t_l-R_A)/s_r}, & \text{for } t - t_l \ \geq R_A \end{cases} \tag{4.13}$$

where c_r is 0.55 and determines the maximum contribution of the refractory period, s_r is 0.8 ms and designates the refraction constant, t is time, t_l is the time of the previous spike and R_A is 0.75 ms and designates the absolute refractory period. For each released vesicle an action potential can be elicited in the fibre if $p(t)$ exceeds a random number between 0 and 1.

Chapter 5

Modifying the Inner Hair Cell and Auditory Nerve Complex Model

This chapter describes work carried out in connection with improving the parameters of the models described in the previous chapter. More specifically, it concerns modelling chinchilla data for the auditory nerve. Part of this work was presented at the Baltic-Nordic Acoustical Meeting, August, 2002, Lyngby, Denmark. The results from that can be found in Appendix B, and will not be presented here.

The intention was to derive some of the model parameters presented in Chapter 4 and Appendix F. A ubiquitous and convenient assumption is that auditory nerves in mammals are similar. Hence the derived parameters could be used in simulating human auditory nerve. This provides the motivation for examination of chinchilla data in the following.

The first Section 5.1 describes the study from which the chinchilla data was taken and the modelling efforts hereof. Section 5.2 describes the final model of the inner hair cell and auditory nerve complex for normal- and impaired hearing employed in this thesis.

5.1 Auditory Nerve Responses to Trains of Clicks

In Wickesberg and Stevens [1998] auditory nerve fibre spikes were recorded from anesthetised chinchillas. The purpose of the study was to examine if the decrease in sensitivity with increasing rate of information presentation in localisation tasks could be explained in terms of the encoding capacity of the auditory nerve. For our purposes, however, the focus is simply modelling the data.

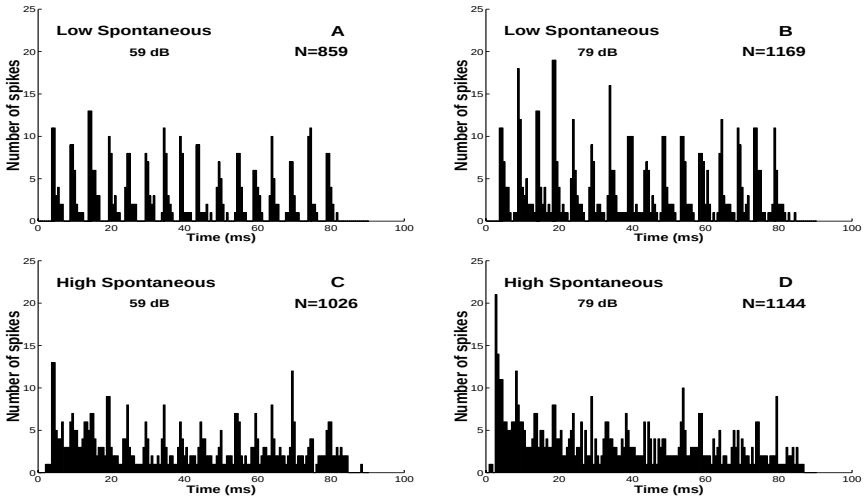


Figure 5.1: Peristimulus histograms of responses to trains of 16 clicks with level 59 and 79 dB peak SPL for two auditory nerve fibres with low (A,B) and high (C,D) spontaneous rate (1.2 spikes/s and 36 spikes/s respectively). The interclick interval was 5 ms and characteristic frequency 2500 Hz. Figures show data collected for 100 presentations. N designates the total number of spikes recorded. Re-plotted from Wickesberg and Stevens [1998]

The stimuli consisted of trains of clicks containing 2, 4, 8, 12 or 16 condensation clicks with a width of 100 μs. The level and the time between the clicks, interclick interval or ICI, was varied. Three distinct levels were tested: 59, 69 and 79 dB peak SPL, where "peak SPL" means that the amplitude of the click was equal to the peak amplitude of a 1000 Hz pure tone at the given SPL. The interclick interval was varied from 1 to 5 ms in steps of 1ms. The sound was delivered to the chinchilla ears by a calibrated insert telephone with a distance of "several millimetres" to the tympanic membrane. Spike times were recorded from the auditory nerve with glass microelectrodes while monitoring

fibre condition. Fibres were classified according to characteristic frequency and spontaneous rate.

Figure 5.1 shows an example of the spike train response to click trains. The high spontaneous rate fibres elicit more spikes than the low spontaneous rate fibre. An increase from a stimulus level of 59 dB peak SPL to 79 dB peak SPL provides a bigger increase in the low spontaneous rate fibres than the high spontaneous rate fibres. The troughs are more visible for the low spontaneous rate fibres. All of these observations are as expected. Wickesberg and Stevens [1998] divides auditory nerve fibres into two categories: high spontaneous rate fibres (> 20 spikes/s) and low spontaneous rate fibres (< 20 spikes/s). This is at odds with the classical division suggested by Liberman [1978], but similar to the classifications originally proposed by Kiang [1965]. Wickesberg's classification will be adopted in the remainder of this chapter.

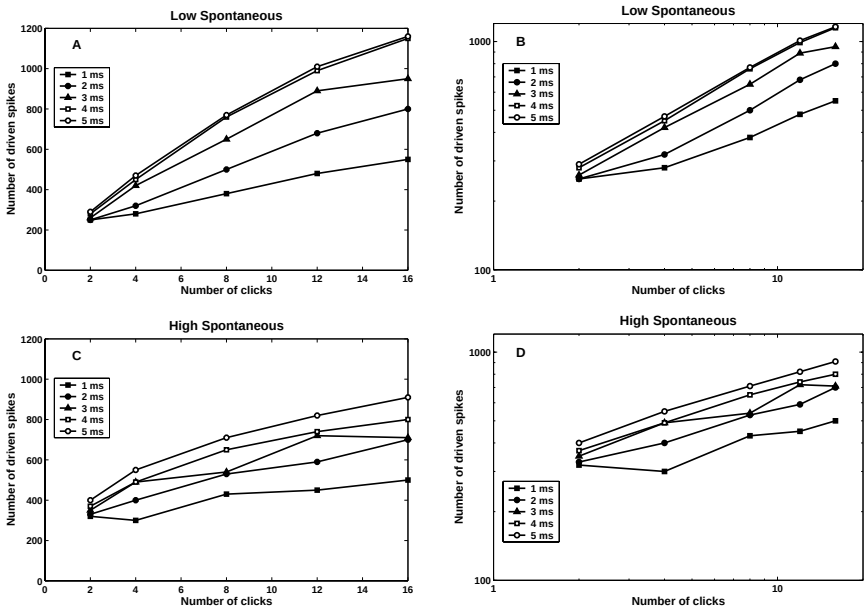


Figure 5.2: The number of driven spikes in the response to 79 dB peak SPL click trains from the fibres introduced in Figure 5.1. The data is shown on linear (A,C) and log-log (B,D) scales respectively. The driven rate is calculated as the total number of spikes minus the number expected from the fibre's spontaneous rate. Re-plotted from Wickesberg and Stevens [1998]

Figure 5.2 shows the number of driven spikes as a function of the number of clicks in the click train. The driven rate is calculated as the total number of spikes minus the number expected from the fibre's spontaneous rate. This shows that the low spontaneous rate fibres show a larger increase in driven spikes with number of clicks than the high spontaneous rate fibres. Further, the shorter the interclick interval the lower the number of driven spikes. The straight lines in plots B and D indicate that the number of driven spikes is a power function of the number of clicks in the click train.

5.1.1 Modelling the Data

Modelling the data presented in the previous section requires a model of the chinchilla middle ear and basilar membrane. Since no published parameters for the DRNL is available for the chinchilla, it was decided to use the guinea pig filter bank employed in Sumner et al. [2003] as an approximation. The middle ear was also taken from Sumner et al. [2003].

Reproducing chinchilla data with these approximations introduced some difficulties. The main one being that the basilar membrane compression knee-point was off by 20 dB. This was resolved by amplifying the stimulus by 20 dB. This is a very crude approximation and most definitely it would have been better to derive the parameters for a chinchilla basilar membrane. However, this was not possible within the given time frame.

The inner hair cell receptor potential model described in Section 4.4.1 and its parameters from Appendix F were adopted without modifications. Instead efforts was focused on fine-tuning synapse parameters as they are responsible for individual nerve fibre characteristics such as spontaneous rate and rate/intensity functions.

Three synapse parameters $G_{\mathrm{Ca}}^{\max}$, $[\mathrm{Ca}^{2+}]_{thr}$, and M (c.f. Section 4.4.2 and Appendix F) were selected for manipulation as recommended in the literature [Sumner et al., 2002]. The parameters represent maximum calcium conductance, calcium concentration threshold and number of free transmitter quanta.

Calcium plays a significant role in controlling the release of neurotransmitter into the synaptic cleft and in turn to the generation of spikes. If the maximum calcium conductance is increased, the number of spikes elicited by the fibre goes up (c.f. Figure 5.3A). The calcium concentration required to elicit spikes is coined as calcium concentration threshold. Increasing this threshold makes the number of elicited spikes go down (c.f. Figure 5.3B). The neurotransmitter

Table 5.1: Best match inner hair cell and auditory nerve complex parameters

			HSR	LSR
$G_{\mathrm{Ca}}^{\max}$	Max Calcium Conductance	10^{-9} Siemens	8.00	4.50
$[\mathrm{Ca}^{2+}]_{thr}$	Ca concentration threshold	$\times 10^{-11}$	2.48	0.00
M	Max free transmitter quanta	Scalar	4	5

is delivered to the nerve fibres in "packets" or vesicles. The number of vesicles available for transportation of neurotransmitter is here called "number of free transmitter quanta". Increasing the number of free transmitter quanta increases the number of elicited spikes (c.f. Figure 5.3C). A detailed discussion on how to manipulate synapse parameters in order to match data for individual fibres is presented in Sumner et al. [2002].

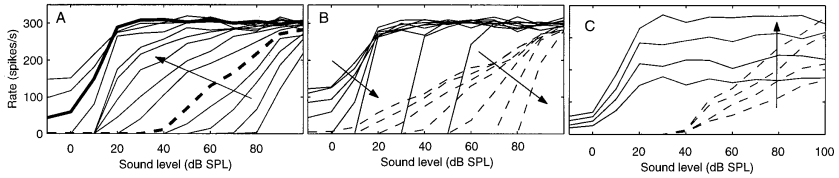


Figure 5.3: Effect on rate/intensity functions of increasing synapse parameters, shown for high spontaneous rate fibres. A: The effect of increasing $G_{\mathrm{Ca}}^{\max}$ is given by the direction of the arrows. B: The effect of increasing $[\mathrm{Ca}^{2+}]_{thr}$ is given by the direction of the arrow. C: The effect of increasing M is given by the direction of the arrow. Redrawn from Sumner et al. [2002]

The starting point for fitting the three synapse model parameters to the chinchilla data was the parameter values published in Sumner et al. [2002] and shown in Appendix F. No specific procedure was used since parameter space for the three parameters is fairly small. The following section presents the results of the modelling efforts.

5.1.2 Results of simulations

The values of the three varied inner hair cell and auditory nerve complex parameters for simulating the high- and low spontaneous rate fibres in Figure 5.1 are shown in Table 5.1. They provided the best match with the fibres in Figure 5.1. Also compare with parameters in Appendix F.

The parameters are within the ranges of the published parameters (c.f. Sumner et al. [2002] and Appendix F). It is unusual that parameter M has a higher value for the low spontaneous fibre than for the high spontaneous rate fibre. However, that is compensated by the higher $G_{\mathrm{Ca}}^{\max}$ and lower $[\mathrm{Ca}^{2+}]_{thr}$. It is also worth noticing that $[\mathrm{Ca}^{2+}]_{thr}$ is at the minimum value of 0.

The simulated spontaneous rate of the high spontaneous rate fibre was 36 spikes/s which matches the actual high spontaneous rate fibre exactly. The simulated spontaneous rate for the low spontaneous rate fibres was 15 spikes/s compared to the 1.2 spikes/s reported for the actual low spontaneous rate fibre. So while the modelled fibre is still a low spontaneous rate fibre the exact spontaneous rate was not modelled. The calculation of spontaneous rate is based on an average of 100 runs since the model has stochastic elements. The spontaneous rate was within ±10 percent of the average.

While it might be possible to obtain a better match of the spontaneous rate it was decided not to pursue the matter further for time reasons.

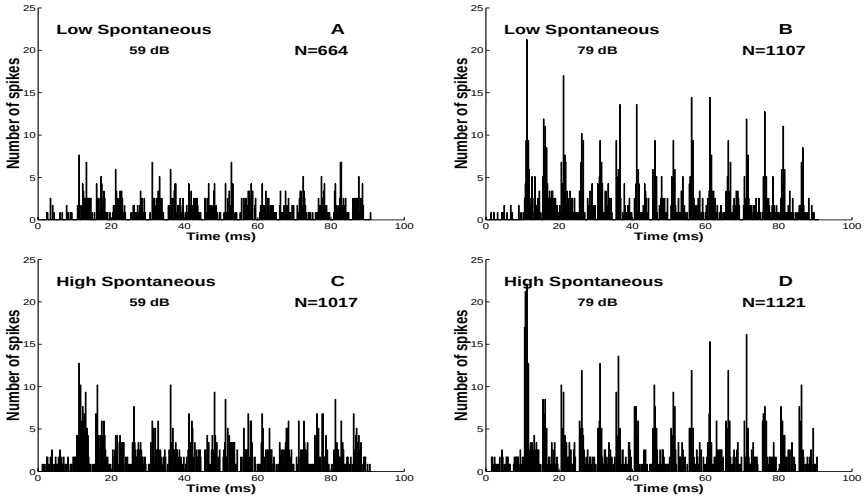


Figure 5.4: Simulated peristimulus histograms of responses to trains of 16 clicks with level 59 and 79 dB peak SPL for two auditory nerve fibres with low (A,B) and high (C,D) spontaneous rate. The interclick interval was 5 ms. The characteristic frequency 2500 Hz for both fibres and their spontaneous rates were 15 spikes/s and 34 spikes/s respectively. Figures show data simulated for 100 presentations. N designates the total number of spikes recorded. Compare with Figure 5.1

Figure 5.4 shows the simulated results for the fibres presented in Figure 5.1. The spike count N show similar trends as the measured fibres for all combinations of fibres. The increase in spike count with level is bigger for the low spontaneous rate fibre. The high spontaneous rate fibre show a very clear response to the first click in Figure 5.4D. This is consistent with Figure 5.1D. Comparing Figure 5.4A and C to Figure 5.1A and C the match does not seem as good. This might be explained by the possibly inaccurate basilar membrane input/output function at 59 dB peak SPL. However, care should be taken not reading too much into the details of the figures.

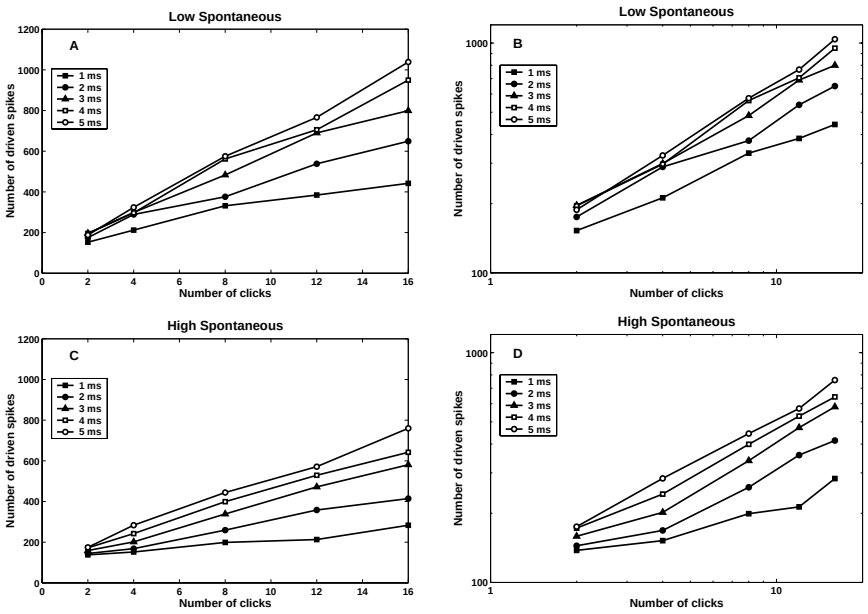


Figure 5.5: Simulation of the number of driven spikes in the response to 79 dB peak SPL click trains from the fibres simulated in Figure 5.4. The data is shown on linear and log-log scales respectively. The driven rate is calculated as the total number of spikes minus the number expected from the fibre's spontaneous rate

Figure 5.5 is the simulated version of Figure 5.2. The most important aspects of the measured data are simulated correctly in that

- The number of spikes increases with number of clicks
- The number of driven spikes is greater for the low spontaneous rate fibres than for the high spontaneous rate fibres
- The number of driven spikes increases with interclick interval.

According to Wickesberg and Stevens [1998] the number of driven spikes should be a power function of the number of clicks. If the log-log plots in Figure 5.5B and D are interpreted as being straight lines this is confirmed by the simulation. The slope should be 0.5.

The model qualitatively matches the measured data as illustrated in the figures shown in this section. Quantitatively the match is not so convincing. As mentioned earlier the fine-tuning of the parameters were stopped for time reasons.

Wickesberg and Stevens [1998] also investigates synchronisation aspects of the auditory nerve fibre data. While some work has been done in this area no conclusions have been reached at this stage. Appendix B reports part of this work.

5.2 Final model for the inner hair cell and auditory nerve complex

The aim of this chapter was to develop a model of the inner hair cell and auditory nerve complex. However, the results only show partial agreement with the data, In comparison to the previously published parameters [Sumner et al., 2002, 2003] the new parameters seem unlikely to provide improvements. Therefor it was decided to proceed with the inner hair cell and auditory nerve parameters from Sumner et al. [2003]. These parameters are shown in appendix F.

The inner hair cell and auditory nerve complex model for normal hearing is presented in Section 5.2.1. Section 5.2.2 describes how the model of inner hair cell loss was developed and presents the end result of this deliberation.

5.2.1 Normal hearing

In order to arrive at a representation of sound stimulus in the auditory nerve a number of assumption must be made. While these assumptions may be “rea-

sonable", many of them do not have direct supporting evidence. This section describes these assumptions.

The inner hair cells properties are assumed to be independent of place in the cochlea, i.e. the same inner hair cell receptor potential model is used across characteristic frequency. This assumption is extended to the synapse model of probability of vesicle release and to the model of spike generation both described in Section 4.4.2

The decision was made to model three different auditory nerve fibre types, namely high, medium and low spontaneous rate fibres. In order to model spread in fibre thresholds it was decided to model three different thresholds within each fibre type. Thus, all together nine auditory nerve fibres were modelled. The exact parameters of these nine fibres are derived in the following section.

According to Liberman [1978] 60 percent of the fibres are high spontaneous, 25 percent are medium spontaneous and 15 percent are low spontaneous rate fibres. It is assumed that this distribution does not change as a function of frequency [Liberman, 1982a]. Also, it was assumed that the three thresholds used within each fibre type were equally represented. Finally, it was assumed that the inner hair cells are equidistantly placed along the basilar membrane.

5.2.2 Impaired hearing

In Section 3.4 a moderately severe hearing loss was introduced. It was argued that part of such a loss should be explained by loss of inner hair cell function. The question now is how such a loss is best modelled. Results should, among other characteristics, show a reduced spike count.

Modelling inner hair cell pathology

From a "strict effective modelling point of view" three ways of modelling a reduced spike count are possible:

1. The entering of ions at the apex of the malfunctioning inner hair cell is reduced resulting in a reduction in receptor potential. This could be simulated by reducing the coupling gain of the hair cell C_{cilia}.
2. The synapses are affected in such a way that fewer spikes are generated in response to stimulus. This can be modelled by manipulating $G_{\mathrm{Ca}}^{\max}$, $[\mathrm{Ca}^{2+}]_{thr}$ or M

3. Part of the inner hair cell population is dead and thus some hair cells do not provide any stimulus for the attached synapses. This can be modelled by simply reducing the number of hair cells.

The above methods all provide a reduction in spike count in the auditory nerve. While not mutually exclusive, evidence supports the later of the three options [Moore et al., 2000]. It was decided to model hearing loss by simulating hair cell loss only and thus assuming neither synapse pathology nor reduction in coupling gain across the hair cell apex.

Conjectured account of absolute threshold

The audiogram in Figure 3.11 on page 76 shows the effective reduction in response caused by dead inner hair cells. In order to model this the number of dead hair cell must be estimated. Models for estimating the *partitioning* of hearing loss into loss produced by outer hair cells and loss produced by inner hair cell has previously been proposed [Moore et al., 1997, 1999]. Moreover, the concept of dead regions, in which entire frequency sections of the inner ear lacks hair cell function, has been investigated [Moore et al., 2000; Huss et al., 2000].

Here, an attempt is made to estimate the percentage of dead inner hair cells in regions of the cochlea, based on the assumption that absolute threshold can be accounted for by spike count exceeding a given threshold for nerve fibres within an 0.445 mm distance of the characteristic frequency place on the basilar membrane. This is further assumed to hold for both normal hearing and impaired hearing listeners.

While this assumption might seem unfounded, it can be justified to a certain degree by the following arguments.

1. A simple spike count cannot account for loudness in general, however, this is particularly evident at higher levels [Relkin and Doucet, 1997]. At lower levels a simple spike count provides a better account of performance limits for intensity discrimination than other proposed intensity coding schemes [Heinz et al., 2001a].

2. Calculating loudness is most commonly done by means of excitation patterns (e.g. Zwicker and Fastl [1990]; Moore et al. [1997]). These excitation patterns are calculated for each critical band or ERB, which in turn are added to yield loudness. In the model context here it is thus appropriate

to calculate the excitation pattern as the spike count in one ERB, which corresponds to approximately 0.89 mm on the basilar membrane [Moore et al., 1997]).

3. The spike count account of absolute threshold presented here is consistent with the fact that loudness can be estimated at, and below, absolute threshold [Buus et al., 1998].

4. The hearing loss modelled here is a sensorineural hearing loss and assumed to originate from dead hair cells. Explicitly, it is assumed not to have a retro-cochlear component. Hence the "central process of counting spikes" is the same in normal hearing and impaired hearing.

Calculating the number of dead inner hair cells

In order to account for the loss of sensitivity corresponding to the audiogram in Figure 3.11 the inner hair cell, the percentages of lost inner hair cells were estimated for selected frequencies. They are shown in Table 5.2.

The method of calculation was as follows. High spontaneous rate auditory nerve fibres generally have lower thresholds than their medium- and low spontaneous rate counterparts [Liberman, 1978, 1982b,a]. So it is assumed that a population of high spontaneous rate fibres account for absolute threshold. Moreover, the population should have thresholds enabling them to account for absolute threshold, i.e. they should be sensitive enough. In order to achieve this sensitivity and in order to be able to use the parameters derived for guinea pig in Sumner et al. [2003] the inner hair cell coupling gain C_{cilia} was adjusted.

The adjustment was done in such a way that threshold of a single fibre with best frequency equal to a given audiogram frequency was at 0 dB HL. Threshold for a single fibre is defined as the level where the spike count per second is 20 spikes above spontaneous rate [Evans, 1972; Sumner et al., 2003]. The rate/intensity function of such a fibre is shown in Figure 5.6. This renders the auditive nerve fibres sensitive enough to provide the nerve firings enabling threshold detection based on spike count.

Next step is to calculate a specific loudness correlate, i.e. produce a spike count for the relevant fibres within 0.445 mm of the characteristic place for the audiogram frequencies. Since we have approximately 3,500 inner hair cells and the length of the basilar membrane is 35.6 mm [Greenwood, 1990; Yost, 2000] this yields 3,500 x 0.89/35.6 = 88 inner hair cells per ERB, again assuming that the hair cells are equidistantly positioned. From each of these inner hair

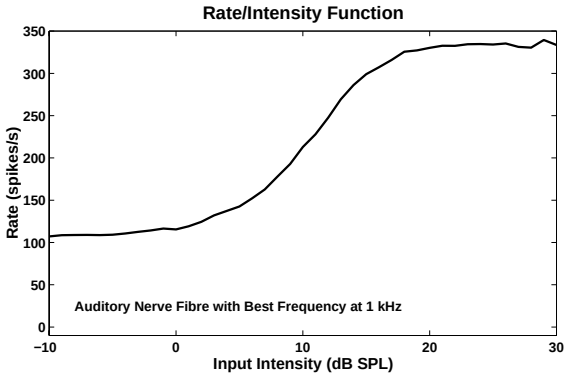


Figure 5.6: Example of the rate/intensity function used as basis for simulating the absolute threshold

cells 30,000 x 0.89/35.6/88 = 9 nerve fibres project. Of these 60 percent are high spontaneous rate fibres. Since not all fibres are sensitive enough to detect absolute threshold it is assumed that only one fibre is actually capable of this. The exact number of fibres is not crucial, but the calculation here shows that the number is small, i.e. 1 to 5 for each hair cell.

Determining the physical threshold of a single fibre is different from determining the absolute behavioural threshold in that the criterion is different. The physical threshold of a single nerve fibre was defined as the intensity at which the firing rate increases by 20 spikes per second. No criterion for directly relating spike count to absolute behavioural threshold is known to the author, however, as mentioned earlier, the literature is rich in loudness models which indirectly address absolute threshold.

One way of modelling absolute threshold in the framework of the model discussed here, is to adopt the criterion employed in Section 3.1.2 for the simulation of the forward masking study done by Lopez-Poveda et al. [2002]. The criterion for this two-interval-forced-choice-procedure was proposed in Plack et al. [2002]. Threshold is defined as the lowest level l for which Equation 5.1 holds.

$$\frac{f(p_l) + N}{f(0) + N} > k \tag{5.1}$$

where p_l is a pure tone with level l and 0 symbolises silence, k is a criterion value, N is a noise floor, and f is the function relating stimulus to spike count for all fibres within the ERB of interest. In this sense f is a specific loudness correlate,

which is calculated using the models of the peripheral models described in the thesis.

Similarly, in Section 3.1.2, the internal variable was calculated as the output of the temporal window plus "an internal noise floor". Here, the internal variable is calculated as the spike count plus "an internal noise floor". In the forward masking simulation the noise floor could be determined based on the absolute threshold of the probe. There is no obvious way of estimating the noise floor for our purpose here, other than simply matching the data.

Because of the stochastic nature of the spike generation model a 3-down-1-up, two-intervals-forced-choice procedure was used to determine the absolute thresholds. If the criterion described above was not met, an interval was chosen at random. The stop criterion was five reversals or three reversals at the same level. Step size started at 8 dB and finished at 2 dB. The initial level was set to 20 dB HL for the simulation of normal hearing and 20 dB SL for the simulated hearing loss. According to Levitt [1971] this procedure provides the 79.4 percent point on the psychometric function, which in this case is a simulated psychometric function.

Thresholds corresponding to normal hearing was achieved with a criterion value k =1 dB and a noise floor corresponding to 5000 spikes per second for the whole ERB. The spike rate was averaged over a 100 ms period thus ignoring any onset effects. 100 ms was used as a compromise between a possible onset detection mechanism and the temporal integration time period. It was found that very short periods could affect simulated threshold, but for periods between 10 - 150 ms thresholds were fairly constant.

Finally, the percentage of dead inner hair cells was estimated for each of the audiogram frequencies by gradually reducing the number of inner hair cells until thresholds shown in Figure 3.11 were reached. The basilar membrane parameters for impaired hearing from Table 3.11 were used. The resulting percentages of dead inner hair cells as a function frequency are show in Table 5.2.

It is important to note that the technique for estimating absolute threshold outlined above, is *not* an attempt of accounting for the actual mechanisms underlying absolute threshold detection, but simply a tractable method of relating fibre threshold to behavioural threshold based on observations from psychophysics and physiology. In this limited study, it was found that absolute threshold *could* be accounted for by this technique.

Table 5.2: Percentage of lost inner hair cells for the simulated hearing loss

125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
0 %	10 %	10 %	10 %	30 %	60 %	75 %

Chapter 6

The Complete Model

The types of synapses were briefly mentioned in Section 5.2.1, but the details of the distribution employed here was not discussed. This is a very speculative endeavour, but mandatory in order to conjecture a complete representation of sound signals in the auditory nerve. Moreover, some model implementation issues remain. A description of these issues is given in Section 6.1. Following this is a summary of the models for normal- and impaired hearing in Section 6.2.

6.1 Finalising the model

The ambitious goal of simulating the great variety in auditory nerve fibres is not going to be reached in this thesis. Instead the basis for *types* of nerve fibres has been established. This section tries to bridge the gap between what is known about how the auditory nerve fibres are distributed along the cochlear partition and what is required by a "complete" model.

For example, we know that approximately 60 percent of fibres are high spontaneous rate fibres. So we could assume that they are distributed evenly along the length of the cochlear partition, We do not know, however, the exact distribution of thresholds and dynamic ranges of these nerve fibres. So in this case some assumptions must be made. Similar arguments hold for medium- and low spontaneous fibres where the variety is apparently even greater. Moreover, little is known about the inter-subject variability of the these properties. What can

be said, is that the encoding enables fairly uniform psychophysical performance across normal hearing subjects in a great number of psychoacoustic experiments.

In the following the frequency-position function shown in Equation 6.1, proposed for humans by Greenwood [1990], is used.

$$f(x) = 165.4(10^{0.06x} - 1) \tag{6.1}$$

where $f(x)$ is the characteristic frequency in Hz for the place on the basilar membrane x millimetres from the apex of the cochlea. This frequency-position function applies for low level tones. In the following absolute basilar membrane positions are specified in terms of distance to the apex, adhering to the convention from Equation 6.1.

As a matter of proposing a tractable distribution it was decided to implement the complete model as follows. Only the section from 4 mm to 30 mm of the basilar membrane was modelled. This corresponds to a characteristic frequency range from 120 Hz to 10 kHz and corresponds to 28 ERB. Within each ERB 36 channels were modelled yielding a total of 1008 channels. This number should be compared to the number of inner hair cells for the modelled section of the basilar membrane, 3500 x 26/35.6 = 2556, yielding approximately $2\frac{1}{2}$ hair cells per channel.

In order to simulate the number of fibres correctly each channel was assigned 30,000 x 26/35.6/1008 = 22 nerve fibres. The well-known distribution of 60 percent high spontaneous rate fibres, 25 percent medium spontaneous rate fibres and 15 percent low spontaneous rate fibres [Liberman, 1978, 1982b] was employed. The prototype fibres, for which the parameters are shown in Table F.4, were used. In order to simulate variation three different thresholds for each of the fibre types were implemented. The three thresholds within each fibre type were equally represented.

Different thresholds were achieved by adjusting the coupling gain C_{cilia} from the inner hair cell receptor potential model. This method is an easy way of implementing the different thresholds, but no claim as to how physiologically plausible this is, will be attempted here.

This gives a total of nine types of nerve fibres for which the rate/intensity functions are shown in Figure 6.1.

The loss of inner hair cells was calculated for sections around the audiogram frequencies. It is assumed that the loss extends to a region of the basilar membrane. Although the absolute thresholds were determined by examining sections

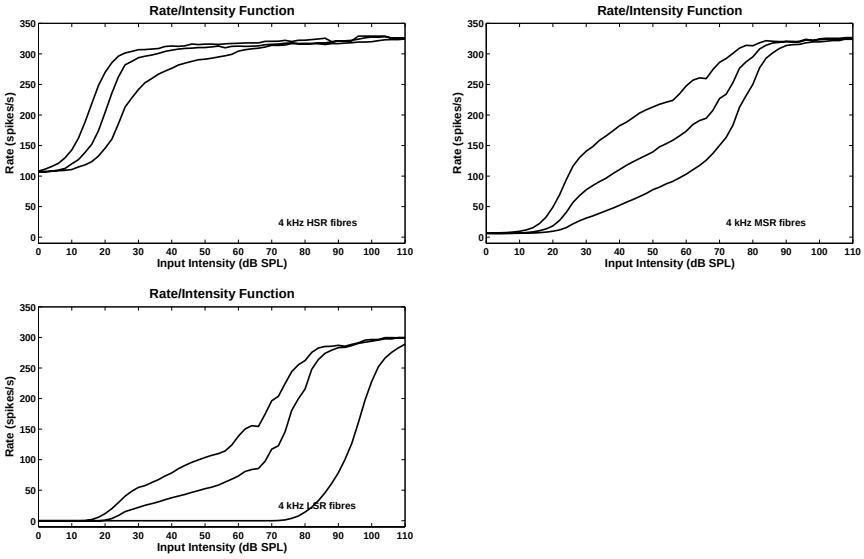


Figure 6.1: The nine types of auditory nerve fibres employed in the final model

of 0.89 mm on the basilar membrane, there is no reason to believe that hair cell loss is confined to the audiogram frequencies. It seems reasonable to assume that each relative loss shown in Table 5.2 extends to the adjacent octave bands. Therefore interpolation between the loss percentages was implemented in order to calculate the loss percentage for any given place.

The interpolation was based on a linear *position* scale on the basilar membrane, i.e. using Equation 6.1. For example, the loss at 1500 Hz, for the hearing loss from Table 5.2, is 22 percent.

6.2 Final model summary

This section gives a brief recap of the model employed in this thesis and how its parameters were derived. The model for normal hearing is described first, followed by the description of the model for impaired hearing.

6.2.1 Complete model for normal hearing

The outer- and and middle ear for normal hearing is modelled as a simple linear phase filter with transfer characteristics as shown in Figure 1.9. It is based on parameters published in Lopez-Poveda and Meddis [2001] which in turn were derived from Goode et al. [1994].

Basilar membrane vibration for normal hearing is modelled using the DRNL [Meddis et al., 2001] with filter bank parameters derived based on forward masking experiments from Lopez-Poveda et al. [2002]. The parameters are shown in Table 3.10.

The inner hair cell receptor potential is modelled by the revised version [Sumner et al., 2002] of the model proposed by Shamma [1986]. The parameters are shown in Table F.1.

The inner hair cell and synapse for normal hearing is modelled as described in Sumner et al. [2002]. Parameters of the model were fitted to individual fibres from a study of chinchilla auditory nerve fibres [Wickesberg and Stevens, 1998]. However, it was decided to use the parameters suggested by Sumner et al. [2003] in the final version of the model. These parameters are shown in Table F.4 and Table F.2

The entire auditory nerve fibre population for normal hearing is modelled as evenly distributed nerve fibres along the cochlear partition. The section modelled has characteristic frequencies ranging from 120 Hz to 10 kHz partitioned into 1008 channels. Three type of fibres are used: high-, medium- and low spontaneous rate fibres in the ratios 12:5:3. Further, each of these fibre types were modelled with three different thresholds as described in Section 6.1. The rate/intensity function for the nine fibres are shown in Figure 6.1.

6.2.2 Complete model for impaired hearing

The outer- and and middle ear for impaired hearing is modelled exactly as for normal hearing, i.e. as a simple linear phase filter with transfer characteristics as shown in Figure 1.9. It is based on parameters published in Lopez-Poveda and Meddis [2001] which in turn were derived from Goode et al. [1994].

Basilar membrane vibration for impaired hearing is modelled using the DRNL [Meddis et al., 2001] with filter bank parameters derived based on forward masking experiments form Lopez-Poveda et al. [2002], but modified to match forward

masking study and audiogram for subject JK's left ear in Oxenham and Plack [1997]. The parameters are shown in Table 3.11.

The inner hair cell and synapse for impaired hearing is modelled as described in Sumner et al. [2002]. The same parameters were used for impaired hearing as were used for normal hearing Sumner et al. [2003]. These parameters are shown in Tables F.1, F.2 and F.4.

The entire auditory nerve fibre population for normal hearing is modelled as evenly distributed nerve fibres along the cochlear partition. The section modelled has characteristic frequencies ranging from 120 Hz to 10 kHz partitioned into 1008 channels. Three type of fibres are used high-, medium- and low spontaneous rate fibres in the ratios 12:5:3. Further, each of these fibre types were modelled with three different thresholds as described in Section 6.1. The rate/intensity function for the nine fibres are shown in Figure 6.1.

The nerve fibres are evenly distributed along the cochlear partition, except for areas with dead inner hair cells where the number of hair cells is reduced. The reduction is calculated for frequencies 125Hz, 250Hz, 500Hz, 1kHz, 2kHz, 4kHz and 8kHz. It is based on a novel method proposed in Section 5.2.2. The modelled reduction in inner hair cells is shown in Table 5.2. It is based on the audiogram for subject JK's left ear in [Oxenham and Plack, 1997]. The reduction is applied to a region on the basilar membrane corresponding to Table 5.2. Interpolation between the percentages are performed according to the frequency-position function proposed by Greenwood [1990].

Chapter 7

Representation in the auditory nerve

The model derived in previous chapters is a very general model of the auditory periphery and as such capable of describing many psychophysical phenomena. The possible applications of the model not explored in the previous chapters include loudness, pitch, suppression, sound quality, speech perception and many others.

The original intention of the thesis was to elaborate and extend the auditory nerve representation of stimuli in general, to specifically accommodate features salient to speech. For time reasons this has not be possible. This chapter does present "a representation of speech signals in normal and impaired ears" as indicated by the thesis title, but the representation is not specific to speech.

This chapter has four sections. The first section introduces the basic presentation form of the auditory nerve discharge patterns used in this thesis. Section 7.2 describes the details of the chosen presentation called the rate diagram. Following this, Section 7.3 present rate diagrams for speech in quiet. Rate diagrams for speech in noise are presented in Section 7.4. Finally, Section 7.5 briefly discusses speech representation in the auditory nerve in the context of the rate diagram.

7.1 Representation in the auditory nerve

In the first part of this thesis a model transforming sound in the air to neural activity patterns was developed. This massive representation forwarded from the 30,000 nerve fibres carries all the information needed for human hearing. The challenge is to unveil how the representation is *presented* to the various parts of the brain. Similarly, it is a matter of discussion how the neural firing patterns is best presented graphically in order to supply information about important features of the stimulus.

One way of presenting auditory nerve firing patterns is simply plotting the number of spikes as a function of time for each nerve fibre. Since it is impossible to do this for all nerve fibres a more compact way of representing this is to group nerve fibres based on their characteristic frequency. A two dimensional plot could then have time on the abscissa, characteristic frequency as the ordinate and a colour code denoting spike count. This representation will serve as the basic representation in the following and will be referred to as the rate diagram.

The representation chosen here does not provide any direct information about periodicity or other temporal aspect of the nerve firings. For time reasons this interesting topic is not addressed in the thesis.

7.2 Developing the rate diagram

The rate diagram has time as abscissa, frequency as ordinate and a colour code as the "third axis". The time resolution, in terms of sample rate, was chosen to be 10 kHz, while the underlying simulation were carried out at 64 kHz[1]. There is no standard length of the x-axis in the rate diagram – it is allowed to vary according to the stimulus.

The choice of y-axis is more complicated. Clearly a linear frequency axis is not a good option. Instead a linear basilar membrane position spacing ranging from 125 Hz to 10 kHz was chosen, as described in Section 6.1. In practice the end-points constitute centre frequencies for one ERB wide bands. So the frequency range covered is actually 125 Hz $-\frac{1}{2}$ERB = 108 Hz to 10 kHz $+\frac{1}{2}$ERB = 10645 Hz .

[1] Even with this fairly low resolution the rate diagrams for 500 ms speech each take up 400 MB of disk space

Since the loss of inner hair cell is modelled as missing channels, simply plotting this would leave "holes" in the rate diagram. In order to compare rate diagrams for normal- and impaired hearing, it was decided to distribute the spikes falling in one "impaired hearing bin" to the corresponding number of "normal hearing bins". Since the number of frequency bands is 30 ERBs by 36 channels pr. ERB = 1080, and the maximum loss modelled here is 70 percent at high frequencies, the resolution on the y-axis is acceptable.

The colour coding was chosen to be an inverted version of gray-scale provided by MatLab. In order to get greater contrast graphically all plots were calculated as the sum of the response to 300 repetitions of the stimulus. Spike counts were then normalised, i.e. divided by a factor of 100 in order to obtain a spike count between 0 and 10 in all time/frequency bins. The standard rate diagram was then defined as the data plotted on this gray-scale going from 0 to 10 spikes. This corresponds to a maximum count of 10 (gray-scale) divided by 300 (repetitions) times 100 (normalising factor) = 30 in each time/frequency bin for one stimulus presentation.

Some stimuli produce too little activity to be visible in a standard rate diagram. Rather than changing the standard colour coding between 0 and 10, it was decided to multiply the spike count by an appropriate factor. This factor is referred to as the gray-scale factor. All such deviations from the standard rate diagram are explicitly stated. It is, of course, desirable to adhere to the standard diagram in order to facilitate comparisons of diagrams.

The rate diagram represents nine fibre types. The standard diagram simply adds the spikes produced by all these fibres into the appropriate frequency/time bins. This way the spikes are lumped together across fibre types, something that is unlikely to occur in the brain. Discerning the contribution in the rate representation into subpopulation responses (e.g. high- , medium- and low spontaneous rate populations) might provide complementary information about encoding in the auditory nerve. Hence, it is possible, but not standard, to produce subpopulation rate diagrams. Again, all such deviations from the standard rate diagram are explicitly stated

An example of the rate diagram is shown in Figure 7.1. It shows the response to a four tone complex. Any tone complex could have been used as an example, but a tone complex similar to the one employed in Buus et al. [1998] was chosen. Each of the tones are four Barks apart from the neighbouring tone. The tone frequencies are 838, 1600, 2915 and 5790 Hz.

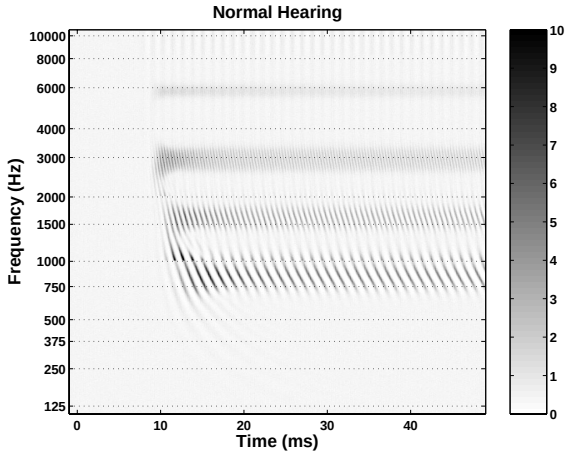


Figure 7.1: An example of rate diagram as defined in this thesis. The stimulus is a four tone complex with frequencies 838, 1600 2915 and 5790 Hz. the overall level is 30 dB SPL. The response is for a normal hearing person. A gray-scale factor of three has been used (see Section 7.2 for definition)

From the diagram it is clear that the individual tone frequencies are represented in the response. Also, the onset response is greater than the later responses. Finally, in addition to the tonotopic classification of frequency, stimulus periodicity is reflected in the rate diagram, in that ripples occur with a period corresponding to stimulus for the given place.

7.3 Rate diagrams for speech signals in quiet

This section presents rate diagrams for normal- and impaired hearing in response to speech under a variety of stimulus configurations. The stimulus used is taken from the Dantale CD [Elberling et al., 1989]. The speech stimulus is presented in quiet. Rate diagrams for speech in noise are presented in the following section.

The Danish word “gas” was chosen for the examples below for several reasons. It is close to the corresponding English word and it has significant energy in the high frequency region where the hearing loss under examination is the greatest.

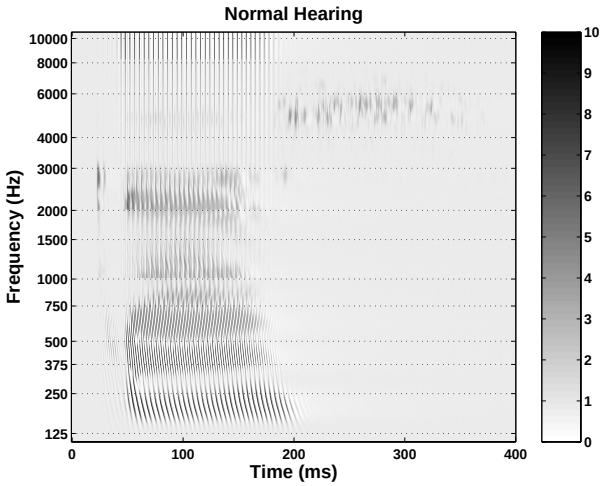


Figure 7.2: Rate diagram for normal hearing in response to the word "gas" at a level of 30 dB SPL with a gray-scale factor of two. The stimulus was taken from Elberling et al. [1989]

Figure 7.2 shows the rate diagram for the normal hearing in response to the Danish word "gas" at an overall level of 30 dB SPL. The relative low level means that a gray-scale factor of two was used in order to enhance contrast.

From the diagram it is evident that the speaker is female, in that the spikes corresponding to the fundamental frequency are in the proximity of 200 Hz, allowing for the travelling wave to continue slightly beyond the characteristic frequency.

Moreover, formant frequencies from the vowel are discernible at approximately 400 Hz, 550 Hz and 2 kHz. The initial "g" is represented as a short burst with most spikes falling between 2 and 3 kHz, followed by a period of silence. Finally, the "s" is visible as energy around 5 kHz with onset time around 200 ms.

Figure 7.3 shows the rate diagram for normal hearing in response to the Danish word "gas" at an overall levels of 60. Comparing to Figure 7.2 the formants are hardly discernible. The majority of the auditory nerve fibres are high spontaneous rate fibres, which saturate at this level, or lower. The silence immediately following the "g" is still present, and the high frequency component representing the "s" is also visible.

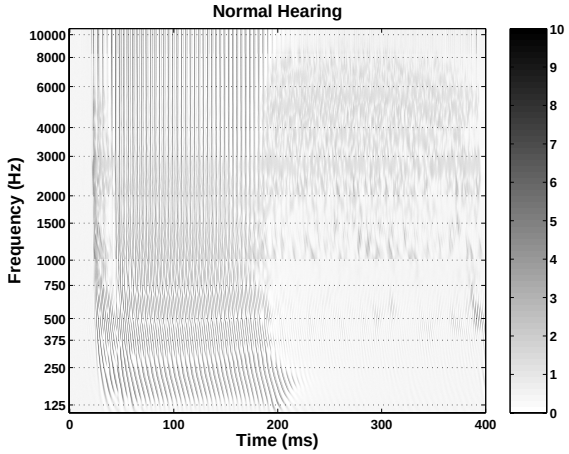


Figure 7.3: Rate diagram for normal hearing in response to the word "gas" at a level of 60. Compare with Figure 7.2

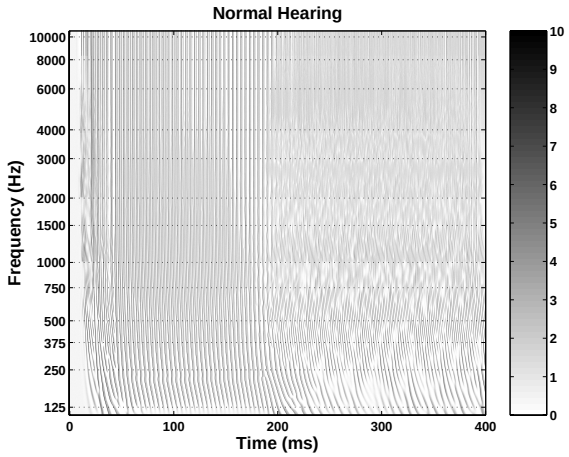


Figure 7.4: Rate diagram for normal hearing in response to the word "gas" at a level of 90.

Figure 7.4 shows the rate diagram for normal hearing in response to the Danish word "gas" at an overall level of 90. At this level the silence after the "g" is not visible from the rate diagram and neither are the formants. A change in pattern at the onset of the "s" occurs even at this level. Both Figure 7.4 and Figure 7.3 show clear periodicity patterns, but there is no obvious interpretation of these. However, contrary to what is apparent from the rate diagram, an increases in level should improve intelligibility.

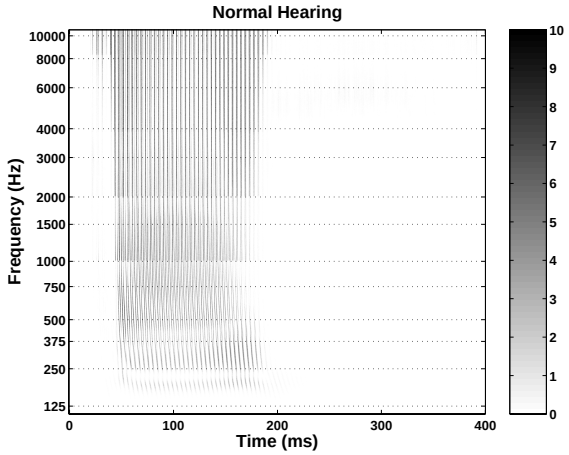


Figure 7.5: Rate diagram for normal hearing in response to the word "gas" at a level of 90. Only the least sensitive low spontaneous rate fibres are shown. A gray-scale factor of 25 was used

Figure 7.5 shows the rate diagram for normal hearing, least sensitive, low spontaneous rate fibres (see Section 6.1, Figure 6.1 for definition), in response to the Danish word "gas" at an overall levels of 90. The interesting part of this diagram is that the fundamental frequency and the first two formants are discernible. Particularly considering the onset responses. This illustrates that even at this level the rate diagram provides information and it points to a possible role of low spontaneous rate fibres. Admittedly, these observations are "well hidden" in the representation and it could be argued that they are over-interpretations, but this is more a limitation of the rate diagrams than the underlying representation.

Figure 7.6 shows the rate diagram for the impaired hearing in response to the Danish word "gas" at an overall levels of 60 and 90 dB SPL respectively. It is

clear that the overall response is lower for the impaired hearing, not surprisingly, particularly at high frequencies.

Figure 7.6 top panel shows very little response to the "g", i.e. the response begins at approximately 50 ms, whereas response can be seen in the lower panel and also in Figure 7.3. Moreover, Figure 7.6 lower panel resembles Figure 7.3 more than Figure 7.6 top panel resembles Figure 7.3, at least for the first 200 ms. This is consistent with the widely held belief that impaired hearing approaches normal hearing at higher levels.

7.4 Rate diagrams for speech signals in noise

Speech in noise is presented in different signal-to-noise ratios and at various overall levels. Only one type of noise is used here. Briefly, this noise has a speech shaped spectrum, i.e. a flat power spectrum up to 500 Hz and a roll-off of 12 dB/octave above 500 Hz. Further, the envelope spectrum of the speech shaped noise matches that of speech. The noise is identical to the speech shaped noise used by Elberling et al. [1989].

Figure 7.7 shows the response to the word "gas" at signal-to-noise ratio of one and an overall level of 60 dB SPL. Based on these rate diagrams the most one can say is that it is remarkable that normal hearing persons are able to perceive anything under these conditions... While some "structured ripples" are present in the top panel, they do not seem to relate to the speech stimulus in that they persist after cessation of the speech stimulus. No structure relating to the speech stimulus appear to be present in the lower panel.

Figure 7.7 shows the response to the word "gas" at signal-to-noise ratio of 30 dB and an overall level of 60 dB SPL. While it is difficult to interpret the content of the diagrams, it is obvious that the 30 dB signal-to-noise ratio make the diagrams very similar to the ones from the "in-quiet" condition (c.f. Figures 7.3 and 7.6 top panel).

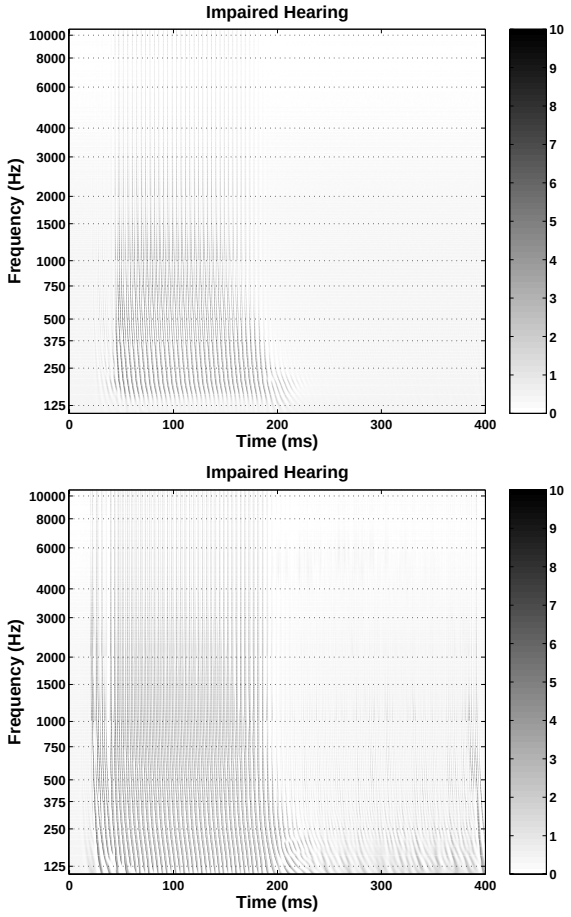


Figure 7.6: Rate diagram for impaired hearing in response to the word "gas" at a level of 60 (top) and 90 dB SPL (bottom)

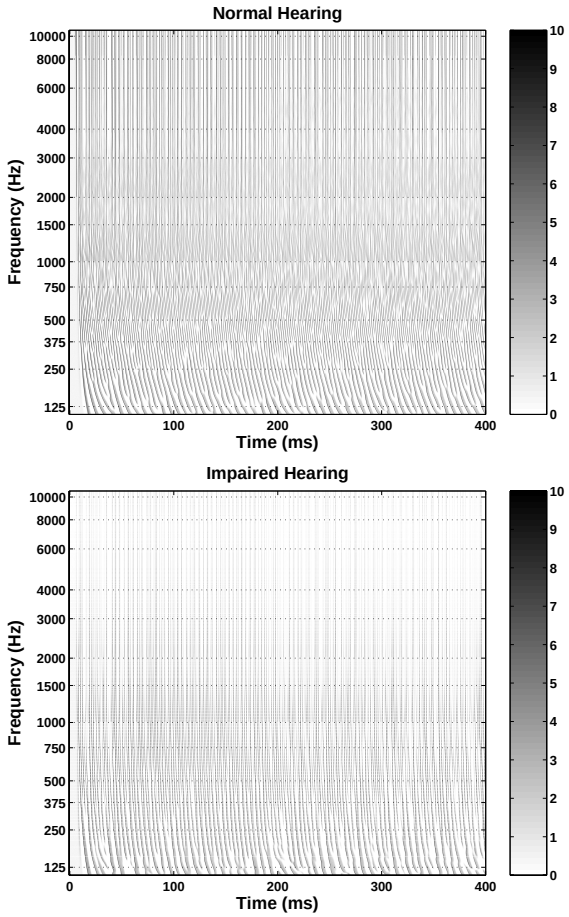


Figure 7.7: Rate diagrams for normal hearing (top) and impaired hearing (bottom) in response to the word "gas" at signal-to-noise ratio of 0 dB. The overall level is 60 dB SPL

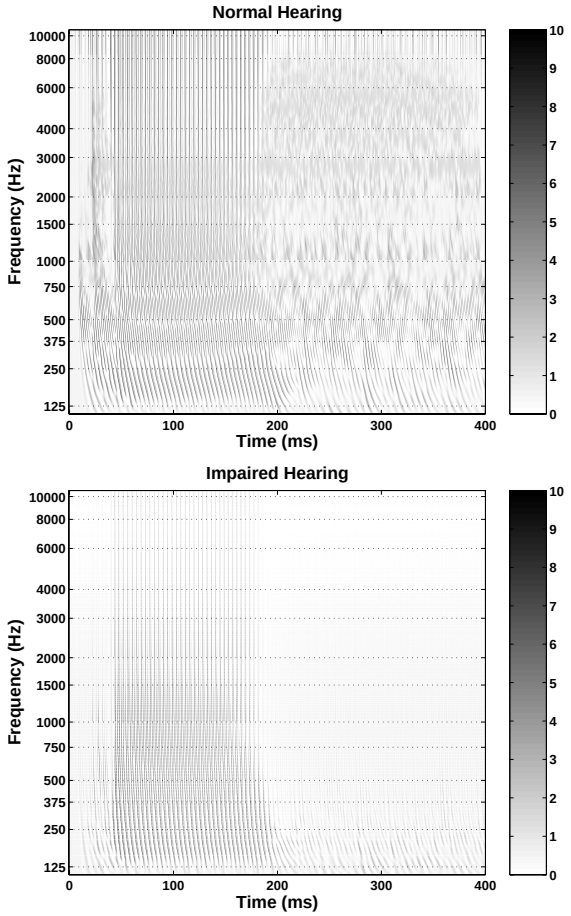


Figure 7.8: Rate diagram for normal (top panel) and impaired hearing (bottom panel) in response to the word "gas" at signal-to-noise ratio of 30 dB. The overall level is 90 dB SPL

7.5 Discussion of speech representation

For low levels the rate diagrams exhibit a clear rate-place representation of the stimulus. As levels go up this representation becomes increasingly obscure, although some features persist in some low spontaneous rate fibres. For the hearing loss examined very little rate-place encoding was evident from the rate diagrams.

The speech in noise rate diagrams provided an even more obscure representation of the stimulus for the 0 dB signal-to-noise condition. In contrast the 30 dB signal-to-noise ratio showed almost the same rate diagrams as the speech in quiet condition.

None of the above observation are surprising, but they clearly indicate that the rate diagram is inadequate as a tool for understanding speech representation in the auditory nerve. It is important to distinguish the rate diagram presented here from the underlying representation – the discharge patterns of the auditory nerve fibres. So while the rate diagrams presented here are inadequate in accounting for the representation of speech, the underlying representation is worth investigating.

Seen in this light the question: "How is speech encoded in the auditory nerve?" has a more interesting counterpart "How is speech decoded from the representation in the auditory nerve?". While this thesis set out to explore the former of the questions, and to some extent has, the latter is the more interesting question.

Many answers to the above question have been suggested over the years. Rather, than answering the question directly it might be better to consider finding some "persistent" or "robust" features of speech, which are also "persistent" and "robust" in in the auditory nerve representation. The model presented in this thesis could serve as the fundamental investigation tool.

Chapter 8

Conclusions and perspectives

This short chapter is divided into three sections. The first gives a brief description of the most notable results achieved in this thesis. Section 8.2 discusses possible future projects from these results. Finally, a conclusion to the entire thesis is drawn.

8.1 Summary of the results

Chapter 3 presented simulation of three forward masking experiments. The data from the most comprehensive experiment was used as the basis for deriving parameters for the DRNL filter model of the basilar membrane. Although some discrepancies from the psychophysical data was observed the errors were for the most part comparable with one standard deviation of the psychophysically measured points. It was suggested that the temporal window must be frequency dependant and that this, at least to some extent, would account for the aforementioned discrepancies. An additional improvement of the results might be achieved if adaptation is modelled.

Chapter 5 presented a novel method of calculating inner hair cell loss from absolute thresholds. The method was based on the conjecture that absolute

behavioural threshold can be modelled based on spike count from appropriately selected auditory nerve fibres.

Based on the results of the basilar membrane model parameters, the calculated inner hair cell loss and results from literature models for the entire auditory nerve fibre population for normal hearing and impaired hearing was suggested.

For time reasons this representation was not investigated in detail. A rudimentary diagram showing a graphical representation of the auditory nerve fibre discharge patterns was developed. It was found that this rate diagram was incapable of showing basic properties of speech encoding in the auditory nerve.

8.2 Future projects

The previous section pointed to the most important results of work conducted in connection with this thesis. This section outlines possible continuations and improvement of these results. First, forward masking is discussed. Next further ideas for models of impaired hearing are presented. Finally, the topic of speech decoding from the auditory nerve is considered in Section 8.2.3.

8.2.1 Modelling forward masking

The simulation of forward masking results described in Chapter 3 were originally carried out in order to derive parameters for the DRNL. However, modelling forward masking is a research topic in its own right as the comprehensive literature shows, e.g. Wilson and Carhart [1971]; Relkin and Turner [1988]; Plack and Moore [1990]; Oxenham and Moore [1994]; Turner et al. [1994]; Oxenham and Plack [1997]; Plack and Oxenham [1998, 2000b]; Oxenham [2001]; Plack et al. [2002].

In this thesis forward masking was only modelled in terms of temporal integration. Adaptation was never considered even though the tools modelling peripheral adaptation were described in Chapter 5. Extending the forward masking model to accommodate adaptation phenomena would only require a minimum of implementation, whereas parameter tuning, specifically nerve fibre parameters, would require the bulk of the effort. However, this would also be an interesting and relatively uncharted research area.

8.2.2 Models of impaired hearing

Outer hair cell loss was modelled as frequency specific adjustment of a single DRNL parameter. It was based on psychophysically measured absolute thresholds and a few forward masking data points. In this way, but with more data, more comprehensive and thus more accurate modelling is possible. In order to get the relevant data it seems mandatory to design and carry out psychophysical experiments. Hence, the present thesis does not provide much of a head start for pursuing this idea.

Calculating inner hair cell loss was done based on the conjecture that absolute threshold can be modelled based on spike count from nerve fibres with the appropriate characteristics. While quite speculative, this attempt to link psychophysical data to data from physiology deserves attention. At this point in time, however, it is not clear how to carry this idea forward in that experimental design providing data appear to be beyond the present day capabilities. Hopefully, this conclusion can be falsified by clever experimental design or technical advances in the near future.

8.2.3 Speech decoding from the auditory nerve

The rate diagram developed in Section 7.2 was an attempt at presenting a simple summary of auditory nerve fibre discharge patterns. As such it was successful. However, it tells us little about how these discharge patterns are processed in the brain stem, let alone the auditory cortex.

As indicated in the discussion of speech representation in Section 7.5 the interesting question is not so much "what the representation in the auditory nerve looks like", but more "how is it decoded". This thesis proposed an answer to the former question and unfortunately the latter question was not addressed for time reasons.

An interesting beginning to an answer to how speech is decoded from the auditory nerve could be an examination of the complex modulation spectrum as described in Greenberg and Arai [2001]. Greenberg argues that preservation of the complex modulation spectrum is the key to preserving speech intelligibility under various types of degradations, such as reverberation and hearing impairment. While the basis of this argument is an ill-defined [Atlas, 2002], yet intuitively well-known, physical property, namely the complex modulation spectrum, the challenge would be to identify a similar representation in the auditory

nerve. The work carried out in connection with this thesis provides the auditory nerve representation required by such a project.

8.3 Conclusion

The goal of developing a model for the representation of speech signals in normal and impaired ears has been reached. Although this representation has undergone no direct tests as to its validity and generality, the intermediate models have all undergone fairly comprehensive testing. Admittedly, parts of the model of hearing impairment rely on conjectures. Moreover, the distribution of auditory nerve fibre types also rely on conjectures. However, at present knowledge in these areas are incomplete and it is not possible to collect direct evidence.

In this situation, proposing and evaluating models is an important tool for examining how underlying mechanisms could work. Seen in this perspective, the present thesis offers "a model for the representation of speech signals in normal and impaired ears" for which much data has yet to be acquired for a proper evaluation.

Bibliography

R. Aibara, J. T. Welsh, S. Puria, and R. L. Goode. Human middle-ear sound transfer function and cochlear input impedance. *Hearing Research*, 152:100–109, 2001.

J.A. Assad, G.M.G. Shepard, and D.P. Corey. Tip-link integrety and mechanical transduction in vetebrate hair cells. *Neuron*, 7:985–994, 1991.

L. Atlas. Acoustic/modulation frequency transform for single-channel talker separation. In *Dynamics of Speech Production and Speech Perception, NATO Advanced Study Institute Meeting, June 24 – July 6, 2002, Il Ciocco, Italy*, 2002. Presentation given at the meeting.

R. Baker, Rosen S., and A.M. Darling. An effective characterisation of human auditory filtering across level and frequencythat is also physiologically reasonable. In A.R. Palmer, A. Rees, A.Q. Summerfield, and R. Meddis, editors, *Psychophysical and Physiological Advances in Hearing*, pages 81–88. Whurr Publisher, London, 1998.

G. von Békésy. *Experiments in Hearing*. McGraw-Hill, New York, 1960.

W.E. Brownell, C.R. Bader, D. Bertrand, and Y. Ribaupierre. Evoked mechanical responses of isolated cochlear outer hair cells. *Science*, 227:194–196, 1985.

S. Buus, M. Florentine, and T. Poulsen. Temporal integration of loudness in listeners with hearing losses of primarily cochlear origin. *Journal of the Acoustical Society of America*, 104:3464–3480, 1999.

S. Buus, H. Müsch, and M. Florentine. On loudness at threshold. *Journal of the Acoustical Society of America*, 104:399–410, 1998.

L. H. Carney and T. C. T. Yin. Temporal coding of resonances by low-frequency an fibres: Single fibre responses and a population model. *Journal of Neurophysiology*, 60:1653–1677, 1988.

L.H. Carney. A model for the responses of low frequency auditory-nerve fibers in cat. *Journal of the Acoustical Society of America*, 99:401–417, 1993.

T.U. Christiansen. Towards modeling basilar membrane response to speech. In L. Tranebjærg, J. Christensen-Dalsgaard, T. Andersen, and T. Poulsen, editors, *Genetics and the function of the auditory system, 19th Danavox Symposium*, pages 393–410, 2002. The Danavox Jubilee Foundation, ISBN 87-982422-9-6.

T.U. Christiansen. Modeling auditory nerve fiber responses to click trains. In *Proceedings of Joint Baltic-Nordic Acoustical Meeting*, volume 48, pages 14–18, 2003. Ultragarsas (ISSN 1392-2114).

D.P. Corey and A.J. Hudspeth. Kinetics of the receptor current in bullfrog saccular hair cells. *Journal of Neuroscience*, 3(5):962–976, 1983.

P. Dallos. Overview: Cochlear neurobiology. In P. Dallos, A.N. Popper, and R.R. Fay, editors, *The Cochlea*, chapter 1. Springer, New York, 1996. ISBN 0-387-94449-4.

T. Dau and A. Püschel, D.and Kohlrausch. A quantitative model of the "effective" signal processing in the auditory system. ii. simulations and measurements. *Journal of the Acoustical Society of America*, 99:3623–3631, 1996.

T. Dau, D. Püschel, and A. Kohlrausch. A quantitative model of the "effective" signal processing in the auditory system. i. model structure. *Journal of the Acoustical Society of America*, 99:3615–3622, 1996.

C. Elberling, C. Ludvigsen, and P.E. Lyregaard. Dantale: A new danish speech material. *Scandiniavian Audiology*, 18:169–175, 1989.

L.F. Elfner and W.E. Caskey. Continuity effects with alternately sounded noise and tone signals as a function of manner of presentation. *Journal of the Acoustical Society of America*, 38:543–547, 1965.

E.F. Evans. The frequency response and other properties of single fibers in the guinea-pig cochlear nerve. *Journal of Physiology*, 226:263–287, 1972.

H. Fastl. Temporal masking effects i. *Acustica*, 35:287–302, 1976.

M. Florentine and S. Buus. An excitation-pattern model for intensity discrimination. *Journal of the Acoustical Society of America*, 70:1646–1654, 1981.

G. Frank, W. Hemmert, and A.W. Gummer. Limiting dynamics of high-frequency electromechanical transduction of outer hair cells. *Proceedings National Academy Science*, 96:4420–4425, 1999.

C.D. Geisler. *From sound to synapse*. Oxford University Press, New York, 1998. ISBN 0-19-510025-5.

C. Giguère and P.C. Woodland. A computational model of the auditory periphery for speech and hearing research: I. ascending path. *Journal of the Acoustical Society of America*, 95:331–342, 1994.

B.R. Glasberg and B.C.J. Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47:103–138, 1990.

J. Goldberg and Browne P. Responses of binaural neurons of dog superior olivary to dichotic tonal stimuli: some physiological mechanisms of sound localization. *Journal of Neurophysiology*, 32:613–636, 1969.

R.L. Goode, M. Killion, K. Nakamura, and S. Nishihara. New knowledge about the function of the human middle ear: development of an improved analog model. *American journal of Otolaryngology*, 15:145–154, 1994.

S. Greenberg. Acoustic transduction in the auditory periphery. *Journal of Phonetics*, 16:3–17, 1988.

S. Greenberg and T. Arai. The relation between speech intelligibility and the complex modulation spectrum. In *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech-2001)*, pages 473–476, 2001.

D.D. Greenwood. A cochlear frequency-position function for several species – 29 years later. *Journal of the Acoustical Society of America*, 87:2592–2605, 1990.

P.S. Guth, A. Aubert, A.J. Ricci, and C.H. Norris. Differential modulation of spontaneous and evoked neurotransmitter release from hair cells: some novel hypotheses. *Hearing Research*, 56:69–78, 1991.

W.M. Hartmann. *Signals, sound, and sensation*. Springer-Verlag, Berlin, 1997.

M.G. Heinz, H.S. Colburn, and L.H. Carney. Evaluating auditory performance limits: I. one-parameter discrimination using a computational model for the auditory nerve. *Neural Computation*, 13:2273–2316, 2001a.

M.G. Heinz, X. Zhang, I.C. Bruce, and L.H. Carney. Auditory-nerve model for predicting performance limits of normal and impaired listeners. *Acoustic Research Letters Online*, 2:91–96, 2001b.

S.D. Holmes. *Segregation Of Concurrent Vowels: An Auditory Model*. PhD thesis, Department of Psychology, University of Essex, 2002.

T. Houtgast. Psychophysical evidence for lateral inhibition in hearing. *Journal of the Acoustical Society of America*, 51:1885–1894, 1972.

A.J Hudspeth and P.G. Gillespie. Pulling strings to tune transduction: Adaptation by hair cells. *Neuron*, 12:1–9, 1994.

M. Huss, B.C.J. Moore, and B.R. Glasberg. Development and evaluation of a noise producing equal masked thresholds at all frequencies as a tool for identification of 'dead regions'. *British Journal of Audiology*, 34:106–107, 2000.

T. Irino and R.D. Patterson. A time domain level-dependant auditory filter: The gammachirp. *Journal of the Acoustical Society of America*, 101:1412–1441, 1997.

R.L. Jenison. A composite model of the auditory periphery for the processing of speech based on the filter response functions of single auditory-nerve fibers. *Journal of the Acoustical Society of America*, 90:773–786, 1991.

R.L. Johnson. The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *Journal of the Acoustical Society of America*, 68:1115–1122, 1980.

D.T. Kemp. Stimulated acoustic emissions from within the human auditory system. *Journal of the Acoustical Society of America*, 64:1386–1391, 1978.

Y.S.K. Kiang. *Discharge patterns of single fibers in the cat's auditory nerve*. M.I.T. press, Cambidge, Massachusetts, 1965.

B. Kollmeier, R.P. Derleth, and T. Dau. Modeling the effective auditory signal processing for hearing-impaired listeners. In A.R. Palmer, A. Rees, A.Q. Summerfield, and R. Meddis, editors, *Psychophysical and Physiological Advances in Hearing*, pages 482–490. Whurr Publisher, 1993.

H. Levitt. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49:467–477, 1971.

M.C. Liberman. Auditory-nerve responses from cats raised in a low-noise chamber. *Journal of the Acoustical Society of America*, 63:442–455, 1978.

M.C. Liberman. Morphological differences among radial afferent fibers in the cat cochlea: An electron microscopic study of serial sections. *Hearing Research*, 3:45–63, 1980.

M.C. Liberman. The cochlear freqeuncy map for the cat: Labeling auditory-nerve of known characteristic frequency. *Journal of the Acoustical Society of America*, 72:1441–1449, 1982a.

M.C. Liberman. Single-neuron labeling in the cat auditory nerve. *Science*, 216: 1239–1241, 1982b.

M.C. Liberman. Physiologigy of cochlear efferent and afferent neurons: Direct comparisons in the same animal. *Hearing Research*, 34:179–192, 1988.

M.C. Liberman, L.W. Dodds, and S. Pierce. Afferent and efferent innervation of the cat cochlea: Quantitative analysis with light and electron microscopy. *The Journal of Comparative Neurology*, 301:443–460, 1990.

E.A. Lopez-Poveda and R. Meddis. A human cochlear filterbank. *Journal of the Acoustical Society of America*, 110:3107–3118, 2001.

E.A. Lopez-Poveda, C.J. Plack, and R. Meddis. Cochlear nonlinearity between 400 and 8000 hz in listeners with normal hearing. *Journal of the Acoustical Society of America*, 113:951–960, 2002.

R.F Lyon. A computational model of filtering detection, and compression in the cochlea. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1282–1285, 1982.

D. McFadden and M.F. Yama. Upward shift in masking pattern with increasing masker intensity. *Journal of the Acoustical Society of America*, 74:1185–1189, 1983.

R. Meddis. Simulation of mechanical to neural transduction in the auditory receptor. *Journal of the Acoustical Society of America*, 79:702–711, 1986.

R. Meddis. Private communication, 2002.

R. Meddis, L.P. O'Mard, and E.A. Lopez-Poveda. A computational algorithm for computing nonlinear auditory frequency selectivity. *Journal of the Acoustical Society of America*, 109:2852–2861, 2001.

S. Mehrgardt and V. Mellert. Traformation characteristics of the external human ear. *Journal of the Acoustical Society of America*, 61:1567–1576, 1977.

B.C.J. Moore. *An introduction to the psychology of hearing*. Academic Press, London, fourth edition, 1997. ISBN 0-12-505627-3.

B.C.J. Moore and B.R. Glasberg. A model of loudness perceptionnd applied to cochlear hearing loss. *Auditory Neuroscience*, 3:289–311, 1997.

B.C.J. Moore, B.R. Glasberg, and T. Baer. A model for the prediction of thresholds, loudness and partial loudness. *Journal of the Audio Engineering Society*, 45:224–240, 1997.

B.C.J. Moore, B.R. Glasberg, C.J. Plack, and A.K. Biswas. The shape of the ear's temporal window. *Journal of the Acoustical Society of America*, 83: 1102–1116, 1988.

B.C.J. Moore, B.R. Glasberg, and D.A. Vickers. Further evaluations of a model of loudness perception applied to cochlear hearing loss. *Journal of the Acoustical Society of America*, 106:898–907, 1999.

B.C.J. Moore, M. Huss, D.A. Vickers, and T. Baer. Psychoacoustics of dead regions. In A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs, and R. Schoonhoven, editors, *Physiological and Psychophysical Bases of Auditory Function*, pages 482–490. Shaker, 2000.

B.C.J. Moore, R.W. Peters, and B.R. Glasberg. Auditory filter shapes at low center frequencies. *Journal of the Acoustical Society of America*, 88:132–140, 1990.

D.C. Mountain and A.E. Hubbard. Analysis and synthesis of cochlear mechanical function using models. In H.L. Hawkins, T.A. Mullen, A.N. Popper, and Fay R.R., editors, *Auditory Computation*, chapter 3. Springer-Verlag, New York, 1996. ISBN 0-387-97843-7.

R.C. Naidu and D.C. Mountain. Measurements of the stiffness map challenge a basic tenet of cochlear theories. *Hearing Research*, 124:124–131, 1998.

A.J. Oxenham. Forward masking: Adaptation or integration? *Journal of the Acoustical Society of America*, 109:732–741, 2001.

A.J. Oxenham and B.J.C. Moore. Modeling the additivity of nonsimultaneous masking. *Hearing Research*, 80:105–118, 1994.

A.J. Oxenham and C.J. Plack. A behavioral meassure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. *Journal of the Acoustical Society of America*, 101:3666–3675, 1997.

R.D. Patterson. Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America*, 59:640–654, 1976.

R.D. Patterson, J. Holdsworth, I. Nimmo-Smith, and P. Rice. *The auditory filterbank*. MRC-APU Report 2341, Cambridge, England, 1991.

R.D. Patterson and I. Nimmo-Smith. Off-frequency listening and auditory-filter asymmetry. *Journal of the Acoustical Society of America*, 67:229–245, 1980.

J.O. Pickles. *An introduction to the physiology of hearing.* Academic Press, London, 1982.

C.J. Plack and B.C.J. Moore. Temporal window shape as a function of frequency and level. *Journal of the Acoustical Society of America*, 87:2178–2187, 1990.

C.J. Plack and A.J. Oxenham. Basilar-membrane nonlinearity and the growth of forward masking. *Journal of the Acoustical Society of America*, 103:1598–1608, 1998.

C.J. Plack and A.J. Oxenham. Basilar-membrane nonlinearity estimated by pulsation threshold. *Journal of the Acoustical Society of America*, 107:501–507, 2000a.

C.J. Plack and A.J. Oxenham. Effects of masker frequency and duration in forward masking: further evidence of peripheral nonlinearity. *Hearing Research*, 150:258–266, 2000b.

C.J. Plack, A.J. Oxenham, and V. Drga. Linear and nonlinear processing in temporal masking. *Acta Acustica*, 88:348–358, 2002.

D. Pralong and S. Carlile. Measuring the human head-related transfer functions: A novel method for the construction and calibration of a miniature "in-ear" recording system. *Journal of the Acoustical Society of America*, 95:3435–3444, 1994.

D. Pralong and S. Carlile. The role of individualized headphone calibration for the generation of high fidelity virtual auditory space. *Journal of the Acoustical Society of America*, 100:3785–3793, 1996.

W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical recipes in C++*. Cambridge University Press, Cambridge, UK, second edition, 2002. ISBN 0-521-75033-4.

S. Puria, J. J. Rosowski, and W. T. Peake. Sound-pressure measurements in the cochlear vestibule of human-cadaver ears. *Journal of the Acoustical Society of America*, 101:2754–2770, 1997.

F. Rattay, I.C. Gebeshuber, and A.H. Gitter. The mammalian auditory hair cell: A simple electric circuit model. *Journal of the Acoustical Society of America*, 103:1558–1565, 1998.

E. M. Relkin and J.R. Doucet. Is loudness simply proportional to the auditory nerve spike count. *Journal of the Acoustical Society of America*, 101:2735–2740, 1997.

E. M. Relkin and C. W. Turner. A reexamination of forward masking in the auditory nerve. *Journal of the Acoustical Society of America*, 84:584–591, 1988.

T. Ren. Longitudinal pattern of basilar membrane vibration in the sensitive cochlea. *Proceedings National Academy Science*, 99:7101–7106, 2002.

W.S. Rhode. Observations of the vibration of the basilar membrane in squirrel monkeys using the mõssbaer technique. *Journal of the Acoustical Society of America*, 49:1218–1231, 1971.

W.S. Rhode and N.P Cooper. Nonlinear mechanics in the apical turn of the chinchilla. *Auditory Neuroscience*, 3:101–121, 1996.

W.S. Rhode and A. Recio. Study of mechanical motions in the basal region of the chinchilla cochlea. *Journal of the Acoustical Society of America*, 107: 3317–3332, 2000.

M.A. Ruggero. Responses to sound of the basilar membrane of the mammalian cochlea. *Current Opinion in Neurobiology*, 2:449–456, 1992.

M.A. Ruggero and N.C. Rich. Application of comercially-manufactured doppler-shift laser velocimetry to the measurement of basilar-membrane vibration. *Hearing Research*, 51:215–230, 1991.

M.A. Ruggero and A.N. Temchin. Middle-ear transmission in humans: wide-band, not frequency-tuned? *Acoustic Research Letters Online*, 4:53–58, 2003.

I.J. Russell, A. Cody, and G. Richardson. The responses of inner and outer hair cells in the basal turn of the guinea pig cochlea and in the mouse cochlea grown in vitro. *Hearing Research*, 22:199–216, 1986.

M.B. Sachs and P.J. Abbas. Rate versus level functionsfor auditory nerve fibres in cats: tone burst stimuli. *Journal of the Acoustical Society of America*, 56: 1835–1847, 1974.

J. Santos-Sacchi and Dilger J. Whole cell currents and mechanical responses of isolated outer hair cells. *Hearing Research*, 35:143–150, 1988.

H.F. Schuknecht. Further observations on the pathology of presbycusis. *Archives of Otolaryngology*, 80:369–382, 1964.

D. Sen and J.B. Allen. Reproducing upward spread of masking and two-tone suppression using a nonlinear model of the cochlea. volume 22. American Otolaryngology, 1999.

S. Seneff. A joint synchrony/mean rate model of auditory speech processing. *JP*, 16:55–76, 1988.

S. Shamma. A biophysical model of cochlear processing: Intensity dependence of pure tone responses. *Journal of the Acoustical Society of America*, 80: 133–145, 1986.

S. Shamma. The acoustic features of speech sounds in a model of auditory processing: vowels and voiceless fricatives. *JP*, 16:77–91, 1988.

B.A. Stach. *Clinical Audiology*. Singular Publishing Group, San Diego, London, 1998. ISBN 1-56593-346-x.

K. Steiglitz. *A digital signal processing primer*. Addison-Wesley Publishing Company, Menlo Park, California, 1996. ISBN 0-8053-1684-1.

C.J. Sumner, Lopez-Poveda E.A., and R. Meddis. A nonlinear filter-bank model of the guinea-pig cochlear nerve: Rate responses. *Journal of the Acoustical Society of America*, 113:3264–3274, 2003.

C.J. Sumner, E.A. Lopez-Poveda, O'Mard L.P., and R. Meddis. A revised model of the inner-hair cell and auditory nerve complex. *Journal of the Acoustical Society of America*, 111:2178–2188, 2002.

J. Tonndorf and S.M. Khanna. Tympanic-membrane vibrations in human cadavers studied by time-averaged holography. *Journal of the Acoustical Society of America*, 52:1221–1233, 1972.

C.W. Turner, E.M. Relkin, and J. Doucet. Psychophysical and physiological forward masking studies: Probe duration and rise time effects. *Journal of the Acoustical Society of America*, 96:795–800, 1994.

M. van der Heijden and A. Kohlrausch. The role of envelope fluctuations in spectral masking. *Journal of the Acoustical Society of America*, 97:1800–1807, 1995.

N.F. Viemeister. Intensity coding and the dynamic range problem. *Hearing Research*, 34:267–274, 1988.

N.F. Viemeister and G.H. Wakefield. Temporal integration and multiple looks. *Journal of the Acoustical Society of America*, 90:858–865, 1991.

L.A. Westerman and R.L Smith. Conservation of adaptating components in auditory-nerve responses. *Journal of the Acoustical Society of America*, 81: 680–691, 1987.

R.E. Wickesberg and H.E. Stevens. Responses of auditory nerve fibers to trains of clicks. *Journal of the Acoustical Society of America*, 103:1990–1999, 1998.

R.H. Wilson and R. Carhart. Forward and backward masking interactions and additivity. *Journal of the Acoustical Society of America*, 49:1254–1263, 1971.

G.K. Yates. Basilar memerane nonlinearity and its influence on auditory nerve rate-intensity functions. *Hearing Research*, 50:145–162, 1990.

G.K. Yates. Cochlear structure and function. In B.J.C. Moore, editor, *Hearing*, chapter 2. Academic Press, san Diego, 1995. ISBN 0-12-505626-5.

G.K. Yates, I.M. Winter, and D. Robertson. Basilar memerane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range. *Hearing Research*, 45:203–220, 1990.

W. A. Yost. *Fundamentals of hearing.* Academic Press, San Diego, California, 2000.

X. Zhang, M.G. Heinz, I.C Bruce, and L.H. Carney. A phenemenological model for the response of auditory-nerve fibers: I. nonlinear tuning with compression and suppression. *Journal of the Acoustical Society of America*, 109:648–670, 2001.

J. Zheng, W. Shen, D.Z. He, K.B. Long, Madison L.D., and Dallos P. Prestin is the motor protein of cochlear outer hair cells. *Nature*, 405:149–155, 2000.

E. Zwicker. Dependence of post-masking on masker duration and its effects on temoral effects in loudness. *Journal of the Acoustical Society of America*, 75: 219–23, 1984.

E. Zwicker and H. Fastl. *Psychoacoustics*. Springer-Verlag, Berlin, 1990.

E. Zwicker, G. Flottorp, and Stevens S.S. Critical bandwidth in loudness summation. *Journal of the Acoustical Society of America*, 29:548–557, 1957.

Appendix A

DRNL parameters

Table A.1: Example of values for the free DRNL parameters for six centre frequencies as described in Section 2.3.3. The values are taken from Lopez-Poveda and Meddis [2001].

	Signal Frequency (Hz)	250	500	1000	2000	4000	8000
Linear-	CF_{lin} (Hz)	244	480	965	1925	3900	7750
path-	BW_{lin} (Hz)	100	130	240	400	660	1450
parameters	g	1150	850	520	410	320	220
	CF_{nl} (Hz)	250	500	1000	2000	4000	8000
Nonlinear-	BW_{nl} (Hz)	84	103	175	´300	560	1100
path-	a	2194	5184	7558	9627	22288	43584
parameters	b	0.450	0.28	0.130	0.078	0.045	0.030
Effective	CF	258	508	998	2006	3978	7720
DRNL	BW_{3dB}	50	68	118	210	415	755

Table A.2: The linear regression version of the parameters for the human DRNL filter bank derived from Table A.1

	Regression coefficients	p_0	m
Linear-path-parameters	CF_{lin} (Hz)	-6.7620 x10^{-2}	1.01679 x 10^{0}
	BW_{lin} (Hz)	3.7280 x10^{-2}	7.85630 x 10^{-1}
	g	4.2041 x10^{0}	-4.7909 x 10^{-1}
Nonlinear-path-parameters	BW_{nl} (Hz)	-3.1930 x 10^{-2}	7.7426 x 10^{-1}
	a	1.4030 x 10^{0}	8.1916 x 10^{-1}
	b	1.6191 x 10^{0}	-8.1867 x 10^{-1}

Appendix B

BNAM article

This appendix contains a copy of Christiansen [2003]. The page content in the present version is scaled down from A4, used in the proceedings, to the regular paper size used throughout the thesis. The original proceedings page numbers have been omitted while thesis page headings are included.

Modelling auditory nerve fibre responses to click trains

Thomas Ulrich Christiansen
Ørsted•DTU, Acoustic Technology, Technical University of Denmark,
Ørsteds Plads, Building 352, DK-2800 Kgs. Lyngby, Denmark

Email: tuc@oersted.dtu.dk

Abstract: Computer models of the chinchilla outer- and middle ear, basilar membrane, inner hair-cell and synapse are established and presented. Moreover, combining the models and adjusting parameters for the auditory nerve fibres only, the response of individual auditory nerve fibres to the click train stimulus used in [R. E. Wickesberg, and H. E. Stevens, "Responses of auditory nerve fibers to trains of clicks", J. Acoust. Soc. Am., **103**, (4), 1990-1999 (1998)] was simulated.

The model results are compared to the experimental results from the before mentioned study. Synchronisation coefficients for individual fibres and fibre populations were examined. A monotonically increasing function was expected when plotting synchronisation coefficients as a function of interclick interval. For fibre populations this was modelled. However, more than 40 percent of the fibres showed non-monotonic behaviour. The model is powerful enough to qualitatively reproduce the synchronisation results from the before mentioned study.

Keywords: Computer models, auditory nerve, cochlear models

1. INTRODUCTION

As part of the Ph.D.-project with the running title "A model for the representation of speech signals in normal and impaired human ears" modelling of auditory nerve fibres was carried out. In a study performed by Wickesberg & Stevens [5] data was recorded from the auditory nerve of chinchillas in response to trains of clicks. The present paper describes the efforts involved in replicating the synchronisation coefficients (SC) [3] of the auditory nerve fibres of the original study.

2. THE ORIGINAL STUDY

In the following, data and figures from the original study are reproduced with permission from the authors.

The stimulus

The stimuli reported here consisted of trains of clicks all containing 16 condensation clicks with a width of 0.1 ms. The level and the time between the clicks (interclick interval, ICI) were the parameters varied. Three distinct levels were tested: 59, 69 and 79 dB peak SPL, where "peak SPL" means that the amplitude of the click was equal to the peak amplitude of a 1000 Hz pure tone at the given SPL. The ICI was varied from 1 to 5 ms in steps of 1ms. For further experimental details please refer to [5].

The results

Examples of the responses are given in Figure 1 by means of peristimulus time histograms. Note that increasing the stimulus intensity increases the number of elicited auditory nerve firings (spikes) for both fibres, but decreases the synchronisation for the high spontaneous rate fibre i.e. the individual peaks are no longer visible.

Examples of individual fibre synchronisation to click trains are shown in Figure 2. The SC was calculated according to [3].
SC is a measure the timing of spikes relative to the clicks. If all spikes were perfectly timed the SC would be 1.0. The SC decreases as the timing of the spikes gets less precise with SC reaching the value of 0.0 when spikes are evenly distributed across the stimulus period.
Finally, in Figure 3 the SC as a function of ICI is shown averaged across fibres. The high spontaneous rate and low spontaneous rate fibre populations are shown separately. The responses from each fibre type are shown for the three levels 59, 69, and 79 dB pe SPL.

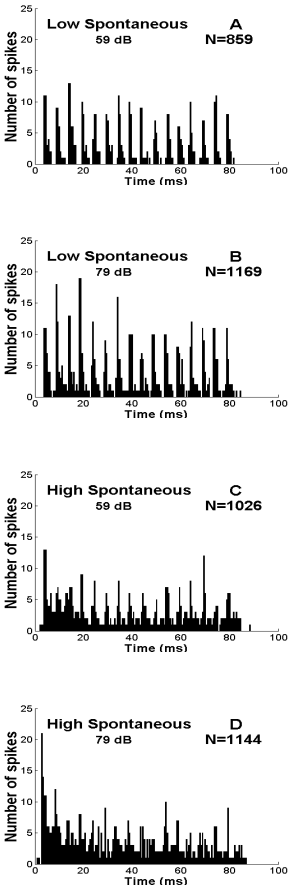


Figure 1: Peristimulus time histograms of responses to 16 clicks for two auditory nerve fibres with (A, B) low (1.2 spikes/s) and (C, D) high (36 spikes/s) spontaneous activity. The interval between clicks were 5 ms and the levels were 59 (A, C) and 79 (B, D) dB peak SPL. Both auditory nerve fibres had characteristic frequencies of 2500 Hz. N=total number of action potentials. Adopted from [4].

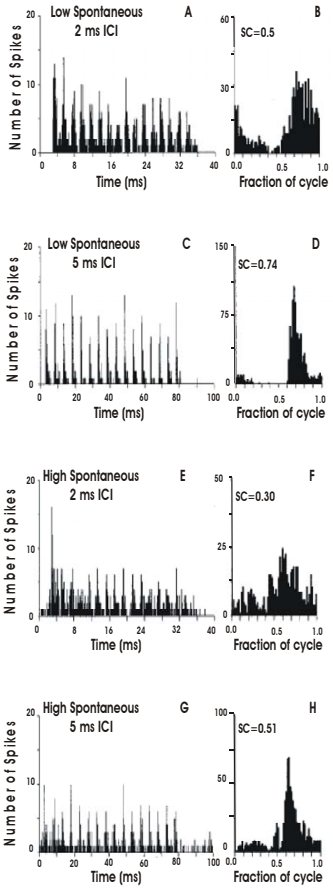


Figure 2: Peristimulus time (A, C, E, G) and period (B, D, F, H) histograms for the responses of low and high spontaneous auditory nerve fibres. The low spontaneous unit (A-D) had a characteristic frequency of 3535 Hz and the high spontaneous rate fibre (E-H) had a characteristic frequency of 3242 Hz. The click trains were presented in interclick intervals of 2 ms (A, B, E, F) or 5 ms (C, D; G, H) at a level of 69 dB pe SPL. The synchronisation coefficients (SC) are displayed for each period histogram. Adopted from [4].

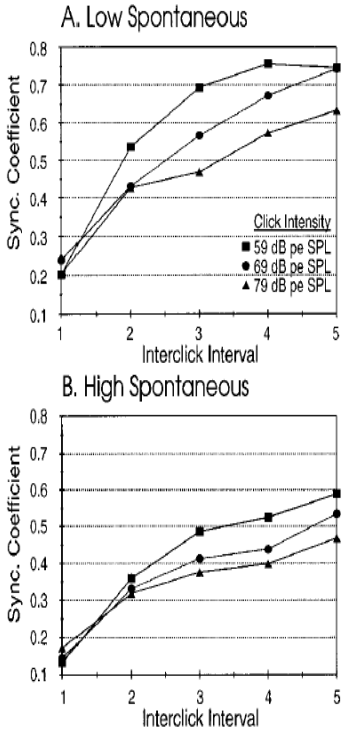


Figure 3: Average synchronisation coefficients determined at the frequency whose period equaled the interclick interval and calculated from the responses to the 16 clicks presented. at levels 59 (■), 69 (●) and 79 (▲) dB pe SPL for low (A) and high (B) spontaneous rate units. Both types of fibres showed a monotonic increase in SC with ICI and a slight decrease in SC with increase in intensity.
Adopted from [4].

3. THE COMPUTER MODEL

The computer model used in this study was developed as a part of the ongoing Ph.D.-project with the running title "A model for the representation of speech signals in normal and impaired human ears". Certain parameters of the model had to be changed in order to reflect chinchilla hearing rather than human hearing. The parameters in question pertain to the middle ear and the basilar membrane.

The middle ear

The sound was delivered to the chinchilla ears by a calibrated insert telephone with a distance of "several millimetres" [5] to the tympanic membrane. In order to simulate the middle ear a second order linear bandpass Butterworth filter with an upper cut-off frequency off 22 kHz and a lower cut-off frequency of 12.5 kHz was employed. This filter is taken from [1]. The output of the filter is converted to stapes velocity and propagated to the basilar membrane model. The stapedial reflex was ignored in the experiment as well as in the modelling efforts.

The basilar membrane

No attempt to model cochlear mechanics was made. Instead the movement of the basilar membrane (BM) was modelled directly by a special filter arrangement: the Dual Resonance Non-Linear filter (DRNL) [2], which shows a remarkably realistic response for a variety of stimuli.

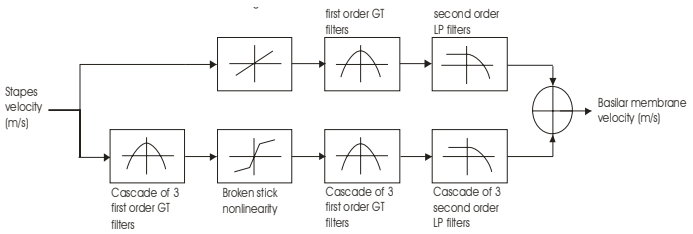


Figure 4: The DRNL construction. Top row shows the linear processing path. The lower row shows the nonlinear path

GT filters = Gamma tone filters.
LP filters= Butterworth low-pass filters

Adopted from [2].

The DRNL has two processing paths: One linear and one nonlinear (cf. Figure 4). The output is the sum of the two processing paths. The linear path consists of a linear gain, a bandpass and a lowpass filter in cascade. The nonlinear path consists of a bandpass filter, a compression function followed by a second bandpass filter. The bandpass filters themselves are cascaded gamma tone filters. The low pass filters consists of four cascaded second order Butterworth filters. Since the centre frequency (CF) of the linear path deviates slightly form the CF of the nonlinear path the filter arrangement shows "dual resonance" hence the name.

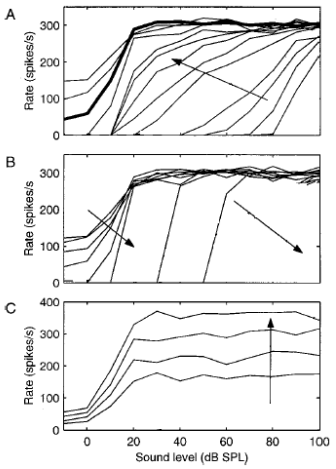
The input to the filter is stapes motion in m/s. The output of the model was chosen to be BM vibration velocity for the given place on the BM. The alternative, BM displacement, could have been used, but it was found that compression threshold changed less with respect to frequency when using the velocity measure [6].

The inner hair cell receptor potential

The BM velocity at the characteristic frequency was propagated to the inner hair cell (IHC) model. The model proposed by Shamma et. al. was used [7]. The IHC receptor potential in volts was calculated. Only one (generic) IHC was used in the modelling efforts i.e. no attempt to model changes in IHC characteristics with characteristic frequency was made.

The IHC-synapse model

The IHC receptor potential was converted into spikes using the IHC-synapse complex model proposed by Sumner et. al. [1].
Each IHC is innervated by multiple nerve fibres each having different "synaptic transmission characteristics". In order to model these different transmission characteristics, parameters of the model was fine-tuned so as to accurately match the data. The bulk of the work pertaining to the present paper lay in the fine-tuning of these synapse parameters.
In the literature fibres are normally characterised by 1) characteristic frequency, 2) spontaneous rate, 3) absolute threshold, 4) saturation level and sometimes also 5) "saturation type" (e.g. sloping saturation). Characteristics 2 through 5 are conveniently displayed as rate/intensity functions (cf. Figure 5).
In the present paper three synapse parameters were selected for manipulation. They were $G_{Ca}^{\max}$ (maximum Calcium conductance), Ca_{thr}^{2+} (Calcium conductance threshold), and M (number of free transmitter quanta).
Calcium plays a significant role in controlling the release of neurotransmitter into the synaptic cleft and in turn to the generation of spikes. If the maximum Calcium conductance is increased, the number of spikes elicited by the fibre goes up (cf. Figure 5 a). The Calcium concentration required to elicit spikes is coined as Calcium concentration threshold. Increasing this threshold makes the number of elicited spikes go down (cf. Figure 5b). The neurotransmitter is delivered to the nerve fibres in "packets" or vesicles. The number of vesicles available for transportation of neurotransmitter is here called "number of free transmitter quanta". Increasing the number of free transmitter quanta increases the number of elicited spikes (cf. Figure 5c).
A detailed discussion on how to manipulate synapse parameters in order to match data for individual fibres is presented in [1].



igure 5: Effect of varying synapse parameters on ite-intensity functions, shown for high spontaneous ite fibres. See text for details.

. The effect of increasing G_{Ca}^{max} in the direction of the arrow.

. The effect of increasing Ca_{thr}^{2+} in the direction of the arrow.

. The effect of increasing M in the direction of the arrow.

dopted from [1]

3. MODELLING THE DATA

The preliminary results of modelling the average SC as a function of ICI for a population of fibres are shown in Figure 6 (original data shown in Figure 3). For convenience the simulations performed and shown here only consist of one single HSR and one single LSR fibre with best frequency 2500 Hz. Thus, the average response of a population is modelled by means of a single fibre,
Qualitatively the results are in good agreement with the results from Figure 3, albeit the SCs for high spontaneous rate fibres are very close for the three levels. Moreover, the low spontaneous rate fibre shows a slight decline in SC from ICI from 2 ms ICI to 3 ms ICI for 79 dB pe SPL.

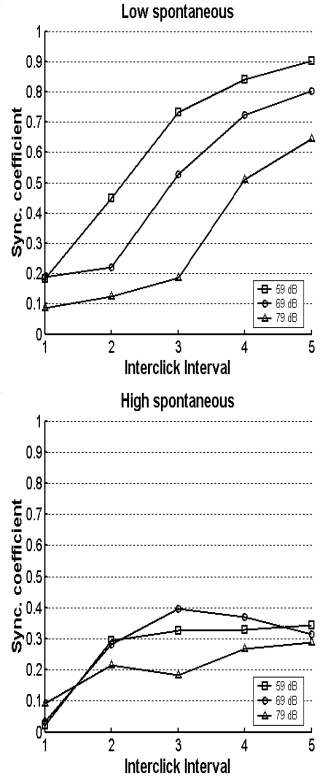


Figure 6: Synchronisation coefficient as function of interclick interval modelled for high spontaneous rate fibres and low spontaneous rate fibres. The results are preliminary.

4. DISCUSSION

In the process of modelling average fibre response it was noticed that some fibres deviate considerably from the average response. An example of such a deviation (taken from the original study) is shown in Figure 7.

The SC/ICI-function for this one fibre is non-monotonic in contrast to the average response of the fibres. Examining the individual fibres from the original study it turns out that more than 40 percent of the fibres show non-monotonic behaviour. It has neither been possible to explain the non-monotonic behaviour nor has it been

possible to find any systematic distribution of local minima for the individual fibres albeit individual fibre responses to click stimuli should be strongly dependant on the characteristic frequency of the fibre [4].

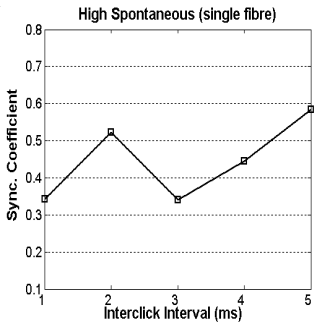


Figure7: Synchronisation coefficients as a function of interclick interval for a single fibre as recorded in [4]. Note the non-monotonic function. Although this fibre represents atypical behaviour the results cannot be readily discarded as erroneous since many fibres show a similar atypical behaviour.
Unpublished data from [4]. Shown with permission from the authors.

5. FUTURE WORK

The modelling results presented here are preliminary in nature. An obvious improvement of the modelling efforts will involve a variety of fibres rather than just the two types (HSR and LSR) applied here. Moreover, a distribution of best frequencies of the fibres would be desirable. The original study [5] only provides some details with respect to fibre characteristics all of which have been applied in the simulations.
The goal of future analysis of the individual fibre data will be to find 1) an explanation of the non-monotonic behaviour SC/ICI-functions and 2) a systematic relation between fibre characteristics and the local minima of the SC/ICI-function. Although there is no guarantee that answers to the above questions exist it seems reasonable to at least investigate the possibility.

REFERENCES

1. C. J. Sumner, E. A. Lopez-Poveda, L. P. O'Mard, and R. Meddis, "A revised model of inner-hair cell and auditory-nerve complex", J. Acoust. Soc. Am. **111** (5), 2178-2188 (2002)
2. E. A. Lopez-Poveda, and R. Meddis, "A human nonlinear cochlear filterbank", J. Acoust. Soc. Am., **110** (6), 3107-3118 (2001)
3. J. M. Goldberg, and P. B. Brown, "Responses of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization", J. Neurophysiol. **32**, 613-616 (1969)
4. N.Y.S. Kiang, T. Watanabe, and L.F. Clarke, "Discharge Patterns of Single Fibers in the Cat's Auditory Nerve", MIT Universiry Press (1965).
5. R. E. Wickesberg, and H. E. Stevens, "Responses of auditory nerve fibers to trains of clicks", J. Acoust. Soc. Am., **103**, (4), 1990-1999 (1998)
6. R. Meddis, L. P. O'Mard, and E. A. Lopez-Poveda, "A computational algorithm for computing nonlinear auditory frequency selecrtivity", J. Acoust. Soc. Am. **109** (6), 2852-2861 (2001)
7. S. A. Shamma, R. S. Chadwick, W. J. Wilbur, K. A. Morrish, and J. Rinzel, "A biophysical model of cochlear processing: Intensity dependence of pure tone responses", J. Acoust. Soc. Am. **80** (1), 133-145 (1986)

Appendix C

Danavox article

This appendix contains a copy of Christiansen [2002]. The page content in the present version is scaled from A5, used in the proceedings, to the regular paper size used throughout the thesis. The original proceedings page numbers have been omitted while thesis page headings are included.

Towards modeling Basilar Membrane Response to Speech

Thomas Ulrich Christiansen
Ørsted•DTU, Technical University of Denmark,
Acoustic Technology,
Building 352, DK–2800 Lyngby, Denmark,
tuc@oersted.dtu.dk

Introduction

Models of the auditory periphery have been proposed addressing either selected parts of the physiology e.g. (Meddis, 1986) or psychoacoustic phenomena e.g. (Kollmeier et. al., 1993).

Today the availability of such models makes it possible, in principle at least, to model the auditory signal processing from the outer ear to the auditory nerve (AN) for normal listeners as well as for impaired listeners.

This paper is part of a Ph.D.-project with the running title "A Model for the Representation of Speech Signals in Normal and Impaired Ears". The aim of the project is to gain a more detailed understanding of the way speech signals are encoded in the AN, and to comprehend the relevance of the encoded signal to speech cues. Moreover, consider a normal hearing person and a person with a sensorineural hearing loss. How does the perception of sound diverge for these two persons? The answers to this fundamental question is sought in the representation of sound in the AN.

This paper examines one of many possible Basilar membrane (BM) models. It is based on results from pulsation threshold experiments performed by Plack and Oxenham (Plack and Oxenham, 2000). The results of this psychoacoustic experiment are used in the design of a Dual-Resonance-Non-Linear filter (DRNL), which mimics the response of the BM in some detail (Meddis et. al., 2001). In particular the DRNL reproduces the nonlinearity with respect to stimulus level, much in line with recent signal processing models e.g. the gamma chirp filter (Irino and Patterson, 1997).

The work previously done (Oxenham and Plack,1997, Plack and Oxenham 2000, Meddis et. al., 2001, Lopez-Poveda and Meddis, 2001), involved simple stimuli *i.e.,* pure tones or clicks. This paper presents tentative results of DRNL-filtering of speech signals. Moreover, the DRNL modeling of BM is enhanced to simulate a sensorineural hearing loss.

Cochlear models: A quick walkthrough

The term cochlear model is most frequently used for models describing the micro- and/or macromechanics of the inner ear. Table I presents an attempt of grouping the various types of cochlear models.

Mechanical							**Conceptual**		
Macro-			**Micro-**		**Other**		**Signal processing**		
1^d	2^d	3^d	Passive	Active	Multi modal	Other	Parallel filter paths	Filters in series	Other

Table I:
An attempt at categorising cochlear models. For details please refer to the text.

Macromechanical models

The macromechanical models focus on the fluids in the cochlear encasement and their interaction with the organ of Corti. Hence they are also known as hydromechanical models. The BM itself is reduced to a single mass with very simple viscoelastic properties.

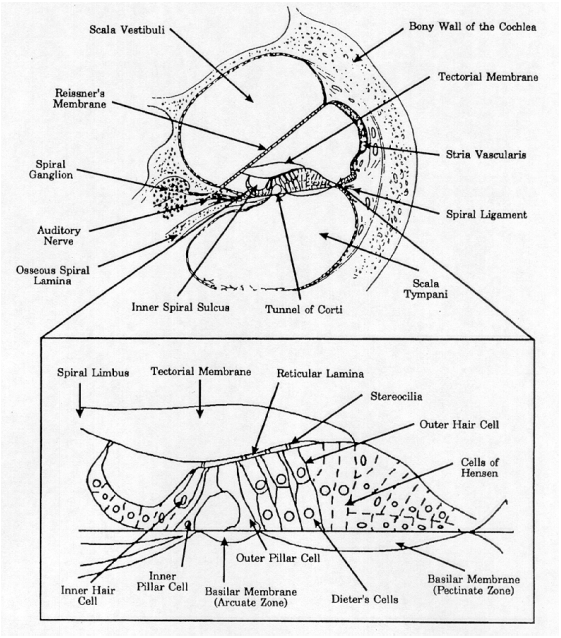


Figure 1:
Schematised drawing of the cross section of one cochlear turn.
Top panel shows the fluid filled scalae and the organ of Corti.
Lower panel shows a magnified view of the Organ of Corti.
Reprinted with permission (Hawkins et al., 1996).
Copyright Springer-Verlag.

One-dimensional models, 1^d in Table I, provide response along the base-to-apex dimension. Two-dimensional models also incorporate displacement in height across the cochlear partition. Three-dimensional models include a description of the radial movement across the BM (cf. Figure 1). Frequently, one-dimensional models are expressed in terms of the pressure difference between the upper- and lower- scalae at the same place on the BM.

As the cochlea narrows from base to apex the scalae-fluid-to-BM-mass ratio decreases. The two-dimensional models take this into account yielding a more accurate description. In particular the steepness of high frequency fall-off is better modeled. Arcuate and pectinate zones along the radial dimension of the BM (cf. Figure 1, lower pane) show differences in compliance. Three-dimensional models have been proposed to take this fact into account. Historically, the success of three-dimensional models has been hindered by the considerable analytical complexity involved. Furthermore, computer power has been a limiting factor.

The discovery of evoked otoacoustic emissions (Kemp,1978) has lead to the emergence of active macromechanical models – models that mimic the active production of kinetic energy in the cochlea. This subdivision of the macromechanical models is not shown in Table I.
Alone the macromechanical models fail to explain significant experimental results (Hawkins et. al., 1996), and today macromechanical models primarily serve a complementary role to the micromechanical models.

Micromechanical models

Turning now to the micromechanical models the focus changes from the fluids in the cochlea to the mechanical properties of the various components constituting the organ of Corti. The complex anatomy of the organ of Corti makes modeling of the mechanical properties an intricate matter subject to ubiquitous speculation.

Micromechanical models are categorised in terms of their “degrees of freedom” rather than their dimensionality. The degrees of freedom are, in turn, determined by the level of detail at which the mechanics of the physiology is modeled. The degrees of freedom are not shown in Table I. As the macromechanical models the micromechanical models can be either passive or active.

Multimodal models allow more than one wave propagation mode. They are either macromechanical or micrmechanical. Multimodal, micromechanical models are “state-of-the-art” in cochlear modeling eg. (Sen and Allen, 1999).

Extensive and detailed experimental animal- as well as human data is necessary for further progress in the area is to occur. Thus progress hinges on improved measuring techniques. Particularly mid-frequency human cochlear responses are in demand, but also simultaneous measurements of pressure and volume velocities would be valuable for shedding light on the active cochlear mechanisms (Hawkins et. al., 1996).

Other mechanical models haven been realised e.g. physical replica of the BM mounted inside a duct as performed at the IBM Zurich, Research Laboratory.

Filter models

Modelling BM response can also be performed at a “conceptual” level *i.e.,* signal processing schemes can be applied in order to produce the desired response irrespective of the intrinsic physiology and mechanics. An approach similar to the one-dimensional model described above is an arrangement of cascaded filters each tuned to a lower frequency than the preceding filter. The travelling wave of the cochlea is modeled as the signal travels along the cascaded filters.

An alternative approach is to present the signal to a filter bank *i.e.,* filters in parallel where each filter represents one place on the BM. The DRNL filterbank presented here falls in this latter category of cochlear models.

The Pulsation Threshold Experiment

The foundation of the DRNL is the pulsation threshold experiment. The pulsation threshold technique first introduced by Houtgast (Houtgast, 1972), can be described as follows:
An interrupted sound is perceived as continuous if another sound fills the interruption period with sufficient energy. Take the example of a pure tone signal (T) and a lower frequency pure tone masker (M). T is alternated with M. Consider the place at the BM with best frequency (BF) equal to that of T. If the BM response to M at this place is equal to or greater than the response to T, the stimulus will be perceived as continuous. For a given frequency and level of T, the pulsating threshold is determined as the level of M at which the perception changes from pulsating to continuous. Frequently a 0.6 T-to-M frequency ratio is used (Plack and Oxenham, 2000).

If T and M are sufficiently far apart in frequency it can be assumed that the BM at the BF of T responds compressive nonlinearly to any increase in the level of T e.g. an increase of 20 dB in the level of T results in (much) less than 20 dB increase in BM response. Moreover, the BM responds liniearly to any increase of the level of M at this place e.g. a 20 dB increase in the level of M yields a 20 dB increase in response. Hence BM compression can be estimated as follows:

For a given level the pulsation threshold is determined. An increase in the level of T is applied and the pulsation threshold is determined for this level. The compression ratio is simply the T-to-M level ratio. By reiterating the method for multiple levels and frequencies the shape of the masking function can be estimated.

The method is remarkable in that it gives an estimate of BM compression from a psychoacoustic experiment. This is particularly interesting since direct measurements are only possible at the apical and basal turns of *in vivo* animal cochlea at present (Plack and Oxenham, 2000). The accuracy of the compression ratio determined using pulsation threshold is quite good, although care must be taken in order to avoid the caveats inherently associated with psychoacoustic experiments.

The DRNL filter

The DRNL models BM response for a given place along the BM. It is based on the same premise as that of the pulsation threshold experiment: At a given place on the BM, the response to a tone, with frequency equal to BF for this place, is compressively nonlinear. In contrast, the BM response to a tone of (much) lower frequency is linear at the same place on the BM.

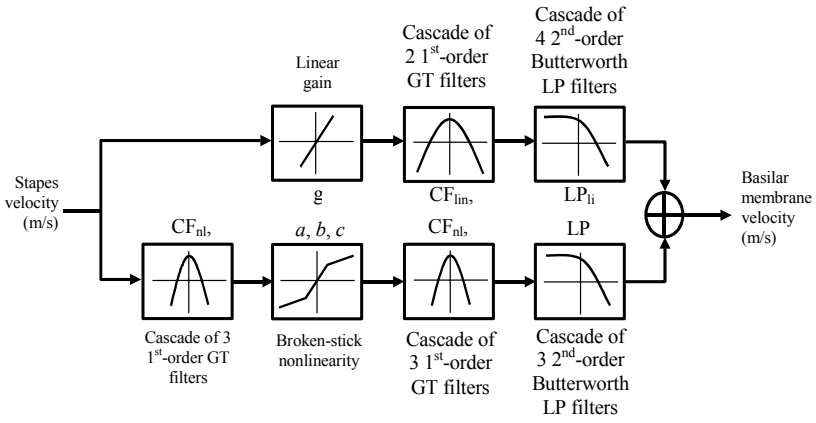


Figure 2:

The DRNL construction. Top row shows the linear processing path.
The lower row shows the nonlinear path.
CF_{lin} = Center frequency of the linear path
GT filters = Gamma tone filters
g = linear gain
CF_{nl} = Center frequency of the nonlinear path
a,b,c = broken stick nonlinearity parameters (see text)
From (Meddis et.al., 2001).

The DRNL has two processing paths: One linear and one nonlinear (cf. Figure 2). The output is the sum of the two processing paths. The linear path consists of a linear gain, a bandpass and a lowpass filter in cascade. The nonlinear path consists of a bandpass filter, a compression function followed by a second bandpass filter. The bandpass filters themselves are cascaded gamma tone filters. The low pass filters consists of four cascaded second order Butterworth filters. Since the center frequency (CF) of the linear path deviates slightly form the CF of the nonlinear path the filter shows “dual resonance” hence the name.

The broken-stick nonlinear function used:

$$y(t) = sign(x(t)) \cdot \min(a|x(t)|, b|x(t)|^c) \qquad (1)$$

where *x(t)* is the input,
*y(t)*is the output,
a,b and *c* are parameters of the model.

The input to the filter is stapes motion hence a middle-ear transfer function is applied. The output of the model was chosen to be BM vibration velocity for the given place on the BM. The alternative, BM displacement, could have been used, but it was found that compression threshold changed less with respect to frequency when using the velocity measure (Meddis et. al., 2001).

Animal data evaluation

In (Meddis et. al., 2001) DRNLs were fitted to chinchilla and guinea pig data. In order to evaluate the quality of the model a number of comparisons to measurements in animals were performed. The comparisons were done for:

1. **Basic BM input/output function**
   The results for the 800 Hz BF for the chinchilla, 10 kHz and 18 Hz BF for the guinea pig where modeled successfully. Remarkably, the notches in BM response reported in animal data at 100 dB SPL was also modeled, at least for the 10 kHz BF.
2. **BM phase response**
   The phase lag is greater in the model than in the animal data for high intensities and frequencies just above BF. Improvement might be achieved by means of broader filters in the nonlinear path.
3. **Two tone suppression (2TS)**
   2TS where not found at the apical site (800 Hz) as expected. Strong 2TS were successfully modeled for the basal sites (10 kHz and 18 kHz).
4. **BM impulse response**
   There are some discrepancies between animal data and model impulse response. The culprit is speculated to be the imperfection of clicks produced by loudspeakers in the animal studies. A simulation of DRNL response to “imperfect clicks” supports the speculation.

5. **Distortion products**
   The distortion products 2f1-f2 and 3f1-2f2 were tested for f1=12.5 kHz and f2=14 kHz for 50 dB SPL stimulus at BF=10 kHz. The results were comparable to the measurements available.

For more details please refer to (Meddis et. al., 2001).

From animal data to a human filterbank

The basic construction of the DRNL was based on animal data. This section presents a description of how the human DRNL filterbank was developed.

DRNL parameter	Description
CF_{lin}	The center frequency of the linear filter path
BW_{lin}	The filter bandwidth for the linear filter path
g	The gain factor of the linear path
CF_{nl}	The center frequency of the nonlinear filter path
BW_{nl}	The filter bandwidth for the nonlinear filter path
a	Broken-stick nonlinearity sensitivity parameter (see equation (1))
b	Broken-stick nonlinearity parameter (see equation (1))
c	Broken-stick nonlinearity compression parameter (see equation (1))

Table II:
The DRNL parameters fitted to human data separately for each of the six BFs.

Pulsation threshold data for BFs at 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz were measured (Plack and Oxenham, 2000). The data were collected for signal levels ranging from 25 dB SPL to 85 dB SPL in steps of 5 dB. From this, six DRNLs corresponding to the BFs of the pulsation threshold data were constructed. The DRNL parameters in Table II were calculated so as to fit the experimental data.

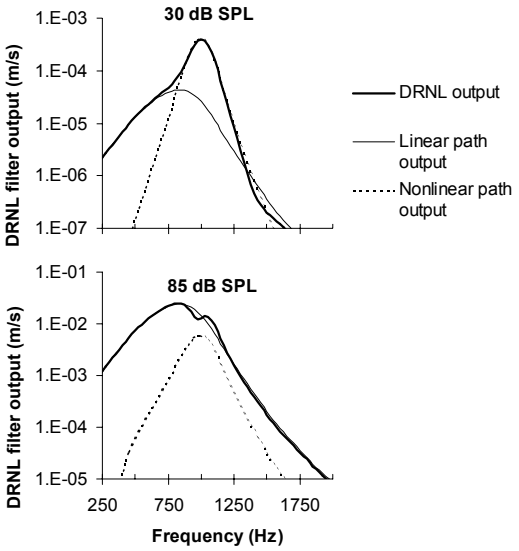


Figure 3
DRNL response for two input levels.
Notice the change in shape and BF.
From (Lopez-Poveda and Meddis, 2001)

At low signal levels the nonlinear path dominates the output, at higher levels the linear path dominates the output.
On the basis of the six DRNLs and the assumption in equation (2) a human filterbank is established. The set of regression coefficients are fitted to match the six filters.

$$\log_{10}(\mathrm{P}_{DRNL}) = p_0 + m\log_{10}(\mathrm{B}) \qquad (2)$$

where P_{DRNL} is a given DRNL-parameter from Table II,
p_0 and m are regression coefficients pertaining to P_{DRNL}
B is the BF at the point of the BM which the filter is calculated for.

Discussion

The DRNL filterbank described here falls in the category of a signal processing model with filters arranged in parallel (cf. Table I).

The advantage of the approach is its ability to reproduce a range of experimental data while maintaining fast computation (Meddis et. al., 2001). The motivation for developing DRNL filters is the necessity to have a processing stage preceding the AN/Inner hair-cell (IHC) complex the eventual goal being AN response to complex stimuli.

The good agreement with experimental data is encouraging and widens the scope for further application of DRNLs. Obvious candidates for future research are hearing impairment modeling and BM response to complex stimuli.

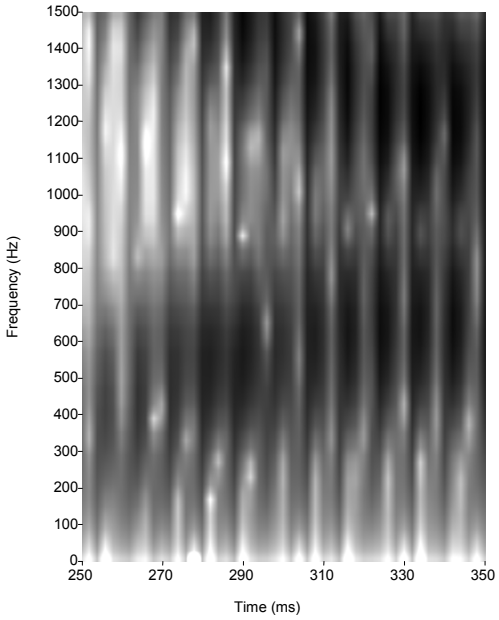


Figure 4:
Spectrogram for the steady-state part of the vowel 'a'
in the English word 'jar' spoken by a male speaker.
Two formants can be seen F_1=600 Hz, F_2=1300 Hz.
The spectrogram was created using the software package 'Praat'.

Figure 4 shows a part of the spectrogram for the vowel 'a' as in the English word 'jar'. The selected part of the spectrogram corresponds to the lower frequencies of the steady-state part of the vowel. This stimulus is used throughout this section.

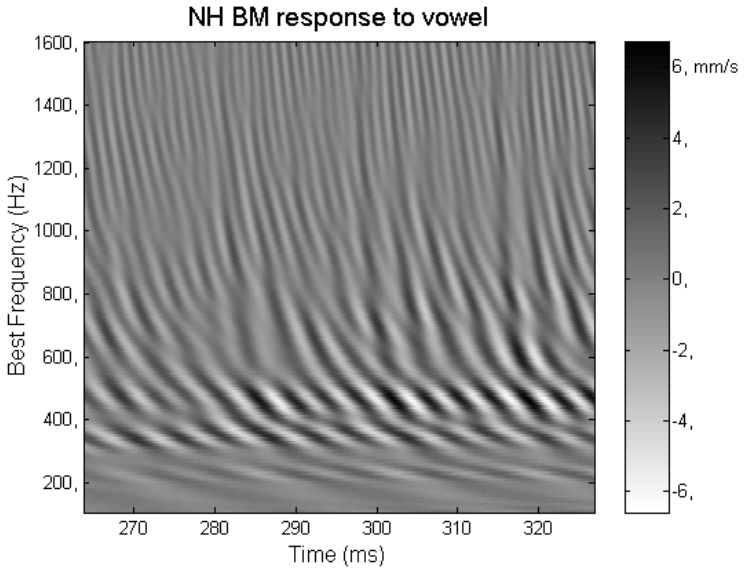


Figure 5:
Simulated BM response for a normal hearing person.
The stimulus is the English vowel 'a' as in the word 'jar' spoken by a male speaker.
Presentation level is 60 dB SPL Only the steady-state part of the response is shown.
Note that the response shown includes contributions from the outer- and middle ear.

The DRNL filterbank modeling human BM response described in this paper was implemented in MatLab. The simulations shown in this section have been performed using this implementation. The DRNL parameters used are adopted from (Lopez-Poveda and Meddis, 2001).

Figure 5 shows the BM response to the steady-state part of the vowel shown in Figure 4. The first formant F_1 at 600 Hz gives a maximum excitation at BF around 500-600 Hz in Figure 5. The response to the second formant F_2 at 1300 Hz is not visible in Figure 5. It is difficult to interpret the results in greater detail at this stage. Note that transfer functions of the outer- and middle ear are incorporated in the response shown in Figure 5.

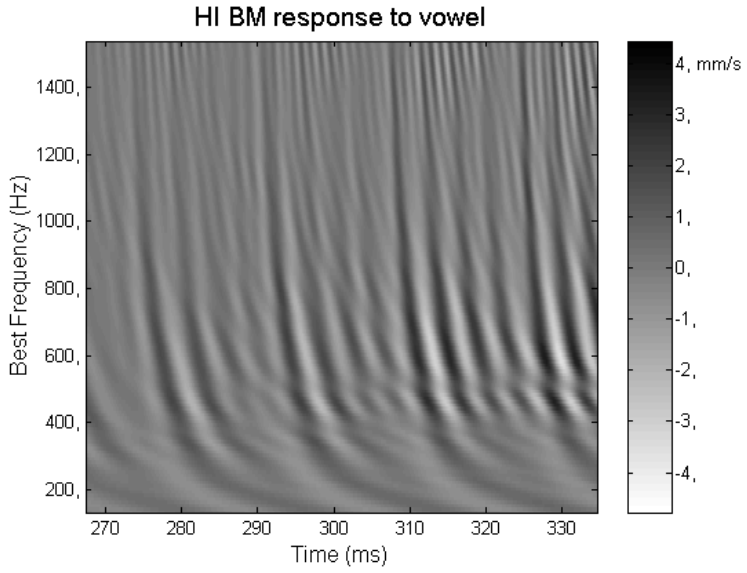


Figure 6:
Simulated BM response for a person with a mild hearing loss.
The stimulus is the English vowel 'a' as in the word 'jar' spoken by a male speaker.
Presentation level is 60 dB SPL. Only the steady-state part of the response is shown.
Note that the response shown includes contributions from the outer- and middle ear.

Modelling sensorineural hearing loss can be achieved by modifying the broken-stick nonlinearity parameters a, b and c of the DRNL (cf. Table II and equation (1)). Indeed this has been successfully demonstrated in (Lopez-Poveda and Meddis, 2001).

Figure 6 shows the simulated BM response for a person with a mild, flat hearing loss to the steady-state part of the vowel shown in Figure 4. The first formant F_1 at 500 Hz has a clear correspondence in Figure 6. The extent of the maximum excitation of the BM, however, is greater than for the normal hearing person (cf. Figure 5). This can be interpreted as the well-known 'broader auditory filters' associated with sensorineural hearing loss. The repsonse to the second formant F_2 at 1300 Hz is not visible in Figure 6.

The simulations presented in this section are based on BM models previously published. They show that it is possible to model BM response to complex stimuli for normal hearing as well as for people with a sensorineural hearing loss. The results are tentative and the accuracy of the model in the context of sensorineural hearing loss and complex stimuli is hard to ascertain. In the course of the Ph.D.-project from which this paper originates, the accuracy will be examined both qualitatively and quantitavely.

Conclusions and future work

This paper presented the DRNL – a signal processing scheme modeling BM response as implemented in (Meddis et. al. 2001, Lopez-Poveda and Meddis, 2001). A DRNL filterbank was developed and fitted to human data. Further, speech stimulus was presented to this human DRNL filterbank. The filterbank parameters were manipulated to mimic a mild sensorineural hearing loss and the response to speech stimulus was simulated. The tentative results are promising although only responses to vowels were considered.

This paper only examines the steady-state part of vowels. The reason is that the transient nature of speech requires a realistic response from the human DRNL filterbank. While DRNLs have proven their effectiveness in response to non-transient stimulus the accuracy of the DRNLs in response to transient stimulus is unclear.

The discrepancies from measurements in phase response combined with the inaccuracies of the impulse response of the DRNLs support the speculation that DRNLs may not be well suited for modeling BM response to speech – at least not in the specific form the DRNLs were presented here.

The DRNL filterbank was applied for modeling sensorineural hearing loss. Arguably this modeling effort was quite crude, but if the hearing loss is characterised only by means of an audiogram, which is quite common, surely the model cannot be expected to provide more detailed results. A more detailed description of the nature of the sensorineural hearing loss viz. in terms of dead zones (Moore et. al., 2000) will pose yet a new challenge for DRNL models.

This paper is part of a Ph.D.-project which seeks to provide a model of the auditory periphery going from the outer ear to the AN. The ambition is to build a model which is realistic in terms of response to complex stimuli *i.e.* realistic in the frequency domain *and* in the time domain. A significant part of the model is the response of the BM as presented in this paper. In the near future work on the temporal aspects of the DRNL filter response will be performed in order to ensure realistic response to speech, in particular the transient parts of speech. Subsequent work related to the IHC/AN complex is planned. Sensorineural hearing loss is an integral part of all the modeling efforts pertaining to the project.

The DRNL was proposed as a component in a larger scale model modeling response to complex stimuli in the auditory periphery (Meddis et. al., 2001). Although some problems with phase response and to some extent impulse response remain open, the DRNL provides a valuable and directly applicable contribution to future models of the auditory periphery or even higher levels of the auditory pathway.

References

Hawkins, H. L, Mullen, T.A., Popper, A.N., Fay R.R., (1996), "Auditory Computation", Springer-Verlag, New York, ISBN 0-387-97843-7, From chapter 3: "Analysis and Synthesis of Cochlear Mechanical Function Using Models" by Mountain D. C, Hubbard, A. E..

Houtgast T., (1972), "Psychophysical evidence for lateral inhibition in hearing", J. Acoust. Soc. Am. **51**, p. 1885-1894

Irino T., Patterson R. D. (1997), "A time domain level-dependant auditory filter: The gammachirp", J. Acoust. Soc. Am. **101,** 412-419.

Kemp, D.T., (1978), "Stimulated acoustic emissions from within the human auditory system", J. Acoust. Soc. Am. **64,** p. 1386-1391

Kollmeier B., Derleth, R.P. & Dau T., (1993), "Modeling the 'effective' auditory signal processing for hearing impaired listeners", Psychophysical and Physiological Advances in Hearing. A.R. Palmer, p.482-490, Whurr Publisher.

Lopez-Poveda E.A., Meddis R., (2001), "A human cochlear filterbank", Unpublished paper, referenced with permission.

Meddis, R. (1986), "Simulation of mechanical to neural transduction in the auditory receptor", J. Acoust. Soc. Am. **79**, 702–711.

Meddis R., O'mard L.P., Lopez-Poveda, E.A. (2001), "A computational algorithm for computing nonlinear auditory frequency selectivity", J. Acoust. Soc. Am. **109**, p. 2852-2861

Moore, M. Huss, D.A. Vickers and T. Baer, (2000), "Psychoacoustics of dead regions", Physiological and Psychophysical Bases of Auditory Function, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs and R. Schoonhoven, ed., Shaker, Maastricht

Oxenham, A.J, Plack, C. J., (1997), "A behavioral meassure of basilar-membrane nonlinearity in listeners with noraml and impaired hearing", J. Acoust. Soc.Am.,**101**, p.3666-3675.

Plack, C. J.and Oxenham, A.J., (2000), "Basilar-membrane nonlinearity estimated by pulsation threshold", J. Acoust. Soc.Am.,**107** (1), p.501-507.

Sen D., Allen, J. B. (1999), "Reproducing upward spread of masking and two-tone suppression using a nonlinear model of the cochlea", ARO, 22(ABST:480):121, February 1999.

Appendix D

Guided trial-and-error method

This appendix describes the method employed when trying to optimise the DRNL parameters to account for the forward data experiment Lopez-Poveda et al. [2002].

Vibration at a given place on the basilar membrane at a time Δt after offset is of a given signal is a function of probe level l, probe frequency f, and time interval Δt;

$$V = E(f, l, \Delta t) \tag{D.1}$$

where V is basilar membrane velocity and E is a "general excitation function". Since the temporal window is essentially a temporal window V can be expressed as

$$V = E_t(f, l)e^{-\Delta t)/\tau} \tag{D.2}$$

where E_t is excitation function, and τ is a time constant related to the temporal window.

Since the probe level is fixed at 14 dB SL, we assume that the excitation produced by the probe is linearly related to the sound level l in Equation D.2 with $\Delta t = 0$. We can thus express the level of the probe:

$$E_t(f, 14dBSPL) = yP \tag{D.3}$$

where P is the sound pressure level an y is the proportionality factor converting to excitation.

Assuming that forward masking thresholds occur when masker and probe produce a "comparable level of excitation" Equations D.2 and D.3 can be combined:

$$E_t(f, 14dBSPL) = yPe^{\Delta t)/\tau} \tag{D.4}$$

The value of τ can be estimated from the data as in Lopez-Poveda et al. [2002]. The proportionality factor y can be estimated based on physiological observations close to the absolute threshold. The values used in the thesis were $\tau = 0.036$ and $y = 10^{-4}$.

With these assumption we can plot an input/output function based on Equation D.4 and the forward masking threshold data form Lopez-Poveda et al. [2002]. An example of this is shown in Figure D.1.

The *modelled* forward masking thresholds is now compared to the calculated I/O-function. This is simply done by calculating the basilar membrane velocity in response to pure tones at the level of the measured forward masking thresholds. Such a comparison is shown in Figure D.1

In this manner the original task of manipulating DRNL parameters to simulate forward masking thresholds has been transformed into a task of matching input output functions. This makes the task considerably easier in that the effects of manipulating the DRNL parameters are more readily understood. For instance increasing the parameter g (linear gain) will move the red line upward. This way of visualising the effects of changing DRNL parameters enables a much faster fitting process. The computations involved are all straightforward and the effects of new DRNL parameters can be estimated very quickly. However, to precisely check and compare forward masking thresholds the time consuming method described in Section 3.2.3 has to be followed.

As indicated above the method provides a visual representation of an *approximation* to the forward masking threshold (sometimes called the "chi-by-eye" approach [Press et al., 2002]). The way the method was employed throughout the thesis was to get a good fit from a visual point view i.e. no data points were considered more important than others.

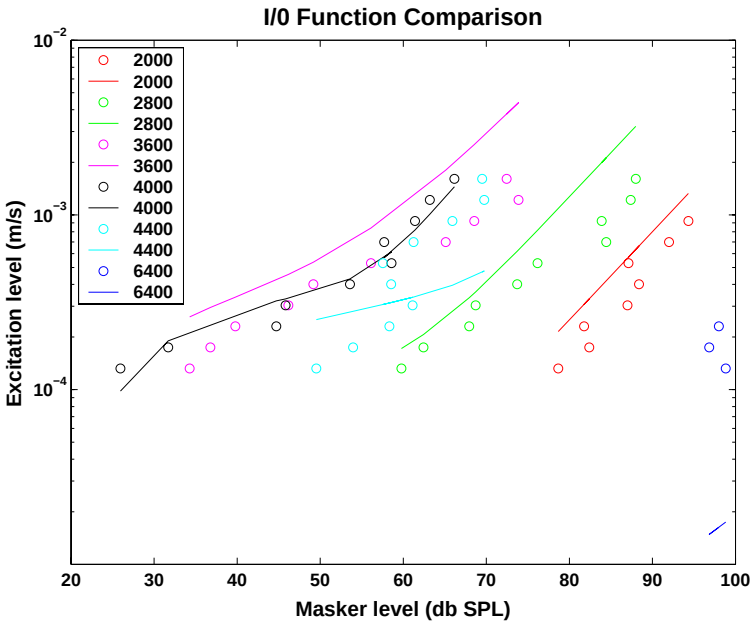


Figure D.1: Comparisons of input/output functions for measure data (points) and simulated data (lines). The colours designate the different probe-to-masker frequency ratios. The abscissa values designate the measured forward masker levels. The ordinate is the excitation level form the model (lines) and the calculated values (points). The example above is for the test subject ELP, centre frequency 4kHz

Appendix E

Forward masking simulation results

This appendix presents the result of the simulations for the three forward masking studies examined in the thesis. While the details of the most important simulations are presented and discussed in the text this appendix shows the "second-best" results.

E.1 Forward masking with varying masker frequencies

This section presents complementary simulation results for the experiment presented in Section 3.2.1.

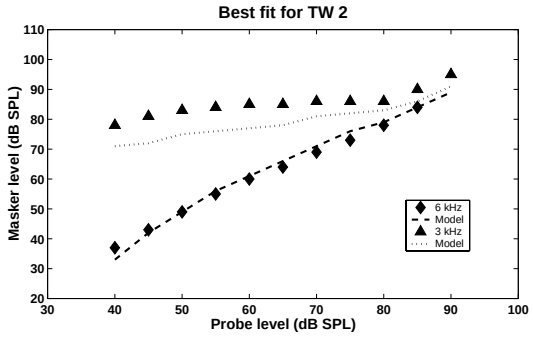


Figure E.1: Forward masking study with varying masker frequencies. The RMS errors was 4.17. The criterion value k was 1.5

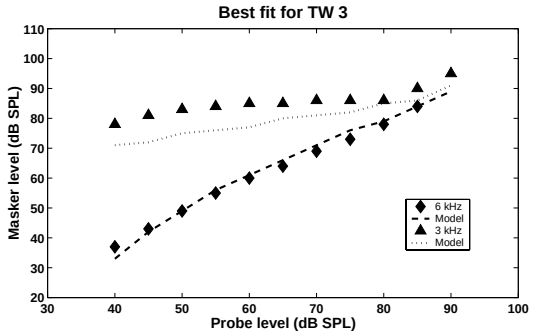


Figure E.2: Forward masking study with varying masker frequencies. The RMS errors was 4.05. The criterion value k was 1.3

E.2 Forward masking with varying probe and gap durations

This section presents complementary simulation results for the experiment presented in Section 3.2.2.

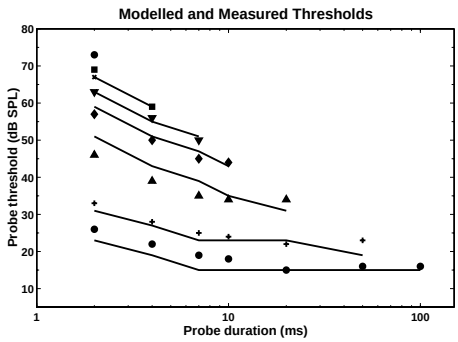


Figure E.3: Forward masking study with varying probe and gap durations. The RMS errors was 2.61. The criterion value k was 1.30

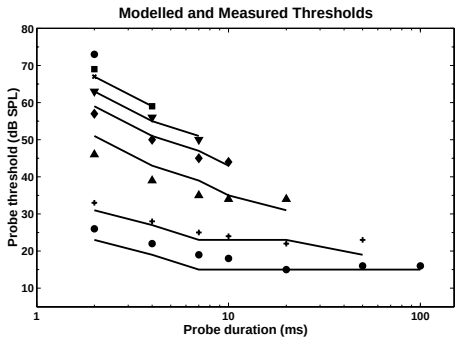


Figure E.4: Forward masking study with varying masker frequencies. The RMS errors was 3.38. The criterion value k was 2.00

E.3 Forward masking with varying frequencies and gap durations

This section presents complementary simulation results for the experiment presented in Section 3.2.3.

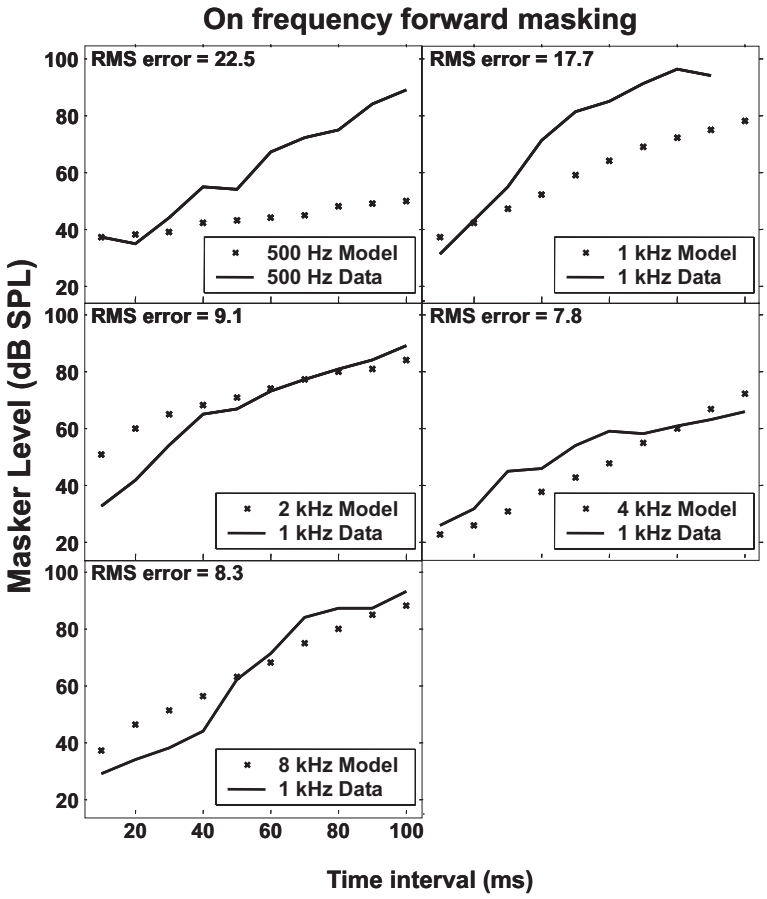


Figure E.5: Comparison of forward masking results for the on frequency results. The criterion value k was varied across frequency. This plot is shows the best fit with the TW 1 temporal window.

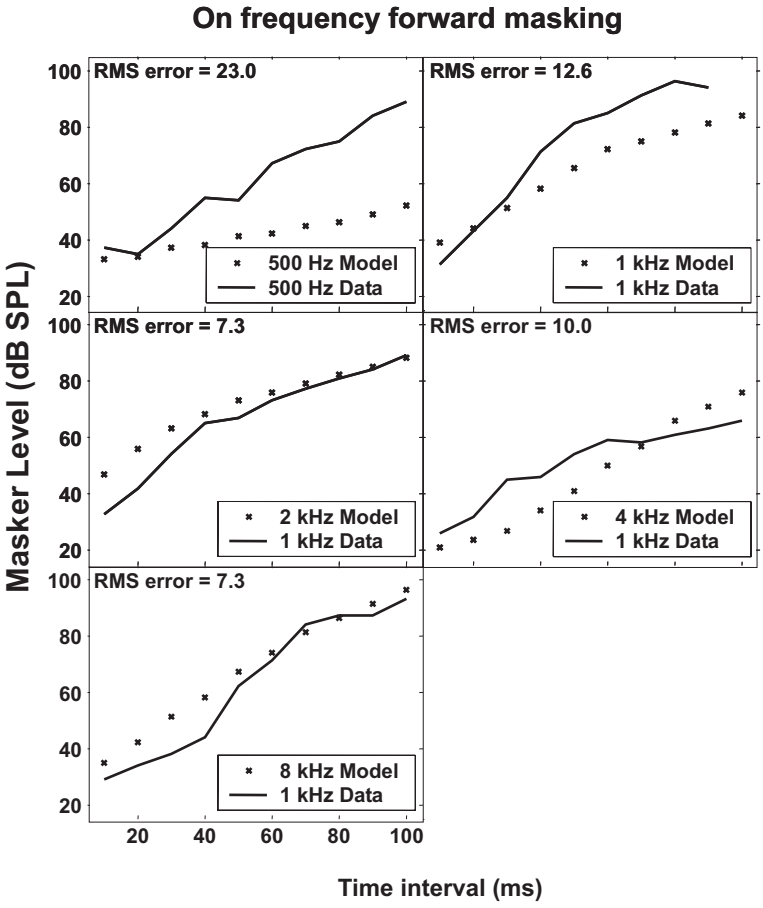


Figure E.6: Comparison of forward masking results for the on frequency results. The criterion value k was varied across frequency. This plot is shows the best fit with the TW 2 temporal window

Appendix F

Inner hair cell and auditory nerve complex parameters

Table F.1: Inner Hair Cell Receptor Potential parameters as used in the model. From Sumner et al. [2002] and Shamma [1986]

E_t	Endocochlear potential	Volt	1.00×10^{-3}
E_k	Potassium reversal potential	Volt	-7.045×10^{-4}
$G0$	Resting conductance	Siemens	1.974×10^{-9}
g_k	Potassium conductance	Siemens	1.8×10^{-8}
$E_k\prime$	Correcttion, $R_p/(R_t + R_p)$	Scalar	0.04
G_{cilia}^{max}	Max mechanical conductance	Siemens	8×10^{-8}
s_0	Displacement sensitivity	meter^{-1}	8.5×10^{-8}
u_0	Displacement offset	meter	7×10^{-9}
s_1	Displacement sensitivity	meter^{-1}	5×10^{-7}
u_1	Displacement offset	meter	7×10^{-9}
C_m	Total capacitance	Farad	6×10^{-12}
τ_c	Cilia/BM time constant	Seconds	2.13×10^{-3}
C_{cilia}	Cilia/BM Coupling gain	dB	16

Table F.2: The Inner Hair Cell and Auditory Nerve Complex Parameters shared for all fibre types. From Sumner et al. [2002]

z	Scalar	$(\text{seconds}[Ca^{2+}]^3)^{-1}$	2×10^{32}
E_{Ca}	Reversal Potential	Volt	6.6×10^{-2}
β_{Ca}		Scalar	400
γ_{Ca}		Scalar	130
τ_m	Calcium current time constant	seconds	1×10^{-4}
τ_{Ca}	Calcium diffusion time constant	seconds	1×10^{-4}
y	Replenishment rate	seconds^{-1}	10
l	Loss rate	seconds^{-1}	2.58×10^3
x	Reprocessing rate	seconds^{-1}	6.63×10^1
r	Recovery rate	seconds^{-1}	6.58×10^3

Table F.3: The Inner Hair Cell and Auditory Nerve Complex Parameters distinguishing fibre types. From Sumner et al. [2002]

			HSR	MSR	LSR
$G_{\text{Ca}}^{\max}$	Max Calcium Conductance	10^{-9} Siemens	8	4.5	2.75
$[Ca^{2+}]_{thr}$	Ca concentration threshold	$\times 10^{-11}$	4.48	3.2	4
M	Max free transmitter quanta	Scalar	10	10	6

Table F.4: The Inner Hair Cell and Auditory Nerve Complex Parameters distinguishing fibre types. From Sumner et al. [2003]

			HSR	MSR	LSR
$G_{\text{Ca}}^{\max}$	Max Calcium Conductance	10^{-9} Siemens	7.2	2.4	1.6
$[Ca^{2+}]_{thr}$	Ca concentration threshold	$\times 10^{-14}$	0	3.35	1400
M	Max free transmitter quanta	Scalar	10	10	10